

RESEARCH ARTICLE

Open Access



# Evaluation of undetected cases during the COVID-19 epidemic in Austria

C. Rippinger<sup>1,2\*</sup>, M. Bicher<sup>1,2</sup>, C. Urach<sup>1</sup>, D. Brunmeir<sup>1</sup>, N. Weibrecht<sup>1</sup>, G. Zauner<sup>1</sup>, G. Sroczynski<sup>3</sup>, B. Jahn<sup>3,4</sup>, N. Mühlberger<sup>3</sup>, U. Siebert<sup>3,4,5,6</sup> and N. Popper<sup>1,2</sup>

## Abstract

**Background:** Knowing the number of undetected cases of COVID-19 is important for a better understanding of the spread of the disease. This study analyses the temporal dynamic of detected vs. undetected cases to provide guidance for the interpretation of prevalence studies performed with PCR or antibody tests to estimate the detection rate.

**Methods:** We used an agent-based model to evaluate assumptions on the detection probability ranging from 0.1 to 0.9. For each general detection probability, we derived age-dependent detection probabilities and calibrated the model to reproduce the epidemic wave of COVID-19 in Austria from March 2020 to June 2020. We categorized infected individuals into presymptomatic, symptomatic unconfirmed, confirmed and never detected to observe the simulated dynamic of the detected and undetected cases.

**Results:** The calculation of the age-dependent detection probability ruled values lower than 0.4 as most likely. Furthermore, the proportion of undetected cases depends strongly on the dynamic of the epidemic wave: during the initial upswing, the undetected cases account for a major part of all infected individuals, whereas their share decreases around the peak of the confirmed cases.

**Conclusions:** The results of prevalence studies performed to determine the detection rate of COVID-19 patients should always be interpreted with regard to the current dynamic of the epidemic wave. Applying the method proposed in our analysis, the prevalence study performed in Austria in April 2020 could indicate a detection rate of 0.13, instead of the prevalent ratio of 0.29 between detected and estimated undetected cases at that time.

**Keywords:** COVID-19, Undetected cases, Agent-based modelling

## Background

Undetected severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) infections play an important role in the spread of coronavirus disease 2019 (COVID-19). A simulation study estimated that, before widespread travel restrictions were put in place (January 10th – 23rd 2020), 86% of all infections in China were undetected and that up to 80% of the documented

infected individuals had been infected by undetected infected individuals [1]. Infections are not recognized due to mild or absent symptoms, limited awareness of the virus in the general population, or lack of testing. The proportion of unreported cases, however, changes over time because of increased awareness, improved testing strategies and established measures changing the spread of the epidemic (isolation, self-quarantine, contact precautions, and travel restrictions). For example, in China, immediately after travel restrictions were in place (after January 23rd 2020), the fraction of all undetected infections decreased to 35% [1].

\* Correspondence: [claire.rippinger@dw.at](mailto:claire.rippinger@dw.at)

<sup>1</sup>DWH Simulation Services, DEXHELPP, Neustiftgasse 57-59, 1070 Vienna, Austria

<sup>2</sup>TU Wien, Institute of Information Systems Engineering, Favoritenstraße 11, 1040 Vienna, Austria

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Although undetected cases of COVID-19 contribute significantly to the spread of the disease, it is still unclear how large the number of unreported cases really is. Clearly, it strongly depends on the predominant testing strategy but also on the proportion of asymptomatic cases. For the latter, estimates have been made, such as that from Iceland where 43% of the positive cases in an overall population screening reported no symptoms [2] and for passengers of the Diamond Princess resulting in an asymptomatic proportion of only 17.9% [3].

Methods commonly used to determine the number of undetected COVID-19 cases include representative studies using reverse transcription polymerase chain reaction (PCR) or antibody tests. In the first, a random sample of the general population is drawn and PCR tests are applied to identify currently infected individuals independent of current symptoms or recent exposure [4–6]. These studies provide an approximation of unreported cases, but only for a specific point in time. On the other hand, studies using antibody tests try to identify people who have been infected by SARS-CoV-2 in the past. Both types of studies are often limited due to sample size and representation of specific populations or regions and results depend on test accuracy and the time point of sample collection (when virus or antibodies are already or still detectable).

In this modeling study, we aim to determine the impact of detection probabilities on the temporal variation of the epidemic, including the fractions of undetected and detected cases of COVID-19 during an epidemic wave. We account for preventive measures in a retrospective analysis of secondary data using an agent-based model representing the entire Austrian population. We simulate different scenarios to account for the uncertainty in the proportion of undetected cases due to contradictory or lacking data. Our results can guide the interpretation of the results of prevalence studies and estimations of the proportion of undetected cases of COVID-19 in Austria.

## Methods

We used a previously published agent-based model to simulate the COVID-19 epidemic in Austria by modeling the different disease stages and events of the patient pathway [7]. Considering age-dependent detection probability, we calibrated the model to the past epidemic curve in Austria. As model results, we considered the ratio between detected and undetected cases of COVID-19 during the different phases of the epidemic.

### Agent-based simulation model

A detailed description of the model is provided by Bicher et al. [7]. Briefly, the agent-based model simulates each member of the Austrian population using statistical

representatives, mapping a contact network based on different locations, such as households, workplaces, schools, and leisure time [8]. The model considers the entire disease-pathway of a COVID-19 patient with its different potential stages and events starting with the infection of healthy individuals. The model considers the time period between infection and symptom onset, as well as a “reaction-time” that covers the time between symptom onset, testing, positive test result, and the time at which the COVID-19 patient is recorded as a confirmed case in the official vigilance system. In an alternative pathway, infected individuals in the model would never be tested for COVID-19 and remain an undetected case throughout the simulation.

The time period from February 19th 2020 to June 10th 2020 is simulated and reported outcomes include the number of detected and undetected cases in each of the different disease stages for each point in time. In the model, four groups of infected individuals are distinguished: *presymptomatic*, *symptomatic unconfirmed*, *confirmed* and *never detected*. *Never detected* individuals will never be tested for COVID-19. Their disease course is modeled over 9.4 days in total, accounting for the time from infection until the infectious period and the time of the infectious period. *Presymptomatic*, *symptomatic unconfirmed*, *confirmed* individuals receive a positive test result at some point in time during the simulation. *Presymptomatic* individuals are still in the incubation period and *symptomatic unconfirmed* individuals have started to experience symptoms but have not yet received a positive test result. Finally, the *confirmed* cases are recorded until their recovery, considering different time periods depending on the severity of the disease. Undetected cases at a given time consist of *presymptomatic*, *symptomatic unconfirmed* and *never detected*. Fatal COVID-19 infections are not investigated for this analysis and are counted as recovered cases. Confirmed cases are classified into several disease severities: mild cases which can recover at home, severe cases which require hospitalization, and critical cases which require treatment at an intensive care unit (ICU). Table 1 provides an overview of the parameter values for the mentioned time periods.

