



OPEN

Country-specific optimization strategy for testing through contact tracing can help maintain a low reproduction number (R_0) during unlock

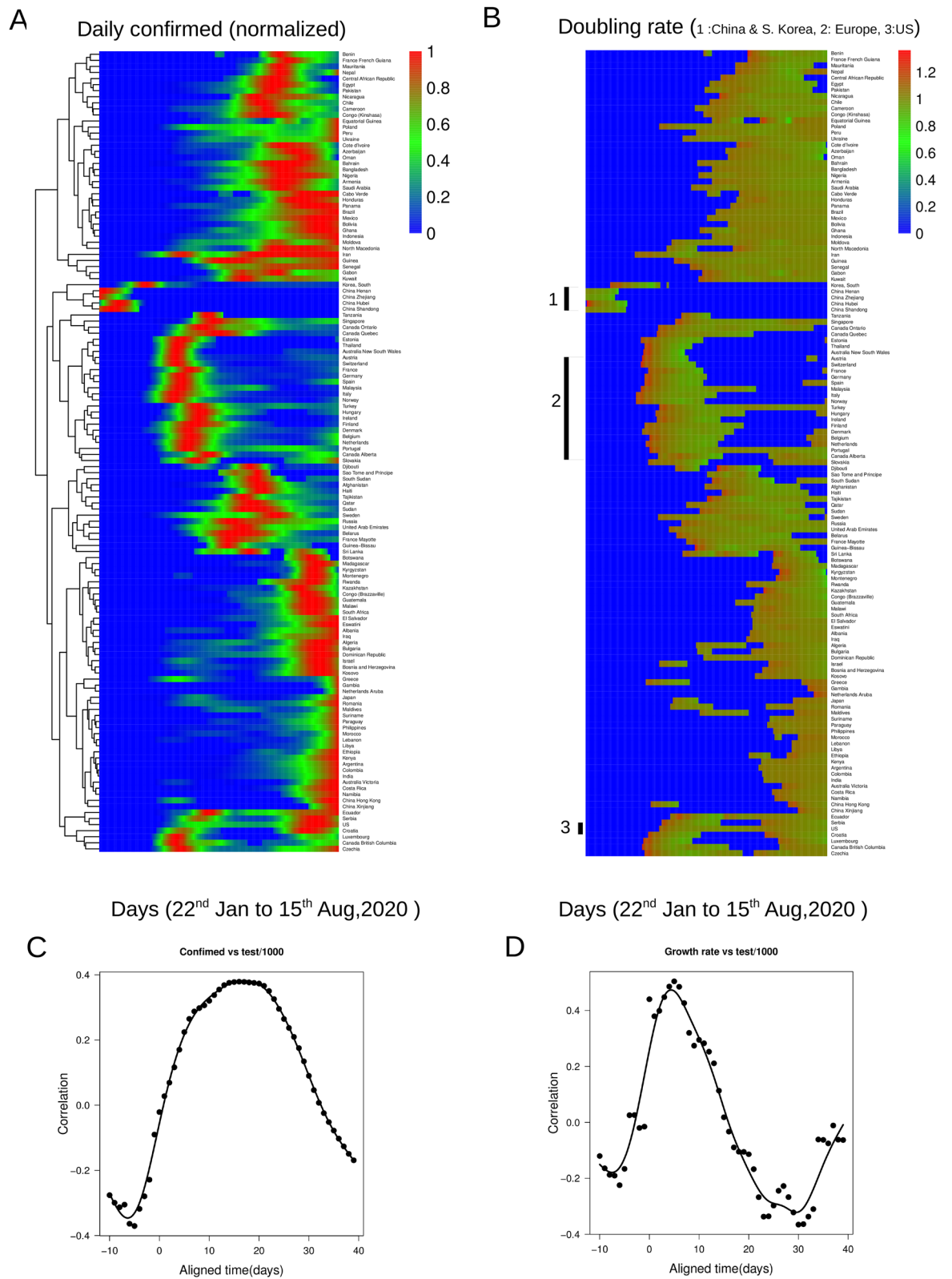
Uddipan Sarma^{1✉} & Bhaswar Ghosh^{2✉}

In response to the COVID19 pandemic, many countries have implemented lockdowns in multiple phases to ensure social distancing and quarantining of the infected subjects. Subsequent unlocks to reopen the economies started next waves of infection and imposed an extra burden on quarantine to keep the reproduction number (R_0) < 1. However, most countries could not effectively contain the infection spread, suggesting identification of the potential sources weakening the effect of lockdowns could help design better informed lockdown-unlock cycles in the future. Here, through building quantitative epidemic models and analyzing the metadata of 50 countries from across the continents we first found that the estimated value of R_0 , adjusted w.r.t the distribution of medical facilities and virus clades correlates strongly with the testing rates in a country. Since the testing capacity of a country is limited by its medical resources, we investigated if a cost–benefit trade-off can be designed connecting testing rate and extent of unlocking. We present a strategy to optimize this trade-off in a country specific manner by providing a quantitative estimate of testing and quarantine rates required to allow different extents of unlocks while aiming to maintain $R_0 < 1$. We further show that a small fraction of superspreaders can dramatically increase the number of infected individuals even during strict lockdowns by strengthening the positive feedback loop driving infection spread. Harnessing the benefit of optimized country-specific testing rates would critically require minimizing the movement of these superspreaders via strict social distancing norms, such that the positive feedback driven switch-like exponential spread phase of infection can be avoided/delayed.

Declaration of the coronavirus pandemic by WHO severely overhauled global economic and social endeavors¹. In response to the initial upsurge of infection spread^{3,4}, many European and East Asian countries were able to contain the first wave successfully by imposing strong mitigation measures through nationwide lock-down coupled with rigorous testing and quarantine strategies. However, even with strict lock-downs, many countries struggled to contain the growth of infection². During the pre-vaccine phase⁵ isolation of the infected population following aggressive testing and ensuring strong social distancing were the two most widely accepted ways globally to contain the infection spread and reduce the fatalities^{6,7}. Despite the recent launching of several vaccine programs worldwide, complete immunization can take several months to years to cover the complete cross-section of susceptible age groups in all countries, especially in the economically fragile countries. Also, there are indications that vaccination alone may not be sufficient to completely contain COVID19⁸; this inevitably rearticulates the importance of occasional lockdown and quarantine strategies in the immediate future.

However lockdowns, also had major financial implications in the livelihood of millions of people, especially (but not restricted to) in low and medium-income countries^{9–12}, calling for immediate plans to open up the economic activities. The unlocking measures then subsequently led to an increase in infection rates which initiated the next wave of infection in several countries who were largely successful in containing their respective first waves. The lockdown-unlock cycles were not successful to the extent desired and there could be scope

¹Vantage Research, Sivasamy St, CIT Colony, Mylapore, Chennai, Tamil Nadu 600004, India. ²Center for Computational Natural Sciences, International Institute of Information Technology, Hyderabad 500032, India. ✉email: uddipans@gmail.com; bhaswar.ghosh@iiit.ac.in



◀ **Figure 1.** The Covid-19 outbreak in different countries and its relation with the testing rate. (A) The heat map displays the clustered dynamics for around 100 different countries for the daily cases normalized to the maximum for each country (Days correspond to 22nd January, 2020 and to 15th August, 2020). The dendrogram is based on hierarchical clustering of the time traces. The horizontal axis represents the time in days while the vertical axis corresponds to different countries. The color code provides the (normalized) number of confirmed cases. (B) The heat map represents the doubling rate as a function of time clustered according to the dynamics of the time traces as shown in (A). The horizontal axis represents the time in days while the vertical axis corresponds to different countries. The color code provides the doubling rates of confirmed cases. The results are shown on an exponential scale. (C) Pearson correlation coefficient of test rate with the daily confirmed cases at different time points along the dynamics are shown. The solid lines represent loess based local regression fits. The time traces are aligned w.r.t the maximum doubling rates. 0 on the x-axis corresponds to the maximum doubling rate for all the countries. (D) The Pearson correlation coefficient of test rate with the doubling rate at different time points along the dynamics aligned as described in (C). The solid lines represent loess based local regression fits.

of improvement in the strategies of testing and quarantining to maintain the basic economic activities while minimizing the spread of infection at the same time. The number of testing conducted per day in a country is limited by the country's resource capacity¹³ but it is also observed that the testing capacity is steadily rising in most of the countries^{14–16}. Intuitively, the allowed extent of unlocking in a country should be a function of testing capacity as the daily testing captures the trend (increase or decrease) of infections per day, but it is not yet well established how testing facilities and extent of unlocking can be connected quantitatively. In this study we show that a country-specific cost–benefit trade-off between testing rate and unlocking extent can facilitate devising quantitative guidelines for unlocking, while considering the maximum resource capacity for testing within a given country.

