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

Permalink
https://escholarship.org/uc/item/5xx6c0sv

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

Publication Date
2020-12-14

DOI
10.1017/ice.2020.1376

Peer reviewed
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 PhD1-2, Steven G. Krantz PhD3, Michael B. Bonsall PhD4, Thomas Kurien MD5, Siddappa N. Byareddy PhD6, David A. Swanson PhD7, Ramesh Bhat PhD8 and Kurapati Sudhakar PhD (retired)9

1Medical College of Georgia, Augusta, Georgia, 2Laboratory for Theory and Mathematical Modeling, Department of Medicine - Division of Infectious Diseases, Medical College of Georgia, Augusta University, Augusta, Georgia, 3Department of Mathematics, Washington University, St Louis, Missouri, 4Mathematical Ecology Research Group, Department of Zoology, University of Oxford, Oxford, United Kingdom, 5Department of Medicine, Pondicherry Institute of Medical Sciences, Puducherry, India, 6Department of Pharmacology and Experimental Neuroscience, University of Nebraska Medical Center, Omaha, Nebraska, 7Department of Sociology, University of California–Riverside, Riverside, California, 8NMIMS 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 pandemic, only a fraction of $Y_i$, say, $Y_i^\prime$ are ever reported (and also diagnosed due to incomplete testing) such that $Y_i^\prime < Y_i$ 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

© The Author(s), 2020. Published by Cambridge University Press on behalf of The Society for Healthcare Epidemiology of America. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.
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

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

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

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

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

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

However, $\tilde{R}_0$ or $\tilde{R}_t$, (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. 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