### Age-dependent detection probability

The detection probability  $\theta$  determines whether an infected individual will be tested positive for COVID-19. It combines the probability of developing symptoms specific enough to get tested for COVID-19, and the probability of being detected because of other testing strategies, such as screening of care homes. We made the simplifying assumption that  $\theta$  remains constant for all age-groups during our considered time interval (March to June 2020) but is in fact age dependent due

**Table 1** Parameter values for the model

	Value	Reference
Incubation period	5.1 days (1.78)	Lauer et al. [9]
Latency period	2 days	Robert Koch Institute [10]
Time until infectious period	3.1 days (1.78)	Incubation period minus latency period
Infectious period	6.3 days (1.25)	Expert estimates in line with Robert Koch Institute [10]
Delay from symptom start to positive test result	3.8 days (2.38)	Hellewell et al. [11]
Time between positive test result and recovery for detected mild cases	13.3 days (2.46)	Based on reported COVID-19 data in Austria (Epidemiologisches Meldesystem [12])
Time between positive test result and recovery for detected severe cases	20.0 days (11.75)	Based on reported COVID-19 data in Austria (Epidemiologisches Meldesystem [12])
Time between positive test result and recovery for detected critical cases	25.0 days (9.09)	Based on reported COVID-19 data in Austria (Epidemiologisches Meldesystem [12])

Data are mean (standard deviation). The latency period is assumed to be a fixed value

to reported evidence of increasing probability to develop more severe symptoms with increasing age [13].

To determine the age-dependent detection probability, we first calculated the cumulative incidence of confirmed cases in Austria  $\sigma = \frac{\#confirmedCases}{\#population}$  as of May 6th 2020. We then assumed that all age groups have been affected equally by the disease and that the mismatch between the age distribution of the confirmed cases and the Austrian age pyramid is caused by an age-dependent detection rate  $\theta_i$ . Considering 10-year age groups, we calculate

$$\tilde{\theta}_i = \frac{c_i \cdot \theta}{p_i \cdot \sigma}, \tag{1}$$

whereby  $c_i$  denotes the cumulative number of confirmed cases of this age group and  $p_i$  denotes the total population of this age group. For high values of  $\theta$ , some age groups, especially the elderly people, are overrepresented, resulting in a value  $\theta_i > 1$ . Therefore, we finally compute

$$\theta_i = \min(\tilde{\theta}_i, 1). \tag{2}$$

**Calibration**

Model calibration is the process of adjusting model parameters to match the data observed in the real world, in this case confirmed COVID-19 cases. To achieve this goal, we adjust the detection probability and the infection probability in case of contact.

In our model, increasing the detection probability parameter has two major consequences: (a) the number of confirmed cases in the model at a given point in time

increases because more cases are detected, and (b) the number of confirmed cases is decreasing over time because there are fewer undetected cases causing uncontrolled infections throughout their whole infectious period without being quarantined. Decreasing the detection probability has the opposite effect. Therefore, the infection probability is adjusted whenever the detection probability is changed to keep the number of confirmed cases at the desired level.

To account for the lack of information considering  $\theta$ , the calibration of the infection probability has been performed for different values of  $\theta$  ranging from 0.1 to 0.9 to reproduce the epidemic curve of COVID-19 cases in Austria [14]. This calibration process has been split into two parts: calibrating the infection probability  $\alpha_1$  and  $\alpha_2$  before and after the peak of the epidemic wave. In the first part, the infection probability  $\alpha_1$  is calibrated using a bisection algorithm with a Monte-Carlo Simulation in the loop to reproduce the officially reported number of confirmed cases at the peak of the epidemic wave. In the second part,  $\alpha_2$  is adjusted using the same bisection algorithm to reproduce the decrease of the officially reported number of positive COVID-19 cases in Austria until June 10th 2020. The lockdown measures are modelled equivalently and as described by Bicher et al.<sup>7</sup> and include school closure, increased home office use, leisure time contact reduction, and increased hygienic measures starting on March 16th 2020 and being gradually lifted from April 16th 2020 onwards.

**Actual detection rate and interpolation**

While the detection probability  $\theta$  is a model input, the observed detection rate  $\vartheta$  can be obtained as a model output by evaluating

$$\vartheta = \frac{\#confirmedCases}{\#allInfectedCases} \tag{3}$$

at the end of the simulation. To clearly differentiate between the input parameter  $\theta$  and the observed model output  $\vartheta$ , we henceforth denote the first as detection probability and the second as detection rate. The detection rate relies heavily on the detection probability. Therefore, the model results for different assumptions of  $\theta$  can be expressed as a function of  $\vartheta$ . Assuming the model output reacts steadily to changes in  $\theta$ , model output for input parameters  $\theta$  in between the directly calibrated input parameters can be interpolated using piecewise linear splines.

The model was developed, calibrated, and analyzed following the International Society for Pharmacoeconomics and Outcomes Research – Society for Medical Decision Making (ISPOR-SMDM) Modeling Good Research Practices guidance [15].

**Results**

Age-dependent detection probabilities  $\theta_i$  given an overall detection probability  $\theta$  calculated via (1) are provided in Table 2. As expected, with increasing  $\theta$ , all age-specific  $\theta_i$  increase as well with younger age groups having lower values than older age groups. It can be seen that  $\theta$  lower than 0.4 leads to no overrepresented age groups, that is age classes  $i$  with  $\tilde{\theta}_i > 1$  (compare (2)).

Calibrated infection probabilities gained by fitting the simulation results to the epidemic wave in Austria from March 11th 2020 to June 18th 2020 for different assumptions of the overall detection probability  $\theta$  are shown in Table 3. Results indicate that the impact of a varied detection probability on the confirmed cases can be adjusted with small variation of the infection probability  $\alpha_1$  and  $\alpha_2$  before and after the observed peak of the confirmed cases.