Here, firstly we built a dynamic epidemic model^{17–20} that captures the spread of COVID19 infection and calibrated the model to country-specific time series data¹ for confirmed, recovered and dead populations. We did this for 50 different countries with various stages of infection. The model simulations and subsequent analysis suggests how the extent of partial unlock (extent of unlock) and quarantine rates can be optimally combined to maintain $R_0 < 1$; this depends on transmission rate, quarantine capacity and associated cost of testing. We also connected the extent of unlock with the frequency of the periodic unlock²¹ time. Although the dynamic epidemic models build with ordinary differential equation (ODE) can quantitatively estimate infection spread parameters with high confidence they can not explicitly capture the relation between R_0 and testing rates, hence, we next developed an agent-based stochastic epidemic model which uses the country specific infection parameters optimized using the ODE model and then incorporates testing through contact tracing of individual infected agents. Using this hybrid modeling approach, we first estimated the testing rate required to maintain $R_0 < 1$ and further studied how superspreaders²² can spread the infection during strict lockdowns; our simulations show that a small fraction of superspreaders can account for the majority of infections observed during the lockdowns which would then eventually be amplified further in the next cycle of unlocking. Indeed, most countries reported a continuous rise in infection during the lockdown, indicating such plausible underlying contribution from superspreaders. A recent study on two Indian states showed that a small fraction of superspreaders were responsible for transmitting ~80% of infection during lockdown²². Analysis of epidemic models specific to different countries shows a switch-like spread of infection can occur as a function of disease transmission rate and quarantine rate stemming from the implicit systems-level positive feedback loop primarily driving the spread of infection epidemics such as COVID19. Our analysis strongly suggests that minimizing the movement of superspreaders during the lockdowns can be very critical to the rapid success of such lockdowns and it can delay/circumvent the onset of the exponential increase in infection in the subsequent unlocks.

Results

Country specific doubling rate of the infection is dependent on the testing rate. To investigate the impact of testing on the epidemic dynamics in different countries, we first took the daily confirmed infection time course data from WHO¹ and clustered the region-wise data according to their dynamic patterns. A hierarchical clustering algorithm (using pheatmap package in R) is used to analyze the dynamics of selected countries (countries with at least 1000 infections per day in their maximum infection spread phase between 15th Jan to 15th August 2020 were selected). Figure 1A shows provinces in China and S. Korea, for instance, are clustered together, as the infection spread there at the earliest times. Subsequently, infection in different parts of Europe/US is followed by infection in other Asian and South American countries (in Fig. 1A the color bar represents daily confirmed cases normalized to the maximum for each country (see Figure S1A for an enlarged version representing certain regions in the heatmap). Next, we calculated the doubling rate (the inverse of the doubling time which shows how much time does the population take to double the number of infections in each country) from the time series data of these countries, as shown in Fig. 1B. Both Fig. 1A, B comparatively show that infection peaks and doubling rates vary across countries. For a more direct comparison of the time evolution of the infection trajectories, we superimposed them. In order to superimpose the trajectories, the doubling rate time courses for the different countries were aligned at the maximum doubling rate (Details in SI). Linear correlations analysis (Methods) between different time points of the aligned trajectories and numbers of tests/10,000 population across these countries are shown in Fig. 1C. We observed that the normalized daily cases showed a positive correlation with the test rate around a time point of 16 days on the aligned time axis (Figs. 1C, S1C, where 0 corresponds to the time of maximum doubling rate), suggesting, countries with higher test rates tend

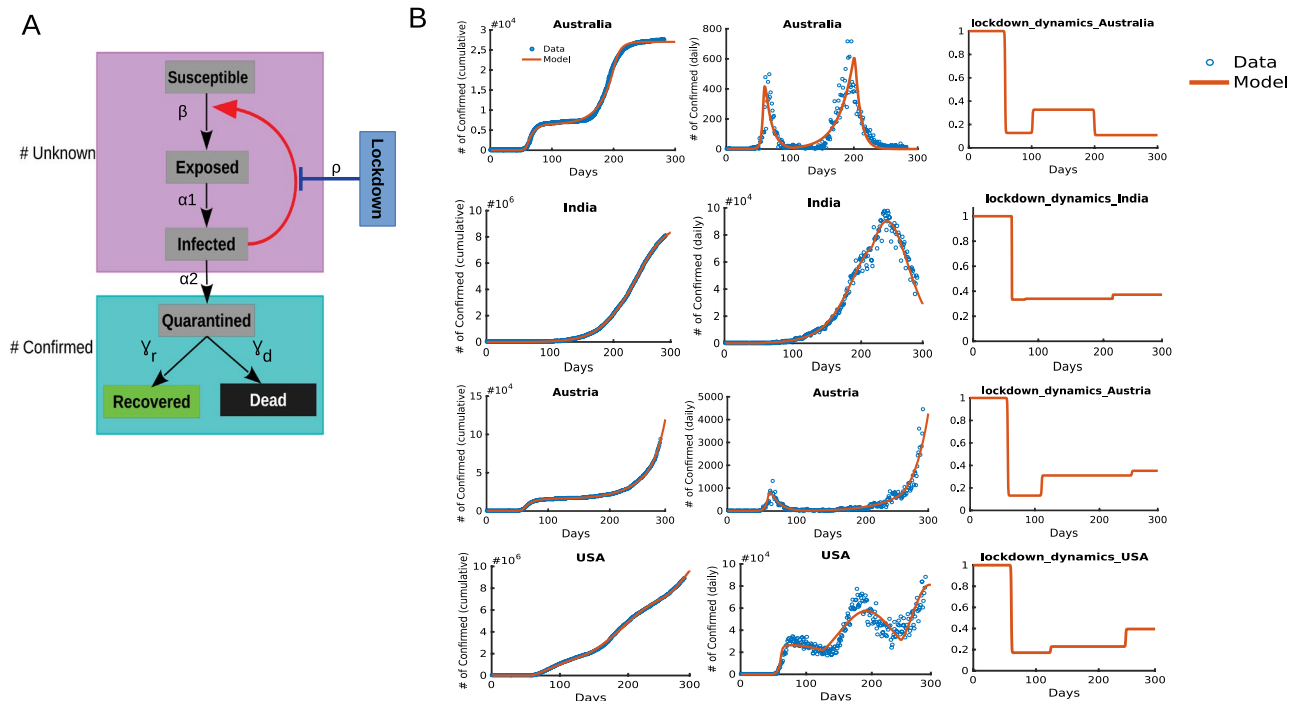


Figure 2. The description of the SEIQR model and calibration for different countries. **(A)** Schematic representation of the SEIQR model which contains susceptible, exposed, infected, quarantined, recovered and dead compartments. The lockdown is implemented through a sigmoid function as shown in the methods section. The quarantined, recovered and dead cases together comprise the confirmed cases. The arrow from infected to susceptible represents the positive feedback that fuels the infection spread in the population which in turn is negatively regulated by the lockdown, as indicated by a blunt headed bar. **(B)** The SEIQR models fit the data for 4 representative countries as indicated for the cumulative confirmed and daily confirmed cases. The number of days in X axis corresponds to the time course data available in JHU CSSE [41] where 0 corresponds to 22nd January, 2020 and the end time point corresponds to 30th October, 2020. The lockdown function in the third column shows the extent of lockdown (also the percentage of lockdown where 1 = 0% and 0 = 100% lockdown) which is also estimated by fitting during the model calibration.

to have higher confirmed cases. On the other hand, the doubling rate after around 26 days displays a negative correlation with test rate (Figs. 1D, S1B lower panel), hence, countries with higher testing rates are also able to reduce the rate of infection spread faster.

During the start of the infection, the doubling rate is usually high for a higher testing rate as indicated by a positive correlation. Thus, a higher testing rate (Figure S1B top panel) during the onset of infection allows more accurate estimation of the spread of infection which can be critical in containment of the infection at its early stages through preventive measures such as quarantine. Next, in order to quantitatively investigate the effect of testing rate and lockdown on R_0 we next took recourse to a quantitative epidemic model and subsequently an agent-based model.

R_0 values during lockdown period exhibit significant dependency on the test rate. To calculate the R_0 for different countries we next constructed a compartmental model to quantitatively capture the dynamics of infection. The model contained five compartments—susceptible (S), exposed (E), Infected (I), Quarantined (Q) and removed (R), where R contains both recovered (Re) and dead (D) population fractions²³. Figure 2A shows the structure of the implemented SEIQR model. A susceptible person is exposed to the infection through transmission from an already infected person (rate parameter β). After exposure, the exposed individual (E) is infected too and moves to compartment I (rate parameter α_1) and it subsequently gets quarantined in compartment Q (rate parameter α_2). The quarantined individual eventually either recovers (rate parameter γ_r) or dies (rate parameter γ_d). In this model, we have not considered reinfection of the recovered individual. Infection dynamics were simulated by a set of 6 coupled ordinary differential equations. The SEIQR model thus constructed is next fitted to the confirmed (Co), recovered (Re) and dead (De) population trajectories obtained from the public data (details in methods and SI, section II).

