

A POOR-MAN'S APPROACH TO THE EFFECTIVE REPRODUCTION NUMBER

A poor-man's approach to the effective reproduction number: the COVID-19 case

Revised 4/22/2020 16:54:00

José Menéndez

Department of Physics, Arizona State University, Tempe AZ 85287-1504
jose.menendez@asu.edu; <https://orcid.org/0000-0001-8739-9197>

It is shown that estimates of the effective reproduction number R_t for COVID-19 using standard packages such as EPIESTIM can be reproduced very accurately using the expression $R_t = c(t)/c(t - \tau)$, where $c(t)$ is the incidence at time t and τ the mean value of the series interval.

The effective reproduction number R_t —a fundamental epidemiological parameter that characterizes the temporal dynamics of an infectious disease—is notoriously difficult to determine without detailed modeling[1-6]. During the ongoing COVID-19 pandemic, daily tallies of the incidence $c(t)$ (new cases on day t), recovered $r(t)$ (individuals who are declared cured on day t) and deceased $d(t)$ (individuals who die on day t) have become available for many countries and regions [7]. With this information, one might naively expect that R_t can be determined as

$$R_t = \frac{c(t)}{x(t)}, \quad (1)$$

with $x(t) = r(t) + d(t)$. This definition is exactly equivalent to the expression for the basic reproduction number in a ‘Susceptible-Infectious-Removed’ (SIR) framework, but can be obviously applied to the observed data without any reference to a specific model. Calculations with Eq. (1) for COVID-19, however, disagree with values of R_t computed from standard packages such as EPIESTIM [4], as seen in Fig. 1.

Part of the reason for the discrepancy is the lack of reliable $r(t)$ data. For some countries, such as the United Kingdom, the published cumulative number of recovered people is *less* than the cumulative number of deceased individuals, a very unlikely scenario. Even globally, the sum total of deceased individuals is a substantial fraction of cured individuals, which is also highly implausible. Furthermore, there is a *relative* time shift between $c(t)$ and $x(t)$, since the former includes the incubation time plus the time to develop symptoms serious enough to be reported. However, attempting to correct for this relative delay does not improve the agreement in Fig. 1 in any significant way.

Standard packages do not suffer from the above problems because they rely only on $c(t)$ data, and therefore are not affected by systematic errors in $r(t)$. For these calculations one requires the infectivity profile of the disease, which is approximated as the distribution of the standard interval [4]. Calculations based on EPIESTIM [4] using a series interval

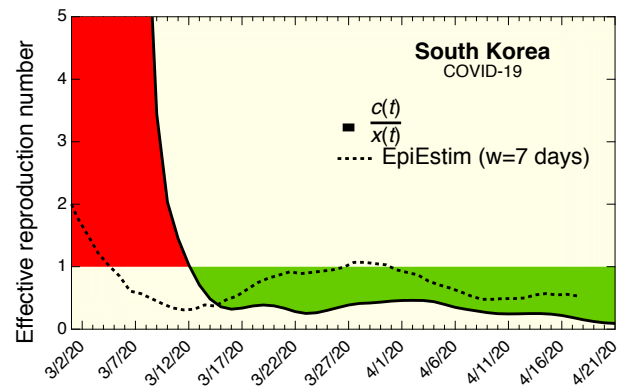


FIG. 1. Solid black line: effective COVID-19 reproduction number R_t for South Korea calculated from Eq. (1). Dotted line: R_t calculated with the package EPIESTIM using a mean infectious period $\tau = 5.8$ days with a standard deviation of 2.9 days and a time window of $w = 7$ days.

mean $\tau = 4.7$ days and standard deviation 2.9 days, from Ref. [8], are updated daily at [9]. A more recent study finds $\tau = 5.8$ days, with 44% presymptomatic transmission [10]. The recovery time, on the other hand, is much longer, and this implies that individuals at advanced stages of the disease have a very low infectivity. Under these conditions Eq. (1) cannot be valid, since its denominator should not be $x(t)$ but the number of people who became infected approximately at time $t - \tau$. The latter is greater than the former during the ramp-up phase of the disease, and this explains why the curve calculated from Eq. (1) shows a higher R_t during this time. Conversely, the number of people who became infected at time $t - \tau$ is *less* than $x(t)$ at later stages, and this explains why the curve calculated from Eq. (1) gives a lower value of the number of R_t at these times.