Figure 1 shows the simulated results for the epidemic wave in Austria from March 11th 2020 to June 18th 2020 for two exemplary values of  $\vartheta$ , distinguishing the four different groups of infected individuals mentioned in section “Methods”. Each group of infected individuals peaks at a different point in time due to different disease durations and sequences within the disease pathway. In particular, the peak of the sum of all groups, which we will denote as the *real peak*, occurred a few days before the peak of the confirmed cases, the *official peak*, on April 2nd. Figure 2 visualizes how the real peak would change with increasing the detection probabilities. The left axis displays the number of cases at the real peak, the right axis its date. It shows that increasing the detection rate leads to a lower and earlier real peak. Figure 3 displays the cumulative number of infected individuals until June 18th 2020 depending on the detection rate, which directly correlates with the cumulative incidence in the population. With unchanged number of cumulative confirmed cases (lower curve) the overall cumulative cases (upper curve) decreases with the detection probability, leading to a lower cumulative incidence.

Figure 4 shows how the ratio between currently infected detected and undetected cases changes during the epidemic. In the upswing phase, the detected cases account for a very small part of all the infected cases, whereas their share increases rapidly with the start of the lockdown policies on March 16th. The share of confirmed cases increases more slowly at the time of the *official peak* and decreases again in the “steady-state phase” at the end of the epidemic wave. Figure 5 shows the same for the cumulative detected and undetected cases. Again, the cumulative detected cases account for only a small proportion of all cumulative infected cases during the upswing phase of the epidemic, before this proportion reaches a constant state a few days after the *official peak*.

**Table 2** Values for  $\theta_i$  for a given  $\theta$

	$\theta_0$	$\theta_{10}$	$\theta_{20}$	$\theta_{30}$	$\theta_{40}$	$\theta_{50}$	$\theta_{60}$	$\theta_{70}$	$\theta_{80}$	$\theta_{90}$
$\theta = 0.1$	0.02	0.04	0.11	0.10	0.12	0.14	0.10	0.10	0.16	0.29
$\theta = 0.2$	0.03	0.09	0.22	0.20	0.24	0.28	0.20	0.21	0.33	0.58
$\theta = 0.3$	0.05	0.13	0.32	0.29	0.36	0.41	0.30	0.31	0.49	0.87
$\theta = 0.4$	0.06	0.17	0.43	0.39	0.48	0.55	0.40	0.41	0.65	1.00 <sup>(1)</sup>
$\theta = 0.5$	0.08	0.21	0.54	0.49	0.59	0.69	0.50	0.51	0.82	1.00 <sup>(1)</sup>
$\theta = 0.6$	0.09	0.26	0.65	0.59	0.70	0.83	0.59	0.62	0.98	1.00 <sup>(1)</sup>
$\theta = 0.7$	0.11	0.30	0.75	0.69	0.83	0.97	0.69	0.72	1.00 <sup>(1)</sup>	1.00 <sup>(1)</sup>
$\theta = 0.8$	0.12	0.34	0.86	0.79	0.95	1.00 <sup>(1)</sup>	0.79	0.82	1.00 <sup>(1)</sup>	1.00 <sup>(1)</sup>
$\theta = 0.9$	0.14	0.39	0.97	0.88	1.00 <sup>(1)</sup>	1.00 <sup>(1)</sup>	0.89	0.93	1.00 <sup>(1)</sup>	1.00 <sup>(1)</sup>

$\theta_i$  denotes the age-dependent detection probability for people within the age group [i, i+9].  $\theta_{90}$  denotes the age-dependent detection probability for people aged 90 or older. Cells marked with <sup>(1)</sup> indicate that the age class is overrepresented with  $\tilde{\theta}_i > 1$  and were truncated to one via (2)

**Table 3** Calibrated values for the infection probability  $\alpha_1$  and  $\alpha_2$  for different detection probabilities  $\theta$

	$\theta = 0.1$	$\theta = 0.2$	$\theta = 0.3$	$\theta = 0.4$	$\theta = 0.5$	$\theta = 0.6$	$\theta = 0.7$	$\theta = 0.8$	$\theta = 0.9$
Infection probability before peak $\alpha_1$	5.960%	5.862%	5.840%	5.840%	5.850%	5.960%	6.030%	6.032%	6.047%
Infection probability after peak $\alpha_2$	5.215%	4.502%	4.526%	4.830%	5.148%	5.394%	5.729%	6.394%	6.555%
Observed detection rate $\vartheta$	0.09	0.18	0.27	0.37	0.45	0.55	0.63	0.70	0.75

Each parameter value triple  $(\theta, \alpha_1, \alpha_2)$  is specified to lead to the same curve for the confirmed infected individuals. The row  $\vartheta$  displays the actual detection rate evaluated by (3) at the end of the simulation

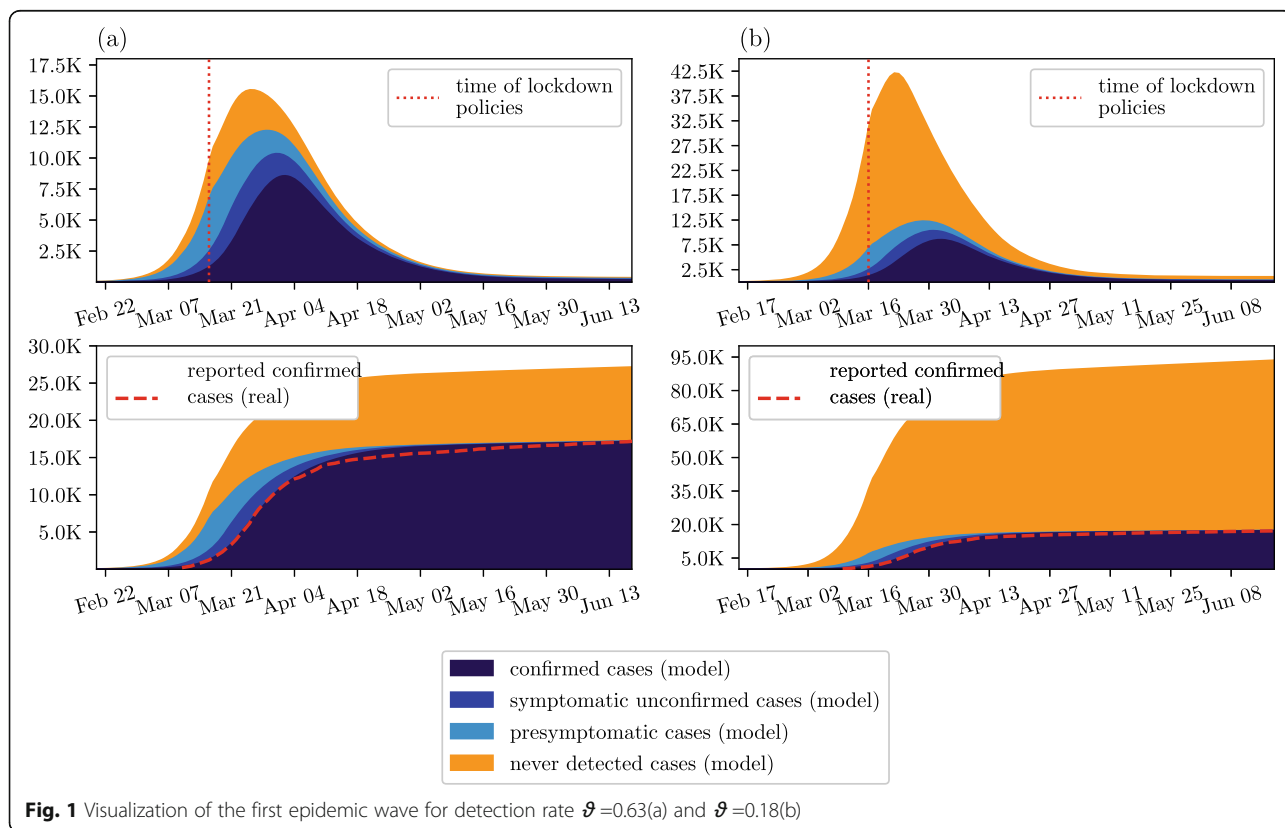
Figure 6 illustrates how the model results can be used to interpret the findings of studies identifying currently infected individuals. It displays the number of active cases of COVID-19 for different dates as a function of the detection rate. Given an estimate of active cases for a specific date, the model results for this date as a function of the detection rate can be interpolated by linear splines and the detection rate matching the estimate of active cases can be determined.