Figure 2B shows the confirmed cumulative and daily cases, as well as the estimated lockdown extent (details of lockdown function in methods sections) and lockdown dynamics for four representative countries with different susceptible population sizes and stages of infection. Model fits for the 50 countries are shown in Supplementary Figure S2. The average incubation time independently estimated from the fitting infection dynamics of all the 50 countries (average of $\frac{1}{\alpha_1}$) is approximately 8.1 days which is close to the observed value²⁵. Our simple SEIQR

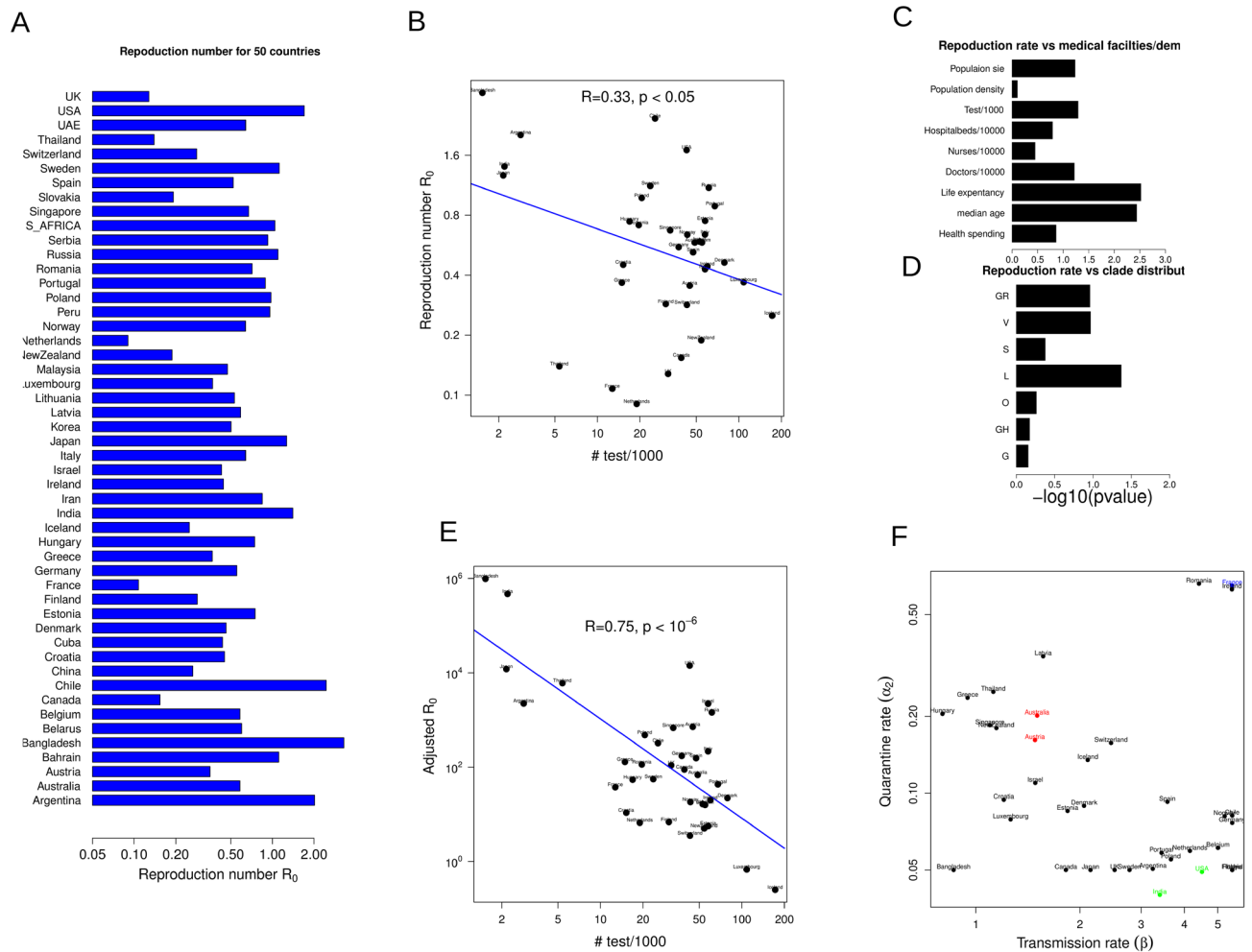


Figure 3. The reproduction number (R_0) displays correlation with the testing rate across countries. (A) The barplot shows R_0 for the fitted 50 countries after lockdown. (B) The scatter plot describes correlation between the test rate and R_0 after lockdown for the 39 countries with at least 10 complete genome sequences as of the end of June, 2020 described in the main text. (C) The statistical significances of the linear dependence of R_0 on different demographic and medical facilities factors are shown based on linear regression for the same 39 countries. (D) The statistical significance of linear dependence of R_0 on the frequencies of the clades (G, GR, GH, L, O, S, V) are shown. (E) The scatter plot indicates the correlation between the adjusted R_0 values and the test rates. R_0 values for each country are adjusted based on the correlation obtained in (C, D) described in the method section. (F) The values of the fitted transmission rates and quarantine rates for all the 50 countries are plotted and the representative countries for consideration of the cost–benefit analysis and agent based model simulation are shown in colors red, green and blue.

model thus quantitatively captured the cumulative infection dynamics of countries with different approaches towards lockdown and subsequent unlock implementation. Parametrizing the model with multiple country specific waves of infection (data until Oct 2020), where each wave is of distinct magnitude and duration (Fig. 2B, shown for four representative countries Australia, USA, Austria and India), resulted in country specific infection parameters of higher confidence. The model fitted to cumulative trajectories of a country also captured well the daily infection numbers in that country (Fig. 2B, 2nd column). We next used these country specific parameters to calculate the respective R_0 values.

R_0 for different countries were calculated just after lockdown (Fig. 3A). The test rate displays a negative correlation with R_0 values (Fig. 3B), however, the correlation is not significant (correlation coefficient ~ 0.33). R_0 presumably depends on many other factors including demographics, medical facility and distribution of virus strains in the population which may interfere with the influence of test rate in spread of infection. In order to systematically adjust for the influence of such factors on R_0 , we collected several publicly available datasets of demographics, medical facilities and genome sequences in a country specific manner¹ (details in methods). R_0 values have a significant correlation with several of these factors (Figs. 3C, S3A). We also observed a statistically significant correlation of R_0 with the frequency of clades in the population; specifically the L and GR strains show significant influence on R_0 (Figs. 3D, S3B). We next linearly adjusted the R_0 values (see methods) using

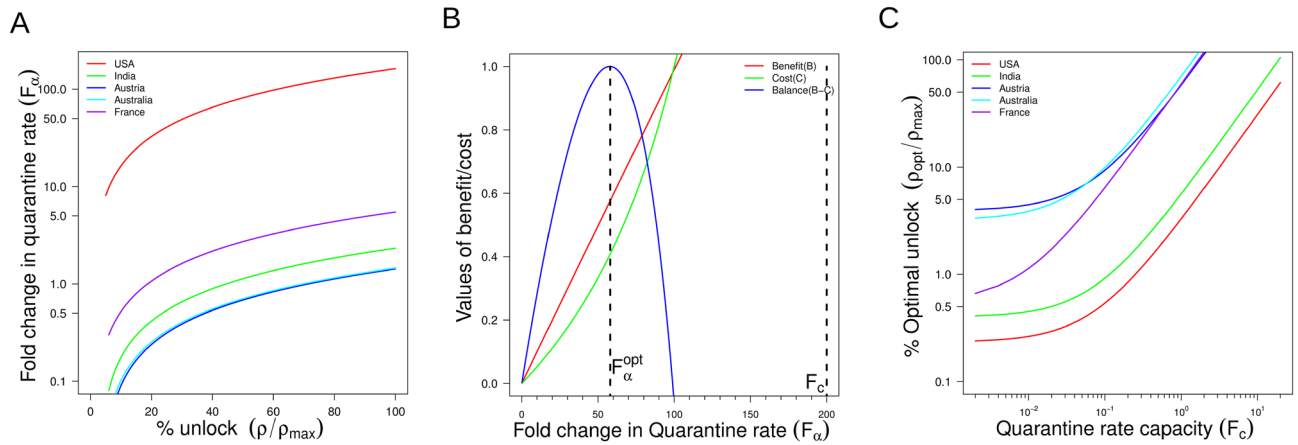


Figure 4. The optimization of trade-off between testing rate and extent of unlock. **(A)** The relationship between the extent of unlock and the corresponding quarantine rate required to keep R_0 values equal to one for five countries as indicated. The equation $F_\alpha = \frac{\beta - \rho}{\alpha_2 \rho_{max}}$ (SI IV) is plotted with F_α on y-axis and $\frac{\rho}{\rho_{max}}$ on x-axis for different country specific (α_2, β) values. **(B)** The cost and benefit curves as a function of quarantine fold change for the countries. The benefit curve $B = \frac{F_\alpha}{F_0 + F_\alpha}$ (SI IV) is plotted at different values of F_α on x-axis for fixed value of $F_0 = 10^4$. The cost curve $C = \lambda F_c \frac{F_\alpha}{F_c - F_\alpha}$ (SI IV) is plotted at different values of F_α for fixed value of maximum quarantine capacity $F_c = 200$ and cost per quarantine capacity $\lambda = 5 \times 10^{-5}$. The cost–benefit is optimized at a value $F_\alpha^{opt} \sim 60$ for a maximum capacity $F_c = 200$ of the fold change in quarantine rate. **(C)** The optimum extent of unlock calculated from F_α^{opt} as a function of the maximum quarantine capacity F_c for different countries as indicated. The Eq. (3) for optimal unlock is plotted with $\frac{\rho_{opt}}{\rho_{max}}$ on the y-axis and F_c on x-axis by fixing the parameters $F_0 = 10^4$ and the cost per quarantine capacity, $\lambda = 5 \times 10^{-5}$ for different country specific (α_2, β) values as indicated. Both the axes are plotted in log-scale.

the most significant factors derived from the correlation analysis—median age, population size, doctors/10,000 patient and frequency of the L clade—which shows that the adjusted R_0 values exhibit much higher correlation with the test rate (Fig. 3E. correlation coefficient 0.75). Here, for the correlation analysis, we choose countries only with at least 10 complete genome sequences till the end of June 2020 which selects 39 countries out of the original set of 50 countries.

