

# What is 80% good enough for real-time analyses during the COVID-19 pandemic?

## Abstract

### Main text

“An 80% right paper before a policy decision is made is worth ten 95% right papers afterwards, provided the methodological limitations imposed by doing it fast are made clear.”

This quote from Whitty et al. [Whitty\_2015] captures much of the academic discourse around real-time analyses conducted to inform COVID-19 health policy. Here we use examples from our policy adjacent COVID-19 work to explore the issues leading to analysis being “80% correct” and highlight the factors that led to these issues. We also discuss what prevents discovery and reporting of methodological limitations and at what threshold a real-time analysis is of a high enough standard to be used to inform health policy.

Collectively, we have published numerous studies related to the COVID-19 pandemic, have contributed studies to government advisory committees, presented results to the WHO and other bodies, and have worked on previous infectious disease outbreaks with similar issues such as the 2013-2016 Ebola outbreak in West Africa. Our work spans non-parametric statistical modelling of transmission, short-term forecasts, scenario modelling using mechanistic compartmental models, and a range of reactive analyses usually conducted with short-time horizons to inform policy-makers.

The first example of our work we discuss is a large-scale project to estimate the effective reproduction number of COVID-19 in real-time [Abbott2020v1; Abbott2020v2]. Starting in February 2020 with a focus on China and growing to include national estimates globally as well as estimates at smaller scales in multiple countries with a range of data sources. This work has fed into the Scientific Pandemic Influenza Group on Modelling (SPI-M) aggregated reproduction number estimate and short-term forecasts [sherratt\_exploring\_2021; funk\_short-term\_2020] and the site where estimates are presented has had over 400,000 unique users. As based on surveillance data, done at large scale, and repeated on a routine basis this work has many obvious limitations. In our published work we reported a range of these including the lack of location-specific data on the time from infection to case report, lack of clarity on the interaction between the generation time and the reproduction number, and the difficulty in extrapolating current reported cases to infer the dynamics of current infections. However, our initial work also included an unknown, and hence unreported, limitation in that it assumed that the delay from infection to report could be treated as reversible when this in fact introduces bias [Abbott2020v1]. This bias has since been investigated extensively and has been shown to result in over-smoothed estimates that lag real-world changes in transmission [gostic]. Though this issue has since been mitigated via new methodological developments [Abbott2020v2] it is likely that it resulted in flawed inferences being drawn from the estimates. An additional unacknowledged limitation of this work is the difficulty in maintaining and running a large scale, compute intensive, project over a prolonged period of time for which the best data sources also differ by location.

A critical issue that influenced the initial flawed implementation was a lack of knowledge of developments in certain areas of the literature (in this case the HIV literature) [gostic]. This was potentially partly the result of a lack of experience within our team (as much of the development work was done by early career researchers) but also the result of the failure of those working in the areas where this had previously been explored to propagate robust tooling and methodology to allow others to mitigate the issue. This was likely exacerbated by the difficulty in developing tools whilst also trying to deploy them. A wider issue with this project was its continuous nature and lack of discrete outcomes which made incentivising substantial

contributions difficult, especially once the initial work had been published. In theory this could have been resolved by building collaborations with those developing new methodology as contributors to the project but in practice this was difficult to negotiate despite significant time and resources being spent on this aim. Unfortunately whilst a range of real-time reproduction number estimations now exist in general these have been only partially evaluated and they are rarely designed with reusability, robustness, or routine usage in mind.

A related example is two linked studies that were conducted in December 2020 and June 2021 to estimate the transmissibility advantage of the Alpha and Delta variants of COVID-19 respectively [Davies2021; AbbottDelta]. Both of these analyses used the same underlying regression framework to model reproduction numbers from the first example as a mixture of the variant of concern and the other variants reproduction numbers (with this only being a small part of the analyses in [Davies2021]). This allowed for a multiplicative transmission advantage to be estimated after adjusting for static and time-varying variables influencing transmission. These analyses were presented to SPI-M and the Scientific Advisory Group for Emergencies 91 (SAGE) with the first being widely reported in the media, often as a point estimate. These analyses reported a range of limitations including the use of the expected reproduction number and proportion of cases that had the variant of concern, the use of S-gene target failure as a proxy for variant status, the lack of UK specific generation time estimates, the lack of modelling spatial auto-correlation explicitly, the usage of non-parametric reproduction number estimates rather than a single model, and the assumption that the generation time was unchanged between variants. These issues were likely to have led to spuriously precise results though the magnitude of this is difficult to quantify. Bias may have been introduced by using the mean estimated reproduction numbers and mean proportion of cases with the variant of concern as this weights all areas equally. Bias may also have been introduced due to the assumption that the generation time was unchanged.

