medRxiv preprint doi: https://doi.org/10.1101/2020.03.21.20040444; this version posted April 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

From a single host to global spread. The global mobility based modelling of the COVID-19 pandemic implies higher infection and lower detection rates than current estimates.

Marian P. Siwiak Ph.D. 1, Pawel Szczesny Ph.D. 1, Marlena M. Siwiak Ph.D. 1
1 Data 3.0, Sevenoaks, TN14 7TG, UK
Corresponding Author:
Marian P. Siwiak Ph.D. 1
3 Forge Way, Sevenoaks, TN14 7TG, United Kingdom
Email address: siwiak@data30.co.uk

Abstract

Background.

Since the outbreak of the COVID-19 pandemic, multiple efforts of modelling of the geo-temporal transmissibility of the virus have been undertaken, but none succeeded in describing the pandemic at the global level. We propose a set of parameters for the first COVID-19 Global Epidemic and Mobility Model (GLEaM). The simulation starting with just a single pre-symptomatic, yet infectious, case in Wuhan, China, results in an accurate prediction of the number of diagnosed cases after 135 days in multiple countries across four continents.

Methods.

We have built a modified SIR metapopulation transmission model and parameterized it analytically according to the literature, and by fitting the missing parameters to the observed dynamics of the virus spread. We compared our results with the number of diagnosed cases in sixteen selected countries which provide reliable statistics but differ substantially in terms of strength and speed of undertaken precautions. The obtained 95% confidence intervals for the predictions fit well to the empirical data.

Results.

The parameters that successfully model the pandemic are: the basic reproduction number R_0, ~4.4; a latent non-infectious period of 1.1. days followed by 4.6 days of the presymptomatic infectious period; the probability of developing severe symptoms, 0.01; the probability of being diagnosed when presenting severe symptoms of 0.6; the probability of diagnosis for cases with mild symptoms or asymptomatic, 0.001.

Discussion.

Parameters that successfully reproduce the observed number of cases indicate that both R_0 and the prevalence of the virus might be underestimated. This is in concordance with the newest research on undocumented COVID-19 cases. Consequently, the actual mortality rate is putatively lower than estimated. Confirmation of the pandemic characteristic by further

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.21.20040444; this version posted April 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

refinement of the model and screening tests is crucial for developing an effective strategy for the global epidemiological crisis.

Introduction

A novel coronavirus SARS-CoV-2 has already spread into 186 countries and territories around the world (as of March 21st, 2020). With over half a million confirmed infections and over 24 thousand deaths (as of March 26th, 2020), it became a global challenge. COVID-19, the disease caused by this coronavirus, was characterised as a pandemic by WHO on 11th of March 2020.

While a number of different measures to contain the virus have been implemented by countries all over the world, their effectiveness remains to be seen. The models used to inform decision-makers are differing significantly in their basic assumptions because it is the first coronavirus of such an impact in terms of the number of fatal cases. Also the existing modelling approaches often use biased data for tuning parameters or assessing models quality. Until an effective treatment is available, the accuracy of these models and the decisions made on their basis are the major factors in reducing the overall mortality in the COVID-19 pandemic.

Multiple efforts of calculating the transmissibility of the SRAS-Cov-2 virus and its geo-temporal modelling have been undertaken, but none of the models succeeded to describe the pandemic at the global level. For those models the estimates of the basic reproduction number of the virus were typically obtained using only Chinese data on the number of diagnosed cases. Additionally the actual prevalence of the virus remains unknown, as many infections are mild, asymptomatic or with atypical symptoms. In fact, many COVID-19 cases pass unnoticed (in China, over 50% according to the research). This hampers successful modelling of the pandemic.

This study presents the first global modelling of COVID-19 pandemic that builds on top of successful modelling framework GLEAM. The basic reproduction number for SARS-CoV-2 used in the simulation is 4.4. It is higher than the value proposed by WHO, but best-fits the observed number of diagnosed cases over 135 days in multiple countries around the globe. Our analysis also provides the estimation of the global rate of total diagnosed to undiagnosed cases of 0.0061. The set of parameters used in our simulation forms a solid foundation for further modelling of the pandemic.

In this study, we present putatively the first global model of SARS-CoV-2 spread that within confidence intervals accurately depicts the current state of diagnosed cases of COVID-19 for multiple countries at once. Implications on the transmissibility and policymaking are also discussed.

Materials & Methods

Modelling software

The model is based on The Global Epidemic and Mobility Model (GLEaM) framework(Balcan et al., 2010), implemented in the GLEAMviz software(Van den Broeck et al., 2011). The GLEaM model integrates sociodemographic and population mobility data in a spatially structured stochastic disease approach to simulate the spread of epidemics at a worldwide scale. It was previously used for a real-time numerical forecast of the global

medRxiv preprint doi: https://doi.org/10.1101/2020.03.21.20040444; this version posted April 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

spreading of A/H1N1 (Tizzoni et al., 2012), and the accuracy of that modelling was later confirmed(Tizzoni et al., 2012).

Data sources

The reference data about the number of SARS-CoV-2 diagnosed patients in the period from Jan 22, 2020, to Mar 26, 2020, was downloaded from the Johns Hopkins University of Medicine Coronavirus Resource Center GitHub repository https://github.com/CSSEGISandData/COVID-19.
Information about the severity of developed symptoms was derived from the worldometer.info website https://www.worldometers.info/coronavirus/.
Information on testing efforts in selected countries comes from https://ourworldindata.org/coronavirus-testing-source-data website.
Other data sources, such as subpopulation selection, commuting patterns, or air travel flows, used during simulation are embedded in the GLEAM software and well described by its developers.

Model parametrization

Below and in (**Table 1**) we present two subsets of model parameters: 1) reliable and evidence-based derived from literature, and 2) knowledge and analysis-based estimations.
The average latency period (***lp***) of 5.6 days is a consensus of different estimations calculated previously (Lauer et al., 2020).
Due to 1) long ***lp***, effectively much longer than reported for other coronaviruses, and 2) known cases of presymptomatic transmission(Woelfel et al., 2020; Tong et al., 2020), for the modelling purposes we decided to split the latency period into two parts: 1) average latent non-infectious period (***lnip***) of *1.1* days (based on the time of infectivity for other viruses(Wallinga & Teunis, 2004)), and 2) average presymptomatic infectious period (***pip***) of *4.5* days. This split produces two parameters used in the model:
1) latency rate for the non-infectious period - non-infectious epsilon (***niε***):

$$ni\varepsilon = 1/lnip\,,$$

and
2) latency rate for the infectious period - latency rate infectious epsilon (***iε***):

$$i\varepsilon = 1/(lp - lnip)\,.$$

As the Republic of Korea provides high quality, reliable data and conducted a large number of tests during the pandemic, we decided to use Korean proportion of severe to diagnosed cases as a base for the probability of developing the severe condition (***pS***) and we set it to *0.01*. We assumed that patients with mild symptoms, in contrast to those in severe condition, are still capable of travelling. For model simplicity, we decided to merge into one compartment all mild and asymptomatic cases.
We decided to set the probability of detection of a severe infection (***pDS***) to *0.6*, in order to accurately mimic two obstacles typically preventing proper diagnosis. Firstly, the majority of patients with a severe course of the disease are either chronically ill or above 60(Zhou et al., 2020)- their symptoms might be mistaken with those caused by their general health condition, and thus not reported on time. Secondly, the model is supposed to reflect the average illness detection around the globe which includes many countries with low quality or underfinanced healthcare where the number of SARS-CoV-2 tests is limited.
Another parameter of the model, ***pDM*** is the probability of being diagnosed with COVID-19 when expressing either mild symptoms or an asymptotic illness course. This parameter