**Discussion**

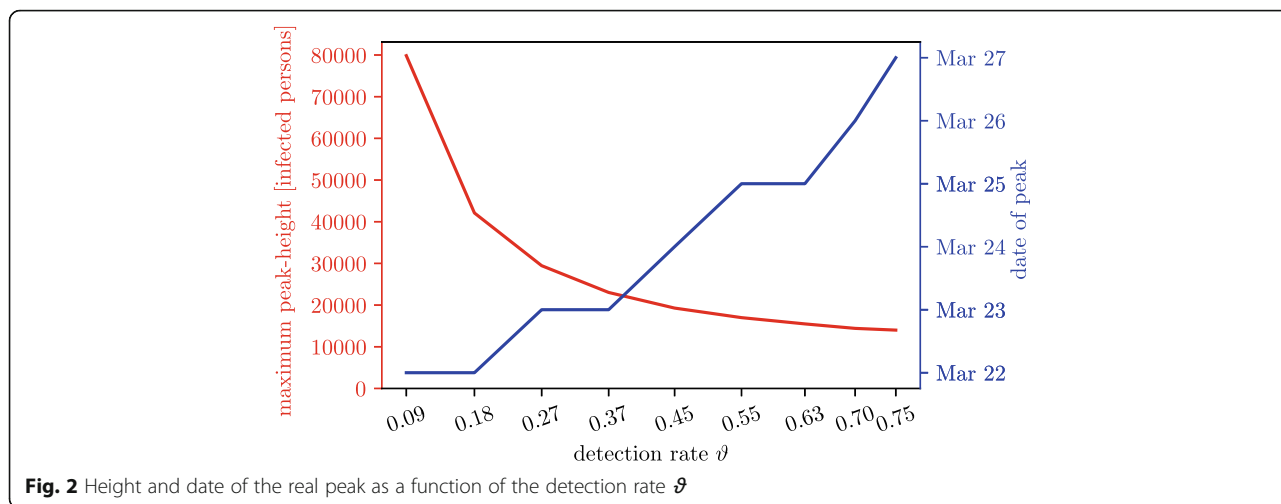
The ratio between detected and undetected cases strongly depends on the phase of the epidemic. Results of randomized screenings to evaluate the detection rate therefore need to be interpreted within the context of the phase of the epidemic. This holds for both studies using antibody tests and PCR tests.

Screening programs using antibody tests are less sensitive but still depend on timing within the epidemic. The different severities between detected and undetected cases lead to different disease lengths. Therefore, the confirmed cases are underrepresented within the total cases as long as the disease wave is still on the upswing (see Fig. 5). Consequently, results of antibody screenings would underestimate the detection rate, if performed too early before the completion of the wave. Similarly, considering the indication of waning antibody, screenings performed too long after an epidemic wave could underestimate the total number of COVID-19 infections, and therefore, overestimate the detection rate.

As for studies using PCR tests to determine the currently infected individuals, the simulation results show that the actual detection rate cannot be determined without additional information about the disease wave. An observed ratio of 60:40 for detected vs. undetected



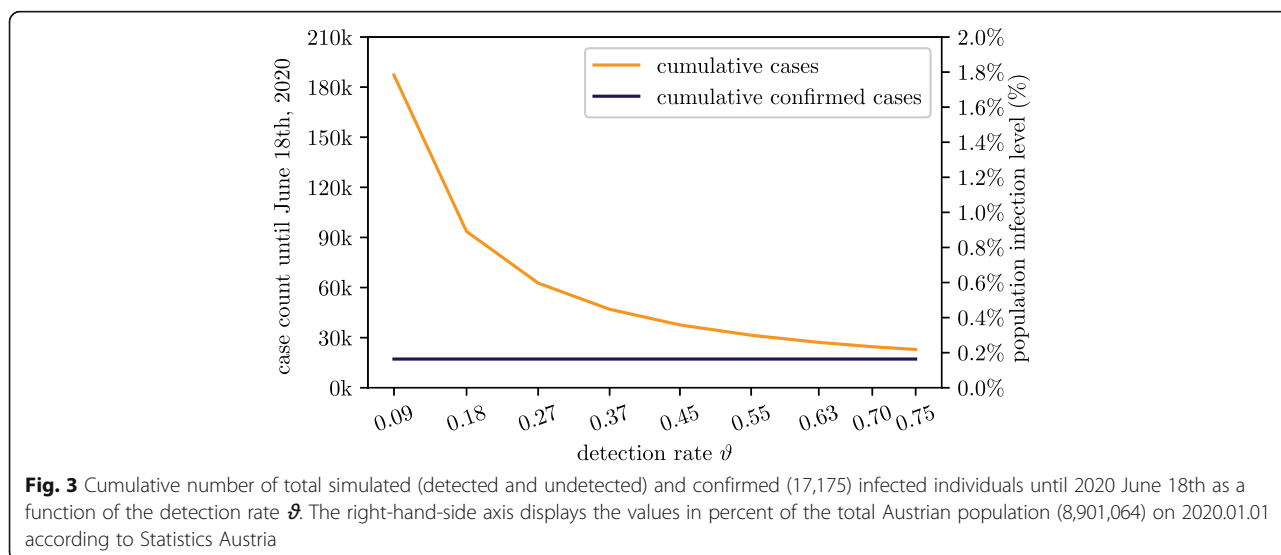
**Fig. 1** Visualization of the first epidemic wave for detection rate  $\vartheta = 0.63$ (a) and  $\vartheta = 0.18$ (b)