Whilst the issues with these studies are perhaps less critical than in the first example as they were unlikely to have negatively impacted health policy decision making. They do, however, highlight a general theme in COVID-19 analysis, quantification of uncertainty[zelner] or lack thereof. It is difficult to assess the impact of these omissions, without replication of these studies with improved methods. In this instance this is a work in progress[Abbott:pr] but in general this is rarely done likely because it is difficult to justify as novel work and because it can require substantially more complex methods. All the limitations present in this example were due to lack of capacity, both in terms of researcher time and technical infrastructure such as compute resources, and lack of researcher experience. It is important to note that the analysis was repeated effectively unchanged when concerns were raised about the Delta variant despite several months having passed in which improvements could have been made. This was again largely due to lack of capacity as other questions of interest had been raised in the interim but given the likely development of new variants in hindsight this may have been an oversight. The limitations of this example were mitigated in the first instance by including the analysis in a collection of other estimates derived using various techniques [Davies2021] and in the second instance by comparison to a range of independent studies when aggregated at SPI-M. Unfortunately however many of these approaches made similar simplifications to streamline analysis and hence were likely biased and spuriously precise in the similar areas. Within these studies some of the limitations were partially addressed by reporting results for a range of models and a range of generation times though this was difficult to communicate to those making use of the estimates.

**Example from Adam** Suggestion: Delta transmission advantage from travelers

**LSHTM SPI-M scenario modeling** Rosie and Sam

**What**

- Renewal equation based method as an extension of the first example with the addition of an adjustment for the proportion of the population that were susceptible.
- Compartmental model approach with significant internal model complexity
- Submitted to SPI-M by independent teams with the renewal equation based method no longer being used from January 2021

## Known limitations

## Unknown limitations

## Causes

- Independent data, computational, and validation pipelines within LSHTM with little to no overlap.
- Independent pipelines across groups

## Mitigation

- Scenarios were modelled by multiple independent teams and then collaboratively discussed potentially reducing the impact of issues. However, it is likely that the issues we experienced were also problems for other groups which may have reduced the benefit of this ensemble approach as all models likely to make similar simplifications.

## Conclusions

- Lack of breadth of knowledge in small research groups and difficulty in collaborating across groups on rapidly developing problems. This lack of knowledge includes both domain area expertise, statistical knowledge, and technical knowledge such as knowledge of software design, cloud computing, and task scheduling.
- As noted elsewhere this issue is exacerbated by the persistent myth of the lone researcher and the encouragement of this ideal by funding bodies [kucharski\_covid-19\_2020]. ECRs are actively encouraged to pursue their own research interests with little reward for working collaboratively as part of a wider effort, we think that this likely negatively impacted the quality of our work.
- A persistent issue in academia more widely is the need for all work to be “novel”. In the context of real-time analyses, this means that little follow-up work is done to evaluate the quality of real-time analyses.
- Many of the currently available, and commonly used, software tools require a high level of user knowledge to use appropriately. For work with short timelines, it may be difficult to produce results that are of sufficient quality without more tools customized to common tasks and a greater emphasis on learning how to use the currently available tools in detail.
- All of the examples in this letter have made a point of providing fully open-source code and in the case of the first two example open-source software to allow others to use the methodology. In theory this should allow other researchers to explore the analysis and mitigate some of the limitations but in practice the incentive structure of science instead favours new independent work and so these incremental lessons are not learnt.
- Though not the case for the example discussed here precise estimates are often reward by publication in higher impact journals and hence fully including uncertainty can be difficult to motivate.

## References