# **Epidemic forecast follies**

Check for updates

## P. L. Krapivsky<sup>1,2</sup> & S. Redner<sup>2</sup>

We introduce a simple multiplicative model to describe the temporal behavior and the ultimate outcome of an epidemic. Our model accounts, in a minimalist way, for the competing influences of imposing public-health restrictions when the epidemic is severe, and relaxing restrictions when the epidemic is waning. Our primary results are that different instances of an epidemic with identical starting points have disparate outcomes and each epidemic temporal history is strongly fluctuating.

Now that the most severe (we hope) manifestations of the Covid-19 epidemic have passed, one can't help but realize that many of the early forecasts of the Covid-19 epidemic toll were wildly inaccurate and inconsistent with each other. Moreover, individual forecasts could change dramatically over a period of few days. For the USA, in particular, the earliest estimates for the Covid-19 epidemic death toll ranged from tens of thousands to many millions, with the current death toll (as of September 2023) reported to be 1.175 million out of a total of 108.5 million cases (all data taken from ref. 1). Perhaps even more striking are the huge fluctuations and the dramatically different time courses in the daily death rate in different countries.

To illustrate these statements, Fig. 1 plots the reported daily death rates for the six countries in the world with populations greater than 60 million and with the largest total death rates. They are: USA (3.507 deaths/1000), UK (3.339/1000), Brazil (3.275/1000), Italy (3.174/1000), Russia (2.743/1000), and France (2.556/1000). For reference, the country with the largest reported total death rate is Peru (6.582/1000), while the world average is (0.887/1000). For many reasons, the accuracy of the data may vary widely from country to country so that some of the numbers reported in ref. 1, such as the suspicious smoothness of the data for Russia, should be interpreted with caution.

One of the many confounding features of Covid-19 is asymptomatic transmission, in which the epidemic may be unknowingly spread by individuals who did not know that they were contagious. Partly because of this feature, a wide variety of increasingly sophisticated multi-compartment models were developed that build on the classic SIR and SIS models of epidemic spread. These models typically attempted to faithfully account for subpopulations in various stages of the disease and recovery, as well as the transitions between these stages. Models of this type gave rise to complex dynamical behaviors that could sometimes mirror reality in a specific setting or over a limited time range. However, embellishments of SIR and SIS-type models still seem to be incomplete because of the difficulty in simultaneously accounting for both the disease dynamics and its interaction with social forces.

The discrepancy between the observed wildly varying features of Covid-19 and supposedly deterministic outcomes of SIR and SIS models is especially striking. In fact, the determinism of the SIR and SIS models is actually illusory. The SIR model, for example, is an inherently stochastic process<sup>2,3</sup> that is characterized by the reproductive number  $R_0$ . This quantity is defined as the average number of individuals to whom a single infected individual transmits the infection before this single individual recovers. In the supercritical regime,  $R_0 > 1$ , it is possible that the outbreak may quickly die out. This happy event occurs with probability  $R_0^{-1}$  if one individual was initially infected. Otherwise, the infection quickly spreads, and the behavior becomes effectively deterministic because the distribution of the epidemic size becomes narrow. In this case, a finite fraction  $c = c(R_0)$  individuals catch the disease, with *c* implicitly determined by the criterion  $c + e^{-cR_0} = 1^4$ .

Conversely, if  $R_0 < 1$ , the outbreak quickly dies out, so while the subcritical SIR process is still manifestly stochastic, it is not a threat to the population at large. The interesting and the most strongly stochastic behavior emerges in critical SIR and SIS models<sup>5–16</sup>. For the SIR model, in particular, the distribution of the number of infected individuals has a power-law tail. For a finite population of size *N*, the critical SIR model does not lead to a pandemic, because the average number of individuals who contract the disease scales as  $N^{1/3}$ .

We argue here that significant forecasting uncertainties are an integral feature of processes caused by the interplay between the dynamics of the disease transmission and the social forces that arise in response to the epidemic. Each attribute alone typically leads to either exponential growth (due to disease transmission at early times) or to exponential decay (due to effective mitigation strategies). Within our model, the competition between these two exponential processes leads to a dynamics that is extremely sensitive to seemingly minor details.