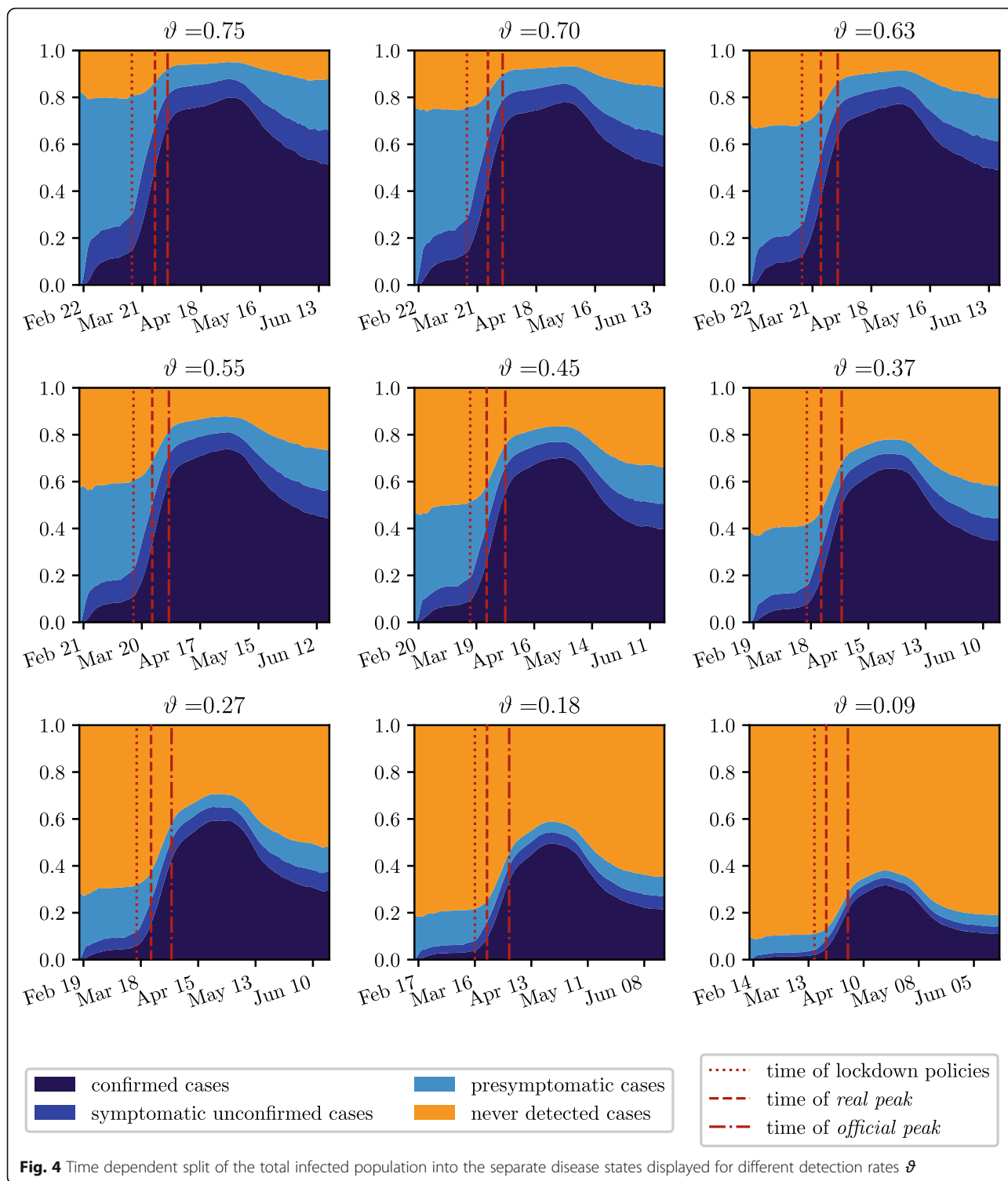


cases in a randomized trial might indicate  $\vartheta = 0.37$  if the study was performed a couple of days after the official peak, or  $\vartheta = 0.75$  if it was performed a couple of days before. Therefore, the disease progression should always be considered retrospectively when evaluating and interpreting a study performed to determine the detection rate.

In Austria, three different nationwide randomized prevalence screening programs have been performed to get an image of current prevalence [6]. The first program was organized by SORA Institute for Social Research and Consulting with tests between April 1st and 6th (sample size  $n = 1544$ , 6 positive tests) [4], the other two programs were executed by the Austrian Statistics Institute between April 22nd to April 25th (sample size  $n = 1432$ , 1 positive test result) and May 26th to May 30th (sample size  $n = 1279$ , no positive test results). As

indicated by Fig. 4, the first and second study were executed at a time with the lowest possible fraction of undetected cases possible: shortly after the *official peak* of the epidemic wave. Consequently, a high sample size would be required to detect a significant number of undetected cases and even then the resulting ratio needs to be treated with care: For example, an actual detection rate of  $\vartheta = 0.18$  results in a prevalence of almost 50% detected cases on April 10th. With respect to the ratio between undetected and detected, the timing of the third study was much better suited. Unfortunately, the total number of infected individuals at the end of May was extremely low which, again, required a large sample size. As a result of the unfortunate timing and the limited number of participants, the results of all three studies in Austria have not been particularly useful in providing conclusions about the detection rate parameter. Yet, the

