Title

How relevant is the basic reproductive number computed during the coronavirus disease 2019 (COVID-19) pandemic, especially during lockdowns?

Journal

Infection Control and Hospital Epidemiology, 43(1)

ISSN

0899-823X

Authors

Rao, Arni SR Srinivasa
Krantz, Steven G
Bonsall, Michael B
et al.

Publication Date

2022

DOI

10.1017/ice.2020.1376

Peer reviewed

Infection Control & Hospital Epidemiology (2021), 1–3
doi:10.1017/ice.2020.1376

Letter to the Editor

How relevant is the basic reproductive number computed during the coronavirus disease 2019 (COVID-19) pandemic, especially during lockdowns?

Arni S.R. Srinivasa Rao PhD[1,2], Steven G. Krantz PhD[3], Michael B. Bonsall PhD[4], Thomas Kurien MD[5], Siddappa N. Byrareddy PhD[6], David A. Swanson PhD[7], Ramesh Bhat PhD[8] and Kurapati Sudhakar PhD (retired)[9]

[1]Medical College of Georgia, Augusta, Georgia, [2]Laboratory for Theory and Mathematical Modeling, Department of Medicine - Division of Infectious Diseases, Medical College of Georgia, Augusta University, Augusta, Georgia, [3]Department of Mathematics, Washington University, St Louis, Missouri, [4]Mathematical Ecology Research Group, Department of Zoology, University of Oxford, Oxford, United Kingdom, [5]Department of Medicine, Pondicherry Institute of Medical Sciences, Puducherry, India, [6]Department of Pharmacology and Experimental Neuroscience, University of Nebraska Medical Center, Omaha, Nebraska, [7]Department of Sociology, University of California–Riverside, Riverside, California, [8]NMIMS University, Mumbai, India and [9](formerly with) Centers for Disease Control and Prevention, World Bank, and United States Agency for International Development

To the Editor—The basic reproductive number R_0 in epidemiology is defined as the average number of secondary infections that will be likely produced by a primary infected person in a predominantly susceptible population. Mathematically, it is an accurate measure of disease spread.[1] However, the value of R_0 is difficult to estimate from epidemiological data, for example, during the ongoing coronavirus disease 2019 (COVID-19) pandemic. In recent studies on COVID-19, for example,[2–4] computed a time-varying R_0 has been computed, which researchers called R_t. They ascertained that the decline in R_t is due to continued lockdowns and nonpharmaceutical interventions. Although the conclusions in those studies are supported by the data, estimates of R_t raise methodological issues that require further consideration. Here, we convey the essential and technical difficulties in estimating either R_0 or R_t from the data, and we discuss how a model-based R_0 may not adequately capture the actual spread of the disease. Although these limitations are generally unavoidable (even after defining appropriate error structures and statistical modeling), the inappropriate use of this metric, especially in the ongoing COVID-19 pandemic, has important implications for infectious disease mitigation planning.

Suppose that Y_0 is the number of infected people at time t_0 who could generate secondary infections between t_0 and t_1, say, Y_1. However, the testing of all the potential infected individuals during this period need not be complete. Y_1 could generate further secondary infections between t_1 and t_2, say, Y_2, and so on. Again, the testing of the samples through contact tracing need not be complete (Fig. 1). That is, Y_{i+1} at t_{i+1} could be generated by Y_i at t_i for $i = 0, 1, \ldots$. In reality, during most epidemics, and especially for the COVID-19 pandemice, only a fraction of Y_i, say, $Y_i^{'}$ are ever reported (and also diagnosed due to incomplete testing) such that $Y_i^{'} < Y_i$ for all i.[5,6] This partial reporting (including partial diagnosis and partial testing) could also be due to lockdowns and lack of proper knowledge regarding COVID-19 (forced or natural behavior changes in the community, eg, lockdowns and use of masks). The average number of secondary infections generated by Y_i individuals is $Y_{i+1}\ /Y_i$. If there is variation in the infected people or a rapid aggregation of infected people, then it is more appropriate that we should use the geometric mean instead of the arithmetic mean approaches to determine expected reproductive numbers. Not only is the former far better suited than the latter to deal both with fluctuations and numbers that are not independent of one another, it also is the only correct mean when using results that are presented as ratios.[7–9]

Suppose that Y_{i+k} is the number of infected people at time t_{i+k} when lockdowns are introduced at k for $k = 0, 1, 2\ \ldots$.