The above considerations suggest an alternative definition of R_t as

$$R_t = \frac{c(t)}{c(t - \tau)}, \quad (2)$$

A POOR-MAN'S APPROACH TO THE EFFECTIVE REPRODUCTION NUMBER

A computation of R_t for different countries using Eq. (2) with $\tau = 5.8$ days is shown in Figs. 2-5, where it is compared with calculations of R_t using EPIESTIM for the same value of τ . Since the $c(t)$ data are quite noisy, a smooth version $c_s(t)$ is used. The smooth $c_s(t)$ is computed using a regularization approach that minimizes the sum of the mean square deviation between $c_s(t)$ and $c(t)$ plus a term proportional to the square of the second derivative $c_s''(t)$. The relative weight of the two terms is controlled by a parameter λ that is defined as in Eq. 3 of Ref. [11], with the matrices \mathbf{B} and $\tilde{\mathbf{B}}$ taken as the identity matrix. The chosen value was $\lambda = 100 \text{ day}^4$. A comparison between $c_s(t)$ and $c(t)$ is shown in the inset of Fig. 2 for the case of South Korea. The regularization method has the advantage that it distorts the shape of a curve less than a typical sliding average, while being superior to Savitzky-Golay[12] smoothing in terms of noise reduction. This was important in the context of the comparison made here between Eq. (2) and the EPIESTIM package, because artificial shape distortions are eliminated. For practical applications of Eq. (2), however, a simple 7-day running average should suffice.

Two calculations with EPIESTIM were performed, both with $\tau = 5.8$ days and a standard deviation of 2.9 days. The first calculation uses a time window $w = 1$ day, and the second calculation a time window $w = 7$ days. In the latter case, the value that appears in the figures at a particular time t corresponds to the window for which t is the middle point. The $w = 1$ day results are very noisy, as expected, but both EPIESTIM curves are in very good agreement with the calculations from Eq. (2). It is important to point out that the incidence data have a significant delay with respect to the actual time of infection, so that all figures should be interpreted at representing the situation between one and two weeks before the dates in the horizontal axes. We made no attempt to correct for reporting delays as in Ref. [9].

In summary, an extremely simple algorithm has been introduced that yields effective reproduction numbers very similar to those obtained from more elaborate approaches as exemplified by the EPIESTIM package. The latter should produce more accurate results if the full epidemiological characteristics of the disease are well known, but for calculations at the present time when little is known about COVID-19, the method proposed here is much easier to use and provides comparable accuracy.

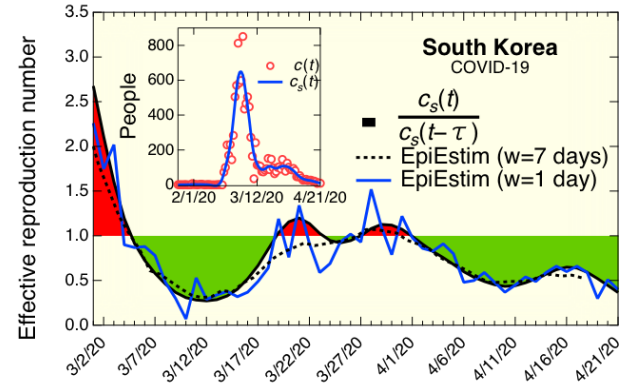


FIG. 2. The COVID-19 effective reproduction number for South Korea. The solid black line is from Eq. (2). The blue and dotted black lines are estimates from EPIESTIM using time windows of 1 day and 7 days, respectively. All calculations used $\tau = 5.8$ days, with a standard deviation (EPIESTIM case) of 2.9 days. The inset shows the reported incidence $c(t)$ (red circles) and the smoothed curve $c_s(t)$.

