





OPEN Application and significance of SIRVB model in analyzing COVID-19 dynamics

Pavithra Ariyaratne, Lumbini P. Ramasinghe, Jonathan S. Ayyash, Tyler M. Kelley, Terry A. Plant-Collins, Logan W. Shinkle, Aoife M. Zuercher & Jixin Chen  

In the summer of 2024, COVID-19 positive cases spiked in many countries, but it is no longer a deadly pandemic thanks to global herd immunity to the SARS-CoV-2 viruses. In our physical chemistry lab in spring 2024, students practice kinetic models, SIR (Susceptible, Infected, and Recovered) and SIRV (Susceptible, Infected, Recovered, Vaccinated) using COVID-19 positive cases and vaccination data from World Health Organization (WHO). In this report, we further introduce virus breakthrough to the existing model updating it the SIRVB (Susceptible, Infectious, Recovered, Vaccinated, Breakthrough) model. We believe this is the simplest model possible to explain the COVID-19 kinetics/dynamics in all countries in the past four years. Parameters obtained from such practice correlate with many indices of different countries. These models and parameters have significant value to researchers and policymakers in predicting the stages of future outbreaks of infectious diseases.

Keywords SIR, SIRV, SIRVB, COVID kinetic models, Global data analysis, Undergraduate teaching

“Pandemic” would have been just a word in modern times, if we all had not experienced the true sense of it until COVID-19 happened. The novel coronavirus, caused by virus SARS-CoV-2, first appeared in Wuhan, China in December 2019¹. The World Health Organization (WHO) announced it a pandemic named COVID-19 in March 2020 and released the alarm in May 2023^{2,3}. During these three years and until this report in the summer of 2024, over 700 million people around the world have been infected with this virus, and over 7 million people died, with an unknown but estimated much larger number of excess deaths by WHO^{4–6}. Among those, over 1 million deaths in the USA⁷, indicating the impact of COVID-19 on the country, which spends the highest percentage on healthcare from GDP⁸. Therefore, it is an interesting study for students to explore how fast COVID-19 spread and recovery using kinetic models.

The COVID-19 pandemic data has been used in our physical chemistry teaching lab course since 2021. In early 2020, when the world was exposed to COVID-19, the kinetic models were used by epidemiologists to predict the impact of COVID-19, way before the world had realized it^{9,10}. The basic reproduction number, R_0 , meaning how many people can get infected by one infected person of COVID-19, was estimated ~ 3.0 in the early stage^{11,12}, higher than seasonal influenza, which is around $0.9–2.1$ ¹³. The incubation period of the SARS-CoV-2 virus was also higher than that of the common flu and society had no natural immunity as it was a novel virus type¹⁴. Until vaccines were invented and properly distributed, the epidemiologists suggested that the best way to slow down the spreading was by enforcing social distancing^{15,16}. Throughout the COVID-19 peak period, kinetic models were used as a main tool for decision making¹⁶. After 4 years since its first outbreak, positive cases are still reported worldwide daily¹⁷, with different variants emerging. Luckily, it is no longer considered deadly due to the cumulative herd immunity achieved by the world. Herd immunity is the point at which the amount of susceptible people is less than $1/R_0$ of the total population¹⁸. Vaccination and getting infected are the two ways to achieve herd immunity. We are interested in analyzing which country takes which way.

Our in-class practice of COVID-19 kinetics analysis began in 2021, it was introduced to teach students how to do kinetic model analysis using an Excel spreadsheet and enabled them to perform lab work remotely¹². Since then, SIR (Susceptible, Infectious, and Recovered) and SIRV (Susceptible, Infectious, Recovered, Vaccinated) models have been used. Kinetics is a special case of dynamics that is studying the quantity change of a system over time, mainly used in physics, chemistry, and engineering. The SIR model has been used in epidemiology as the standard deterministic epidemic model by government organizations such as U.S. CDC¹⁹. It was originally developed by biochemist William Ogilvy Kermack and physician and epidemiologist Anderson Grey McKendrick in 1927 as a meaning to use the simplest possible mathematical model to predict spreading

Department of Chemistry and Biochemistry, Nanoscale and Quantum Phenomena Institute, Ohio University, Athens, OH 45701, USA. ✉email: chenji@ohio.edu

speed²⁰. Modifications have been added over the past years²¹, such as the SIRV model²², SIS model²³, diffusion model^{24–26}. In the past three years, we have been performing this lab experiment, with each year exploring more data. Our first lab in 2021 spring focused on two states' data in the USA using the SIR model¹², and our second lab in 2022 focused on eight states in the USA using the SIR and SIRV model²⁷. Then our 3rd lab in 2023 focused on different countries across all continents²⁸. Our data analysis of SIR model is consistent with those reported. This year, in the spring 2024 lab, the COVID-19 data was downloaded from the Our World in Data website^{29,30}. At the beginning of the pandemic, SIR model is proper because there are small number of infected population and no vaccination. Then vaccination is introduced, and the number of recovered people and vaccinated people became significant and cannot be approximated to be negligible. Since the SIR and SIRV models showed difficulty in explaining the reproduction number in the later stages in our previous data analysis, we are introducing the SIRVB (Susceptible, Infectious, Recovered, Vaccinated, Breakthrough) model to analyze the COVID-19 data. Finding the next simplest model is an essential intention in research and education. It also has practical significance for policymakers to predict stages and to suggest proper social restrictions. The key achievement of these models is to obtain a central parameter R_t , real-time reproduction number, that is independent of testing and can be predicted from social parameters that are readily available³¹. We believe that the stringency index²⁴ and R_0 alone are enough to estimate R_t values in these models, providing a fast estimation of COVID-19 spreading speed.

Method

Using Microsoft Excel and SIR or SIRV model (Fig. 1a, b) with the forward Euler method to extract the real-time effective reproduction number R_E and real-time reproduction number R_t has been explained in our previous publications with a step-by-step instruction described^{12,27,28}. The SIRVB model used in this report has the same structure as the SIRV model with one additional infection pathway (Fig. 1c). The once immune people may still get infected with a breakthrough rate simplified to a single number b . This number is difficult to find and may change over time after the day when a person has recovered or been vaccinated³². The meanings of the parameters are shown in Fig. 1, where the boxes represent the number of people in each category, and the variables on the arrows represent the rate of transfer to another category. The discrete SIR and SIRV models have been detailed and explained in our previous publications^{12,27,28}. We are introducing the data analysis procedure of the Discrete SIRVB model (Fig. 1c) as follows. The analysis can be done in Microsoft Excel, which is relatively time-consuming and is suitable if a relatively larger class shares the load. In this report, we have coded MATLAB codes to speed up the process by automatizing the Excel process we have been doing before.

Briefly, from the raw data downloaded from Our World in Data²⁹, we selected the smoothed new tested positive cases (n) per million population data with an additional 20-day window Gaussian smooth, and fully vaccinated (V) per million population data to analyze the reproduction numbers over time. We set the discrete time step $\Delta t = 1$ day, whose effect on convergence is simulated to be properly small for a million people^{12,27,28}. Thus, the initial values are set to be total population $N = 1,000,000$, $S = N$, daily new case $n = 0$, active infectious population $I = 1$, recovered population $R = 0$, vaccinated population $V = 0$, infection rate constant $\beta = 0$ (day^{-1}), recovery rate constant $\gamma = 0.2$ (day^{-1}), vaccination rate constant $r_v = 0$ (day^{-1}), and breakthrough rate constant $b = 0.2$. N , γ , and b are simplified to be fixed values throughout the calculations, n and V are from the raw data, and the rest are calculated variables^{12,27,28}. The average breakthrough rate value of the $R + V$ population is difficult to find, although it is definitely smaller than $1/R_0$, otherwise, the pandemic would have never ended. As an example, we simplify the value to be fixed over time, $b = 20\%$ based on the literature values^{32–34}, and a few different values will be tested later.