While a variety of models have been proposed to incorporate these competing effects and to understand how they give rise to significant uncertainties in the outcome of an epidemic<sup>17–19</sup>, here we present a different perspective to account for forecasting uncertainties. Our approach is based on mimicking the inherent stochasticity in the development of epidemics through a stochastic dynamics in the reproductive number  $R_0$ . The basic mechanism in our modeling is that  $R_0$  can sometimes decrease, due to the imposition of public-health measures, such as social distancing, vaccinations, etc., and sometimes increase, because of the relaxation of these measures. Focusing only on the dynamics of the reproductive number serves as a useful proxy for the myriad of influences that control the true epidemic dynamics. The central variable in our model is the number of newly infected

<sup>&</sup>lt;sup>1</sup>Department of Physics, Boston University, Boston, MA 02215, USA. <sup>2</sup>Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA. 🖂 e-mail: redner@santafe.edu



Fig. 1 | Covid death rates. The reported daily Covid death rates (7-day moving average) for (a) the USA, (b) UK, (c) Brazil, (d) Italy, (e) Russia, and (f) France. These data cover the period from Feb. 15, 2020 until July 29, 2023 and are all taken from ref. 1.

individuals in each incubation period. Within this framework, we will determine the duration of an epidemic, the time dependence of the number of infected individuals, and the total number of individuals infected when an epidemic finally ends. All three quantities exhibit huge fluctuations that are reminiscent of the actual data.

## Results

#### Systematic mitigation

In this section we investigate what we term as the systematic mitigation strategy. Here, increasingly stringent controls are imposed as soon as an outbreak is detected, in which the reproductive number  $R_0$ exceeds 1, to reduce  $R_0$  to less than 1. The condition  $R_0 = 1$  defines the peak of the epidemic because the number of newly infected individuals reaches a maximum at this point. Once  $R_0$  becomes less than 1, progressively fewer individuals are infected after each incubation period and the epidemic begins to disappear. The number of individuals that become infected after  $R_0$  has been reduced to less than 1 decays exponentially with time and constitute a small contribution to the total number of infections.

Because society is a complicated, with many competing social forces in play, we posit that it is not possible to reduce  $R_0$  instantaneously, but rather, the reduction happens gradually. We therefore assume that after each successive incubation period  $R_0$  is decreased by a random number r whose average value  $\langle r \rangle$  is less than 1. Let us define  $R_k$  as the reproductive number in the  $k^{\text{th}}$  period. Then  $R_k$  is given by

$$R_k = r_k R_{k-1} = r_k r_{k-1} \dots r_2 r_1 R_0, \tag{1}$$

where  $r_k$  is the value of the random variable r in the kth period. The typical number of periods k until  $R_0$  reaches 1 is determined by  $R_0 \langle r \rangle^k = 1$ . In what follows, we assume that when the epidemic is first detected, the reproductive number  $R_0 = 2.5$ , and we take  $\langle r \rangle = 0.95$  for illustration. Using these values,

$$k = \frac{\ln(1/R_0)}{\ln\langle r \rangle} = \frac{\ln(1/2.5)}{\ln(0.95)} \approx 17.86$$

Thus the epidemic typically reaches its peak after 18 periods. However, because of the inherent randomness in the mitigation, with  $R_0$  sometimes decreasing by less than 0.95 and sometimes by more than 0.95 after each incubation period, the true epidemic dynamics can be very different, as illustrated in Fig. 2.

We simulate the systematic mitigation strategy by starting with a single infected individual and reproductive number  $R_0 = 2.5$ . We then choose a set of random numbers  $r_1, r_2, r_3, ...$ , each of which are uniformly distributed between 0.9 and 1, so that  $\langle r \rangle = 0.95$ . We first measure how long it takes until  $R_k$  is reduced to 1, which signals the epidemic peak. We perform this same measurement for  $5 \times 10^6$  different choices of the set of random numbers  $r_1, r_2, ..., r_k$ . As shown in Fig. 2a, the probability q(k) that the epidemic reaches its peak in the  $k^{\text{th}}$  period has a maximum at roughly k = 18 periods, in agreement with the above naive estimate. If one is lucky, that is, if most of the reduction factors  $r_i$  are close to 0.9, the epidemic reaches its peak in as little as 11 periods. If one is unlucky (many of the  $r_i$  close to 1), the epidemic can can continue to grow for more than 30 periods.