medRxiv preprint doi: https://doi.org/10.1101/2020.03.21.20040444; this version posted April 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

depends on previously defined ***pS*** and ***pDS***, as well as the rate of total diagnosed to undiagnosed cases (***tDR***):

$$pDM = (tDR - pS * pDS) \div (1 - pS).$$

Knowing the limitations of previous modelling attempts(Cowling et al.; Ganyani et al., 2020; Zhang et al., 2020; Chen et al., 2020; Wu, Leung & Leung, 2020; Lin et al., 2020; Kucharski et al., 2020), we decided to test a radically different COVID-19 epidemiologic paradigm, i.e. a significantly lower ***tDR***. This means that in our model we assume a higher proportion of undetected cases in comparison to other models proposed so far. Taking into account that none of them was capable of providing a plausible global simulation of the pandemic, plus the fact that the potential low detectability has already been discussed in the literature(Li et al., 2020), we decided to test such a possibility in simulation by setting the lowest possible ***tDR***. Its relation to ***pDM*** sets its minimum to:

$$tDR > pS * pDS.$$

For previously set ***pS*** and ***pDS*** values, ***tDR*** must be greater than 0.006, thus the value used in our simulation was set to 0.0061.

Another important and deeply interconnected parameters required by the model are as follows: the effective contact rate, $\boldsymbol{\beta}$; its reduction level for patients who developed severe symptoms of the disease but were not diagnosed, $\boldsymbol{r\beta}$; and average recovery time since symptoms development $\boldsymbol{\mu}$.

The parameter $\boldsymbol{\beta}$ is derived from the time a host remains infectious, ***d***, and the basic reproduction number of the virus, $\boldsymbol{R_0}$:

$$\beta = R_0 \div d,$$

where:

$$d = \mu + pip.$$

The estimation of R_0 is a topic widely discussed in the literature, with values ranging from 1.4 to 6.49("Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV)"; Majumder & Mandl; Zhao et al.; Imai et al., 2020; Read et al., 2020; Liu et al., 2020). However, following the assumption of much higher than the currently suspected rate of undiagnosed and mild/asymptomatic cases, we decided to use in our model a higher rate of transmissibility, yet well within the range of 2-5, modelled for SARS(Wallinga & Teunis, 2004). The assumed R_0 value leading to presented results is 4.4.

In our study $\boldsymbol{\mu}$ is derived from a safe quarantine period for diagnosed cases(Woelfel et al., 2020). As the safe quarantine time is estimated to be 10 days(Woelfel et al., 2020), we assumed $\boldsymbol{\mu}$ to last on average for 7 days from symptoms development to recovery. The sum of $\boldsymbol{\mu}$ and previously estimated ***pip*** (presymptomatic infectious period) results in ***d*** equal to *11.5 days*, and $\boldsymbol{\beta}$ equal to *0.38261*.

We decided to set $\boldsymbol{r\beta}$ to 0.5, following the assumption for this parameter used in GLEaM modelling of the 2009 influenza outbreak(Balcan et al., 2010). Patients who were diagnosed with COVID-19 are assumed isolated and as such not spreading the disease any further.

Model compartmentalization

To model the virus spread, we modified the compartmental SIR metapopulation transmission model to represent the nature of the COVID-19 epidemic.

In our model, we use seven different population compartments (**Figure 1**).

1. Susceptible population - equal to the general global population. We assume no existing immunity to infection.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.21.20040444; this version posted April 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

2. Latent non-infectious - infected population in the first incubation stage, not yet infectious.
3. Presymptomatic infectious - infected population already infectious, but without developed symptoms.
4. Mild symptoms - joint populations of asymptomatic cases and those with inconspicuous symptoms.
5. Severe symptoms - population infected by SARS-CoV-2 with symptoms affecting their travel ability.
6. Diagnosed - population identified as infected with the SARS-CoV-2 virus. This is the reference line for the model accuracy.
7. Recovered - joint populations of recovered and fatal cases.

The prepared model served as an input for 10 runs (a maximum available in free tier) of GLEaM Monte Carlo analysis based on human mobility, integrating population and two (local and air) mobility layers.

Results

The simulation was started on Nov 12, 2019, with a single presymptomatic individual located in Wuhan, China, and the development of the pandemic spread was modelled for 135 days. The model did not include any information on already implemented movement restrictions and preventive measures undertaken by different governments. As overall data on the pandemic dynamics around the globe is likely to be biased by regions, often considerable in size and population, for which official statistics might be inaccurate, we decided not to compare overall model results with global data. Instead we limited the analysis of results to thirteen countries across four continents (see **Table 2**) which are, in our belief: a) divergent in the proportion of the tested population, quality of healthcare, and strength of undertaken preventing measures; b) likely to provide the public with real data; c) reporting number of cases high enough to assume their population exchange with the rest of the world did not significantly change the pandemic dynamics. Two countries which fulfill the above criteria, were excluded from the analysis: Canada, due to its lack of coherency in reporting COVID-19 cases (as reported in https://ourworldindata.org/coronavirus-testing-source-data); and Australia, as its geographical isolation and early overtaken precautions seem to successfully hamper the spread of the disease.

The obtained 95% confidence intervals of predicted numbers of diagnosed patients were compared with empirical data. In **Figure 2** we present a percentage difference over time between the number of reported confirmed cases and confidence intervals limits for modelled predictions. Positive values state that the model overestimates the number of diagnosed cases; negative values indicate the underestimations of the model; for the observed numbers of diagnosed cases that are within the model's CIs the percentage difference is equal to 0. For selected countries the model predictions fit well to the observed data, and the observed discrepancies are explained in captions to **Figures 4 - 16** showing results for individual countries.

There are two main reasons for the discrepancies between model predictions and reported number of COVID-19 cases. The first is the fast governmental response and early introduced precautions, which significantly influence the pace of the disease spread, but are not reflected in the modelling. For such countries (e.g. Japan, the Republic of South Korea), the model overestimates the number of detected cases. The second reason is the increase of the virus detectability in countries where the proportion of tested individuals is larger, leading to

medRxiv preprint doi: https://doi.org/10.1101/2020.03.21.20040444; this version posted April 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

higher ***tDR*** than the one assumed in our model. This is illustrated by the fact that the spread of the model accuracy between the countries is negatively proportional to the estimated number of tests performed per million citizens reported as of March 26, 2020 (see **Figure 3**). Spearman correlation coefficient calculated for the estimated number of performed tests and the average percentage difference between modelled and reported numbers of diagnosed cases is -0.697 (95% CI: [-0.92, -0.12], n=10).

Figures 4 - 16 confront the number of actual confirmed COVID-19 cases with confidence intervals for the modelled number of diagnosed cases. Some countries present epidemic dynamics different from the model, however, the direction of these deviations may be explained by the measures undertaken by their governments, their societal response, or the number of tests carried per million of citizens (discussed in captions to figures). We believe that further modelling efforts, including careful parameters' modifications that reflect local response, would greatly improve the accuracy of the model, but it is outside of the scope of this work.