These models resemble the logistics problem of self-pumped water/current flows and can be solved mathematically using various methods, among which we will use the Euler method. According to the Euler

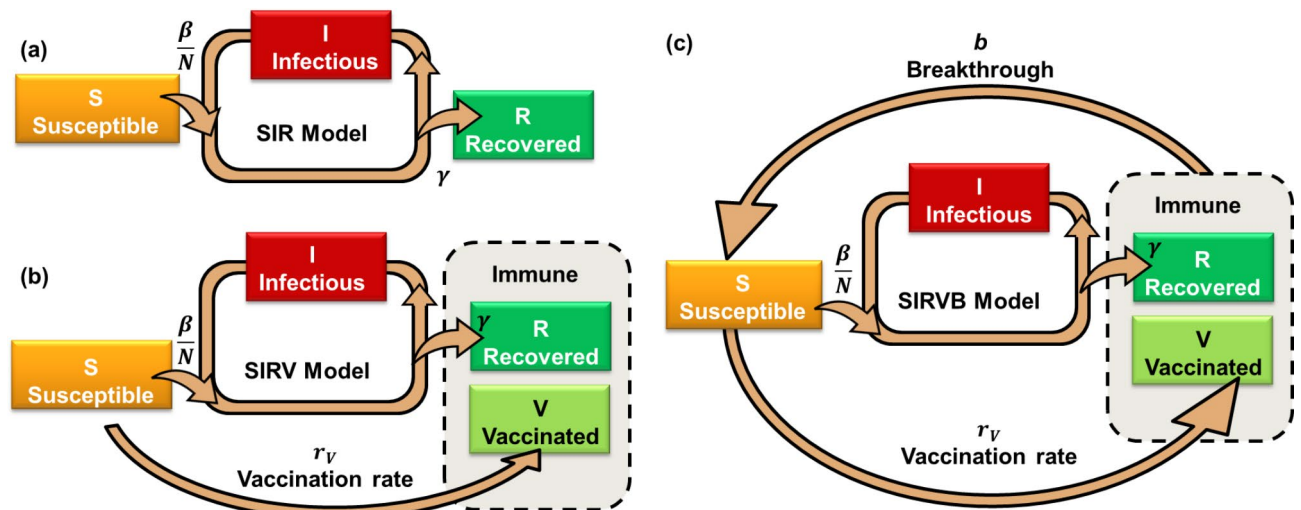


Fig. 1. Scheme of (a) SIR, (b) SIRV, and (c) SIRVB models.

forward method, at time $t = i$, the i th day after day 1 of choice from the raw data, on Jan 5th, 2020, the second-order “rate constant” of infection is calculated from the number of new cases n_i from the raw data.

$$\frac{\beta_i}{N} = \frac{n_i}{S_{eff,i-1} I_{i-1}} = \frac{n_i}{(S_{i-1} + b(R_{i-1} + V_{i-1})) I_{i-1}} \quad (1)$$

The daily fully vaccinated people v_i can be directly pulled out from the raw data or calculated from the accumulated cases of the fully vaccinated population V ,

$$v_i = V_i - V_{i-1} \quad (2)$$

Thus, each day, the susceptible population is reduced by the newly infected and vaccinated population. They are simplified to be no overlap in this report instead of an overlap assumed in the previous classes²⁸,

$$S_i = S_{i-1} - n_i - v_i \quad (3)$$

The infectious population gaining from daily infection and losing to daily recovery,

$$I_i = I_{i-1} + n_i - \gamma I_{i-1} \quad (4)$$

And the recovered population,

$$R_i = R_{i-1} + \gamma I_{i-1} \quad (5)$$

During the calculation, S and I are kept > 1 , and R is kept $< N - S - I - V$ to avoid zeros in the denominators.

Finally, the real-time effective reproduction number and the actual reproduction number are calculated,

$$R_{Ei} = \frac{R_{ti} S_i}{N} = \frac{n_i}{I_i \gamma} \quad (6)$$

$$R_{ti} = \frac{\beta_i}{\gamma} \quad (7)$$

These two curves are then smoothed using the moving median method (15 days) to exclude the spikes caused by irregular data collection.

The purpose of disease control is to reduce the effective reproduction number closer to 1 to flattening-the-curve (FTC)³⁵, or keep it < 1 long enough for the epidemic to fade. For a set of experimental data, R_E values are the same, and R_t values are dependent on the three models in Fig. 1. If no action is taken and no virus mutation, R_t is expected to remain constant throughout the pandemic. With social regulations, R_t is expected to be inversely proportional to social distance and thus, they carry very useful information³¹. It can be used to quantify the overall effect of government stringency and social restrictions. Thus, the calculated values in the SIRVB model are further analyzed and correlated to various social data and indices, to find the significance of the SIRVB model in predicting and evaluating COVID-19 kinetics.

Results and discussion

Assuming a basic reproduction number $R_0 = 3.0$ for COVID-19^{11,12}, the SIR model predicts that herd immunity will be achieved in 45 days with $1 - 1/R_0 = 67\%$ of the population infected and recovered if no restrictions are carried out in each million population (Fig. 2). The infection period can be roughly described as early stage/phase (left part in Fig. 2), middle stage (the sand-clock region), and later stage (right part). R_t remains constant in all three stages, similar to the rate constants in a chemical reaction, and R_E decreases over time. Herd immunity is reached at $R_E = 1$. This 2-month simulated lifetime is way faster than the > 3 -year pandemic. This is because social regulations have reduced the otherwise constant R_t thus the spreading is slowed down, a social practice named flattening-the-curve (FTC)³⁵. The significance and meanings of these parameters have been detailed in our previous publications with literature reviewed^{12,27,28}. In the COVID case, when social responses are triggered, both R_t and R_E are ~ 1 at the quasi-equilibrium stage of the pandemic.

Our models consistently pulled out the effective reproduction number R_E (SIR model) from the raw data Mathieu et al. published on Our World in Data, almost exactly the same but a little noisier than what they had calculated (Fig. 3)²⁹. However, the SIR model ignores the vaccination, and the SIRV ignores the breakthrough, thus both give unreasonable estimations of R_t at later stages of the pandemic, which are supposed to return to R_0 when social restrictions have been removed, as indicated by the social stringency index returning to zero after the pandemic²⁹. We have observed significant different curves between countries probably due to different social environments³⁶, social responses (stringency), and also data collection accuracy.

In these three models, the n and V values are fixed from the testing data and vaccination data reported to WHO. Fixed n values and the γ value yield the same I values for the three models and thus the same effective reproduction numbers R_E (Eq. 6). These values are consistent with what Mathieu et al. have calculated (Fig. 3a). The true I values are an epidemiological and statistical challenge to find. The consistency between our R_E values and those reported by Mathieu et al. suggested that we have chosen a similar γ value with a proper smooth method and analysis procedure/algorithm.