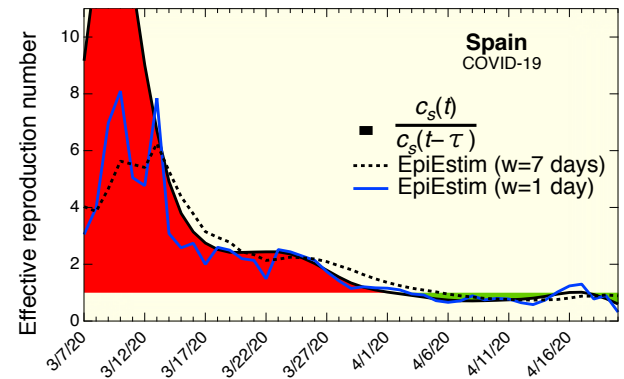


FIG. 3. The COVID-19 effective reproduction number for Spain. The solid black line is from Eq. (2). The blue and dotted black lines are estimates from EPIESTIM using time windows of 1 day and 7 days, respectively. All calculations used $\tau = 5.8$ days, with a standard deviation (EPIESTIM case) of 2.9 days.

A POOR-MAN'S APPROACH TO THE EFFECTIVE REPRODUCTION NUMBER

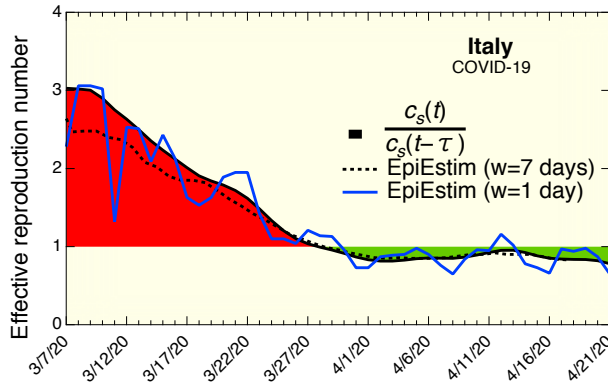


FIG. 4. The COVID-19 effective reproduction number for Italy. The solid black line is from Eq. (2). The blue and dotted black lines are estimates from EPIESTIM using time windows of 1 day and 7 days, respectively. All calculations used $\tau = 5.8$ days, with a standard deviation (EPIESTIM case) of 2.9 days.

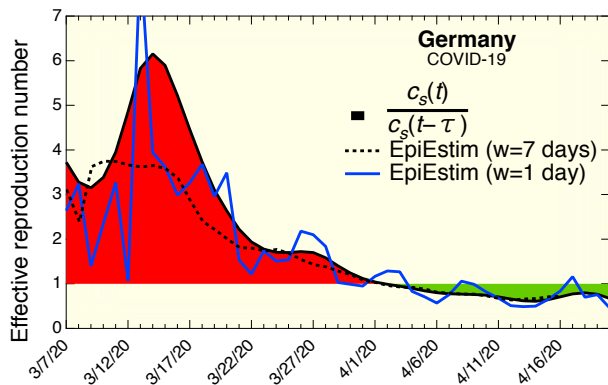


FIG. 5. The COVID-19 effective reproduction number for Germany. The solid black line is from Eq. (2). The blue and dotted black lines are estimates from EPIESTIM using time windows of 1 day and 7 days, respectively. All calculations used $\tau = 5.8$ days, with a standard deviation (EPIESTIM case) of 2.9 days.