Data sharing

The model and the results of the simulation underlying the presented results is freely available at https://github.com/freesci/covid19.

Discussion

The presented model has multiple implications concerning the major characteristics of the COVID-19 pandemic, such as the basic reproduction number of the virus R_0 (higher than previously assumed, yet not above the values estimated for other coronaviruses), and the rate of diagnosed cases ***tDR*** (much lower than assumed so far, especially for cases expressing mild symptoms and asymptomatic). This would indicate that the vast majority of the COVID-19 infections are so mild that they pass unnoticed. This is not implausible, considering the fact that there are 1.9 billion children aged below 15 years in the world (27% of the global population) and predominantly (ca. 90%) the course of their infections is mild or asymptomatic(Dong et al., 2020). Additionally, they gather in large groups at schools on a regular basis which facilitates further disease transmission. Also, some COVID-19 cases may show atypical symptoms (e.g. diarrhoea)(Gao, Chen & Fang, 2020) which hinders correct diagnosis. Taking all this into account, plus the results of our model, one may risk a hypothesis that the virus is already more prevalent in the global population than shown in official statistics at the moment, and consequently, its mortality rate is much lower.

To verify this hypothesis further actions are required. At first, the model should be simulated with a larger number of iterations, which will narrow the obtained confidence intervals and allow further refinement of the parameters. Also, a simulation with the ***tDR*** parameter increasing over time or geographically diverse might better reflect the actual virus detectability in the course of the pandemic. Finally, the real spread of the virus should be assessed empirically by conducting a sufficient number of tests on fully random samples (currently most tests are limited to individuals with strong and typical symptoms). Only after obtaining a solid measurement of the actual prevalence of the virus, one might draw conclusions about its true mortality rate.

We emphasize that our conclusions are a hypothesis based on a single computational model and without empirical verification, they may serve as a platform for further research. At this stage, by no means should they be used as a reason for governmental decisions on lifting the precautions. Even if the true mortality of the virus is indeed lower than announced by the

medRxiv preprint doi: https://doi.org/10.1101/2020.03.21.20040444; this version posted April 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

media, many people remain in the high-risk group. Lack of population resistance facilitates their contact with the virus and may lead to a rapid increase of severe cases in a short period of time (as seen in Italy) leading to the collapse of the healthcare system, which affects the entire society and results in many additional deaths not related to the virus itself. Careful use and tuning of non-drug intervention methods, constant balancing of the disease spread and healthcare capacity, protecting the most vulnerable individuals, farsighted anticipation and agility in decision making may altogether be able to minimize the number of deaths without resulting in the global economic breakdown.

Conclusions

Our model implies that the current consensus on the basic reproduction number of SARS-CoV-2 and its prevalence are misestimated. The overall global data on the pandemic dynamics seems strongly biased by large regions where official statistics may not reflect accurately the actual state of the epidemic, and by the fact that many COVID-19 cases may go unnoticed. The basic reproduction rate of the virus should be confirmed on the basis of reliable data, and its prevalence determined by conducting properly designed screening tests. Our model, if confirmed, could be used as a tool for forecasting and optimizing non-drug interventions and policymaking.

References

Balcan D, Gonçalves B, Hu H, Ramasco JJ, Colizza V, Vespignani A. 2010. Modeling the spatial spread of infectious diseases: the GLobal Epidemic and Mobility computational model. *Journal of computational science* 1:132–145.

Chen T-M, Rui J, Wang Q-P, Zhao Z-Y, Cui J-A, Yin L. 2020. A mathematical model for simulating the phase-based transmissibility of a novel coronavirus. *Infectious diseases of poverty* 9:24.

Cowling BJ, Ali ST, Ng TWY, Tsang TK, Li JCM, Fong MW, Liao Q, Kwan MYW, Lee SL, Chiu SS, Wu JT, Wu P, Leung GM. Impact assessment of non-pharmaceutical interventions against COVID-19 and influenza in Hong Kong: an observational study. DOI: 10.1101/2020.03.12.20034660.

Dong Y, Mo X, Hu Y, Qi X, Jiang F, Jiang Z, Tong S. 2020. Epidemiological Characteristics of 2143 Pediatric Patients With 2019 Coronavirus Disease in China. *Pediatrics*. DOI: 10.1542/peds.2020-0702.

Ganyani T, Kremer C, Chen D, Torneri A, Faes C, Wallinga J, Hens N. 2020. Estimating the generation interval for COVID-19 based on symptom onset data. *medRxiv*.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.21.20040444; this version posted April 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Gao QY, Chen YX, Fang JY. 2020. 2019 Novel coronavirus infection and gastrointestinal tract. *Journal of digestive diseases*. DOI: 10.1111/1751-2980.12851.

Imai N, Cori A, Dorigatti I, Baguelin M, Donnelly CA, Riley S, Ferguson NM. 2020. Report 3: transmissibility of 2019-nCoV. *Reference Source*.

Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, Eggo RM, Centre for Mathematical Modelling of Infectious Diseases COVID-19 working group. 2020. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The Lancet infectious diseases*. DOI: 10.1016/S1473-3099(20)30144-4.

Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, Azman AS, Reich NG, Lessler J. 2020. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Annals of internal medicine*. DOI: 10.7326/M20-0504.

Lin Q, Zhao S, Gao D, Lou Y, Yang S, Musa SS, Wang MH, Cai Y, Wang W, Yang L, He D. 2020. A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. *International journal of infectious diseases: IJID: official publication of the International Society for Infectious Diseases* 93:211–216.

Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, Shaman J. 2020. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science*. DOI: 10.1126/science.abb3221.

Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. 2020. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *Journal of travel medicine* 27. DOI: 10.1093/jtm/taaa021.

Majumder M, Mandl KD. Early Transmissibility Assessment of a Novel Coronavirus in Wuhan, China. *SSRN Electronic Journal*. DOI: 10.2139/ssrn.3524675.

Read JM, Bridgen JRE, Cummings DAT, Ho A, Jewell CP. 2020. Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. *medRxiv*.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.21.20040444; this version posted April 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). *Available at https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov)* (accessed March 20, 2020).

Tizzoni M, Bajardi P, Poletto C, Ramasco JJ, Balcan D, Gonçalves B, Perra N, Colizza V, Vespignani A. 2012. Real-time numerical forecast of global epidemic spreading: case study of 2009 A/H1N1pdm. *BMC medicine* 10:165.

Tong Z-D, Tang A, Li K-F, Li P, Wang H-L, Yi J-P, Zhang Y-L, Yan J-B. 2020. Potential Presymptomatic Transmission of SARS-CoV-2, Zhejiang Province, China, 2020. *Emerging infectious diseases* 26. DOI: 10.3201/eid2605.200198.

Van den Broeck W, Gioannini C, Gonçalves B, Quaggiotto M, Colizza V, Vespignani A. 2011. The GLEaMviz computational tool, a publicly available software to explore realistic epidemic spreading scenarios at the global scale. *BMC infectious diseases* 11:37.

Wallinga J, Teunis P. 2004. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American journal of epidemiology* 160:509–516.

Woelfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Mueller MA, Niemeyer D, Vollmar P, Rothe C, Hoelscher M, Others. 2020. Clinical presentation and virological assessment of hospitalized cases of coronavirus disease 2019 in a travel-associated transmission cluster. *medRxiv*.

Wu JT, Leung K, Leung GM. 2020. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet* 395:689–697.

Zhang S, Diao M, Yu W, Pei L, Lin Z, Chen D. 2020. Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A data-driven analysis. *International journal of infectious diseases: IJID: official publication of the*

medRxiv preprint doi: https://doi.org/10.1101/2020.03.21.20040444; this version posted April 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

International Society for Infectious Diseases 93:201–204.

Zhao S, Lin Q, Ran J, Musa SS, Yang G, Wang W, Lou Y, Gao D, Yang L, He D, Wang MH. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. DOI: 10.1101/2020.01.23.916395.

Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, Xiang J, Wang Y, Song B, Gu X, Guan L, Wei Y, Li H, Wu X, Xu J, Tu S, Zhang Y, Chen H, Cao B. 2020. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet*. DOI: 10.1016/S0140-6736(20)30566-3.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.21.20040444; this version posted April 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Tables and figures

Parameter	Assumed value	Description
R_0	*4·4*	Reproduction number for SARS-CoV-2
β	*0·38261*	Transmission rate
$r\beta$	*0·5*	Reduction in transmission rate resulting from the undiagnosed development of severe COVID-19 symptoms
lp	*5·6 days*	Average latency period
lnip	*1·1 days*	Average latent non-infectious period
pip	*4·5 days*	Average presymptomatic infectious period
$ni\varepsilon$	*0·9(09)*	Probability of transition from ***lnip*** to ***pip*** state
$i\varepsilon$	*0·2(2)*	Probability of transition from presymptomatic to symptomatic state
ps	*0·01*	Probability of developing severe COVID-19 symptoms
pDS	*0·6*	Probability of being diagnosed when expressing severe COVID-19 symptoms
tDR	*0·0061*	Rate of diagnosed SARS-CoV-2 infected individuals
pDM	*0·00(01)*	Probability of being diagnosed when presenting mild or none COVID-19 symptoms
μ	*7 days*	Average recovery time since symptoms development

Table 1: Summary of all the parameters used in the deployed model.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.21.20040444; this version posted April 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Country	Continent	Average percentage difference	Tests per million people
Argentina	South America	*0·034*	*NA*
Austria	Europe	*0·007*	*1778*
Bahrain	Asia	*0·010*	*10968*
Brazil	South America	*0·616*	*14*
France	Europe	*0·319*	*559*
Germany	Europe	*0·087*	*2023*
Israel	Asia	*0·001*	*1247*
Italy	Europe	*0·070*	*3499*
Japan	Asia	*15·410*	*276*
Spain	Europe	*0·433*	*646*
The Republic of Korea	Asia	*1·722*	*6148*
United Kingdom	Europe	*1·016*	*960*
United States of America	North America	*1.503*	*314*

Table 2: Calculated per-country average percentage difference of modelled and observed number of diagnosed cases juxtaposed with testing density in population. Data used to generate **Figure 2**.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.21.20040444; this version posted April 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

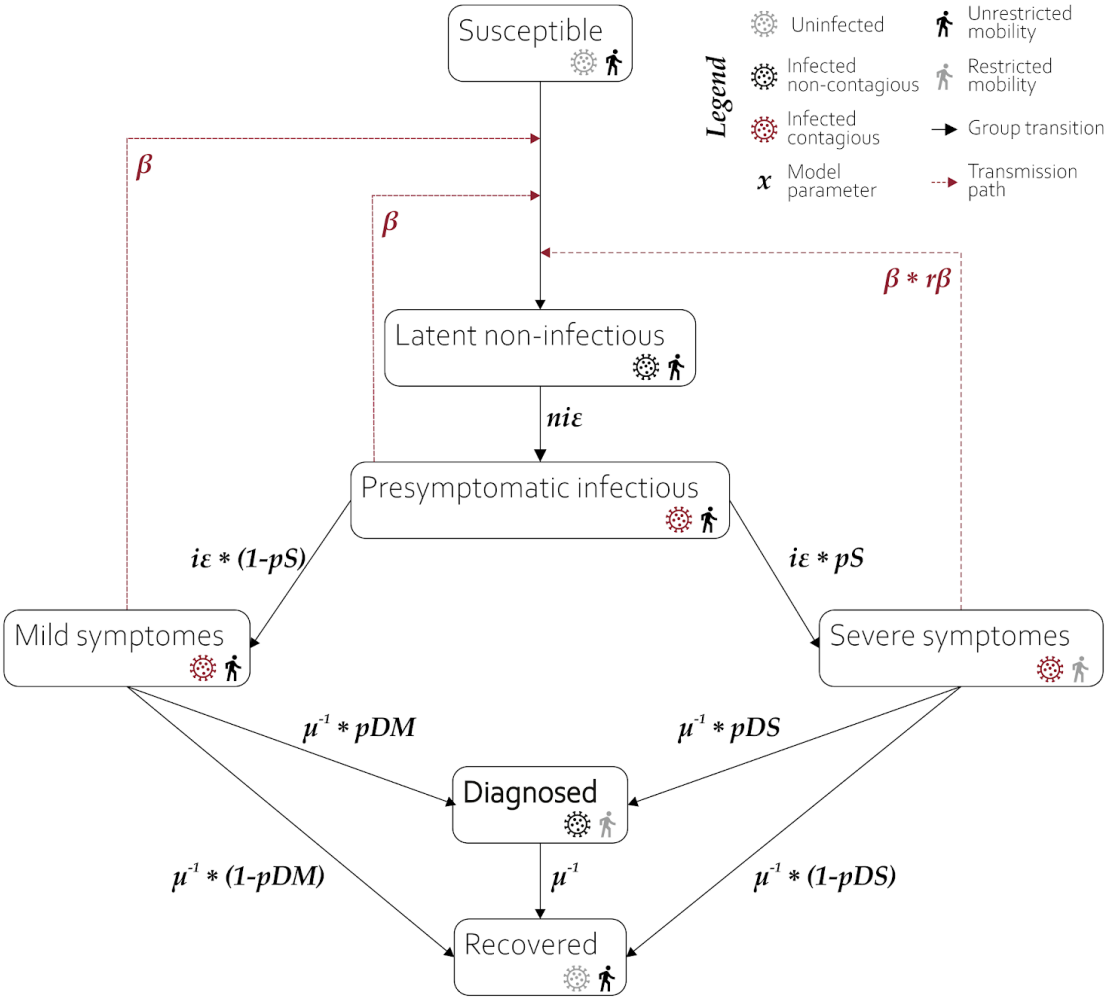

Figure 1: Structure of compartments used in modelling. A susceptible individual in contact with a person: a) presymptomatic, b) who developed mild symptoms, or c) who developed severe symptoms, may contract the infection at rate $\boldsymbol{\beta}$, $\boldsymbol{\beta}$ or $\boldsymbol{r\beta}*\boldsymbol{\beta}$, respectively, and enters the latent non-infectious compartment where he is infected but not yet infectious. During the non-infectious period, each individual has a probability of $\boldsymbol{ni\varepsilon}$ of becoming presymptomatic infectious. The presymptomatic cases have probability $\boldsymbol{i\varepsilon}$ of developing severe or mild COVID-19 symptomes, with probabilities $\boldsymbol{pS}$ and $\mathbf{1}\text{-}\boldsymbol{pS}$ respectively. Transition from symptomatic groups occurs at $\boldsymbol{\mu}$ rate. Individuals who developed severe symptoms do not travel within and between modelled subpopulations and may be either diagnosed with probability $\boldsymbol{pDS}$, or recover with probability of $\mathbf{1}\text{-}\boldsymbol{pDS}$. Individuals whose mild (or non-existent) symptoms are not stopping them from traveling may be diagnosed with probability $\boldsymbol{pDM}$ or recover with probability $\mathbf{1}\text{-}\boldsymbol{pDM}$**.** The diagnosed individuals are considered isolated and effectively non-contagious and recover with rate $\boldsymbol{\mu}$. The recovery does not discriminate between true recovery and fatal cases.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.21.20040444; this version posted April 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