The three models, SIR, SIRV, and SIRVB, calculate the S and R populations very differently, yielding very different R_t values (Fig. 3a). The SIR and SIRV models have difficulties explaining the R_t values of the later stages,

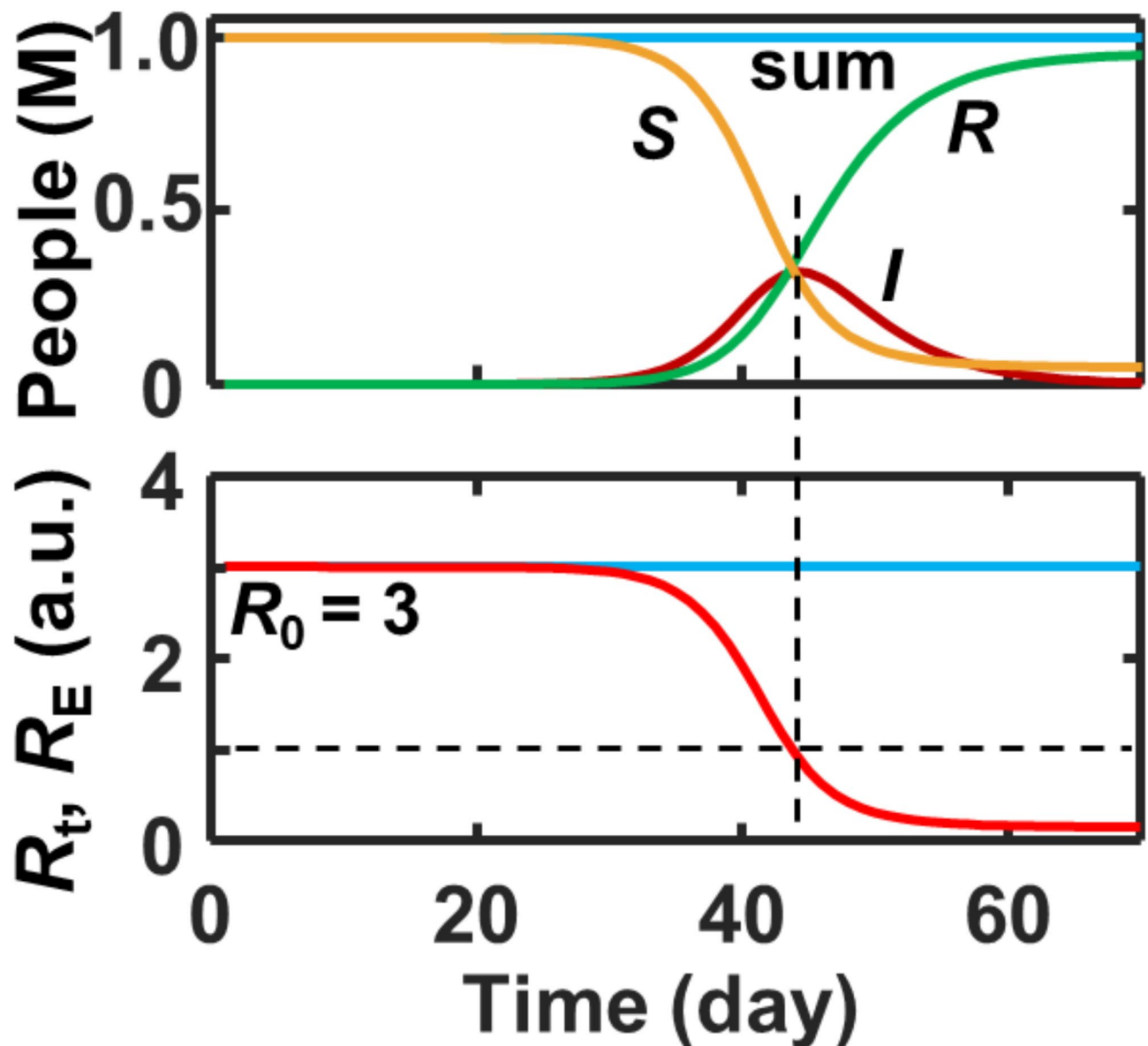


Fig. 2. Simulation of SIR model with $R_t = R_0 = 3.0$. Detail algorithm available in the literature and an Excel file running the simulation can be found in our previous publications^{12,27,28}.

namely 600–1600 days after the first day of the raw data (Jan 5th, 2020). During this period, social restrictions have been lifted and stringency has lowered in most countries. Thus, the R_t values are supposed to be similar to R_0 if the mutations have a similar R_0 or raised to a plateau if they have a larger R_0 .

The SIR model ignores the effect of vaccination in reducing susceptible populations, and typically limited testing may have lower estimated the I population, thus, it yields a R_t very close to R_E over the whole pandemic and after. We have little interest in using the SIR model to explain the raw data.

The SIRV model seems to work properly at first glance for the world average data with reasonable shape and a reasonable R_t plateau value ~ 4 , but it produces an unreasonably large R_t in the later stages for the countries, especially the developed countries that have a high vaccination rate (Fig. 3b). This is reasonable because, in the SIRV model, S can reach 0 and even negative if not restricted, especially when both vaccination and testing rates are high. R_t and β will go to infinity if $S = 0$. Our analysis in Fig. 3b has already restricted S and I to be > 1 and maintained the sum of $S, I, R, V = N$ but still has trouble. Thus, the SIRV model is not suitable for these countries in the late phase of the COVID-19 pandemic. In our past practice, we have assigned an overlap of the R and V population but the over-shooting of R_t still existed in the later stage.

The SIRVB model correctly reflects these expectations in both the world average and most individual countries. However, a dependence on incomes (Fig. 3c) and continents (Fig. 3d) is observed. Individual countries also show different R_t plateau values. Africa data do not have an uprising plateau in the later phase, which is particularly different from the other continents. We believe this is because of the low testing ratio in developing

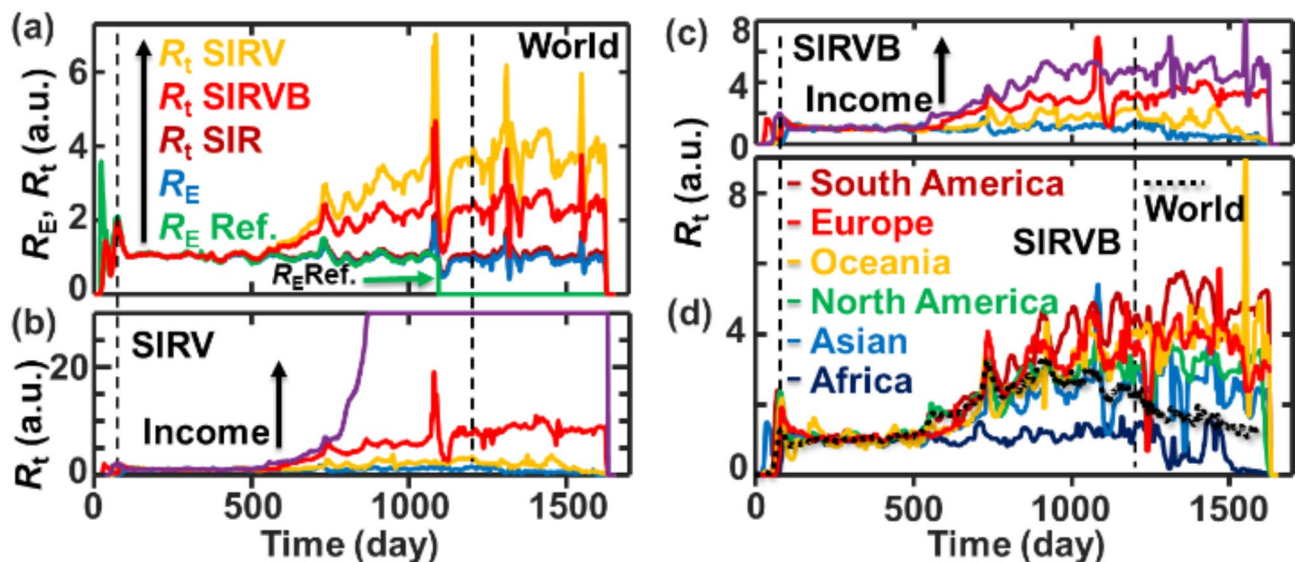


Fig. 3. Reproduction number analysis of the world average data grouped and summarized in the raw data by Mathieu et al.²⁹ (a) R_E and R_t values of different models for the world average data. The reference R_E values is taken from the raw data. R_t values of (b) SIRV model, (c) SIRVB model for the average of four groups of countries ordered from low to high incomes, and (d) SIRVB model for the countries in the six continents. Data are smoothed. The dashed lines are when WHO announced the start (March 11, 2020) and the end of the pandemic (May 5, 2023).

countries^{37,38}. Low statistics will directly affect the calculation of the R_t and R_E values. It has been reported that getting COVID data from some countries in Africa has been challenging³⁹, thus we suspect the difference between Africa and other continents is a low statistical problem of the raw data. The average R_t curve of all 242 countries listed in the raw data is consistently aligned with the average of the curves in Fig. 3c and d. The decay of R_t on the average curves after 900 days is because many countries stop reporting test results, especially after the pandemic. We assume that vaccination data are more accurate because it is easier to track than the other factors thus, we believe uncertainty in vaccination is not the major source of error in this model. The breakthrough b value potentially has a big effect on the model and thus is further analyzed.

