Challenges of modelling COVID-19

ROBERTO A. KRAENKEL

Instituto de Física Teórica – Universidade Estadual Paulista-UNESP, São Paulo, Brazil
Observatório COVID-19 BR
E-mail: roberto.kraenkel@unesp.br

Abstract. This contribution discusses the challenges involved in modelling the COVID-19 epidemic based on the Brazilian experience. Data availability, its treatment, and integration into models will be briefly described. Challenges in building indicators used in modelling an ongoing epidemic will also be analysed and its use will be critically discussed.

Keywords. Mathematical models; epidemics; COVID-19.

PACS Nos 87.10.+e; 87.19.Xx; 05.90.+m

1. Introduction

COVID-19 has been a burden for public health on a global scale with consequences reaching economic, social and human dimensions unseen over the last hundred years. Brazil is one of the hardest hit countries by the pandemic with more than 100,000 deaths so far. The public health system has been under severe stress, and responses have been decided at regional and local levels, creating different scenarios according to state or municipality.

A particular challenge of COVID-19 is the modelling and quantitative analysis in general of an epidemic as it unfolds, that is, outbreak modelling. In such a situation, analysis has to address problems related to immediate data availability and the construction of indicators that can be updated in quasi real-time. In what follows I will discuss several issues connected with this context of modelling an epidemic on the go arising from the Brazilian context but also valid for many other countries, in special LMIC.

2. Data

Data availability is a major problem in many countries, but one should know beforehand what data are we looking for. Browsing websites will easily take you to many data plots on the number of new cases per day in a given place. Most of the times the date associated with a new case will be the notification date. Here begin the problems, as this depends not only on the dynamics of the epidemic but also on the notification system. The standard rule is to use the date of onset of symptoms, but if you rely on aggregated data, you might not have it. The best situation is to have access to a dataset where each line is a patient and you have all the information about the course of his/her infection, except for name and address.

Number of cases notoriously depend on the testing effort. More test, more cases. So, how can one distinguish between a real increase of incidence from an increase of testing efforts? One way is to have the data on the number of tests performed every day on a given population. In the Brazilian case, this data is just not available. Another way to face this situation is to restrict the analysis to severe hospitalized cases if these patients are all tested. That being the case, analysis can be performed on this restricted dataset, which provides a good proxy of the dynamics of the epidemic in the population.

3. Indicators

During an epidemic it is important to get a grip on the progression of the infection in the population. The first indicator would be just the number of new cases per day. But if you are using the date of onset of symptoms as the date associated with a case, you lack the information of the last days, as they have not been reported. The importance of this problem will depend on the delay between onset of symptoms and notification. If it is small, say, two or three days, it represents no big problem, but if it is long – like the Brazilian case, where it can take weeks to report a case – the use of raw numbers will invalidate any analysis. In order to solve this problem one resorts to a technique called nowcasting, a Bayesian statistical technique that produces estimates of the number of
cases at a given date. This kind of technique is now well developed for infectious diseases dynamics [1].

Another common indicator is the reproductive number of the epidemic. This is the average number of secondary infections arising from a primary one. It is denoted by $R_{\text{eff}}$ or $R(t)$. If $R_{\text{eff}} > 1$ the infection is spreading, and if $R_{\text{eff}} < 1$, it is receding. There are several ways to calculate it. One is to fit an epidemic model to the data and infer it from the model’s equations. The disadvantage is that this depends largely on the model parameters and fitting. However, there exists a way to calculate $R_{\text{eff}}$ from data without resorting to compartmental models due to Wallinga and Teunis [2] and further developed by Gostic et al. [3]. The ingredients are the time series of the number of new cases per day and the distribution of the serial intervals. Serial interval is the time lapse between the onset of symptoms in the primary infector and the onset of symptoms in the infected. The number of new cases per day (by onset of symptoms date) is obtained by nowcasting the raw data. The distribution of sequential times is usually taken as a standard one, obtained by studies at the beginning of the pandemic.

4. Models

Compartmental models that usually go under acronyms like SIR, SEIR, ... are of widespread use in infectious disease modelling. What is their use in COVID-19 and what can they provide? The answer will depend on the purpose of the model. If we aim to explore a concept as a preliminary step for a more realistic model, a simple one can be informative. A recent discussion about the effect of heterogeneities in susceptibilities and infectivities provides a good example. A simple model gives a kind of proof of concept, while more elaborate models rely upon providing fits to real data [5, 6].

Models that are used in outbreak dynamics have to be quite detailed. This includes at least taking into account the age structure of the population. In the case of COVID-19 a separation of the infectious compartment into several others is usual: asymptomatic, mild symptoms, severe cases are all different compartments. If hospital and ICU availability is an issue, the severe cases should be further divided into sub-compartments separating cases where ICU is available or not. An example of a model of the kind is the CoMo model developed by The COVID-19 International Modelling Consortium [4]. The reason to increase the complexity of the model comes from the need to provide answers to some important question. For instance, if one aims to explore the effect of school closures, the model has to include age structure as the intervention is age dependent. If one looks for the effects of increasing ICU availability on lethality, one needs compartments like the ones described above. Fitting the model to data also requires adequate compartments. Fitting is usually performed with the number of deaths and hospitalizations, which have to be present in the model.

5. Conclusion

A final word should be said about the use of the indicators and models in advising public health decision makers and the public in general. Indicators such as $R_{\text{eff}}$ are never to be taken in an isolated way. Taking into accounts trends and other indicators (number of new cases and deaths, for example) is crucial. The same precaution is to be taken with model results. Models do not predict the future, they can give us comparative scenarios under different conditions, such as non-pharmaceutical interventions. COVID-19 teaches us to be cautious with too strong statements that risk being discredited the next week, causing harm to the public perception of science in general.

References