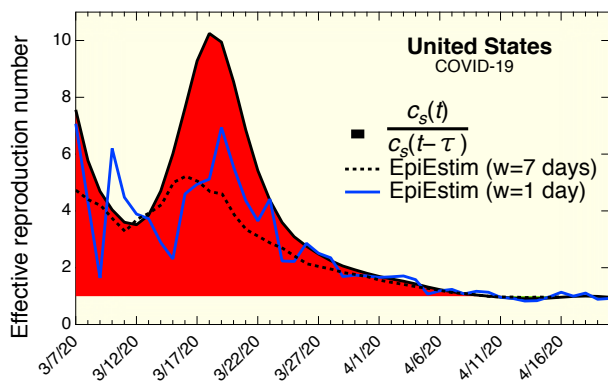


FIG. 6. The COVID-19 effective reproduction number for the United States. The solid black line is from Eq. (2). The blue and dotted black lines are estimates from EPIESTIM using time windows of 1 day and 7 days, respectively. All calculations used $\tau = 5.8$ days, with a standard deviation (EPIESTIM case) of 2.9 days.

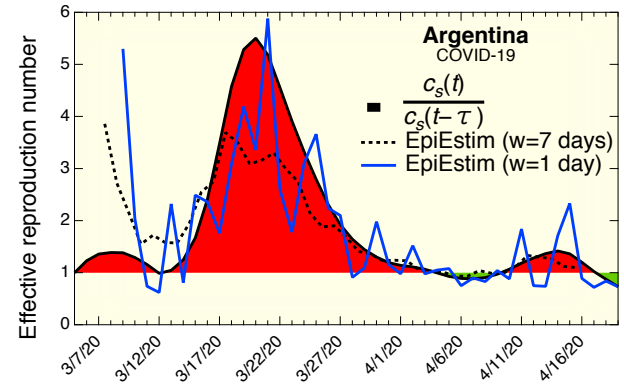


FIG. 7. The COVID-19 effective reproduction number for Argentina. The solid black line is from Eq. (2). The blue and dotted black lines are estimates from EPIESTIM using time windows of 1 day and 7 days, respectively. All calculations used $\tau = 5.8$ days, with a standard deviation (EPIESTIM case) of 2.9 days.

The numerical simulations shown here were carried out on IGOR PRO 8.0 (Wavemetrics, Inc). The code is available upon request.

REFERENCES

- [1] L. M. Bettencourt and R. M. Ribeiro, PLoS One 3, e2185 (2008).
- [2] A. Cintron-Arias, C. Castillo-Chavez, L. M. Bettencourt, A. L. Lloyd, and H. T. Banks, Math Biosci Eng 6, 261 (2009).
- [3] J. Li, D. Blakeley, and R. J. Smith, Comput Math Methods Med 2011, 527610 (2011).
- [4] A. Cori, N. M. Ferguson, C. Fraser, and S. Cauchemez, Am J Epidemiol 178, 1505 (2013).
- [5] R. P. Sanches and E. Massad, Infect Dis Model 1, 88 (2016).
- [6] Q. H. Liu, M. Ajelli, A. Aleta, S. Merler, Y. Moreno, and A. Vespignani, Proc Natl Acad Sci U S A 115, 12680 (2018).
- [7] Data for this work were collected from Protezione Civile (Italy, <http://opendatadpc.maps.arcgis.com/apps/opsdashboard/index.html#/b0c68bce2cce478eaa-c82fe38d4138b1>); Ministerio de Sanidad (Spain, <https://covid19.isciii.es>); Ministerio de Salud (Argentina, <https://github.com/SistemasMapache/Covid19arData>), and aggregator sites such as COVID19info.live
- [8] H. Nishiura, N. M. Linton, and A. R. Akhmetzhanov, medRxiv, doi:10.1101/2020.02.03.20019497 (2020).
- [9] <https://epiforecasts.io/covid/posts/global/>
- [10] X. He et al., Nat Med doi:10.1038/s41591-020-0869-5 (2020).
- [11] J. J. Stickel, Computers & Chemical Engineering 34, 467 (2010).
- [12] A. Savitzky and M. J. E. Golay, Anal. Chem. 36, 1627 (1964).