By changing the average breakthrough rate b in the SIRVB model, the equilibrium R_t plateau can be adjusted (Figs. 3c and 4a–d). The average breakthrough rate of COVID-19 in immune people ($V + R$) must be $< 1/R_0$ for the pandemic to end. But $b > 0.3$ (30%) can be used in the SIRVB model. The plateau R_t values decrease over the increase of the b value, and at $b = 1$, R_t converges to R_E . Note, the SIRV model could have provided a reasonable plateau of R_t if we had factored in ratios of S , I , R , and V to their hidden true values and/or a sophisticated overlap model of different populations.

The SIRVB model is more straightforward than the SIRV model in that it simplifies all complicated factors into a single parameter b and maintains the other values unchanged. Because the protection/immune effect of both infected and vaccinated fades over time^{40–42}, we select $b = 0.2$, a 20% breakthrough rate in the next analysis based on estimations in the literature^{32–34}. In the SIRV model, the cumulative $(I^* \gamma)_{\text{cumu}} = R$. Introducing a breakthrough rate in the SIRVB model allows $(I^* \gamma)_{\text{cumu}} > R$ or even the total population N while maintaining $S + I + R + V = N$ over time. This $b = 0.2$ setting generates R_t values in 2–4 after 1000 days on the average data of the world and five continents (Fig. 3). These values are consistent with the estimations of $R_0 \sim 3$ in the literature¹¹, with no need to assume much bigger R_0 values for the later mutations of SARS-CoV-2. It has been challenging to find the R_0 values for different virus strings^{11,43}. Thus, keeping R_0 the same in a model can be strategically beneficial until solid data are obtained to update its value.

To test the significance of the SIRVB model, we extracted several parameters for each country from the raw data as well as other reports on the Our World in Data website and compared them to the results obtained from the SIRVB model⁴⁴.

First, R_t can be used to confirm/detect herd immunity status with high-quality global or local data, e.g. one city. This is because unlike R_E , which changes over time even if no action is taken, R_t is a constant only dependent on the nature of the virus and the intensity of social interactions⁴⁵. Thus, when social stringency is back to normal, R_t should be close to R_0 , which can be used to back-calculate the true immune populations and judge or predict the herd immunity status. For after-pandemic analysis of the raw data, we can simply use the testing data and find the turning point of R_t at the beginning of the later stage R_t plateau, which is ~ 900 days for the data shown in Figs. 3 and 4. We propose to use this date as the day of herd immunity i.e. $\sim 67\%$ of the population is truly immune to SARS-CoV-2 at the middle point of the sand clock region in Fig. 2.

The consistency of this assignment with other parameters is found in many countries, e.g., for France data (Fig. 5a). R_t is calculated from the raw data (Fig. 5a), and the herd immunity time of France is assigned using R_t at ~ 800 days (Fig. 5b). This is significantly earlier than the 900 days of the world data. This earliness makes

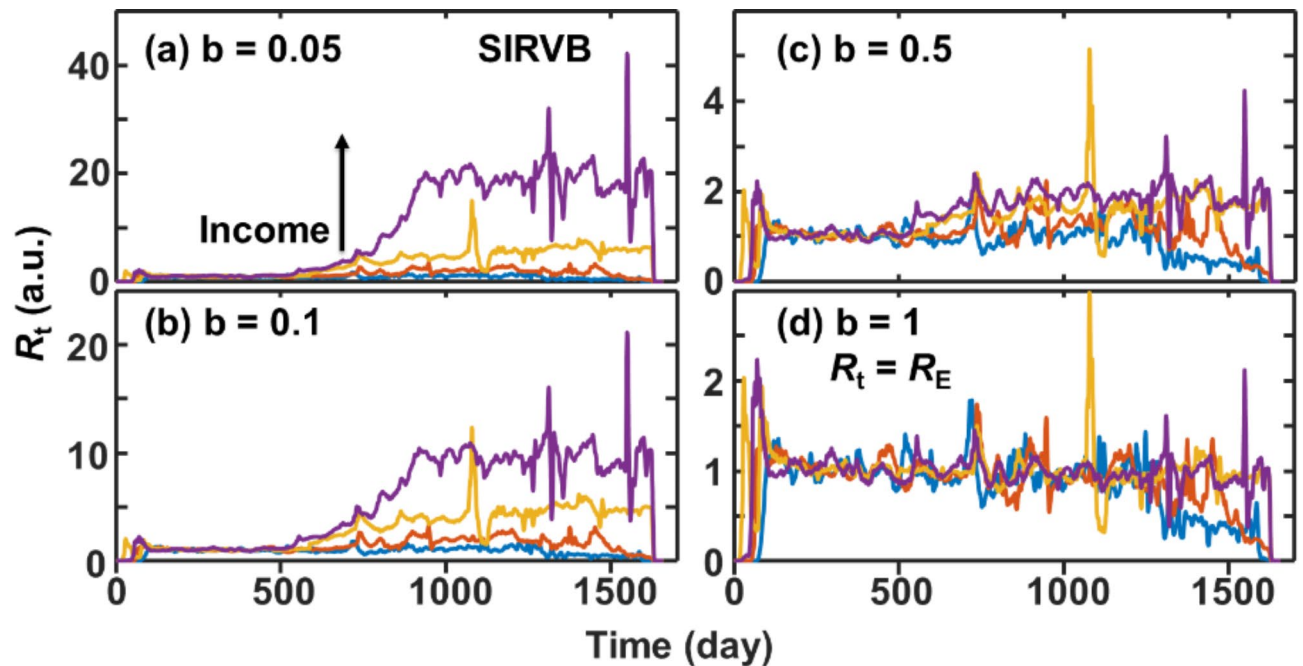


Fig. 4. (a–d) The effect of breakthrough rate on the calculations of R_t values in the SIRVB model. The four groups of countries in the raw data with low to high incomes are used as examples. Curves of $b = 0.2$ have been shown in Fig. 3c.