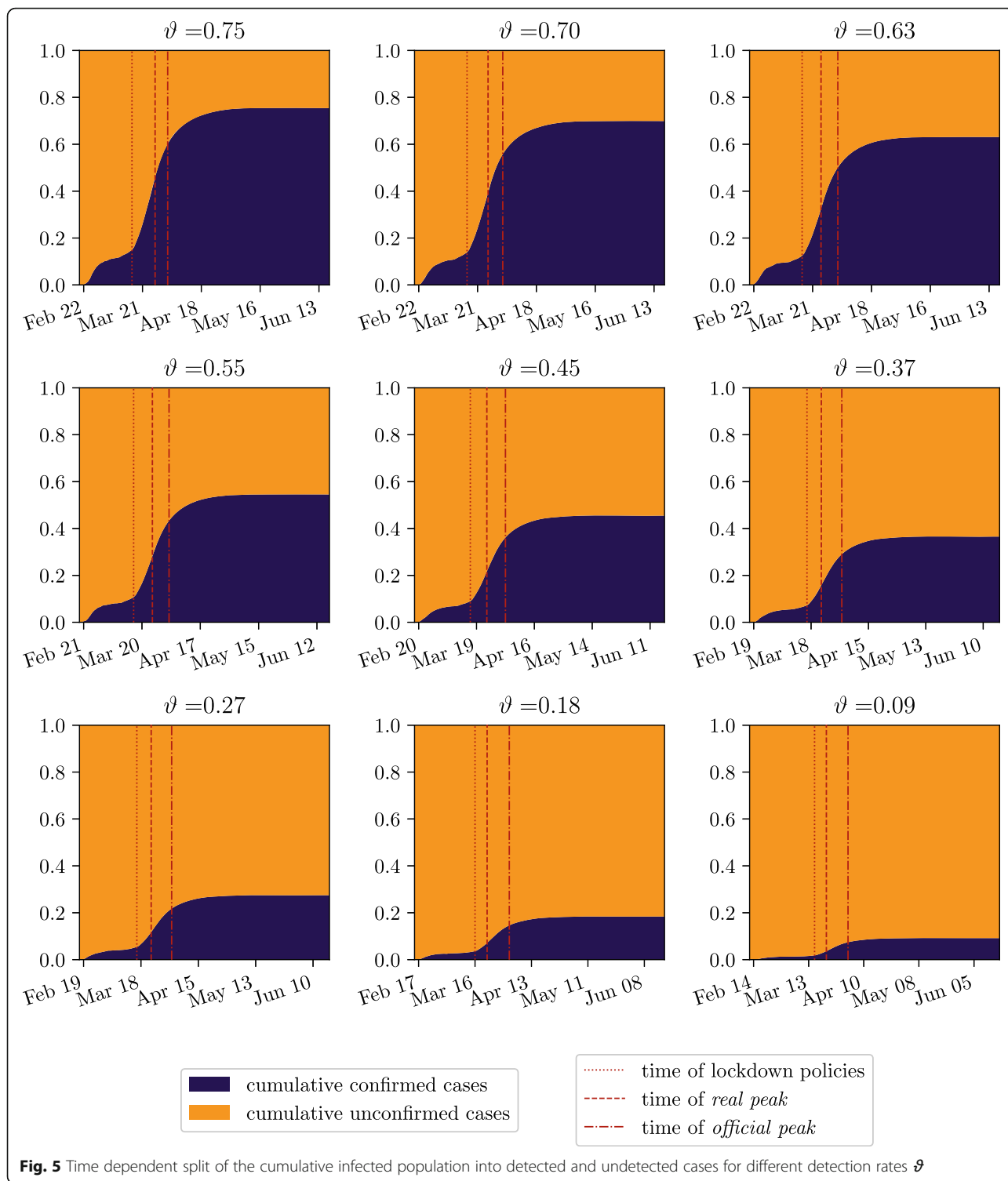




combination of the connected surveys brought very interesting insights into the actual wellbeing of the population.

Nevertheless, taking the dynamic of the epidemic wave into consideration, the first PCR study in Austria

performed by SORA in April (sample size  $n = 1544$ ) [4] could be used to determine feasible values for the detection probability. The study estimates a prevalence of 28,500 infected persons in Austria for April 4th. As demonstrated in Fig. 6, this leads to  $\vartheta = 0.13$ . Note that this

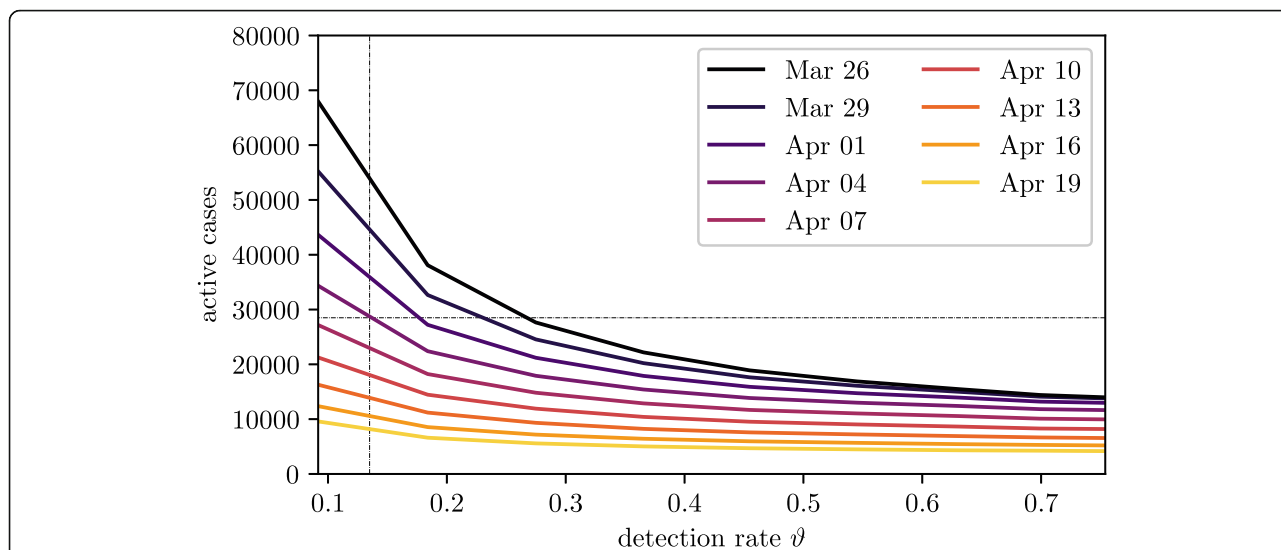


**Fig. 5** Time dependent split of the cumulative infected population into detected and undetected cases for different detection rates  $\vartheta$

number is considerably lower than the prevalent ratio between detected and undetected  $8358/28,500 \approx 0.29$  on this day [14]. Consequently, a large error is made when estimating the detection rate based on this naïve formula.