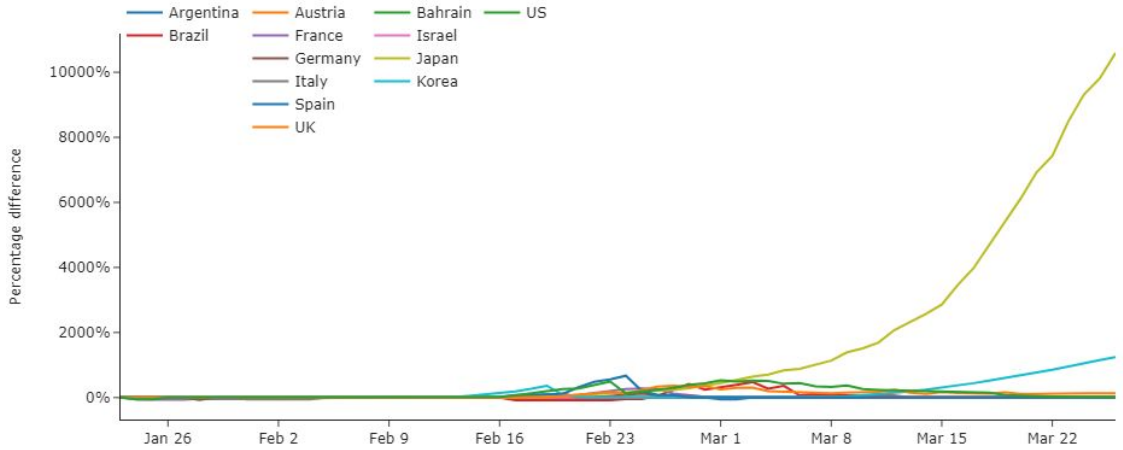

Figure 2: Percentage difference over time between the number of reported confirmed cases and confidence intervals limits for modelled predictions. Positive values state that the model overestimates the number of diagnosed cases, negative values indicate the underestimations of the model. Observed numbers of cases that are within the model CIs are equal to 0. The large discrepancies for Japan and the Republic of South Korea are putatively caused by the fast and pronounced reaction of their governments and early introduced precautions (these are not included in our model). The discrepancies for individual countries are discussed in captions to **Figures 4 - 16**.

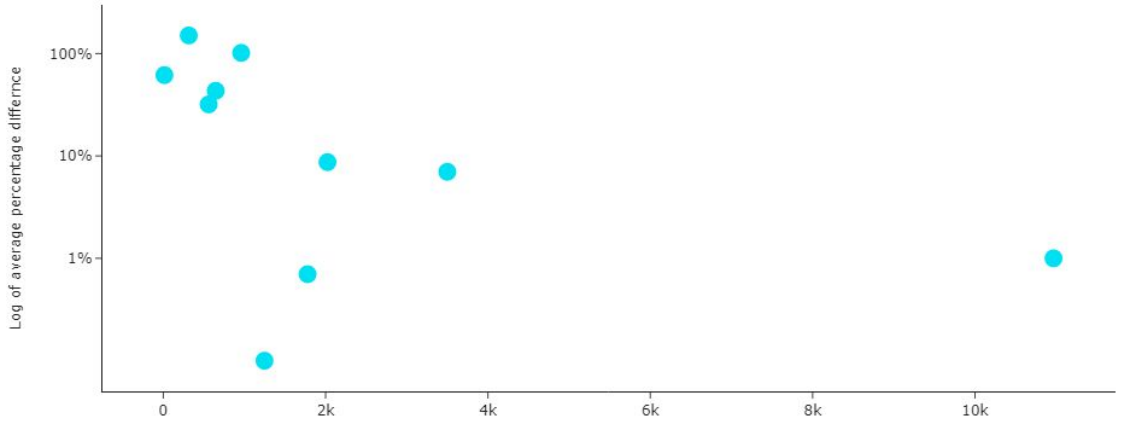

Figure 3: Average percentage difference (logged) between modelled and observed number of detected cases juxtaposed with the number of reported tests per million people in selected countries of known number of tests (see **Table 2** for details). These values are negatively correlated with Spearman correlation coefficient of 95% CI ranging from -0·12 to -0·92 (n=10). Japan and the Republic of Korea were excluded from this analysis, for Argentina the number of tests is not known.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.21.20040444; this version posted April 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

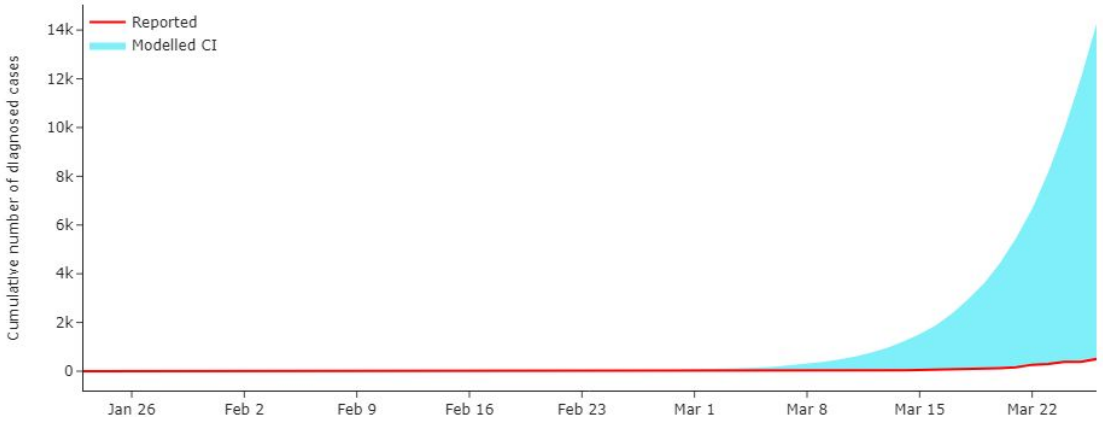

Figure 4: An overlay of modelled confidence interval for the number of diagnosed cases and reported values in Argentina. The observations agree well with model predictions, however, are in its lower bounds, putatively reflecting the low number of tests carried (data for Argentina were not available in our sources).