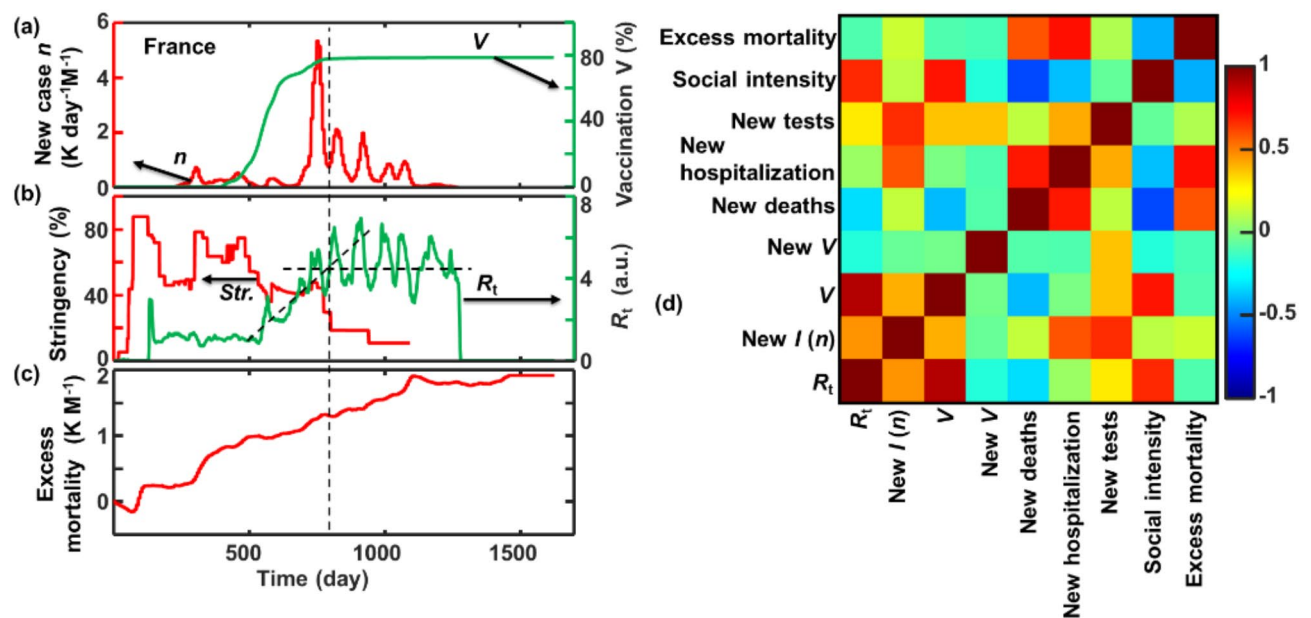


Fig. 5. Example data analysis of France. (a) Raw data of n and V . (b) Social stringency estimated by Mathieu et al.²⁹ in the raw data and R_t calculated in this report. Dashed lines indicate our estimation of the herd immune date. (c) Excess mortality estimated in the raw data. (d) Correlation of R_t with other time dependent parameters for COVID-19 in days 1–1200. If a curve goes up with the increase of another curve, the correlation will be positive, 0 if no correlation, and negative if goes down with the increase of another curve. Please see the supporting information for the detailed data tables.

sense because of the fast and high vaccination rate of France compared to the rest of the world, reaching ~ 80% at this time (Fig. 5a). It is also consistent with the time of the biggest wave in France (Fig. 5a), right before the government relieved social restrictions, and ~ 65% of total excess mortality due to COVID-19, which is estimated in the raw data by Mathieu et al.²⁹.

Zoom in France's data and correlate R_t with other time-resolved raw data, we can also find reasonable correlations and anti-correlations of the evolution of R_t with other time-resolved data e.g. excess mortality (Fig. 5c, d). R_t is calculated from n and V , and a positive correlation is observed for the France data. R_t has a stronger correlation with V because social recovery is timely correlated with vaccination in France (Fig. 5a, b). R_t has low to no correlation with new vaccinations, new deaths, new hospitalizations, and new tests. These low correlations make sense because these parameters are indirectly affect social behavior and may have a delay effect, thus the simple correlation calculation cannot correctly reflect their true correlations.

The positive correlations between R_t , n (new daily I), V , new tests, and the social intensity obtained (inversed) from stringency in the raw data estimated by Mathieu et al.²⁹, beautifully explain the flattening-the-curve practice. The same data can be seen from the anti-correlation between R_t and stringency in Fig. 5b. If a community halves social intensity/frequency/"temperature" by reducing activities or increasing social distance, it halves the disease transmission and thus halves the R_t . The weak correlations of R_t with other parameters, especially new V and new deaths are good signs because R_t is a constant in the SIRVB model if there are no social intensity changes. Because of the large R_0 value of COVID-19, it takes less than two months for it to infect the whole population at normal social intensity. Thus, assigning the turning point of the R_t curve as the herd-immunity date is a reasonable choice.

The other significant correlations or anti-correlations in France data (Fig. 5d) can be explained. The positive correlations between n , new hospitalization, and new tests are expected. The correlations of V with many parameters are due to the time correlation not because of causality correlation as indicated by the weak correlation of the new V (differential) column. New deaths, new hospitalizations, and new tests are correlated with each other, which are expected. They are also expected to be anti-correlated with social intensity when people adjust their social activities with the correct information of the spreading status. New tests have no effect on true R_t in the model, but it changes the value of R_t and social behaviors, so its positive correlations with R_t and other parameters are observed. But its weak to no anticorrelation with social intensity is surprising. Even if we expect a stronger anticorrelation, this weakening could be due to the wave structure of the new cases and a delay in social responses. Finally, in this figure, the excess mortality is positively correlated with n , new death, new hospitalization, and new tests, and anticorrelated with social intensity, which is very reasonable and expected. From the above analysis, we can conclude that the France data is statistically significant and well-behaved in the framework of the SIRVB model.

The R_t curves of the world average and France both have the laydown '5' shape so do many other countries. Thus, we use this shape to fit data from all countries using a stochastic fitting algorithm named "jcfir", which we have previously developed to show better efficiency than gradient descent methods for wave fitting (Fig. 6a)⁴⁶. In countries with high testing and vaccination rates, the summation of the infected and vaccinated population (assuming no overlap for simplicity of the model) has exceeded the total population, for example, France. Thus, the R_t in the later phase is controlled by the breakthrough factor at $R_t \approx 1/b$. The turning point occurs on the 802 days. The world average contains data from countries that have low testing and vaccination rates, thus all $R_t < 1/b$. The global turning point (t_4) is observed on the 940 day (Aug. 2, 2022), 276 days before WHO announced the end of the pandemic⁴⁷ (May 5, 2023), about the average time between the last three big global waves. The slope between t_3 and t_4 reflects the social recovery after the most significant peak of COVID-19 in the area and