Assume that

$$Y_{i+k} < Y_{i+k+1} \text{ for } k = 0, 1, 2, 3, 4. \quad (1)$$

The percentage of growth in the number of infected people during the 4 time intervals (t_{i+k}, t_{i+k+1}) for $k = 0, 1, 2, 3, 4$, are, say, $\gamma_{i+k}\%$ for $k = 0, 1, 2, 3, 4$, respectively. These growth percentages are computed as

$$\gamma_{i+k}\% = \left(\frac{Y_{i+k+1} - Y_{i+k}}{Y_{i+k}} \times 100\right)\% \text{ for } k = 0, 1, 2, 3, 4.$$

The secondary infections caused by an infected individual (Fig. 1) are the people who were not traced by the system. This step assumes that all of the infected people who were identified by the system were either quarantined or were controlled not to spread the virus further. Only a proportion of infected people who were tested and identified during lockdowns was reported, and others were either not diagnosed or not reported. Asymptomatic individuals could be anywhere in the process; that is, they were part of the identified and reported group or were among those who had not been contact traced or diagnosed.

Author for correspondence: Arni S.R. Srinivasa Rao E-mail: arrao@augusta.edu
Cite this article: Srinivasa Rao ASR, *et al.* (2021). How relevant is the basic reproductive number computed during the coronavirus disease 2019 (COVID-19) pandemic, especially during lockdowns?. *Infection Control & Hospital Epidemiology*, https://doi.org/10.1017/ice.2020.1376