While the distribution of epidemic durations is fairly narrow, the total number I of people who were infected during the course of an epidemic can vary by several orders of magnitude. The number of people infected in the k<sup>th</sup>

**Fig. 2** | **Systematic mitigation. a** The probability q(k) that the epidemic reaches its peak after k periods. **b** The probability p(I) that I people have been infected when the epidemic reaches its peak (under the assumption that the initial epidemic size is one person).



period,  $I_k$  is given by  $I_k = R_{k-1}I_{k-1}$ . Thus according to the dynamics of the reproductive number in (1), the total number of infected individuals is

$$I = 1 + I_1 + I_2 + I_3 + I_4 + I_5 \dots$$
  
= 1 + R<sub>0</sub> + R<sub>0</sub>R<sub>1</sub> + R<sub>0</sub>R<sub>1</sub>R<sub>2</sub> + R<sub>0</sub>R<sub>1</sub>R<sub>2</sub>R<sub>3</sub> + R<sub>0</sub>R<sub>1</sub>R<sub>2</sub>R<sub>3</sub>R<sub>4</sub> + ...  
= 1 + R<sub>0</sub> + r<sub>1</sub>R<sub>0</sub><sup>2</sup> + r<sub>2</sub> r<sub>1</sub><sup>2</sup>R<sub>0</sub><sup>3</sup> + r<sub>3</sub> r<sub>2</sub><sup>2</sup>r<sub>1</sub><sup>3</sup>R<sub>0</sub><sup>4</sup> + r<sub>4</sub> r<sub>3</sub><sup>2</sup> r<sub>2</sub><sup>3</sup> r<sub>1</sub><sup>4</sup>R<sub>0</sub><sup>5</sup> + ...  
(2a)

Thus the average number of infected individuals is

$$\langle I \rangle = 1 + R_0 + \langle r \rangle R_0^2 + \langle r \rangle \langle r^2 \rangle R_0^3 + \langle r \rangle \langle r^2 \rangle \langle r^3 \rangle R_0^4 + \langle r \rangle \langle r^2 \rangle \langle r^3 \rangle \langle r^4 \rangle R_0^5 + \dots$$
(2b)

This expression converges because the  $k^{th}$  term quickly decreases with k for an arbitrary distribution of r with support on [0, 1). It is important to point out that the number of newly infected people at each incubation period is based on the assumption that this number is small compared to the total population size, so that the growth in the number of new infections is truly exponential. As shown in Fig. 2b, while the most probable epidemic size is  $\approx 10^4$  (again starting with a single infected individual), there is a nonvanishing probability that the outbreak size can be as small as a few hundred or greater than  $10^7$ . This large disparity in outbreak sizes illustrates how small changes in the way that the epidemic is mitigated can lead to huge changes in the outbreak size.

More dramatically, suppose that the mitigation strategy is slightly less effective and that the reproductive number is reduced at each period by a uniform random variable that lies between [0.95, 1] rather than between [0.9, 1]. Now the peak of the epidemic can occur between 22 and 55 periods, with a most probable duration of 36 periods. However, the epidemic size when the peak of the epidemic is reached ranges between roughly  $10^5$  and  $10^{12}$ , with a most probable size of roughly  $7 \times 10^7$ . The upper value is much larger than the world population and the finiteness of the population would now provide the upper bound. Although the peak of this second epidemic occurs a factor 2 longer as the first one, it typically infects 7000 times more people! We emphasize that the stochastic nature of the random variables  $r_j$  plays a decisive role. Very different behaviors emerge in the deterministic case<sup>20</sup>.

#### Vacillating mitigation

During the acute period of the pandemic in 2020–2021, there was considerable and even vitriolic debate about the efficacy of various mitigation strategies, or even about the utility of any mitigation. If the epidemic is severe, as quantified by the reproductive number  $R_k$  in the *k*th period being substantially greater than 1, people may be more likely to accept restrictions on their behaviors, such as isolating, masking, vaccinating, etc., to reduce their risk of getting sick. These adaptations will reduce the reproductive number. If, however, the reproductive number becomes less than 1, then people will want to relax their vigilance and may also advocate for the opening of various public venues, such as schools, theaters, stadiums, etc. We model this tug-of-war between increased and decreased restrictions by what we term as the vacillating mitigation strategy. This perspective of treating the competition between epidemiology and social behavior was previously treated in more sophisticated models<sup>21,22</sup>. We emphasize that our model merely a proxy for the two competing influences of epidemiology and social behavior.

