In context: **Exploring surveillance data biases when estimating the reproduction number: with insights into varying subpopulation transmission in the first Covid-19 outbreak in England**

*Katharine Sherratt, Sam Abbott, Sebastian Funk*

*Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine*

In the UK, public policy and the media have prominently used the effective reproduction number (Rt) of Covid-19 to summarise ongoing pandemic transmission. Several teams in the UK have been contributing estimates of Rt that are aggregated into a consensus range, but the methods, approaches, and data sources for estimating transmission have varied among teams and over time. For example, data sources could, amongst others, include counts of test-positive cases, hospital admissions, or deaths due to Covid-19. In our team’s submissions to the Scientific Pandemic Influenza Group on Modelling (SPI-M) from March onwards, we saw that even when using a consistent method, Rt estimates were not a single, clear-cut number, but varied depending on the source of data.

In late May, we started to explore whether these differences in transmission estimates from each data source could be a policy-relevant indicator of biased data sampling and subpopulation epidemics. We first presented a summary of the differences in our team’s Rt estimates by data source to SPI-M as a short note in early June. From June onwards we used all three data sources to estimate Rt and contributed them separately to the weekly reproduction number estimates published by SPI-M and considered by the Scientific Advisory Group for Emergencies (SAGE). Over this time, we have adapted our work to support the changing UK policy context. This has meant there are several differences in available data, methods, and implications of this work between the time we first generated the SPI-M report and the time of this publication.

As Covid-19 data became more openly accessible, we started to publish a daily comparison of UK Rt estimates by data source (epiforecasts.io/covid/posts/national/united-kingdom). This had initially been impossible as there were very few sources of public subnational data. Thanks to the Public Health England dashboard (coronavirus.data.gov.uk), public data sources for England increased in both quantity and quality and from October we were able to produce subnational Rt estimates using a variety of public data sources. We felt that presenting these estimates publicly would be useful given the high level of interest in the government’s claimed use of Rt as a policy decision tool.

Between generating the original SPI-M submission and this publication, we significantly developed and improved the software we have built to estimate Rt (“EpiNow2”). We continue to refine our methods for estimating Rt, although the improved methods did not substantially change the trend or direction of differences between estimates and our resulting conclusions.

Our interpretation of the differences in Rt estimates has changed over time as we saw new evidence for concentrated transmission in subpopulations. In the earliest paper presented to SPI-M, discussion centred on the likely effects of hospital-acquired infection and testing availability on differences between Rt from test-positives compared to admissions or deaths over March and May. However, increasing evidence for a widespread and severe epidemic in care homes provided an alternative explanation for such differences. We realised that, even without disaggregated data by age or residence, simply identifying the differences in Rt estimates could have been an early indicator of the epidemic in this vulnerable subpopulation. We therefore continued to track these differences, which once again became wider over the summer as transmission moved between age groups after restrictions were lifted and mass testing became available.

Most importantly, we continue to find new insights into the state of the UK pandemic from comparing Rt estimates. One of the clearest trends we have seen in varying Rt estimates by data source has followed from the National Health Service vaccination campaign. Rt estimates from deaths are now consistently below those from hospitalisations and cases. This is a strong indicator of the positive impact of vaccination, and an encouraging further use for this work.