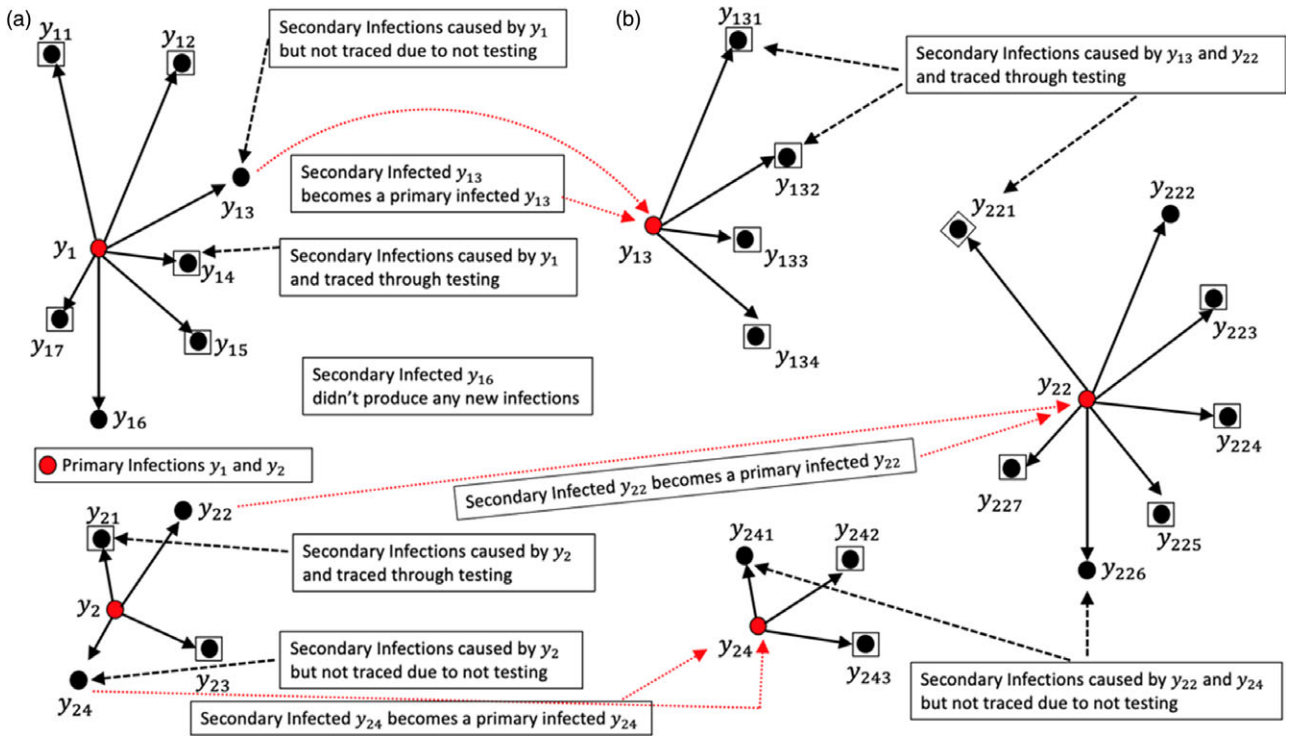

Fig. 1. Demonstration of average number of secondary infections observed through tracing and diagnosing. In **(a)**, let y_1 and y_2 be the two primary COVID-19 infected, where the individual y_1 had generated 7 secondary infections out of which 5 were traced and diagnosed. The individual y_2 had generated 4 secondary infections out of which 2 were traced and diagnosed. The observed arithmetic average secondary infected by $\{y_1, y_2\}$ in (a) was $\frac{5+2}{2} = 3.5$, but the true average by them was $\frac{7+4}{2} = 5.5$. In **(b)**, the third secondary infection in **(a)**, say, y_{13} becomes a primary infected that generates 4 secondary infections out of which all were traced and diagnosed. In **(b)**, the second secondary infection in **(a)**, say, y_{22} becomes a primary infected that generates 7 secondary infections out of which only 5 were traced and diagnosed. Finally, in **(b)**, the fourth secondary infection in **(a)**, say, y_{24} by primary infected y_2 becomes a primary infected that generates 3 secondary infections out of which only 2 were traced and diagnosed. The observed arithmetic average secondary infections by $\{y_{13}, y_{22}, y_{24}\}$ was $\frac{4+5+2}{3} = 3.67$, but if every COVID-19 patient was diagnosed, then the true average secondary infections by them was $\frac{4+7+3}{3} = 4.67$. Note that the total traced and tested could be many fold more than the actual positive cases found. Suppose 22 secondary infections generated during the third generation, then the mean number of secondary infections (geometric) obtained during three generations of spread is $\sqrt[3]{3.61} = 1.53$.

The mean (geometric) number of secondary infections would be appropriate because we were considering proportionate secondary infections. Hence, the mean number of secondary infections during $(t_i, t_i + 4)$ is given by

$$\sqrt[4]{\prod_{k=0}^{3}(1+\gamma_{i+k}\%)}. \tag{2}$$

Similarly, the trend in eq. (1) continues for $k = 0, 1, \ldots n$, then the mean number of secondary infections during the lockdown period $(t_i, t_i + n)$ is given by

$$\sqrt[n]{\prod_{k=0}^{n-1}(1+\gamma_{i+k}\%)}. \tag{3}$$

This point applies to several studies in which the reporting over time of the study is not constant. Even if the testing numbers and testing patterns are constant over a period, the proportion of underreported cases may not be constant. Thus, the estimation of R_0 is likely to be highly variable in any given situation. For the practical purposes of computing R_0 or R_t we usually have data on $Y_i^{'}$, the number tested.