The two competing steps of the vacillating strategy are the following: Militation if R > 1 decrease R by a factor r that is uniformly dis

- Mitigation: if  $R_k > 1$ , decrease  $R_k$  by a factor r that is uniformly distributed in [a, 1], with a < 1.
- Relaxation: if  $R_k < 1$ , change  $R_k$  by a factor *s* that is uniformly distributed in [a, 3-2a].

The first option is the same as in the systematic mitigation strategy. We construct the second option by requiring that  $\langle s \rangle = 1 + \frac{1}{2}(1 - a)$  and  $\langle r \rangle = 1 - \frac{1}{2}(1 - a)$  are symmetrically located about 1. That is, the average decrease in  $R_k$  in a mitigation step equals the average increase in  $R_k$  in the relaxation step. This symmetrical construction seems appropriate to probe the long-term influence of vacillation on the dynamics. If the vacillation strategy was biased towards relaxation,  $R_0$  would remain greater than 1 and the entire planet would be infected. If this strategy was biased towards mitigation, the epidemic would be similar to that in systematic mitigation. Neither of these cases is interesting from the viewpoint of probing long-time behaviors.

In this vacillating strategy,  $R_k$  varies between values greater than 1 and values less than 1. This would lead to an eternal epidemic. To avoid this unrealistic outcome, the other important feature of the relaxation step is that the value of  $R_k$  could still decrease during a relaxation step because a < 1. This possibility ensures that eventually less than one person will be infected in the current incubation period. We now define this event as signaling the end of the epidemic.

Figure 3a–d shows a few representative trajectories of the number of people infected I(t) as a function of time (incubation periods) from the same initial condition of a single infected person and  $R_0 = 2.5$ . While there are some qualitative differences between the trajectories of Fig. 1 and the model outcomes, the important points that are common to the real data and the simulation results are the disparities in the individual trajectories and the strongly fluctuating temporal behavior.

For the vacillating strategy and for the choice a = 0.9, the most likely duration of the epidemic is roughly 400 periods (Fig. 4a), compared to 18 periods for the systematic strategy. The probability that the epidemic lasts much longer than the most likely value decays exponentially with time. An even more dramatic feature of the vacillating strategy is the number of people that are ultimately infected. The most probable outcome is that  $3 \times 10^5$  people are infected when the epidemic ends (Fig. 4b). However, the size of the epidemic can range from  $10^4$  to  $10^8$ . Compared to the systematic mitigation strategy with a reduction factor uniformly in the range [0.9, 1], the epidemic now lasts roughly 20 times longer and infects a factor 30 more individuals.

#### Discussion

This work should not be construed to mean that public-health measures should be ignored. Indeed, the extremely rapid development of a vaccine Fig. 3 | Time histories in vacillating mitigation. a-d Representative trajectories for the number of people I(t) infected at time t for the vacillating mitigation strategy when starting with  $R_0 = 2.5$  and a single infected person. The four realizations shown illustrate the highly unpredictable outcomes of individual epidemics.

Q(k) that the epidemic lasts k periods. **b** The probability P(s) that the epidemic ultimately infects s people starting with  $R_0 = 2.5$  and a single infected person.

that is effective against Covid-19 is an outstanding triumph of modern medical science. It should also be pointed out that some of the many forecasting models for Covid-19 were useful during the early stages of the pandemic. However, when social influences with competing viewpoints began to dictate individual and collective policy decisions, much of the predictive power of forecasting models was lost.

We also emphasize that our simplistic model has little connection to the actual epidemiological and social processes that determine the spread of the epidemic and the changes in individual and collective behaviors in response to the epidemic. Nevertheless, our model seems to capture the tug of war between public-health mandates to control the spread of the disease and the social forces that often advocate for a more laissez-faire approach. Our main message is that there are huge uncertainties in predicting the time course of an epidemic, its ultimate duration, and the final outbreak size. This unpredictability seems to be intrinsic to the dynamics of epidemics where epidemiological influences occur in concert with social forces. In this setting, forecasting ambiguity is unavoidable.

### Data availability

All data needed to evaluate the conclusions are presented in the paper. This data may be requested from the authors.