Owing to the asymptomatic component of the COVID19 infection, testing the population is essential in order to quarantine the asymptomatic people. Thus, the quarantine rate would critically depend on the testing rate. Ideally a high testing rate would allow a country to unlock its economies and at the same time maintain a low infection rate. On the other hand, increasing the testing rate imposes extra financial and resource costs. To study the nature of this cost–benefit trade-off between quarantine and unlock, we specifically selected 5 representative countries with different transmission rates for further investigation: high transmission rates (USA, India), low transmission rates (Australia, Austria) and high transmission rate and high quarantine rate (France) (Fig. 3F shows countries with distinct transmission rates and quarantine rates).

Optimum changes in country-specific quarantine rates can effectively reduce the infection post lockdown. Typically, when lockdown is removed, the transmission rate (β) is expected to increase. For our SEIQR model the R_0 value is given by $R_0 = \frac{\beta - \rho}{\alpha_2 \rho_{max}}$ (see SI section III for derivation) where ρ captures the extent of lockdown ($\rho = 1$ and $\rho = \rho_{max}$ represents full lockdown and full unlock respectively). From this expression, we can calculate the extent by which the quarantine rate (α_2) must be increased in order to compensate for the effect of partial or full unlock such that $R_0 < 1$ can be maintained during unlock. The fold change in α_2 to keep the R_0 value close to one is given by (SI for derivation)

$$F_\alpha = \frac{\beta - \rho}{\alpha_2 \rho_{max}} \tag{1}$$

Here, $\rho = 1$ corresponds to a full lockdown case and a gradual increase in value of ρ depicts partial unlock finally moving towards full unlock at $\rho = \rho_{max}$. The equation shows that the required fold change to compensate for the effect of lockdown removal would reduce as the extent of unlock reduces (Fig. 4A). In fact, our model fitting shows a higher transmission rate for the USA compared to Austria or Australia, asking for a higher quarantine rate in the USA for the same extent of unlock (Fig. 4A). However, for instance for France, in spite of a very high value of β , the required increase in quarantine is low due to its already high quarantine rate (Fig. 4A). Better control on infection spread during unlocks would require an increased quarantine rate which can only be achieved by rigorous contact tracing and testing. We studied this scenario as a cost–benefit tradeoff where on one hand removal of lockdown benefits economic activities but at the same time more testing needs to be conducted to circumvent the enhanced rate of infection spread. If we assume that the maximum quarantine capacity (of fold

change) in α_2 is F_c and investment cost in testing is λ per testing, the optimal fold change in α_2 would be (see SI section IV for details)

$$F_\alpha^{opt} = F_c \frac{\left(\lambda \frac{F_c^2}{F_0}\right)^{\frac{3}{2}} + 1}{\left(\lambda \frac{F_c^2}{F_0}\right)^{\frac{1}{2}} + 1} \quad (2)$$

Figure 4B shows an illustration of the optimality in quarantine rate given a maximum capacity of quarantine rate $F_c = 200$. Both the benefit (B) and cost (C) increases with F_α which leads to an optimal $F_\alpha^{opt} \sim 60$ where (B–C) is maximized (Fig. 4B). The optimal value is much below the maximum quarantine capacity of 200. Finally from Eq. (1) the optimal lockdown removal (ρ_{opt}) to maintain $R_0 < 1$ is given by

$$\frac{\beta}{\alpha_2} \frac{\rho_{opt}}{\rho_{max}} = F_c \frac{\left(\lambda \frac{F_c^2}{F_0}\right)^{\frac{3}{2}} + 1}{\left(\lambda \frac{F_c^2}{F_0}\right)^{\frac{1}{2}} + 1} \quad (3)$$

Hence as the quarantine capacity F_c is low, the ρ_{opt} would be low (Fig. 4C) which illustrates: as the quarantine capacity is low, the extent of unlock should remain optimally low. Notably, higher value of unlock ρ is possible by increasing F_α beyond F_α^{opt} to maintain $R_0 < 1$, but the overall benefit would be low given the cost of testing. In order to increase optimal extent of unlock ρ_{opt} , the maximum quarantine capacity F_c is also required to be increased (Fig. 4C) by augmenting the maximum testing capacity. In addition, the optimal benefit of testing would also depend on the intrinsic transmission rate (β). For instance, owing to higher value of β , given a maximum capacity (F_c , the optimally beneficial unlock extent would be low for the USA (Fig. 4C) compared to Austria or Australia with much lower values of β who could afford a greater degree of unlocking for same F_c . Similarly, Eq. (3) also shows that the optimal extent of lockdown (ρ_{opt}) would increase if the cost of testing (λ) increases. One effective way to manipulate the extent of lockdown is by varying the frequency of full unlock in a periodic lockdown-unlock²⁰ (time interval of full unlock T_1 followed by a time interval of full lockdown T_2) within a time cycle of period $T = T_1 + T_2$ (SI section V and Figure S4 for details).

Agent-based contact tracing model quantitatively connects testing rate to desired unlock extents.

The ODE modeling frameworks used to study the dynamics of infection (including ours) cannot model contact tracing explicitly, hence it cannot connect the number of testing required to achieve a certain quarantine rate. To address this gap we have next developed an agent-based model (ABM) to introduce testing through contact tracing, such that a quantitative understanding on the required number of testing per unit population can also be derived (details in methods), in a country specific manner. In this hybrid modeling approach the agent-based stochastic version of the SEIQR model was simulated using country specific infection parameters obtained via calibrating the ODE model to country specific data (methods). In the ABM model infection spreads outward, starting with one infected agent at the center (Figure S5). The selected model for the parameter values of the USA, for instance, shows as the testing rate increases the maximum confirmed cases also increases (Figure S6A, qualitatively in agreement with data shown in Fig. 1C), but this also leads to the containment of infection much earlier causing a reduction in the reproduction number during different partial unlocks (Fig. 5A). The results are comparable for other countries as well (Figure S6B–D). We next used an interpolation method based on cubic spline fitting (smooth.spline from R) to determine the minimum test rate (N_{test}) required to keep the $R_0 < 1$ during each partial unlock as indicated in Fig. 5A (methods). This calculation results in a monotonic increase in the required testing rate as the extent of unlock gradually progresses from full lockdown towards full unlock (Fig. 5B). The exact values of required testing rates vary among countries for different partial unlocks depending on country specific transmission and quarantine rates (Fig. 5B). It is noticeable that due to low transmission rate before lockdown, countries like Austria, Australia are estimated to be able to unlock by 80% with their current testing rates, whereas for the USA, the current testing rates need to be ramped up at least by 3 times to achieve an unlock of 80%. In fact, this apparently projected connection between the agent-based model and the analytical calculation is justified by the observation that the quarantine rate required to maintain R_0 at a value of 1 from the analytical calculation follows a linear scaling relation with the corresponding testing rate required to maintain R_0 at 1 from the agent-based model (SI V for details).

A small fraction of superspreaders can dramatically increase infection spread during lockdown.

The burden on the testing rate capacity can be essentially reduced by keeping the transmission rate low (e.g. Austria) implemented through strict standard measures of social distancing, wearing face masks, etc. However, even during the full lockdown, a small fraction of superspreading individuals²² can substantially undermine the diligent efforts exercised by the majority of the population. The superspreaders are defined as having a low quarantine rate and high transmission rate compared to the regular spreaders. A recent detailed contact tracing analysis focusing on a couple of states in India demonstrates that during the full lockdown period a small fraction of superspreaders was responsible for transmitting 80% of reported infection, so even with rigorously planned full lockdown $R_0 < 1$ could not be achieved²². Such contributions of superspreaders may generally explain why in most countries full lockdowns could not contain the infection effectively. Using our agent-based contact-tracing model we next simulated different scenarios where a population comprises different proportions of superspreaders.

