On the challenges of generating epidemic metrics from mathematical models

Dr Jasmina Panovska-Griffiths
The Big Data Institute, Nuffield Department of Medicine, University of Oxford, Oxford

The effective reproduction number R, as a headline epidemic metric since the onset of the COVID-19 pandemic, measures the number of secondary infections arising from an existing infection. At the onset of a new disease, in a fully susceptible population, R is the basic reproduction number R0 that describes the number of secondary infections stemming from an initial case. As the epidemic progresses, R reflects the number of secondary infections generated in a population consisting of susceptible, exposed and immune individuals. The growth rate r represents the rate at which the epidemic is growing during the exponential phase of epidemic growth. While R is reflective of the current level of transmission, r is reflective of the transmission speed.

Models fitted to data have been widely used to generate metrics to inform the epidemic status. R increasing above 1, and analogously, r above 0, suggest that the epidemic is growing exponentially with the emerging virus spreading fast. Considering these thresholds can inform if the epidemic is growing or shrinking, or the impact of imposed control measures.

Generating epidemic metrics from models while useful in informing epidemic status, has challenges:
Challenge 1: Understand how to interpret R and r across different models
Challenge 2: Understand how R and r are statistically corelated within and across different models
Challenge 3: Understand whether R and r are the most reliable metrics as the epidemic progresses and different interventions are employed

On the first challenge, although R and r describe broadly similar model outcomes, their exact definition depends on the model structure. For example, in agent-bases models (ABMs) R can be directly counted, in population based SEIR-type models R is the largest eigenvalue of the generating matrix and hence represents a more abstract concept, while in non-mechanistic renewal equation based models R is typically calculated from the incidence level and the generation time of the circulating variant. In fact within models, r and R are related via the generation time of the epidemic: the longer the generation time and the higher the epidemic growth rate, the higher the value of R.

On the second challenge, when generating a combined probability distribution, what are the most relevant representative ranges to use (e.g. 10th to 90th, 5th or 95th or 25th to 75th percentiles?). Or when combining different models, some of which use different data streams or have different model structures, how does the combined estimate relatively weight their contribution? Finally, when combining model outcomes, what meta-analysis statistical approach do we use e.g. fixed or random effects modes?

Finally, on the third challenge, should we in future consider additional metrics, and specifically account for hospitalisation rate in combination to growth rate (via r) and transmissibility of the emerging virus (via R)? And separately, should the gold-

standard epidemic metrics be different if we are considering a number of local/regional epidemics that merge to produce a large national epidemic, compared to having a slower growing but geographically large epidemic? It is important to address these, and similar challenges, as we continue utilising mathematical models to study and inform the pandemic status.