Received: 2 October 2023; Accepted: 24 April 2024; Published online: 07 June 2024

# References

k

- 1. COVID-19 Coronavirus pandemic. https://www.worldometers.info/ coronavirus (2023).
- 2. Bailey, N. T. A simple stochastic epidemic. Biometrika 37, 193-202 (1950).
- З. Bailey, N.T.J. The mathematical theory of infectious diseases (Oxford University Press, Oxford, 1987).
- Hethcote, H. W. The mathematics of infectious diseases. SIAM Rev. 4. 42, 599-653 (2000).
- 5. Ridler-Rowe, C. On a stochastic model of an epidemic. J. Appl. Prob. 4, 19-33 (1967).
- 6. Grassberger, P. On the critical behavior of the general epidemic process and dynamical percolation. Math. Biosci. 63, 157-172 (1983).
- Martin-Löf, A. The final size of a nearly critical epidemic, and the first 7. passage time of a Wiener process to a parabolic barrier. J. Appl. Prob. 35, 671-682 (1998).
- Ben-Naim, E. & Krapivsky, P. L. Size of outbreaks near the epidemic 8. threshold. Phys. Rev. E 69, 050901 (2004).
- 9. Kessler, D. A. & Shnerb, N. M. Solution of an infection model near threshold. Phys. Rev. E 76, 010901 (2007).
- 10. Kessler, D. A. Epidemic size in the SIS model of endemic infections. J. Appl. Prob. 45, 757-778 (2008).
- Gordillo, L. F., Marion, S. A., Martin-Löf, A. & Greenwood, P. E. 11. Bimodal epidemic size distributions for near-critical SIR with vaccination. Bull. Math. Biol. 70, 589-602 (2008).



- Greenwood, P. E. & Gordillo, L. F. Stochastic epidemic modeling. In Mathematical and Statistical Estimation Approaches in Epidemiology, 31–52 (Springer, 2009).
- Van der Hofstad, R., Janssen, A. & Van Leeuwaarden, J. S. Critical epidemics, random graphs, and brownian motion with a parabolic drift. *Adv. Appl. Probab.* 42, 1187–1206 (2010).
- Antal, T. & Krapivsky, P. L. Outbreak size distributions in epidemics with multiple stages. J. Stat. Mech. 2012, P07018 (2012).
- Ben-Naim, E. & Krapivsky, P. L. Scaling behavior of threshold epidemics. *Eur. Phy. J. B* 85, 1–9 (2012).
- Krapivsky, P. L. Infection process near criticality: influence of the initial condition. J. Stat. Mech. 2021, 013501 (2021).
- Castro, M., Ares, S., Cuesta, J. A. & Manrubia, S. The turning point and end of an expanding epidemic cannot be precisely forecast. *Proc. Natl. Acad. Sci. USA* **117**, 26190–26196 (2020).
- Williams, B. J., St-Onge, G. & Hébert-Dufresne, L. Localization, epidemic transitions, and unpredictability of multistrain epidemics with an underlying genotype network. *PLOS Comput. Biol.* **17**, e1008606 (2021).
- Vallée, A., Faranda, D. & Arutkin, M. COVID-19 epidemic peaks distribution in the United-States of America, from epidemiological modeling to public health policies. *Sci. Rep.* **13**, 4996 (2023).
- 20. Bianconi, G. & Krapivsky, P. L. Epidemics with containment measures. *Phys. Rev. E* **102**, 032305 (2020).
- Manrubia, S. & Zanette, D. H. Individual risk-aversion responses tune epidemics to critical transmissibility (R=1). *R. Soc. Open Sci.* 9, 211667 (2022).
- Tkachenko, A. V. et al. Stochastic social behavior coupled to COVID-19 dynamics leads to waves, plateaus, and an endemic state. *eLife* 10, e68341 (2021).

# Acknowledgements

We thank Jean-Mark Luck for helpful discussions and NSF grant DMR-1910736 for partial financial support of this research.

# Author contributions

P.L.K. and S.R. conceived the idea, performed the research, and wrote the manuscript.

# **Competing interests**

The authors declare no competing interests.

# **Additional information**

**Correspondence** and requests for materials should be addressed to S. Redner.

Reprints and permissions information is available at http://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) 2024