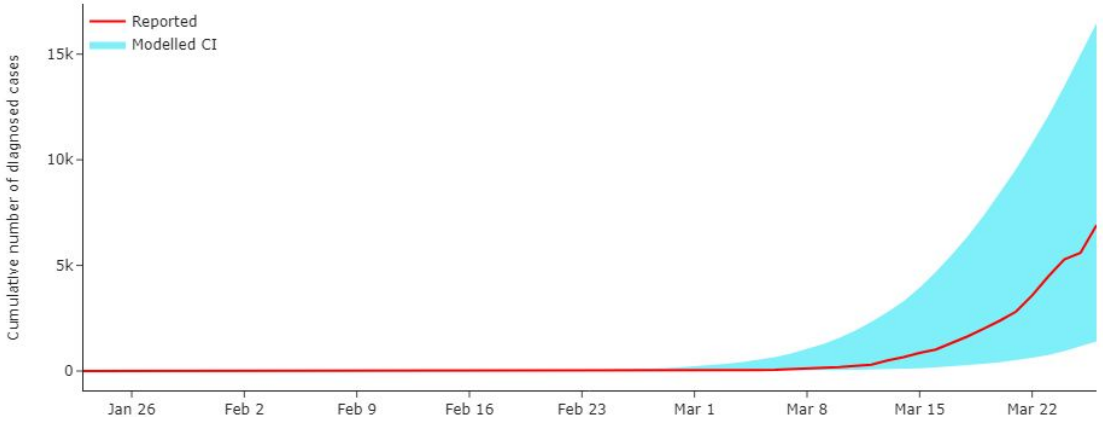

Figure 5: An overlay of modelled confidence interval for the number of diagnosed cases and reported values in Austria. The predictions of the model agree with observations.

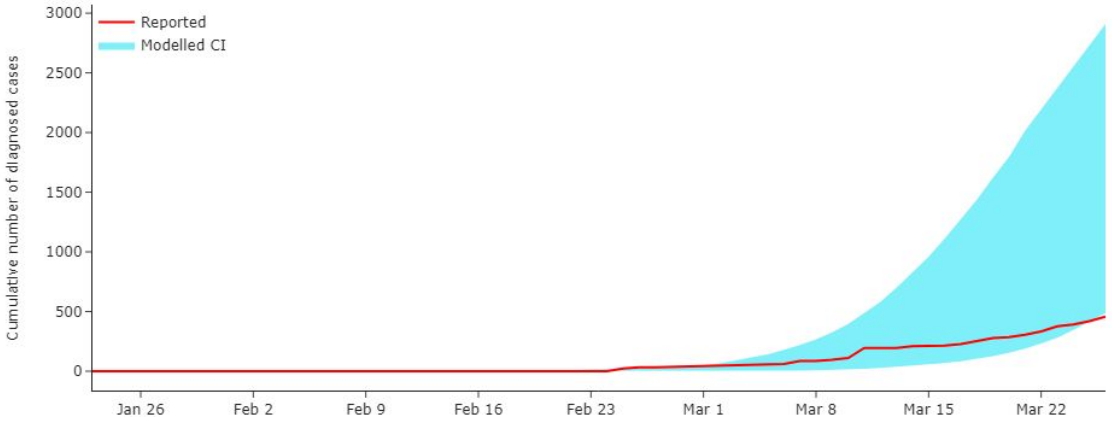

Figure 6: An overlay of modelled confidence interval for the number of diagnosed cases and reported values in Bahrain. The predictions of the model agree with observations.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.21.20040444; this version posted April 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

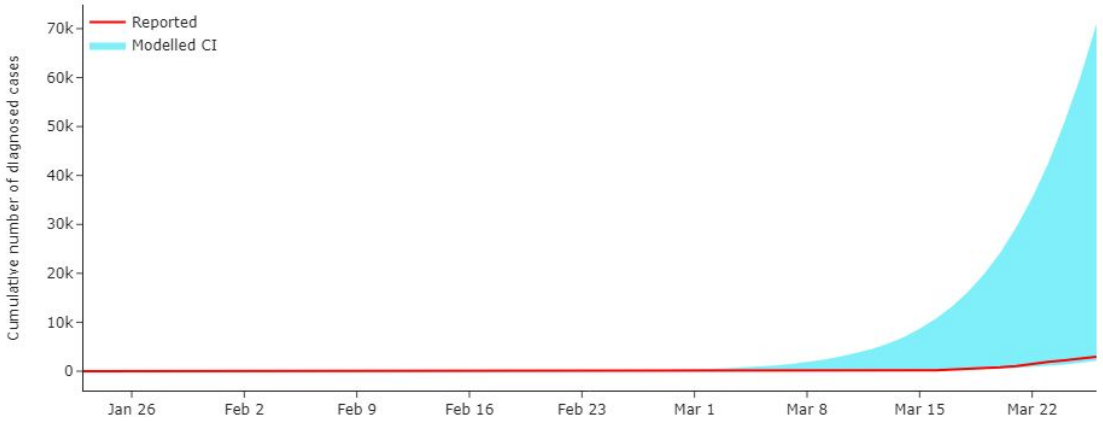

Figure 7: An overlay of modelled confidence interval for the number of diagnosed cases and reported values in Brazil. The observations agree well with model predictions, however, are in its lower bounds, putatively reflecting the low number of tests carried (14 per million people).

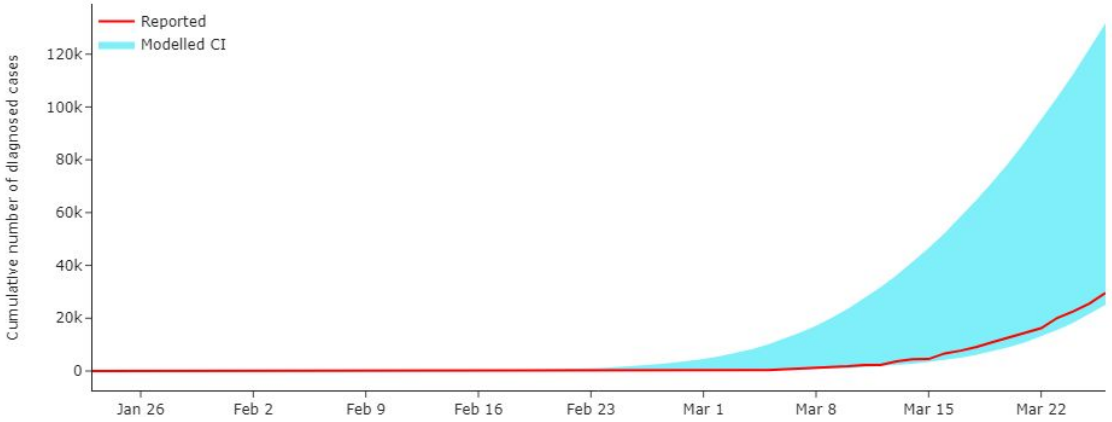

Figure 8: An overlay of modelled confidence interval for the number of diagnosed cases and reported values in France. The predictions of the model agree with observations.