The graphs in Fig. 4 in combination with Fig. 1 indicate that there are points in time much better suited for randomized prevalence studies, namely the time of the *real peak* which occurred about 5 to 10 days after the lockdown measures and 5 to 10 days before the *official*





**Fig. 6** Number of active cases for different dates as a function of the detection rate. The dashed line represents the 28,500 cases on April 4th as estimated by SORA Institute for Social Research and Consulting, leading to  $\vartheta = 0.13$

peak. At that time, the overall number of infected individuals is high, thus the required sample size is reduced. But the detection rate is still comparably low which increases the chances of including undetected individuals in the screening sample. Unfortunately, this is a rather theoretical result, as the time of the *real peak* can only be evaluated retrospectively. Nevertheless, it indicates that the optimal time for prevalence studies lies during the upswing of the epidemic wave, and that they should not be performed too late.

Moreover, the data preprocessing required for parametrization provides new insights into the level of detection probabilities. As displayed in Table 2, probabilities higher than 0.4 cannot be age-stratified using the proposed algorithm indicating this value as an upper bound for the detection probability in Austria between March and June 2020. These insights concur with several publications [5, 16, 17] who all suggest a comparably low detection probability far smaller than 50%, but contradicts the results of several screening programs performed earlier in China [3] and Iceland [2] which all resulted in higher detection rates. Yet, it is not surprising that detection probabilities differ between different countries. Although evaluated for Austria, the qualitative results of our study can be applied to any country or region worldwide if we assume that the testing strategy regarding the target groups, the sensitivity and specificity of the tests, and the testing and reporting delays are similar.

The calculation of the age-dependent detection probability also has an influence on the detection rate  $\vartheta$  as a model output. One would expect that the detection rate  $\vartheta$  should be identical to the detection probability  $\theta$ . Instead,  $\vartheta < \theta$  can be observed, with the difference

increasing with larger values for  $\theta$  (see Table 3). This effect occurs because the simulation model is not able to reproduce the large number of infected elderly people observed in reality. For large values of  $\theta$ , this is mainly caused by capping the age-dependent detection probability at 1. But it is also a result of the implemented contact network for retired people. They often live in smaller households and do not have any workplace contacts, making them less likely to be in contact with an infected individual.

Still, the contact network is one of the key features of our simulation model. It is based on the POLYMOD study [8], a large survey on social contact behavior relevant to the spread of infectious diseases. In addition, the location-based contact network allows us to implement specific lockdown policies such as school closures directly without the need of “guessing” parameter values for their implications. Even if this model feature is not directly needed for evaluating the number of undetected cases of COVID-19, it allows the simulation of the epidemic wave as realistically as possible. A feature much more important for the analysis presented in this study is that the agent-based model includes undetected cases of COVID-19 in the spread of the disease, and a snapshot can be taken of all the infected individuals for every timestep, separating them into their respective disease states.

Our study has several major limitations. First, the calculation of the age-dependent detection probability relies on the simplifying assumption that people of all age groups are equally susceptible to getting infected and that the mismatch between the age distribution of the confirmed cases and the overall Austrian age

distribution is solely due to an age-dependent probability to develop symptoms and subsequently getting tested for the disease. This subject is strongly linked to the debate on whether schools play a major role in the spread of COVID-19. Since the scientific community could not yet reach a consensus on the latter [18, 19], we did not assume that children are less susceptible to becoming infected.

Second, we kept the age-dependent detection probability constant over time. Thus, we did not account for screening programs performed in care homes or the time during which schools were closed and children could only be infected by their immediate family. Moreover, we did not model the sensitivity and the specificity of the current PCR test for SARS-CoV-2 or the increased usage of antigen tests whose performances vary greatly regarding their sensitivity and specificity. Our model does not provide the possibility for an uninfected individual to receive a false positive test result, and an infected individual who gets tested for the disease will always receive a positive test result in our analysis. This way, in our model, the sensitivity of the test is implicitly incorporated into the detection probability  $\theta$  with a lower test sensitivity producing a lower detection probability.

Furthermore, as for all current simulation models for COVID-19, the parameters used in our model are associated with uncertainty. In particular, the recovery time for undetected cases – that is, the time until these mostly asymptomatic and mildly symptomatic individuals are no longer infectious – is highly uncertain. This is certainly an influential parameter in the model but real data and evidence for its values are diverse and hard to find. We also have not modelled a varying infectiousness depending on the severity of the disease or the time elapsed since the start of the infectious period. Instead, all agents are assumed to be equally infectious throughout the infectious period. Current studies suggest that this is likely not the case, and therefore our assumption constitutes a major simplification. Although our simulation model allows us to include infectiousness varying across individuals, we decided against the implementation of infectiousness heterogeneity, as respective data are lacking. On the same note, our model does not specifically consider super-spreader events caused by an overly infectious person or an event where one infected person has contact with an extraordinarily high number of people. However, the detection probability is independent of the origin of the infection (caused by a super-spreader event or not) and the ratio of detected and undetected cases of COVID-19 remains unaffected. Therefore, not specifically considering super-spreader events from the simulation does not bias the results of our study.

Lastly, our simulation model does not give an estimate for the fatal cases. The case fatality ratio (CFR) of COVID-19 strongly depends on the need and the availability of medical resources, namely the ICU-beds and medical personnel. Since this number of total ICU-beds and human resources available for COVID-19 patients in Austria is not fixed but can be slightly adapted to the current need, our model does not consider a threshold for the available medical resources and a subsequential change in the CFR as soon as the maximal capacities are exceeded. Therefore, an estimation of the fatal cases is associated with great uncertainty, and even more importantly, provides no added value to the model results of the current research question of our study.

In conclusion, we presented the outcomes of a retrospective modeling study in which we compared different assumptions about the probability of detecting an individual infected by SARS-CoV-2, leading to different “dark figures”, which allowed us to establish possible value ranges for this relevant disease parameter.

Moreover, our simulation outcomes display that the ratio between confirmed and unconfirmed cases for the Austrian COVID-19 epidemic heavily changes with time. Consequently, the results of randomized screenings are usually interpreted too naively. This study gives guidance on when to setup and how to interpret screening results in a more accurate way.

#### Abbreviations

SARS-CoV-2: Severe acute respiratory syndrome coronavirus-2; COVID-19: Coronavirus disease 2019; PCR: Reverse transcription polymerase chain reaction test; ICU: Intensive care unit; ISPOR-SMDM: International Society for Pharmacoeconomics and Outcomes Research – Society for Medical Decision Making; SORA: Institute for Social Research and Consulting; CFR: Case fatality ratio

#### Acknowledgements

Not applicable.

