Discussion comments – Adam Kucharski

Understanding the extent of disease transmission – and how it changes over time – is a deceptively difficult problem, and these papers highlight some key considerations in how we should estimate and interpret COVID transmission dynamics.

Lourenzo Pellis and colleagues nicely outline the conceptual choices that must be made in estimation of reproduction numbers. Even if we were to be omniscient, and see every contact and every transmission event, we would still have to make conceptual decisions about whether we are looking back at transmission that has already occurred, or predicting ongoing transmission from current infectious individuals, which is subject to right censoring. We would also have to average across these events, implicitly giving a weighting to the sub-outbreaks that make up the wider epidemic. As the authors note, in reality, we have the further challenge of partial observation – as illustrated by contact tracing studies that typically report a much lower R than the observed epidemic trajectory implies [1], and in turn discrepancies between observed case numbers and likely underlying epidemic patterns [2].

Then there is the question of how R estimates are used – are they designed to track the current epidemic, predict its near-future trajectory, or evaluate historical changes in transmission? In practice, a model might be used to do all three, but the same tool won't necessarily be optimal for all of these applications. Both papers note that the transmission process implicit in R estimation can make it slower to respond to sudden changes. R has a clear epidemiological interpretation, which can help in interpretation of control impact – especially if we separate effects of immunity and other factors reducing R, as Pellis et al describe with their R_c breakdown. But this reliance on an inherent transmission assumption can lead to challenges in interpretation, particularly in rapidly changing epidemics that take several generations of infection to settle to equilibrium. The example of schools is highlighted in both papers – opening and closing schools may in the short-term lead to transient effects that don't reflect the underlying transmission that would be observed after convergence to the dominant eigenvector, especially if the epidemic is growing from low levels. Analogous dynamics have been observed for influenza pre-COVID [3], with new strains imported by older age groups travelling internationally before spreading locally in wider segments of the population.

In their analysis of school closures, Bekker-Nielsen Dunbar and Held use case time series directly, rather than a transformed R estimate. This has the advantage of reducing the smoothing incurred during R estimation, but also creates the challenge of interpreting coefficients, particularly if case reporting varies over time – should rapidly varying case numbers be interpretated as the result of sudden changes in transmission, or variability in how cases are reported? Although we can use wider data to characterise the observation process to some extent – and this is something that is typically under-studied compared to transmission modelling – the lack of a 'ground truth' for R or force of infection means we are reliant on our choice of process and observation model. In turn, that raises the issue of interpretability versus predictive power. If our aim is to make mechanistic predictions comparing hypothetical control options, we need to understand how various components of transmission are changing. However, if our aim is simply to make reliable predictions about impacts of past interventions from available data, we may be more willing to do so at the expense of interpretability.

The two papers emphasise the challenge of communicating both statistical and conceptual uncertainty. To cite the example given by Pellis et al, should a 51% estimated probability that R is above 1 be sufficient to make a policy decision? Additional metrics such as infection prevalence, hospital occupancy, growth rates, superspreading events, and predicted cases are suggested by the papers' authors as ways to provide further context. However, as the rise of the Delta variant in the UK during May 2021 has illustrated, the potentially most valuable information in real-time is often the most recent, and hence the least certain. What's more, the combination of heterogeneous immunity, changing contact patterns and variation in subpopulations of spread can limit the extent to which early data allows estimation of a pure biological transmission advantage (i.e. larger R_0), rather than just a simple a growth advantage (i.e. r) or a multiplicative increase in R_e in areas where variants are spreading.

Decisions around how to combine metrics are likely to become harder as COVID-19 vaccination decouples the relationship between available datasets. Ultimately this a good problem to have, with social contacts no longer leading to hospitalisations and deaths. But it will also create further challenges for generalisability – the papers focus on UK and Switzerland respectively, and the spread of variants and vaccines globally will increasingly create a need for tailored local analysis of transmission, whether tracking current spread or evaluating interventions.

References:

1. PHE Technical Report 14, 3 June 2021
2. Pung et al. Relative role of border restrictions, case finding and contact tracing in controlling SARS-CoV-2 in the presence of undetected transmission. MedRxiv, 2020
3. Bedford et al. Global circulation patterns of seasonal influenza viruses vary with antigenic drift. Nature, 2015