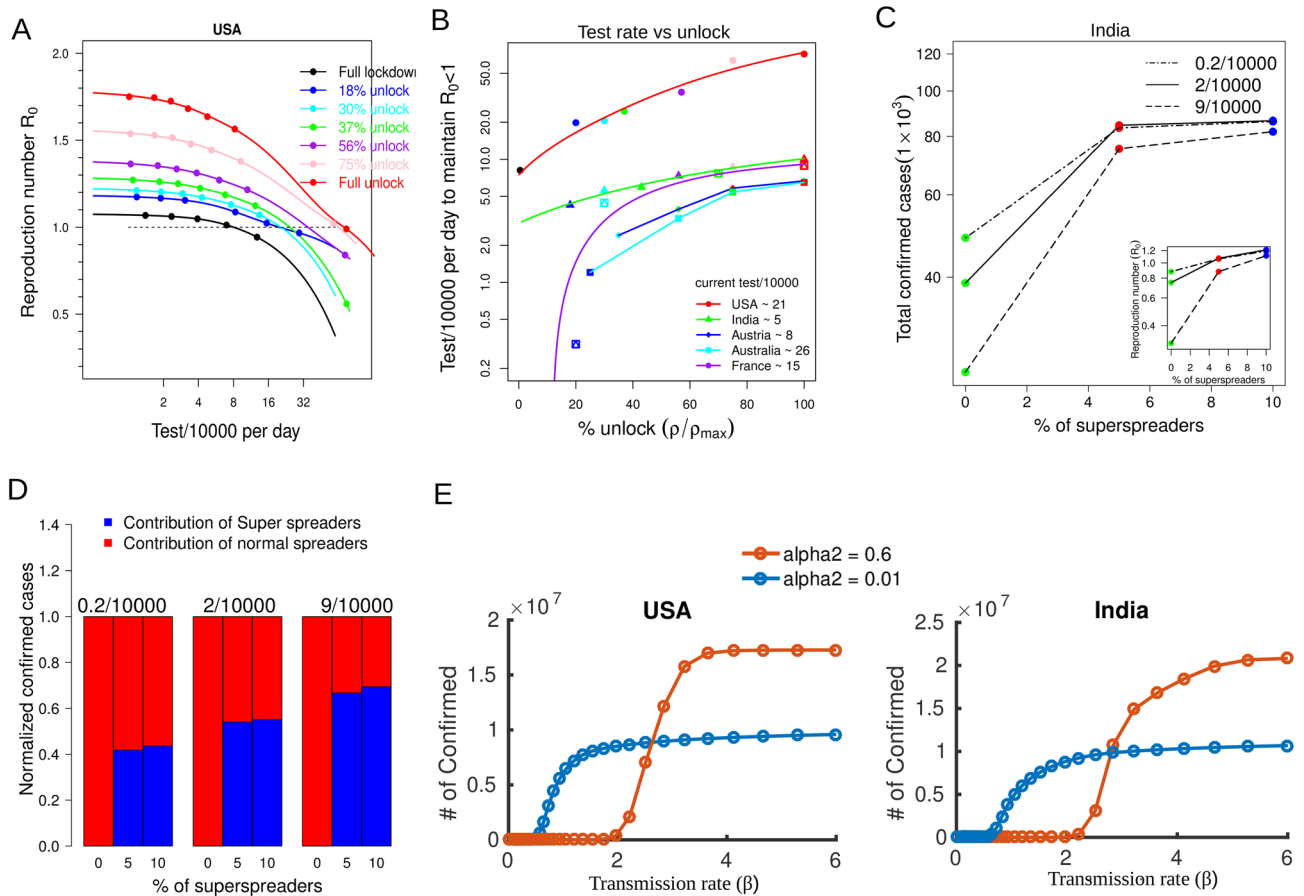


Figure 5. The testing rate and fraction of superspreaders in the population influence the infection rate simulated by the agent based model. **(A)** The R_0 values based on the agent based model with parameter values of the USA as a function of test/10,000 for different extents of unlock as indicated. The other four countries are shown in supplementary figure S4. The dashed horizontal line exhibits the $R_0 = 1$ line, intersection of the line with the curves quantifies the test/10,000 required to maintain $R_0 = 1$. **(B)** The allowed extent of unlock and the corresponding test/10,000 required to keep the R_0 value less than one calculated from the agent based simulation for the different countries with indicated current test rate. **(C)** The total confirmed cases as a function of percentage of superspreaders simulated for parameter values of India. Different lines indicate simulation conducted at different testing rates as indicated. (The inset shows the corresponding R_0 values at different percentages of superspreaders). **(D)** The barplot indicates the contribution of superspreaders and normal spreaders over the total number of confirmed cases as a function of the percentage of superspreaders at 3 different testing rates 0.2/10,000, 2/10,000 and 9/10,000, as indicated. **(E)** The relationship between the transmission rate and the total confirmed cases at two different values of quarantine rate as indicated for two representative countries, the USA and India.

Our simulations suggest that a small increase in the fraction of superspreaders can drastically increase the number of confirmed cases (Fig. 5C & Figure S6F, simulated with parameter sets specific to India) and also the R_0 (Fig. 5C-inset) when the testing rate is kept high to best track the spread. Notably, after the superspreader fraction is more than ~5% the impact of further increase in superspreader fraction is minimal, suggesting, a sharp switch-like saturating phenomenon driving the superspreader effect which suggests that it is critical to maintain a minimal number of superspreaders during the lockdown, for $R_0 < 1$ (Fig. 5C). For instance, our simulations suggest that a small number of superspreaders (5%) contribute to the spreading of the majority of the infection (~70% when the testing rate is 9/10,000, Fig. 5D) and the fraction of total infection for 5% superspreader fraction is marginally less compared to the condition when superspreader composition is 10% in the population (Fig. 5D). The total number of confirmed cases for 5% or 10% superspreaders are also comparable and are distinct from the scenario when the superspreader fraction is 0% (Figure S6F, shown for India). In fact, we observed that the fraction of superspreaders required to have a 50% contribution in total infection over a period of time can be as low as 1% for India, whereas for Austria 3% superspreaders are needed for the same effect (data not shown). Here, Austria owing to a lower intrinsic transmission rate is able to tolerate more superspreaders whereas a very small superspreader fraction in countries like India (or the USA) with high transmission rate can result in a dramatic increase in overall infection. In order to better understand the mechanism for this switch-like behavior, especially at a high quarantine rate, we conducted deterministic simulations at different values of transmission rates (β).

The total confirmed cases increase in an ultrasensitive switch-like manner at a high quarantine rate α_2 (where α_2 is the mechanistic equivalent of testing rate in the agent-based model) while the increase is more gradual at low α_2 (Fig. 5E). This switch can be understood as a consequence of the inherent systems-level positive feedback loop that drives the infection spread in epidemic outbreaks where testing and quarantining are implicit components. Typically the infected and yet not quarantined fraction of the population is the driver of the positive feedback loop (by exposing more subjects to infection) where high quarantine reduces and low-quarantine enhances the strength of this loop. While rigorous testing and quarantining is a necessary measure to contain the infection, it can only delay the onset of the exponential phase of infection spread and movement of superspreaders helps reach the exponential phase faster by fueling the strength of the positive feedback loop to reach the switch threshold. In the ODE model, the superspreaders and regular spreaders would have β values above and below the switching threshold (Fig. 5E), respectively, during the lockdown. This will lead to an overall dramatic increase in the number of infected populations owing to the superspreaders' contribution. This analysis thus suggests that in order to leverage the benefit of augmented testing capacity, tracing the movement of superspreaders appears as crucial as the implementation of other standard social distancing measures.

Discussion

The COVID19 epidemic has disrupted normal life in an unprecedented manner in almost every corner of the world. Similar outbreaks, with different infective capacity, were reported earlier, for instance, the basic reproduction number or R_0 for COVID19 is comparable to the SARS-cov^{26,27} and much higher than the MERS infection^{28,29}. In this study, we quantitatively explored the COVID19 outbreak and its relation with quarantine measures in several countries using both deterministic and agent-based models coupled to analysis of country specific metadata. We first employed a deterministic (ODE based) SEIQR epidemic model and fitted the trajectories from 50 countries at different stages and starting times (day) of infection. We then built an agent-based stochastic model³⁰ to implement contact tracing by utilizing the best-fit parameters from the ODE based SEIQR model and implement them in the agent-based model while ensuring a quantitative scaling between both models. The primary goal of utilizing the two distinct modeling approaches is to connect the testing rate to the unlock process since contact tracing can not be implemented in ODE based epidemic models. Earlier studies discuss the limitations of ODE based models and agent-based models³², but studies connecting both modeling approaches especially in the context of the COVID19 epidemics to understand the relation between testing rate and unlock measures are not reported in the literature. Here, feeding the country specific epidemic parameters from the ODE model into the agent-based model, we derive country specific optimal testing rates through contact tracing. Such analysis aims to guide different degrees of unlocking to open up economies in different countries since long-term lockdown is impractical and detrimental to the economy of any country³¹.

Typically, the actual infected number of people is expected to be higher than the sampled ones and limitations like that pose challenges for devising accurate mathematical models³². The actual susceptible population size is thus mostly unknown and it could often be different from the tested population size, so the number of infected people from the data only captures the tested sample when the total number of tests is less than the total population of a region/country. This is also one of the reasons why increased testing rate is critical in better capturing the magnitude of the infection (and not only the nature of the dynamics, the dynamic features of the infection spread can be robustly captured even with a relatively smaller tested population size) that can lead to devising more accurate epidemic models with better predictive capacities in early stages of the infection. In addition, because of the asymptomatic nature of the infection³³, the need to quarantine the maximum number of infectious people through testing is equally important to successfully contain the infection spread. Apart from the testing rate, other factors like immunization³⁴, age³⁵, sex³⁶ medical facility and specific virus strains³⁷ may play important roles both in transmission and fatality but testing rate is the only variable amongst all these critical factors that can be manipulated immediately post infection via human intervention.

In countries like India where population density and size outmatch the healthcare infrastructure, an early implementation of lockdown was critical in slowing down the spread of the infection. Similarly, in the early days of COVID19 spread, delay in lockdown implementation had a catastrophic impact in Italy or Spain despite adequate health infrastructure per capita, highlighting the importance of implementation of early and rapid measures post infection irrespective of healthcare infrastructure. This is further highlighted by countries like India and the USA where unlock measures immediately witnessed a dramatic increase in infection spread.