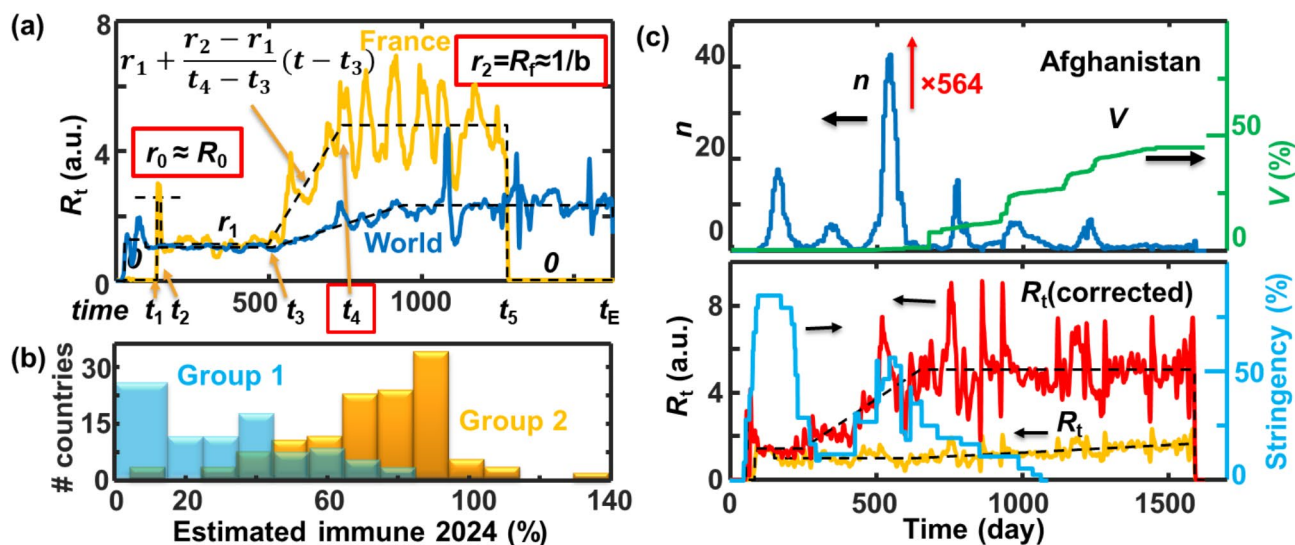


Fig. 6. (a) Example fitting of R_t curve using the laydown '5' shape for France and world average. (b) Histogram of Group 1 countries that do not follow (a) and Group 2 countries that follow the laydown '5' pattern in (a) with today's immune ratios. (c) Example data treatment of Group 1 countries with low COVID-19 testing rates, briefly, multiplying a testing rate correction factor estimated using world average timing and recalculating the R_t values. Please see the supporting information for the detailed data tables.

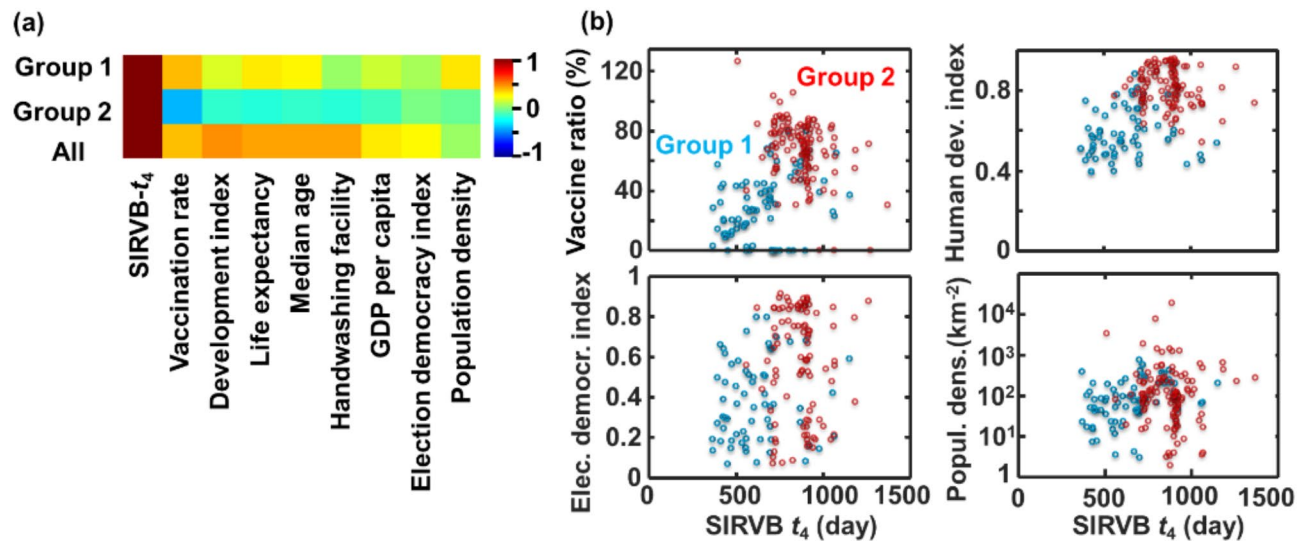


Fig. 7. (a) Correlation of the turning time t_4 in the SIRVB model with parameters of countries from Our World in Data. (b) Example plots of t_4 with four parameters. Full data available in the SI.

is correlated with the release of stringency. We assign Aug. 2, 2022, the global herd immunity day on the SIRVB model.

In the next step, we manually check the R_t curves of the 254 records in the raw data to group them into three groups: 87 countries that do not follow (Group 1), 121 countries follow the laydown '5' pattern (Group 2), and the rest, being the averages of groups of countries, or countries missing a significant amount of data.

Simplifying the total immune population as cumulated n times γ plus cumulative fully vaccinated population,

$$\text{Immune}_t = R_t + V_t \approx \sum_{i=1}^t n_i + V_t \quad (8)$$

The 87 Group 1 countries have a relatively low average immune population (0.30 ± 0.20) compared to that of the Group 2 countries (0.73 ± 0.21) at the end date of the raw data accessed in July 2024 (Fig. 6b). These values are correlated with GDP per capita, which makes sense because both testing and vaccination cost money. The average GDP per capita of Group 1 countries is \$6000, and Group 2 countries is \$20,000. Many Group 1 countries have relatively low testing rates, thus a smaller R_t at the later phase in the SIRVB model compared to Group 2 countries. Some countries also have low vaccination rates. To simplify our analysis, we assume that at the global herd immunity day, these countries also reach herd immunity with an assumed $\sim 67\%$ immune population. We assume a constant correction factor to be,

$$F_c = \frac{N \left(1 - \frac{1}{R_0}\right) - V_{940}}{\gamma \sum_{i=1}^{940} n_i} \quad (9)$$

And the corrected new case per day is

$$n_{ci} = F_c n_i \quad (10)$$

The average correction factor for the 87 Group 1 countries is 775 (1-9000), suggesting a very low testing ratio for many countries.

After the daily case correction, the R_t values are recalculated, and a laydown 5 shape is observed for most of the Group 1 countries. For example, the first country on the list, Afghanistan, has a low detected cases per day per million people, peaking ~ 40 (Fig. 6c). It also has relatively low vaccination at $\sim 20\%$ around day 940. Thus, its R_t is almost flat over time. This data suggests it is still at an early phase till the last day of the raw data in the SIRVB model, which is unlikely true. We apply a correction factor of 564 to the data and a new R_t curve showing a laydown 5 shape. Its trend anti-correlates with the stringency index very well. The herd-immunity date for Afghanistan is re-defined to the 620th day on the revised R_t curve, right after the last major social restriction peak and the most intensive outbreak peak (Fig. 6c).

After we have re-defined the herd immunity date for the Group 1 countries in the model, we correlate these dates with other social parameters to check for a pattern among them (Fig. 7). We leave the Group 2 countries with low testing ratio unchanged, resulting in some having a later herd-immunity date than the rest of the countries. The average day to reach herd immunity (SIRVB model) for Group 1, mostly developing countries, is 655 ± 184 days, and for Group 2, including most developed countries is 873 ± 134 days. After analyzing the correlations of the herd-immunity days with other parameters (Fig. 7a), we suspect this difference is due to the

two different strategies of achieving herd immunity, mainly through infection vs. mainly through vaccination. Also, many developing countries have lower median ages⁴⁸. Because COVID-19 is less toxic to young people⁴⁹, choosing infection to gain herd-immunity for young people is a natural choice.

A few correlations are further expanded in Fig. 7b. Our assigned herd-immunity time (t_4 in Fig. 6) is positively correlated with vaccination for Group 1 and all countries. This correlation makes sense because I and V determine t_4 in the SIRVB model. Interestingly, a negative correlation is observed within Group 2 countries. We suspect it to be the cost of time to wait for the vaccination to be available. Group 1 countries have significantly shorter time than Group 2 because they achieve herd-immunity mainly via infection with an average median age of 24 ± 8 ; thus, there is less need to wait for the vaccines as Group 2 countries with an average median age of 36 ± 7 . Average population density should not matter because it is the local or median population density that determines the R_0 and R_t values. Thus, a significant correlation between this parameter and the herd-immunity day is not observed for all countries. The median age, life expediency, handwashing facility, and election democracy index all contribute to the calculation of the human development index. Thus, their distribution vs. t_4 shows a positive correlation with a binary pattern between the two groups.