#### Authors' contributions

CR participated in the planning of the study, made substantial contributions to designing and programming the model, carried out the result analysis and drafted the manuscript. MB, CU, and DB participated in the planning of the study, made substantial contributions to designing and programming the model, acquisition of information, analysis of results. NW and GZ made substantial contributions to the acquisition of information and analysis and interpretation of results. GS, BJ, NM, and US made substantial contributions to the acquisition of information and interpretation of results and revised the manuscript critically for important intellectual content. NP conceived the study and made substantial contributions to the study design and interpretation of results. All authors have read and approved the final manuscript.

#### Funding

This study has been partially funded by The Austrian Research Promotion Agency (FFG) Covid19 Emergency Call, the WWTF-COVID-19 Rapid Response Funding, the Medizinisch-Wissenschaftlicher Fonds des Bürgermeisters der Bundeshauptstadt Wien, and the Society for Medical Decision Making (SMDM) Covid-19 Decision Modeling Initiative. The funding bodies did not have any influence on the design of the study, the collection, analysis, and interpretation of data, the decision to publish, and preparation, review or approval of the manuscript.

**Availability of data and materials**

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Author details**

<sup>1</sup>DWH Simulation Services, DEXHELPP, Neustiftgasse 57-59, 1070 Vienna, Austria. <sup>2</sup>TU Wien, Institute of Information Systems Engineering, Favoritenstraße 11, 1040 Vienna, Austria. <sup>3</sup>Institute of Public Health, Medical Decision Making and Health Technology Assessment, Department of Public Health, Health Services Research and Health Technology Assessment, UMIT - University for Health Sciences, Medical Informatics and Technology, Eduard-Wallnoefer-Zentrum 1, 6060 Hall i.T., Austria. <sup>4</sup>Division of Health Technology Assessment, ONCOTYROL - Center for Personalized Cancer Medicine, Karl-Kapferer-Straße 5, 6020 Innsbruck, Austria. <sup>5</sup>Institute for Technology Assessment and Department of Radiology, Massachusetts General Hospital, Harvard Medical School, 101 Merrimac St, Boston, MA 02114, USA. <sup>6</sup>Center for Health Decision Science, Departments of Epidemiology and Health Policy & Management, Harvard T.H. Chan School of Public Health, 718 Huntington Avenue, Boston, MA 02115, USA.

Received: 5 October 2020 Accepted: 26 December 2020

Published online: 13 January 2021

**References**

- Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science*. 2020;368(6490):489–93.
- Gudbjartsson DF, Helgason A, Jonsson H, Magnusson OT, Melsted P, Norddahl GL, et al. Spread of SARS-CoV-2 in the Icelandic Population. *N Engl J Med*. 2020;382(24):2302–15.
- Mizumoto K, Kagaya K, Zarebski A, Chowell G. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Eurosurveillance*. 2020;25(10):2000180.
- SORA Institut: COVID-19 Prävalenz. [cited 2020 Aug 13]. Available from: <https://www.sora.at/nc/news-presse/news/news-einzelansicht/news/coronavirus-dunkelziffer-1006.html>.
- Streck H, Schulte B, Kuemmerer B, Richter E, Hoeller T, Fuhrmann C, et al. Infection fatality rate of SARS-CoV-2 infection in a German community with a super-spreading event. *medRxiv*. 2020; 2020.05.04.20090076.
- BMBWF. Corona-Studien. [cited 2020 Aug 13]. Available from: <https://www.bmbwf.gv.at/Themen/Forschung/Aktuelles/Corona-Studien.html>.
- Bicher MR, Rippinger C, Urach C, Brunmeir D, Siebert U, Popper N. Agent-Based Simulation for Evaluation of Contact-Tracing Policies Against the Spread of SARS-CoV-2. *medRxiv*. 2020; 2020.05.12.20098970.
- Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. POLYMOD social contact data. 2017;.
- Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Ann Intern Med*. 2020;172(9):577–82.
- SARS-CoV-2 Steckbrief zur Coronavirus-Krankheit-2019 (COVID-19). Available from: [https://www.rki.de/DE/Content/InfAZ/N/Neuartiges\\_Coronavirus/Steckbrief.html#doc13776792bodyText7](https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Steckbrief.html#doc13776792bodyText7).
- Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Glob Health*. 2020;8(4):e488–96.
- Austrian Ministry of Health. Information about the Austrian COVID-19 reporting system. Erläuterungen zum Amtlichen Dashboard COVID19. [cited 2020 Aug 27]. Available from: <https://info.gesundheitsministerium.at/Basisinformationen.pdf>.
- Ludvigsson JF. Systematic review of COVID-19 in children shows milder cases and a better prognosis than adults. *Acta Paediatr*. 2020;109(6):1088–95.
- Amtliches Dashboard COVID19. [cited 2020 Aug 13]. Available from: <https://info.gesundheitsministerium.at/>.
- Caro JJ, Briggs AH, Siebert U, Kuntz KM. Modeling Good Research Practices—Overview: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1. *Value Health*. 2012;15(6):796–803.
- Hao X, Cheng S, Wu D, Wu T, Lin X, Wang C. Reconstruction of the full transmission dynamics of COVID-19 in Wuhan. *Nature*. 2020; [cited 2020 Jul 28]; Available from: <http://www.nature.com/articles/s41586-020-2554-8>.
- Medical University of Innsbruck. Ischgl-Studie: 42,4 Prozent sind Antikörper-positiv - myPoint. [cited 2020 Jul 6]. Available from: <https://www.i-med.ac.at/mypoint/news/746359.html>.
- Jones TC, Mühlemann B, Veith T, Biele G, Zuchowski M, Hoffmann J, et al. An analysis of SARS-CoV-2 viral load by patient age. *Infectious Diseases (except HIV/AIDS)*; 2020 [cited 2020 Aug 13]. Available from: <http://medrxiv.org/lookup/doi/10.1101/2020.06.08.20125484>.
- Viner RM, Mytton OT, Bonell C, Melendez-Torres GJ, Ward JL, Hudson L, et al. Susceptibility to and transmission of COVID-19 amongst children and adolescents compared with adults: a systematic review and meta-analysis. *Public and Global Health*; 2020 [cited 2020 Aug 13]. Available from: <http://medrxiv.org/lookup/doi/10.1101/2020.05.20.20108126>.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