Reopening the economy was also an impending necessity in all countries under lockdown. Here, utilizing quantitative knowledge on infection and lockdown dynamics from the ODE model and feeding the country-specific best-fit parameters into the agent-based model, we propose strategies for optimal testing and quarantining of the infected subjects to maintain $R_0 < 1$ during the unlock periods. We quantitatively demonstrate the quarantine and testing that can be optimally adjusted (post lockdown) to minimize the infection spread. We demonstrate that the effective increase in testing mediated quarantine measures has to be country specific, primarily depending on the transmission and quarantine rates of a country. Already, many countries have taken initiatives to accentuate the testing capacities either through developing more testing facilities¹⁴⁻¹⁶ or through designing efficient pool testing algorithms^{38,39}. Our study can additionally help facilitate more informed planning of such testing processes where, for instance, if a goal of 30% unlock of economy is set by a given country (or a province in a country), then the respective increase in testing rates to achieve such unlock goals can be predefined as the minimal testing rate target. In an unprecedented efficiency, multiple vaccines are rolled out now, which will eventually aim to contain the COVID19. But it is still relatively unclear how the rapidly evolving variants of the virus will impact the efficacy of different types of vaccines as the vaccines have distinct mechanisms of target engagement. Further, many developing or underdeveloped countries may not have access to such widespread immunization programs to vaccinate their entire population till at least 2022⁴⁰. Thus, systematic planning of

testing and designing of lockdown-unlock measures remain key factors in containing the current spread of COVID19 in specific and perhaps in case of similar such epidemics in the future.

Periodic lockdown²¹ is another efficient way to implement partial lockdown which best works if the R_0 value during the lockdown is much lower than 1 maintained typically via rigorous testing and quarantining. However, in many developing as well as developed countries, it became almost impossible to limit the mobility of people even during lockdown^{11,25}, thus, maintaining a low R_0 during lockdown remains a challenge. A small unidentified fraction of the exposed population during the unlocks can potentially remain unidentified due to the long incubation period characteristic of COVID19. Indeed, we saw a higher surge of a second wave all over Europe that qualitatively resembles the second wave of Spanish flu pandemic of 1918⁴¹. A recent landmark study highlights how a small fraction of unidentified superspreaders dramatically accelerated the spread of infection in two states of India even during nationwide lockdown²² and undermined India's advantage of early lockdown. Looking at the trend of infection dynamics across the world, this could be true to multiple other countries as infection numbers kept on growing constantly even during the lockdowns. Thus, the discovery of superspreader mediated infection spread via contact tracing can be a critical component of lockdown implementation itself which further underscores the advantages of our proposed strategy of optimally augmenting the test rate through contact tracing. Additionally, the emergence of a few more new strains may also have accelerated the upsurge of confirmed cases⁴⁵. In this study, using an approach of cost-benefit optimization we derived a data-driven strategy that aims to achieve $R_0 < 1$ during unlocks which will perhaps also reduce the severity of uncontrolled future outbreaks. A general property of epidemic outbreaks such as COVID19 is that they are fueled by the strength of an implicit positive feedback loop connecting the infected population to the susceptible population. Such positive feedback can promote switch-like responses especially when the testing/ quarantine rates are high. Our simulations suggest a small fraction of superspreaders (with high β) can exploit the switch-like response capacity of the system to trigger a major increase in infection number during the lockdowns. As the superspreader fraction increases beyond a threshold even for a high testing rate scenario a sudden increase in the number of confirmed cases can arise due to overall change in the strength of the positive feedback loop. Thus, the containment of superspreaders during the lockdown would be critical to maintain a low R_0 such that subsequent unlocks can be implemented in an optimal manner, especially in densely populated cities where the transmission probability is higher. In fact, the recent upsurge of infection spread in several regions even after the launching of vaccination programs⁸ reiterates the warning that our complacency in obeying all social distancing and quarantine rules out of desperation to return to pre-pandemic life may incur fatal consequences. Finally, if this COVID19 pandemic is successfully contained in the coming months due to improvements in the efficacy of vaccines, the strategy we propose here may still facilitate better containment of future such outbreaks.

One of the limitations of our study or similar such studies used commonly to model infection spread in epidemics like COVID19 can be the assumption that the population is homogeneously distributed over a region/country with an equal number of contacts for every individual. In the real world the contact network in many cases follows a scale-free structure⁴⁶. In fact, a recent study has shown that incorporating such detail indeed improves model prediction accuracy⁴⁷. However, it is also suggested that such a real-world network would slow down the spread of infection⁴⁶. Thus, in our study, the prediction accuracy is expected to be less with a homogeneous contact network and the required testing rate would plausibly be overestimated (but not underestimated) due to this assumption. Different realizations of the contact network can easily be introduced in our ABM framework, but it will also increase the complexity of the current model with inclusion of additional parameters. Future extensions can systematically explore the effects of different types of real-world contact networks and their meaningful combinations on infections spread as well as testing rate required to maintain $R_0 < 1$. Another model assumption pertains to the fact that all the infected population will eventually be quarantined in the ODE model. This assumption is derived from the observations that during the COVID19 outbreak the capacity for asymptomatic people to spread the infection is much less compared to the symptomatic people³³ hence the positive feedback loop connecting the infected to the susceptible individuals is primarily driven by the infected and symptomatic individuals, an assumption also commonly used in epidemic models used to quantitatively model COVID19 infection²⁴. Here, as the quarantine rate would determine the R_0 value, a high quarantine rate would prevent the infected population from spreading the infection within a relatively shorter time scale while a low quarantine rate would allow the infected individuals to keep on spreading the infection. Based on the country specific quarantine rates the infected (and symptomatic) individuals will be quarantined at different rates, but in the duration between being infected and before being quarantined the individuals will continue to infect others, country specific dynamics of which is quantitatively captured by our country specific models. However, in a scenario where the asymptomatic subject driven infection spread is significantly higher, the models built with the current assumptions (including ours) will not capture the true dynamics of infection spread and would require extension to explicitly incorporate the asymptomatic compartment. In such scenarios understanding the true dynamics of infection would require excessive testing through contact tracing⁴⁸.

For the ABM model, since we are considering testing through contact tracing, if we start with a very high testing rate, the infection rate would be low which would require a small number of infected agents to be traced and tested. On the other hand, if we start with a very low testing rate, many exposed/infected individuals will remain untested. Hence in both low and high testing rate scenarios the total population will not be tested. In addition, since we also have a quarantine rate of infected population without any contact tracing and testing, similar to the ODE model, all the infected individuals will be quarantined eventually albeit slowly in the ABM model as well.

Materials and methods

Calculation of doubling rate. The doubling rate is calculated at each time point over the trajectory by taking the daily cases at that time and the next day. For an exponential growth of the infection, $N(t) = N(t + 1)exp(r_d)$ where $N(t)$ depicts the number of infected people at time t

$$r_d = \log\left(\frac{N(t + 1)}{N(t)}\right)$$

Here $t = 1$ day. The values in the heatmap in Fig. 1B are displayed in an exponential scale of doubling rate to avoid negative values. So a doubling rate of 1.4 in exponential scale means the doubling time of $\frac{1}{\log(1.4)} = 3$ meaning it takes around 3 days to double the number of infected people.

Correlation analysis. For the correlation and significance of the correlation, we used `cor.test()` function in R. The function uses t-distribution statistics to calculate the p-value of the correlation. The t value of a pair of random numbers with correlation r and number of points n is give by,

$$t = r \frac{\sqrt{n - 2}}{\sqrt{1 - r^2}}$$

The t-distribution provides the probability of t values for the null hypothesis that the mean correlation of the two random variables is zero. If the calculated correlation value falls on the tail of the t-distribution, the correlation is significant. The exact p-value is calculated from the t-distribution table of n degrees of freedom. The calculations are all done by the `cor.test()` package in R.

The SEIQR model. The model comprises of susceptible (S), Exposed (E), Infected (I), Quarantined (Q), Removed (R which contains two compartments 'recovered' and 'dead').

The equations are

$$\begin{aligned}\frac{dS(t)}{dt} &= -\beta * \rho(t) * \omega(t) * S(t) * I(t) \\ \frac{dE(t)}{dt} &= \beta * \rho(t)\omega(t)S(t)I(t) - \alpha_1 E(t) \\ \frac{dI(t)}{dt} &= \alpha_1 E(t) - \alpha_2 I(t) \\ \frac{dQ(t)}{dt} &= \alpha_2 I(t) - \gamma_d Q(t) - \gamma_r Q(t) \\ \frac{dR(t)}{dt} &= \gamma_r Q(t) \\ \frac{dD(t)}{dt} &= \gamma_d Q(t)\end{aligned}$$

where $S(t)$, $E(t)$, $I(t)$, $Q(t)$, $R(t)$ and $D(t)$ are the susceptible, exposed, infected, quarantine, recovered and dead population at time t respectively. $\rho(t)$ is the lockdown function and $\omega(t)$ is the time when infection starts in a given country after the first day of detection of the infection date as reported in WHO website (hence simulation start time corresponds to 22nd Jan, 2020 based on the WHO report).

The lockdown function $\rho(t)$ is given as

$$\rho(t) = \frac{1}{1 + \left[\frac{K}{1 + \exp(tc - t)}\right]^S}$$

Here $K = \text{Effect_lockdown_Country}$ which determines the maximal effective lockdown for a given country, for instance $K = 1, 2$ and 3 would result in 50%, 66.6% and 75% maximal lockdown, respectively, during the lockdown. Here $tc = \text{time of lockdown start}$ and $S = \text{strength of lockdown}$ that determines the steepness of the lockdown implementation time from no lockdown to the maximal lockdown.

$\rho(t)$ varies between 1 (no lockdown) and 0 (full lockdown). $\omega(t)$ ensures that the model for a specific country is switched on when the infection begins in that country, hence if the detected case in a country is 40 days after Jan 22nd, the model for that country is switched on 35 days post Jan 22nd (assuming mean incubation time of 5 days), during the calibration.

The lockdown is opened by modifying the $\rho(t)$ function such that $\rho(t)$ returns to 1 from its lockdown status to no-lockdown (1) status in a designated time.