Conclusions

We believe the SIRVB model is the simplest mathematical model to explain the COVID-19 kinetic data of all countries in Our World in Data (raw data) from the World Health Organization (WHO). The calculated R_t values from the model are strongly anti-correlated with the stringency index estimated in the raw data. Thus, it is possible to calculate the R_t values of the SIRVB model just using stringency-index and R_0 in a future outbreak of an epidemic or pandemic, providing valuable prediction power to researchers and policymakers, especially in regions where testing is rare.

In summary, there is an epidemiological significance of SIR, SIRV, SIRVB models, and numerous modifications, and choosing a proper model and finding out the proper parameters could be essential in predicting trends and suggesting policy changes for a future epidemic or pandemic. A modification of the SIRVB model allows the adjustment of low-statistics data to obtain otherwise hidden R_t values, taking advantage of parameters fitted from the data of high-statistics regions. Thus, the central disease monitoring department, such as WHO or the disease control department of a country can have an estimation of the stages of the spreading in low statistics regions. At a super-simple condition, when people do nothing, the kinetics could follow the SIR model, and it could be easy to predict its dynamics. The super complicated social response is all factored in our simple model of a few parameters such as the time-dependent infection rate, the vaccination rate, and the breakthrough rate. Each parameter has its own strategy of social response to change, thus these simple models provide valuable guidance to properly fight the outbreak.

A few weak points of the simplified SIRVB model may need further improvement to a more sophisticated model. The post-pandemic data analysis we have carried out is a mathematical practice based on naïve assumptions rather than a restricted epidemiological study. We believe the parameters of the SIRVB model can be obtained experimentally during a pandemic and used to predict its trend. Due to our limited knowledge of epidemiology, we will leave this task to others who are more suitable. The model does not have an internal overlap of populations; neither does the model have imported cases. Time-dependent variables such as γ and b are assumed constants. Estimation of testing and vaccination rates have significant uncertainties. Also, the model treats the whole entity, either a city, a state, a country, or the whole world, uniformly. These assumptions introduce errors in R_t calculation, which is also part of why the after-pandemic R_t approaches $1/b$ rather than (should be) R_0 . The number of susceptible populations will be estimated to be 0 at a later phase, and all cases are breakthrough which is not true. Thus, any tested positive case, even 1 in the whole population, will give a calculated $R_t = 1/b$ losing its significance.

Data availability

All raw data is downloaded from Our World in Data that is publicly accessible. Data in figures are also included in the supporting information. MATLAB (2024a) source code will be available upon request from J.C. Thus, all results in this paper should be reproducible.

Received: 11 October 2024; Accepted: 11 February 2025

Published online: 12 March 2025

References

- Li, J., Lai, S., Gao, G. F. & Shi, W. The emergence, genomic diversity and global spread of SARS-CoV-2. *Nature* **600**(7889), 408–418. <https://doi.org/10.1038/s41586-021-04188-6> (2021).
- WHO Director-General's opening remarks at the media briefing—5 May. <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing---5-may-2023>. Accessed 30 July 2024 (2023).
- Northwestern Medicine. COVID-19 Pandemic Timeline. <https://www.nm.org/healthbeat/medical-advances/new-therapies-and-drug-trials/covid-19-pandemic-timeline>. Accessed 30 July 2024.
- 14.9 million excess deaths associated with the COVID-19 pandemic in 2020 and 2021. <https://www.who.int/news/item/05-05-2022-14.9-million-excess-deaths-were-associated-with-the-covid-19-pandemic-in-2020-and-2021>. Accessed 05 Aug 2024.
- Global excess deaths associated with COVID-19 (modelled estimates). <https://www.who.int/data/sets/global-excess-deaths-associated-with-covid-19-modelled-estimates>. Accessed 07 Aug 2024.
- COVID-19 cases | WHO COVID-19 dashboard. datadot. <https://data.who.int/dashboards/covid19/cases>. Accessed 09 Aug 2024.
- Cumulative confirmed COVID-19 cases and deaths. Our World in Data. <https://ourworldindata.org/grapher/cumulative-deaths-and-cases-covid-19?country=~USA>. Accessed 08 Aug 2024.
- Government health expenditure as a share of GDP. Our World in Data. <https://ourworldindata.org/grapher/public-health-expenditure-share-gdp>. Accessed 08 Aug 2024.

