

Briefing Note: Calculation of Effective Reproduction Number, R

Contact: Robert Holmes, Ph.D., Presbyterian Health Services

Date: 5/9/2020

Introduction

The base reproduction number R_0 is defined as the expected number of secondary cases produced by a typical primary case in an entirely susceptible population. However during an epidemic it is more convenient to work with the effective reproduction number, R , which is the actual average number of secondaries per primary case. R reflects the impacts of control measures—such as social distancing—and the depletion of a susceptible population as the epidemic progresses.

Methodology

The effective reproduction number is calculated using the methodology of Wallinga and Teunis (1). The methodology is specifically designed for the situation where the time of onset of symptoms is the only known case data; that is where reliable and wide-spread tracing is absent and R cannot be inferred directly from patient case notes.

Briefly, the methodology estimates R by evaluating the likelihood that case A infected case B given a known distribution of generation time (the average time between onset of symptoms in a primary and the secondaries they infect). The effective reproduction number for case A is the sum over all cases B , weighted by the relative likelihood that the particular case B was infected by case A . The effective reproduction number $R(d)$ for any given day d is then the arithmetic mean of the reproduction numbers $R(i, d)$ of those individual cases i who became symptomatic on that day:

$$R(d) = \text{sum}(R(i, d)) / \text{count}(i)$$

Key advantages of this methodology are:

- It makes no assumptions about the shape of the epidemic. Specifically, it does not assume exponential growth; nor does it involve fitting a specific model of the disease's epidemiology.
- It explicitly allows R to vary over time.
- It is computationally tractable.

Implementation

The model uses an implementation of this methodology in the “R” statistical programming language by Boelle and Obadia (2).

- We assume generation time of 4.7 days (standard deviation 2.9 days) as reported by Nishiura *et al.* (3).
- The model is run separately for the state and each of the five regions.
- The model is always run from the start of the epidemic (first positive cases).

Quality Assurance

The following processes have been adopted to ensure the accuracy and reproducibility of results:

- Initial data entry (cases by county) is a two-person process: one enters data, one checks.
- Throughout the analysis pipeline aggregation checks are in place to ensure that cases are not gained or lost.

New Mexico Medical Advisory Team (MAT)

- The analysis pipeline is built using “defensive programming” principles; namely, it throws an error and halts if any of the aggregation or logic checks fail.
- The daily calculations of R are accompanied by a “quality report”, which is reviewed by a PHS Lead Scientist before the results are disseminated.
- All data and code is archived so that any analysis can be “wound back” in time to any previous reported result, and that result reproduced for audit purposes.

Discussion

Are the results reliable or is R sensitive to assumed parameters?

This methodology—like all other methodologies for measuring R —is sensitive to the assumed distribution of the generation time.

The epidemiology of COVID-19 is not yet well-established, so it is possible that the assumed value of generation time will vary from the actual value in a given population; however if R is being used to track the progress of the pandemic—and particularly if it is being used for gating purposes—it is more important to be consistent in the parametrization of the model.

How long does the model take to respond to interventions?

If a social distancing intervention has an immediate effect on the transmission of the disease, it will take approximately one generation period (4.7 days) for this to be reflected in the calculated R .

References

1. Wallinga J, Teunis P. Different epidemic curves for the Severe Acute Respiratory Syndrome reveal similar impacts of control measures. *Am J Epidemiol* 2004;160:509-516.
2. Boelle P, Obadia T. R_0 : Estimation of R_0 and Real-Time Reproduction Number from Epidemics. 2015. R package version 1.2-6. <https://CRAN.R-project.org/package=R0>
3. Nishiura H, Linton NM, Akhmetzanov AR: Serial interval of novel coronavirus (COVID-19) infections [https://www.ijidonline.com/article/S1201-9712\(20\)30119-3/pdf](https://www.ijidonline.com/article/S1201-9712(20)30119-3/pdf)