When the ratios $Y_{i+k+1}\ /Y_{i+k}$ for $k = 0, 1, \ldots n$ are considered, then the geometric mean of these growth rates would be

$$\sqrt[n]{\prod_{k=0}^{n}\frac{Y_{i+k+1}}{Y_{i+k}}} = \sqrt[n]{\frac{Y_{i+n+1}}{Y_i}}. \tag{4}$$

However, $\widehat{R}_0$ or $\widehat{R}_t$, (the estimated basic and time-varying reproductive numbers at the start or ongoing through an epidemic, respectively) may not be at all close to R_0 or R_t even if the Y_i values are generated from a mathematical model for a period $i > 0$ that uses data on susceptible, exposed, infected, and recovered in which the underlying epidemiological processes are time varying. This factor will introduce bias to estimates of model-based basic reproductive rates and time-varying reproductive rates. Some other limitations in various studies arise due to computing R_t after lockdowns were relaxed. Possibly, heterogeneity exists in the data that could have masked R_t measures due to the computation of subnational and regional parameters in several COVID-19–affected countries.

The lesson here is that mathematical models must be used with care. They must be fitted to the data, and their accuracy must be carefully monitored and quantified.[10] Any alternative course of action could lead to wrong interpretation and mismanagement of the disease with disastrous consequences.

Acknowledgments. We thank Dr Natasha Martin, University of California San Diego, and Dr Chris T. Bauch, University Waterloo for providing useful comments on our original draft and pointing us to critical literature.

Authors' contributions. All authors contributed in writing. ASRSR and SGK designed the study, and ASRSR wrote the first draft and conceptualized the study. MBB, TK, SB, DS, RB and SK have contributed in writing, editing and discussions. All authors approved the manuscript.

Financial support. No financial support was provided relevant to this article.

Conflicts of interest. All authors report no conflicts of interest relevant to this article.

References

1. Eisenberg J. R_0: How scientists quantify the intensity of an outbreak like coronavirus and its pandemic potential. School of Public Health, University of Michigan website. https://sph.umich.edu/pursuit/2020posts/how-scientists-quantify-outbreaks.html. Published 2020. Accessed December 14, 2020.
2. Hong HG, Li Y. Estimation of time-varying reproduction numbers underlying epidemiological processes: A new statistical tool for the COVID-19 pandemic. *PLoS One* 2020;15(7):e0236464.
3. Laxminarayan R, Wahl B, Dudala SR, *et al.* Epidemiology and transmission dynamics of COVID-19 in two Indian states. *Science* 2020. doi: 10.1126/science.abd7672.
4. Abbott S, Hellewell J, Thompson RN, *et al.* Estimating the time-varying reproduction number of SARS-CoV-2 using national and subnational case counts. *Open Res* 2020;5:112. doi: 10.12688/wellcomeopenres.16006.1.
5. Gibbons C. Mangen MJJ, Plass D, *et al.* Measuring underreporting and under-ascertainment in infectious disease datasets: a comparison of methods. *BMC Public Health* 2014. doi: 10.1186/1471-2458-14-147.
6. Krantz SG, Polyakov P, Rao ASRS. True epidemic growth construction through harmonic analysis. *J Theoret Biol* 2020;494. doi: 10.1016/j.jtbi.2020.110243.
7. Hayes A. Geometric mean definition. *Investopedia*, September 5, 2020. https://www.investopedia.com/terms/g/geometricmean.asp.
8. Fleming P, Wallace J. How not to lie with statistics: the correct way to summarize benchmark results. *Commun ACM* 1986;29:218–221.
9. Rao ASRS, Krantz SG, Bonsall MB, *et al.* Time varying basic reproductive number computed during COVID-19, especially during lockdowns could be questionable, eLetters. *Science*. 2020. Sciencemag.org (Hyperlink).
10. Krantz S, Rao ASRS. Level of underreporting including underdiagnosis before the first peak of COVID-19 in various countries: preliminary retrospective results based on wavelets and deterministic modeling. *Infect Control Hosp Epidemiol* 2020;41:857–859.