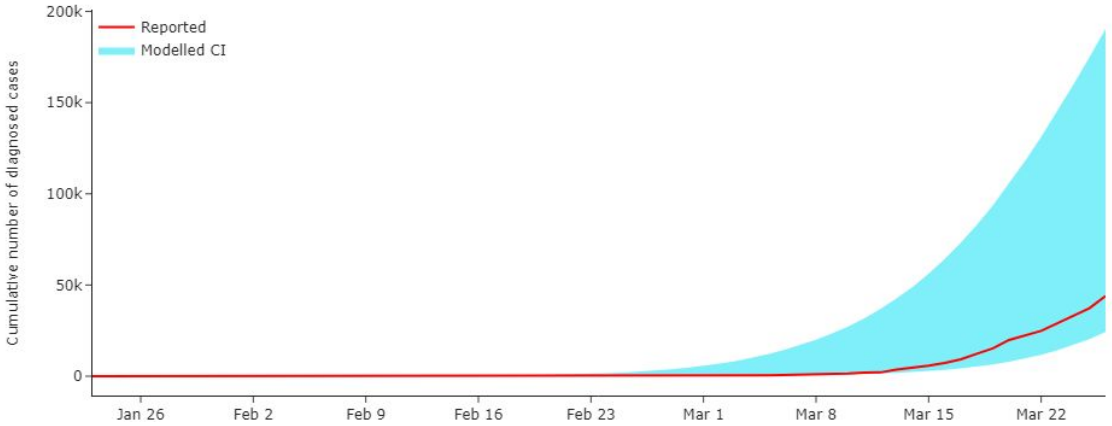

Figure 9: An overlay of modelled confidence interval for the number of diagnosed cases and reported values in Germany. The predictions of the model agree with observations.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.21.20040444; this version posted April 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

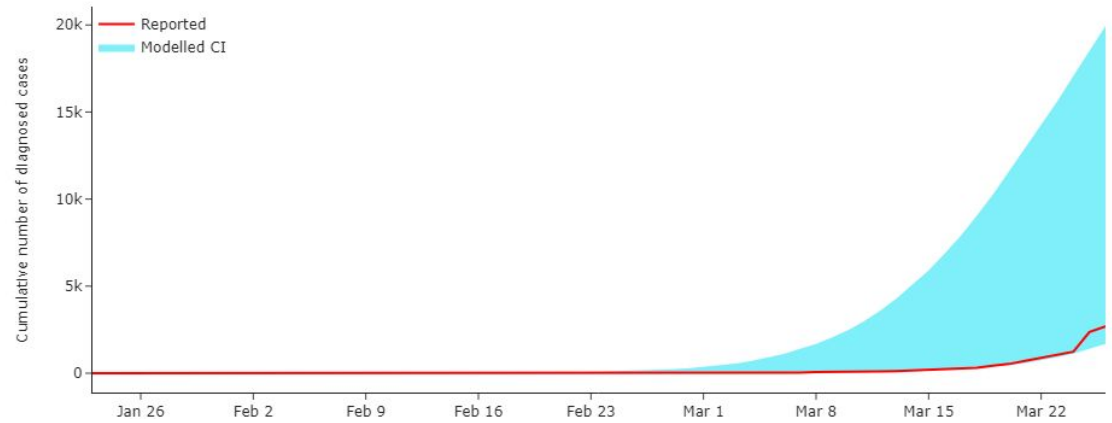

Figure 10: An overlay of modelled confidence interval for the number of diagnosed cases and reported values in Israel. The predictions of the model agree with observations.

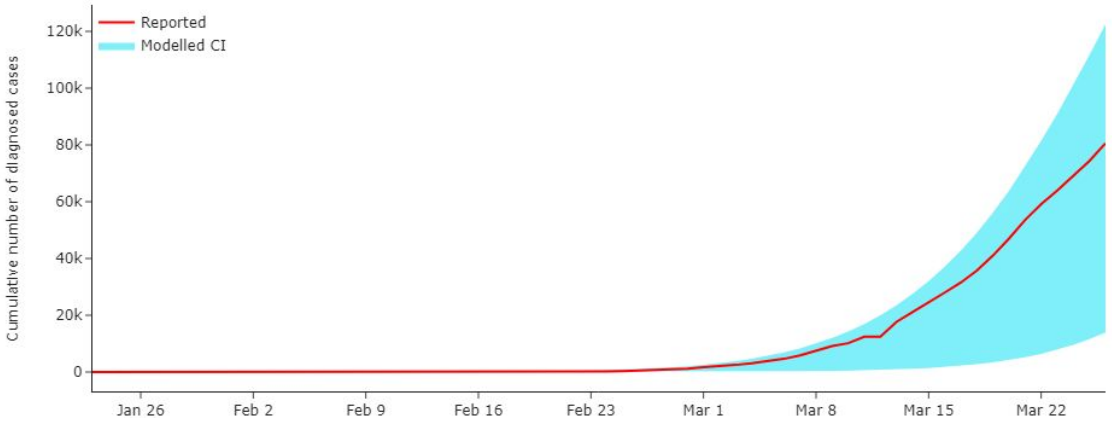

Figure 11: An overlay of modelled confidence interval for the number of diagnosed cases and reported values in Italy. The observations agree well with the model and are close to its upper bounds putatively reflecting the large number of tests carried recently in Italy (3299 per million of people).

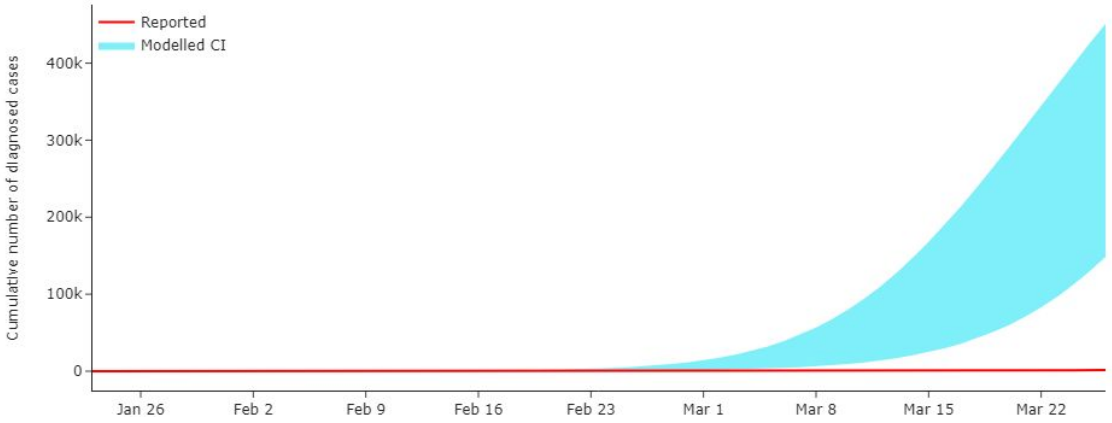

Figure 12: An overlay of modelled confidence interval for the number of diagnosed cases and reported values in Japan. The relatively small number of tests carried in the country (276 per 1 million citizens) leads to locally reduced ***tDR*** (proportion of detected to undetected cases), which possibly is smaller than the ***tDR*** value assumed in the model (0.0061) and caused the

medRxiv preprint doi: https://doi.org/10.1101/2020.03.21.20040444; this version posted April 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

observed overestimations of the model. Additionally, Japan quickly introduced strong precautions which could result in significant reduction of the disease spread.