9. Tirnakli, U. & Tsallis, C. Epidemiological model with anomalous kinetics: Early stages of the COVID-19 pandemic. *Front. Phys.* **8**, 613168. <https://doi.org/10.3389/fphy.2020.613168> (2020).
10. *Mathematical Models of Diseases—The Essential Elements Fall 2021* | Stockton University. <https://stockton.edu/sciences-math/ezin/fall2021/mathematical-models.html>. Accessed 09 Aug 2024.
11. D'Arenzo, M. & Coniglio, A. Assessment of the SARS-CoV-2 basic reproduction number, R_0 , based on the early phase of COVID-19 outbreak in Italy. *Biosaf. Health* **2**(2), 57–59. <https://doi.org/10.1016/j.bsheat.2020.03.004> (2020).
12. Smith, D. K. et al. Teaching undergraduate physical chemistry lab with kinetic analysis of COVID-19 in the United States. *J. Chem. Educ.* **99**(10), 3471–3477. <https://doi.org/10.1021/acs.jchemed.2c00416> (2022).
13. Eisenberg, J. *R0: How Scientists Quantify the Intensity of an Outbreak Like Coronavirus and Its Pandemic Potential* | The Pursuit | University of Michigan School of Public Health | Coronavirus | Pandemic. <https://sph.umich.edu/pursuit/2020posts/how-scientists-quantify-outbreaks.html>. Accessed 09 Aug 2024.
14. CDC. *Similarities and Differences between Flu and COVID-19*. Centers for Disease Control and Prevention. <https://www.cdc.gov/flu/symptoms/flu-vs-covid19.htm>. Accessed 09 Aug 2024.
15. Qureshi, A. I., Suri, M. F. K., Chu, H., Suri, H. K. & Suri, A. K. Early mandated social distancing is a strong predictor of reduction in Peak Daily New COVID-19 cases. *Public. Health* **190**, 160–167. <https://doi.org/10.1016/j.puhe.2020.10.015> (2021).
16. Ghosh, A., Roy, S., Mondal, H., Biswas, S. & Bose, R. Mathematical modelling for decision making of lockdown during COVID-19. *Appl. Intell.* **52**(1), 699–715. <https://doi.org/10.1007/s10489-021-02463-7> (2022).
17. *What to Know About COVID FLIRT Variants* | Johns Hopkins | Bloomberg School of Public Health. <https://publichealth.jhu.edu/2024/what-to-know-about-covid-flirt-variants>. Accessed 09 Aug 2024.
18. Worth, T. *Herd Immunity*. WebMD. <https://www.webmd.com/covid/what-is-herd-immunity>. Accessed 09 Aug 2024.
19. CDC. *Current Epidemic Trends (Based on Rt) for States*. CFA: Modeling and Forecasting. <https://www.cdc.gov/cfa-modeling-and-forecasting/rt-estimates/index.html>. Accessed 11 Dec 2024.
20. Kermack, W. O. & McKendrick, A. G. Contributions to the mathematical theory of epidemics. *Proc. R. Soc. Lond. Ser. Contain. Pap. Math. Phys. Charact.* **115**(772), 700–721. <https://doi.org/10.1098/rspa.1927.0118> (1927).
21. Zanella, M. Kinetic models for epidemic dynamics in the presence of opinion polarization. *Bull. Math. Biol.* **85**(5). <https://doi.org/10.1007/s11538-023-01147-2> (2023).
22. Schlickeiser, R. & Kröger, M. Analytical modeling of the temporal evolution of epidemics outbreaks accounting for vaccinations. *Physics* **3**(2), 386–426. <https://doi.org/10.3390/physics3020028> (2021).
23. Djilali, S., Bentout, S. & Tridane, A. Dynamics of a generalized nonlocal dispersion SIS epidemic model. *J. Evol. Equ.* **24**(4), 83. <https://doi.org/10.1007/s00028-024-01013-1> (2024).
24. Bentout, S. Analysis of global behavior in an age-structured epidemic model with nonlocal dispersal and distributed Delay. *Math. Methods Appl. Sci.* **47**(9), 7219–7242. <https://doi.org/10.1002/mma.9969> (2024).
25. Mahroug, F. & Bentout, S. Dynamics of a diffusion dispersal viral epidemic model with age infection in a spatially heterogeneous environment with general nonlinear function. *Math. Methods Appl. Sci.* **46**(14), 14983–15010. <https://doi.org/10.1002/mma.9357> (2023).
26. Djilali, S., Bentout, S. & Chen, Y. Dynamics of a delayed nonlocal reaction–diffusion Heroin Epidemic Model in a heterogenous environment. *Math. Methods Appl. Sci.* <https://doi.org/10.1002/mma.10327> (2024).
27. Hart, K. D. et al. Remote learning of COVID-19 kinetic analysis in a physical chemistry laboratory class. *ACS Omega* **6**(43), 29223–29232. <https://doi.org/10.1021/acsomega.1c04842> (2021).
28. Athapaththu, D. V. et al. Physical chemistry lab for data analysis of COVID-19 spreading kinetics in different countries. *J. Chem. Educ.* **101**(7), 2892–2898. <https://doi.org/10.1021/acs.jchemed.4c00015> (2024).
29. Mathieu, E. et al. Coronavirus Pandemic (COVID-19). 2020 OurWorldinData.org. <https://ourworldindata.org/coronavirus>. Accessed 09 Aug 2024.
30. Mathieu, E. et al. A global database of COVID-19 vaccinations. *Nat. Hum. Behav.* **5**(7), 947–953. <https://doi.org/10.1038/s41562-021-01122-8> (2021).
31. *Current Epidemic Growth Status (Based on Rt) for States and Territories*. <https://www.cdc.gov/forecast-outbreak-analytics/about/rt-estimates.html>. Accessed 09 Aug 2024.
32. Amanatidou, E. et al. Breakthrough infections after COVID-19 vaccination: Insights, perspectives and challenges. *Metab. Open* **14**, 100180. <https://doi.org/10.1016/j.metop.2022.100180> (2022).
33. Gupta, R. K. & Topol, E. J. COVID-19 vaccine breakthrough infections. *Science* **374**(6575), 1561–1562. <https://doi.org/10.1126/science.abl8487> (2021).
34. Birhane, M. et al. COVID-19 vaccine breakthrough infections reported to CDC—United States, January 1–April 30, 2021. *Morb. Mortal. Wkly. Rep.* **70**(21), 792–793. <https://doi.org/10.15585/mmwr.mm7021e3> (2021).
35. Mat Daud, A. A. Five common misconceptions regarding flattening-the-curve of COVID-19. *Hist. Philos. Life Sci.* **44**(3), 41. <https://doi.org/10.1007/s40656-022-00522-x> (2022).
36. Djilali, S., Bentout, S., Kumar, S. & Touaoula, T. M. Approximating the asymptomatic infectious cases of the COVID-19 disease in Algeria and India using a mathematical model. *Int. J. Model. Simul. Sci. Comput.* **13**(04), 2250028. <https://doi.org/10.1142/S1793962322500283> (2022).
37. Shams, S. A., Haleem, A., Javaid, M. & Analyzing COVID-19 pandemic for unequal distribution of tests, identified cases, deaths, and fatality rates in the top 18 countries. *Diabetes Metab. Syndr.* **14**(5), 953–961. <https://doi.org/10.1016/j.dsx.2020.06.051> (2020).
38. Giri, A. K. & Rana, D. R. Charting the challenges behind the testing of COVID-19 in developing countries: Nepal as a case study. *Biosaf. Health* **2**(2), 53. <https://doi.org/10.1016/j.bsheat.2020.05.002> (2020).
39. Lewis, H. C. et al. SARS-CoV-2 infection in Africa: A systematic review and meta-analysis of standardised Seroprevalence studies, from January 2020 to December 2021. *BMJ Glob. Health* **7**(8), e008793. <https://doi.org/10.1136/bmjgh-2022-008793> (2022).
40. *Study shows effectiveness of updated COVID-19 vaccines wanes moderately over time, is lower against currently circulating variants*. UNC Gillings School of Global Public Health. <https://sph.unc.edu/sph-news/study-shows-effectiveness-of-updated-covid-19-vaccines-wanes-moderately-over-time-is-lower-against-currently-circulating-variants/>. Accessed 09 Aug 2024.
41. *Had COVID recently? Here's what to know about how long immunity lasts, long COVID, and more*. AAMC. <https://www.aamc.org/news/had-covid-recently-here-s-what-know-about-how-long-immunity-lasts-long-covid-and-more>. Accessed 09 Aug 2024.
42. Bates, T. A. et al. The time between vaccination and infection impacts immunity against SARS-CoV-2 variants. *medRxiv*. <https://doi.org/10.1101/2023.01.02.23284120>.
43. Karimizadeh, Z., Dowran, R., Mokhtari-azad, T. & Shafiei-Jandaghi, N. Z. The reproduction rate of severe acute respiratory syndrome coronavirus 2 different variants recently circulated in human: A narrative review. *Eur. J. Med. Res.* **28**(1), 94. <https://doi.org/10.1186/s40001-023-01047-0> (2023).
44. Herre, B., Rodés-Guirao, L. & Ortiz-Ospina, E. *Democracy index*. Our World in Data. <https://ourworldindata.org/grapher/democracy-index-eiu>. Accessed 07 Aug 2024.
45. Gostic, K. M. et al. Practical considerations for measuring the effective reproductive number, R_t . *PLoS Comput. Biol.* **16**(12), e1008409. <https://doi.org/10.1371/journal.pcbi.1008409> (2020).
46. Chen, J. Structured stochastic curve fitting without gradient calculation. *J. Comput. Math. Data Sci.* **12**, 100097. <https://doi.org/10.1016/j.jcmds.2024.100097> (2024).
47. Cheng, K. et al. Declares the end of the COVID-19 Global Health Emergency: Lessons and recommendations from the perspective of ChatGPT/GPT-4. *Int. J. Surg.* **109**(9), 2859–2862. <https://doi.org/10.1097/JS9.0000000000000521> (2023).

48. Median Age by Country 2024. <https://worldpopulationreview.com/country-rankings/median-age>. Accessed 09 Aug 2024.
49. Sara Berg, M. S. What doctors wish patients knew now about COVID-19 risk and age. *Am. Med. Assoc.*, <https://www.ama-assn.org/delivering-care/public-health/what-doctors-wish-patients-knew-now-about-covid-19-risk-and-age>. Accessed 09 Aug 2024.

Acknowledgements

We thank the National Human Genome Research Institute of the National Institutes of Health (award number 2R15HG009972) for supporting our teaching practices. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author contributions

J.C. designed the experiments; all authors analyzed data; P.A., L.R., and J.C. wrote and all authors edited the paper.

Declarations

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/s41598-025-90260-4>.

Correspondence and requests for materials should be addressed to J.C.

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-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

© The Author(s) 2025