Model calibration. Model calibration involves minimizing an objective function that gives best fit parameter sets for confirmed, recovered and dead populations for a given country, simultaneously. We fitted the time series provided by JHU CSSE at github⁴³ to the SEIQR model developed in the study and minimized the objective function using the nonlinear least square fitting algorithm (`lsqnonlin`) from MATLAB.

It returns $x = \text{lsqnonlin}(\text{fun}, x_0, \text{lb}, \text{ub})$, where x is a vector/matrix of variables whose values are to be determined, for instance, in our case x comprises the country specific infection parameters to be estimated via the

fitting the model to the observed infection dynamics for a given country of interest. The function ‘fun’ contains the residuals or sum of squares of the difference between model value and data point for a given time point; the residuals are evaluated for each time point for which data point is available. $\times 0$ is the initial parameter vector that comprises the initial guesses of the model variables and ub and lb respectively represent upper bound and lower bound vectors corresponding to each value in $\times 0$. For instance, to calibrate the confirmed population trajectory for the USA the fitting algorithm minimizes the sum of residuals for [Quarantined_USA + Recovered_USA + Dead_USA] between both model and data. This exercise was carried out for all the 50 countries individually to obtain the country specific infection parameters (a more detailed description is provided in SI II).

Analysis of the virus genome sequences for clade distribution. We first collected the individual sequenced genome from individual countries and their corresponding clades as reported⁴³. Each genome in a particular country belongs to one of the seven clades (G, GH, GR, L, S, V, O) as described⁴³. From the data, we determine the frequency of each clade in individual countries by finding the number of sequences belonging to that clade divided by the total number of sequences. To calculate the correlation, out of the analysed 50 countries, we only considered countries having at least 10 sequences in the database.

Adjusted R_0 values. The R_0 values obtained following the ODE SEIQR model fitting are adjusted with respect to the significant parameters from the correlation analysis described in Figs. 3C, D and S3A,B. The adjustment factor is characterised as

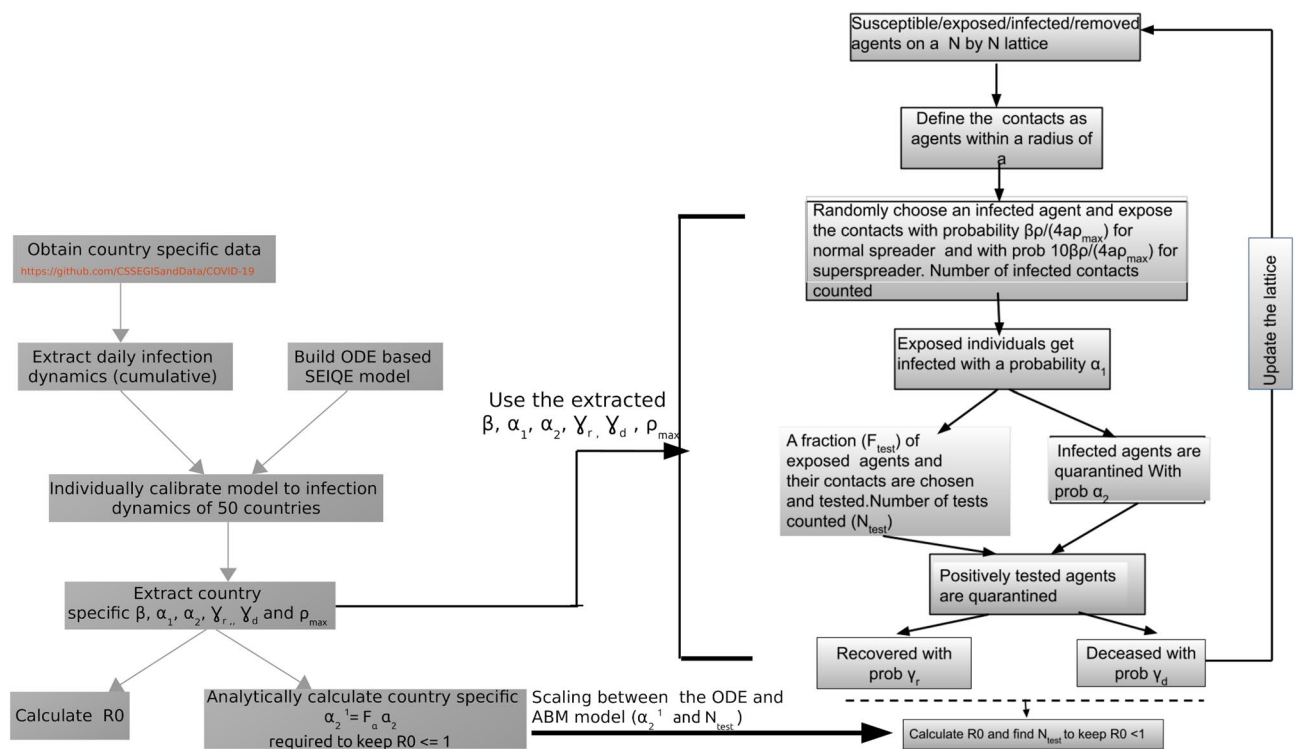
$$A = \frac{1}{\text{populationsize}} * \text{median} \left(\frac{\text{age} * \text{doctors}}{10000} * \frac{\text{nurses}}{10000} * \frac{1}{\text{frequencyofGRclade}} \right)$$

The factors which are positively correlated with R_0 (population size, frequency of GR clade) are placed in the denominator and factors having negative correlation values are placed in the numerator. The adjusted R_0 values are quantified as

$$R_0^{\text{adjusted}} = \frac{R_0}{A_n} \text{ where } A_n = \frac{A}{\max(A)}$$

Agent-based stochastic SEIQR model. We utilized the model structure and calibrated parameters in the deterministic SEIQR model, as described in the previous section to build an agent-based stochastic version of the same³⁰. In this model, (the flowchart below) the individual agents are assumed to be located on a two-dimensional lattice of dimension 300×300 . The points on the lattice represent susceptible agents in the population and each individual on the lattice is surrounded by four nearest neighbors at a minimum distance of one. Here, we considered the agents within a radius of seven as the contacts of every agent on the lattice. Thus, every individual agent would be in contact with a maximum of 28 neighbors. The initial patient at zero time is located exactly at the center having a location of (150,150) on this two-dimensional lattice. The infection spreads through persistent close contact with the nearest neighbors. At every time step, all the exposed and the infected agents are selected and 28 neighbors of the individual get exposed to the infection with a probability β . The value β_{ODE} in the ODE model is defined as the number of individuals a particular person infects per unit time. In the stochastic model each individual would infect $28 \times \beta$ people on an average per unit time. Thus, the conversion between transmission rates of the ODE and stochastic model is given by $\beta = \frac{\beta_{ODE}}{28} \frac{\rho}{\rho_{\max}}$. Here we assumed that both the asymptomatic/exposed and symptomatic patients are capable of spreading the infection. The exposed patient in turn exhibits symptoms with a probability α_1 and the symptomatic infected patient would be further quarantined with a probability α_2 . The quarantined patient finally either recovers or dies with a rate γ .

The susceptible, exposed, infected, quarantined and recovered agents on the lattice are assigned values 0, 1, 2, 3 and 4 respectively which allow us to track the dynamics of each component separately (Detail flowchart of the algorithm is provided below). To simulate country specific infection dynamics the parameters values for the ABM model were taken as the respective fitted values from the ODE based SEIQR model for a particular country. In order to introduce diagnostic testing, we identified all the symptomatic as well as the quarantined individuals from the pool at a particular time point and selected a fraction (F_{test}) of the identified patients for contact tracing. The contacts of each selected individual were traced and diagnostic tests were conducted on them. The close contacts showing positive results were finally quarantined. This procedure is repeated for all the selected individuals. The status of the whole population is updated and the next time step is continued. The simulation was performed for 200 time points which corresponds to 200 days in the real time as the fitted model parameters used in the ABM model (from the ODE model) have units in 1/day. In order to increase the number of testing, the value of F_{test} was increased. The parameter ρ is varied to change the extent of unlocking. However, we assumed that the testing procedure only commences after at least 100 individuals are already infected in the population, in order to introduce an initial delay in responding to the situation. For every infected person, we also keep track of the number of individuals who are getting infected from that particular infected person until that person is quarantined. The number of people infected during the simulation time defines the reproduction number for that individual. The R_0 value for the whole population is quantified by taking the average over reproduction numbers for all the agents (persons) in the population. The number of tests (N_{test}) is also counted at each time point. For a particular values of ρ , the N_{test} and corresponding R_0 value are recorded which enables us to calculate the minimum N_{test} value required to keep $R_0 < 1$. The scaling relation connecting the minimum N_{test} and quarantine rate (α_2) obtained from the ODE model (Eq. (1)) is discussed in SI-VI in detail. A flowchart describing our workflow that also connects the ODE model and the ABM model is given below.



Data availability

The data analysis and simulation were performed in R software [R version 3.4.4 (2018-03-15)]. <https://cran.r-project.org/bin/windows/base/old/3.4.4/>.