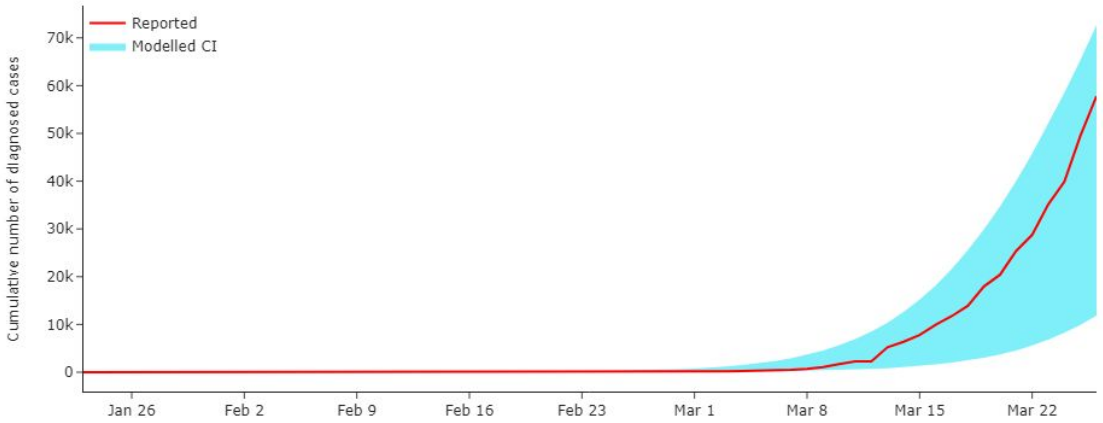

Figure 13: An overlay of modelled confidence interval for the number of diagnosed cases and reported values in Spain. The predictions of the model agree well with observations.

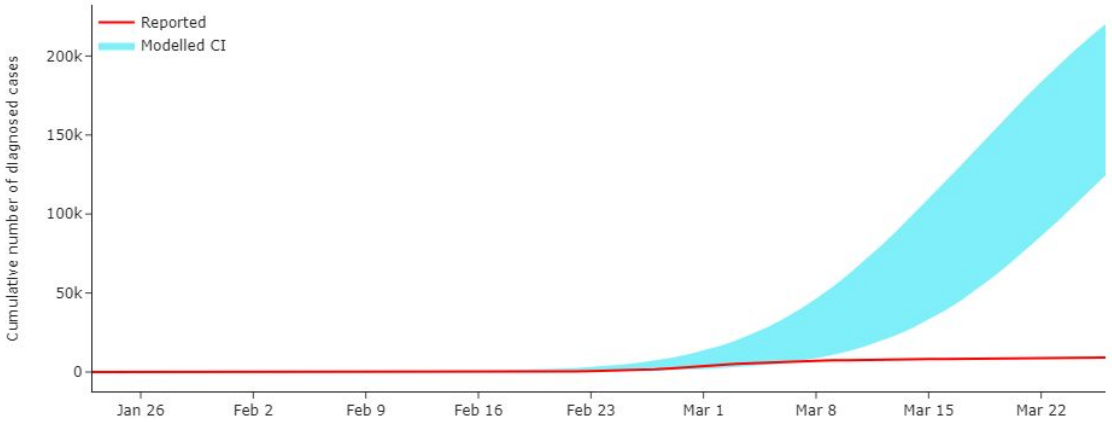

Figure 14: An overlay of modelled confidence interval for the number of diagnosed cases and reported values in the Republic of Korea. The model overestimates the number of diagnosed cases at the end of the tested period, even though the Republic of Korea excels in the number of carried tests (6148 per million people) and should have locally increased ***tDR*** (proportion of diagnosed to undiagnosed cases). Most likely, this discrepancy is caused by the fact that the model does not take into account precautions overtaken by the Korean government that seem to successfully slow down the spread of the disease.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.21.20040444; this version posted April 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

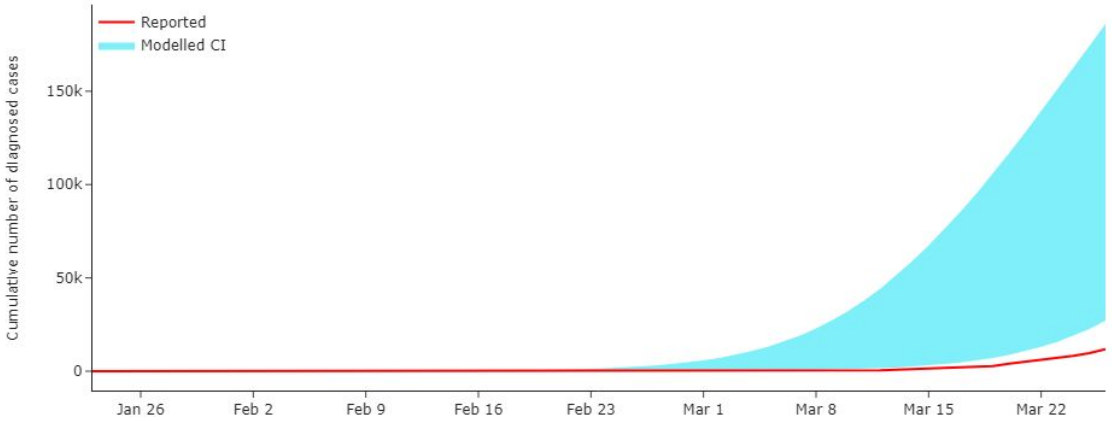

Figure 15: An overlay of modelled confidence interval for the number of diagnosed cases and reported values in the UK. The model overestimates the number diagnosed cases since March 15th. This may be caused by flight-related interventions introduced since January, which in case of the geographically isolated country like the UK may have stronger impact than similar restrictions in countries from the continental part of Europe.

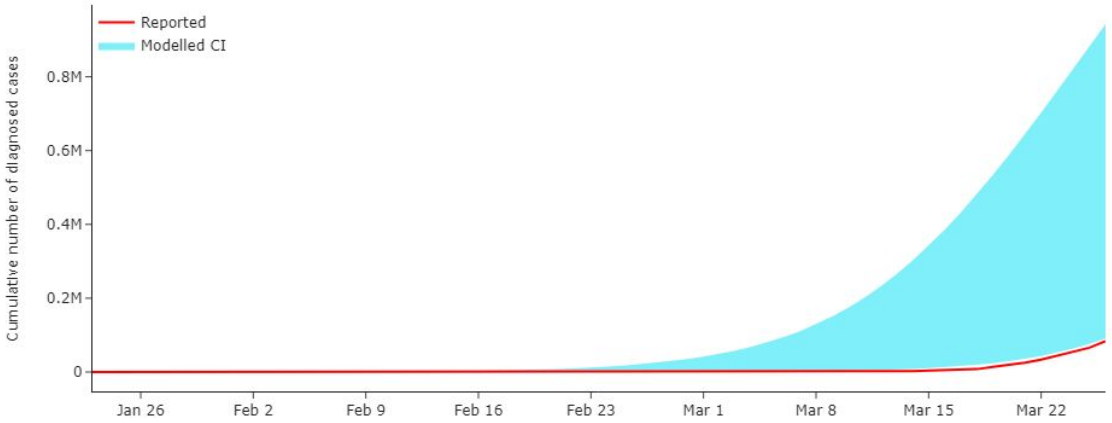

Figure 16: An overlay of modelled confidence interval for the number of diagnosed cases and reported values in the US. The observations are below lower bounds of the model, putatively reflecting the low number of tests carried (314 per million people), which leads to lower ***tDR*** value than assumed in our model.