Received: 13 May 2021; Accepted: 2 December 2021

Published online: 07 January 2022

References

- World Health Organization. Coronavirus Disease 2019 (COVID-19) Situation Report-66 (WHO, 2020). https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200326-sitrep-66-COVID19.pdf?sfvrsn=9e5b8b48_2.
- Li, R., Rivers, C., Tan, Q., Murray, M. B., Toner, E., & Lipsitch, M. The demand for inpatient and ICU beds for COVID-19 in the US: lessons from Chinese cities. medRxiv 2020.03.09.20033241 [Preprint].
- “Not a wave, a tsunami: Italy hospitals at virus limit” AP NEWS, 13 March 2020.
- “COVID-19 infections rise in New York with peak weeks away.” AP NEWS, 25 March 2020. <https://apnews.com/7c7563cb82626a4042797c6aa6da260a>. R. M. Anderson.
- DRAFT landscape of COVID-19 candidate vaccines—20 April 2020. <https://www.who.int/blueprint/priority-diseases/key-action/novel-coronavirus-landscape-ncov.pdf>.
- Hollingsworth, T. D. How will country-based mitigation measures influence the course of the COVID-19 epidemic? (2020). [https://doi.org/10.1016/S0140-6736\(20\)30567-5](https://doi.org/10.1016/S0140-6736(20)30567-5) pmid:32164834.
- Bi, Q., Wu, Y., Mei, S., Ye, C., Zou, X., Zhang, Z., et al. Epidemiology and Transmission of COVID-19 in Shenzhen China: Analysis of 391 cases and 1,286 of their close contacts. medRxiv 2020.03.03.20028423 [Preprint]. <https://doi.org/10.1101/2020.03.03.20028423>.
- Moore, S., Hill, E. M., Tildesley, M. J., Dyson, L. & Keeling, M. J. Vaccination and non-pharmaceutical interventions for COVID-19: a mathematical modelling study. *Lancet Infect. Dis.* **18**, S1473–3099(21)00143–2 (2021).
- Bartik, A., Bertrand, M., Cullen, Z., Glaeser, E. L. & Luca, M. The impact of COVID-19 on small business outcomes and expectations. *Proc. Natl. Acad. Sci. USA* **117**(30), 17656–17666 (2020).
- Walker, P. G. T. The impact of COVID-19 and strategies for mitigation and suppression in low- and middle-income countries. *Science* **369**, 413–422 (2020).
- Weilla, J. A., Stigler, M., Deschenesc, O. & Springbornd, M. R. Social distancing responses to COVID-19 emergency declarations strongly differentiated by income. *Proc. Natl. Acad. Sci. USA* **117**(33), 19658–19660 (2020).
- Bonaccorsia, G. et al. Economic and social consequences of human mobility restrictions under COVID-19. *Proc. Natl. Acad. Sci. USA* **117**(27), 15530–15535 (2020).
- Roser, M., Ritchie, H., Ortiz-Ospina, E., & Hasell, J. Coronavirus pandemic (COVID-19). Published online at OurWorldInData.org; 2020. Retrieved from: <https://ourworldindata.org/coronavirus>.
- Tromberg, B. J. et al. Rapid scaling up of COVID-19 diagnostic testing in the United States: the NIH’s RADx Initiative. *N. Engl. J. Med.* **383**, 1071–1077 (2020).
- <https://www.who.int/india/news/feature-stories/detail/how-india-scaled-up-its-laboratory-testing-capacity-for-covid19>.
- <https://news.cgtn.com/news/2020-06-24/China-s-daily-COVID-19-testing-capacity-hits-3-78-million-RAA1yHsBYk/index.html>.
- Marsland III, R. & Mehta, P. Data-driven modeling reveals a universal dynamic underlying the COVID-19 pandemic under social distancing. 2020. [arXiv:2004.10666](https://arxiv.org/abs/2004.10666)[q-bio.P].

18. Kissler, S. M., Tedijanto, C., Goldstein, E., Grad, Y. H. & Lipsitch, M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science* **14**, 5793 (2020).
19. Huang, Y., Yang, L., Dai, H., Tian, F., & Chen, K. Epidemic situation and forecasting of COVID19 in and outside China. 2020. <https://doi.org/10.2471/BLT.20.255158>.
20. Singh, R., & Adhikari, R. Age-structured impact of social distancing on the COVID19 epidemic in India; 2020. [arXiv:2003.12055](https://arxiv.org/abs/2003.12055).
21. Karin, O. et al. Adaptive cyclic exit strategies from lockdown to suppress COVID-19 and allow economic activity. 2020. medRxiv 2020.04.04.20053579v1.
22. Laxminarayan, R. et al. Epidemiology and transmission dynamics of COVID-19 in two Indian states. *Science* **370**, 691–697 (2020).
23. Jumpen, J. et al. A SEIQR model for pandemic influenza and its parameter identification. *Int. J. Pure Appl. Math.* **52**(2), 247–265 (2020).
24. Dehning, J. et al. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science* **369**(6500), eabb789 (2020).
25. Li, Q. et al. Early Transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N. Engl. J. Med.* **382**, 1199–1207 (2020).
26. Riley, S. et al. Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions. *Science* **300**(5627), 1961–1966 (2003).
27. Lipsitch, M. et al. Transmission dynamics and control of severe acute respiratory syndrome. *Science* **300**(5627), 1966–1970 (2003).
28. Chowell, G. et al. Transmission characteristics of MERS and SARS in the healthcare setting: A comparative study. *BMC Med.* **13**, 210. <https://doi.org/10.1186/s12916-015-0450-0> (2021).
29. Kucharski, A. J. & Althaus, C. L. The role of superspreading in Middle East respiratory syndrome coronavirus (MERS-CoV) transmission. *Euro Surveill.* **20**(25), 14–18 (2015).
30. Cuevas, E. An agent-based model to evaluate the COVID-19 transmission risks in facilities. *Comput. Biol. Med.* **121**, 10382 (2020).
31. <https://www.economist.com/leaders/2020/04/30/life-after-lockdowns>.
32. Bertozzi, A. L., Franco, E., Mohler, G., Short, M. B. & Sledge, D. The challenges of modeling and forecasting the spread of COVID-19. *Proc. Natl. Acad. Sci. USA* **117**(29), 16732–16738 (2020).
33. Moghadas, S. M. et al. The implications of silent transmission for the control of COVID-19 outbreaks. *Proc. Natl. Acad. Sci. USA* **117**(30), 17513–17515 (2020).
34. Luis, E. Escobar, Alvaro Molina-Cruz, and Carolina Barillas-Mury (2020) BCG vaccine protection from severe coronavirus disease 2019 (COVID-19). *PNAS* **117**(30), 17720–17726 (2020).
35. Ioannidis, J. P. A., Cathrine Axfors, B. C. & Contopoulos-Ioannidis, D. G. Population-level COVID19 mortality risk for non-elderly individuals overall and for nonelderly individuals without underlying diseases in pandemic epicenters. *Environ. Res.* **188**, 109890 (2020).
36. Takahashi, T. et al. Sex differences in immune responses that underlie COVID-19 disease outcomes. *Nature* **588**, 315–320 (2021).
37. Korber, et al. Tracking changes in SARS-CoV-2 spike: Evidence that D614G increases infectivity of the COVID19 virus. *Cell* **182**, 812–827 (2020).
38. Ghosh, S., et al. Tapestry: A single-round smart pooling technique for COVID-19 testing. medRxiv; 2020. <https://doi.org/10.1101/2020.04.23.20077727>.
39. Aragón-Caqueo, D., Fernández-Salinas, J. & Laroze, D. Optimization of group size in pool testing strategy for SARS-CoV-2: A simple mathematical model. *J. Med. Virol.* **92**(10), 1988–1994 (2020).
40. Irwin, A. What it will take to vaccinate the world against COVID-19. *Nature* **592**, 176–178. <https://doi.org/10.1038/d41586-021-00727-3> (2021).
41. Barry, J. M., Viboud, C. & Simonsen, L. Cross-protection between successive waves of the 1918–1919 influenza pandemic: epidemiological evidence from us army camps and from Britain. *J. Infect. Dis.* **198**(10), 1427–1434 (2008).
42. <https://www.hindustantimes.com/india-news/fake-news-led-to-migrants-exodus-mha/story-AEYej7C7jaLAkuQ2dCicML.html>.
43. Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Inf. Dis.* **20**(5), 533–534 (2018).
44. Elbe, S. & Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Global Chall.* **1**, 33–46 (2017).
45. Weisblum, Y. et al. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *Elife* **9**, e61312 (2020).
46. Karaivanov, A. A social network model of COVID-19. *PLoS ONE* **15**(10), e0240878 (2020).
47. Pizzuti, C., Socievole, A., Prasse, B. & Van Mieghem, P. Network-based prediction of COVID-19 epidemic spreading in Italy. *Appl. Netw. Sci.* **5**(1), 91 (2020).
48. Cui, Y., Ni, S. & Shen, S. A network-based model to explore the role of testing in the epidemiological control of the COVID-19 pandemic. *BMC Infect. Dis* **21**, 58 (2021).

Acknowledgements

Authors thank the Department of Biotechnology (No. BT/RLF/Re-entry/32/2017), Government of India for funding.

Author contributions

U.S. and B.G. conceived the study. U.S. built the epidemic model, developed model fitting and prediction strategies and performed optimization of country specific parameters. B.G. developed the agent-based model, performed the stochastic simulations, analytical calculations, genome sequence analysis and regression analysis. U.S. and B.G. analyzed the results and wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-03846-z>.

Correspondence and requests for materials should be addressed to U.S. or B.G.

Reprints and permissions information is available at www.nature.com/reprints.

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



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 Author(s) 2022