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.[1] captures much of the academic discourse[2] 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 an analysis being "80% correct" and discuss the factors that led to these issues.

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[3,4]. 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[5,6] and the site where estimates are presented has had over 400,000 unique users. As based on surveillance data, done at a 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 introduces bias[3]. This bias has since been investigated extensively and has been shown to result in over-smoothed estimates that lag real-world changes in transmission[7]. Though this issue has since been mitigated via new methodological developments[4] 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 which led to substantial simplifications and approximations being made.

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)[7]. This was potentially 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. A wider issue with this project was its continuous nature and lack of discrete outcomes which made incentivising 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[8,9]. Both of these analyses used the same underlying regression framework to model reproduction numbers 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[8]). 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 (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 between variants.

The issues with these studies are perhaps less critical than in the first example. They do, however, highlight a general theme in COVID-19 analysis, quantification of uncertainty[10] 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[11] 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 statistical expertise. 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[8] 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 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 3

Example 4

The examples we have highlighted all share several common themes with the most common being limited breadth of knowledge in small research groups. This lack of knowledge extends from domain knowledge to statistical and technical skills, such as software engineering, required to handle data. These skills are especially important for real-time analyses in which results are required quickly and data sources can be complex. This issue is compounded by the lack of well-developed software tools designed to be used flexibly for real-world analysis questions. Many of the currently available generic open-source tools, whilst very high quality, require a substantial level of user knowledge and for relatively common domain-specific problems can require significant researcher time to produce results. Most of the available domain-specific tools have significant limitations or are not continuously supported. A related issue is the difficulty in collaborating across groups on rapidly developing problems and on method development more generally.

It is plausible that the issues highlighted by the examples discussed here are not generalisable to other researchers and indeed the conclusion that should be drawn may be that we as researchers should not be contributing real-time analysis. However, we feel that the issues highlighted are more systematic than this and point to persistent structural flaws within the academic system as discussed in Kucharski et al[12]. Many of the issues we have highlighted are exacerbated by an academic culture that promotes individual scientific

achievement over collaborative incremental development, does not promote a culture of ongoing technical and statistical training, and indeed does not reward or value these skills in applied researchers. All of the examples in this letter were published with open-source code. In theory, this should allow other researchers to explore the analysis and mitigate some of the limitations but in practice, incentive structures favour new work rather than incremental improvements. We feel it is vital that ourselves and others re-evaluate real-time analysis so that lessons can be learnt for the future. This can most easily be done when researchers are open and honest about the limitations of their work and the causes of these limitations and funders are supportive of this goal.

References

1. Whitty, C. J. M. (2015). What makes an academic paper useful for health policy? *BMC Medicine*, *13*(1), 301. https://doi.org/10.1186/s12916-015-0544-8
2. Brooks-Pollock, E., Danon, L., Jombart, T., & Pellis, L. (2021). Modelling that shaped the early COVID-19 pandemic response in the UK. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1829), 20210001. https://doi.org/10.1098/rstb.2021.0001
3. Abbott, S., Hellewell, J., Thompson, R. N., Sherratt, K., Gibbs, H. P., Bosse, N. I., Munday, J. D., Meakin, S., Doughty, E. L., Chun, J. Y., Chan, Y.-W. D., Finger, F., Campbell, P., Endo, A., Pearson, C. A. B., Gimma, A., Russell, T., Flasche, S., Kucharski, A. J., . . . Funk, S. (2020). Estimating the time-varying reproduction number of SARS-CoV-2 using national and subnational case counts [version 1]. *Wellcome Open Research*, *5*, 112. https://doi.org/10.12688/wellcomeopenres.16006.1
4. Abbott, S., Hellewell, J., Thompson, R. N., Sherratt, K., Gibbs, H. P., Bosse, N. I., Munday, J. D., Meakin, S., Doughty, E. L., Chun, J. Y., Chan, Y.-W. D., Finger, F., Campbell, P., Endo, A., Pearson, C. A. B., Gimma, A., Russell, T., Flasche, S., Kucharski, A. J., . . . Funk, S. (2020). Estimating the time-varying reproduction number of SARS-CoV-2 using national and subnational case counts [version 2]. *Wellcome Open Research*, *5*, 112. https://doi.org/10.12688/wellcomeopenres.16006.2
5. Sherratt, K., Abbott, S., Meakin, S. R., Hellewell, J., Munday, J. D., Bosse, N., null, null, Jit, M., & Funk, S. (2021). Exploring surveillance data biases when estimating the reproduction number: With insights into subpopulation transmission of COVID-19 in England. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1829), 20200283. https://doi.org/10.1098/rstb.2020.0283
6. Funk, S., Abbott, S., Atkins, B. D., Baguelin, M., Baillie, J. K., Birrell, P., Blake, J., Bosse, N. I., Burton, J., Carruthers, J., Davies, N. G., Angelis, D. D., Dyson, L., Edmunds, W. J., Eggo, R. M., Ferguson, N. M., Gaythorpe, K., Gorsich, E., Guyver-Fletcher, G., . . . Investigators, I. (2020). *Short-term forecasts to inform the response to the Covid-19 epidemic in the UK* (p. 2020.11.11.20220962). https://doi.org/10.1101/2020.11.11.20220962
7. Gostic, K. M., McGough, L., Baskerville, E. B., Abbott, S., Joshi, K., Tedijanto, C., Kahn, R., Niehus, R., Hay, J. A., Salazar, P. M. D., Hellewell, J., Meakin, S., Munday, J. D., Bosse, N. I., Sherrat, K., Thompson, R. N., White, L. F., Huisman, J. S., Scire, J., . . . Cobey, S. (2020). Practical considerations for measuring the effective reproductive number, Rt. *PLOS Computational Biology*, *16*(12), e1008409. https://doi.org/10.1371/journal.pcbi.1008409
8. Davies, N. G., Abbott, S., Barnard, R. C., Jarvis, C. I., Kucharski, A. J., Munday, J. D., Pearson, C. A. B., Russell, T. W., Tully, D. C., Washburne, A. D., Wenseleers, T., Gimma, A., Waites, W., Wong, K. L. M., Zandvoort, K. van, Silverman, J. D., Diaz-Ordaz, K., Keogh, R., Eggo, R. M., . . . Edmunds, W. J. (2021). Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science*, *372*(6538), eabg3055. https://doi.org/10.1126/science.abg3055
9. Abbott, S., Kucharski, A. J., & Funk, S. (2021). *Estimating the increase in reproduction number associated with the Delta variant using local area dynamics in England.*
10. Zelner, J., Riou, J., Etzioni, R., & Gelman, A. (2021). Accounting for uncertainty during a pandemic. *Patterns*, *2*(8). https://doi.org/10.1016/j.patter.2021.100310
11. Abbott, S., & Funk, S. (n.d.). *covid19.sgene.utla.rt: Recast to a non-linear model.* https://github.com/epiforecasts/covid19.sgene.utla.rt/pull/42
12. Kucharski, A. J., Funk, S., & Eggo, R. M. (2020). The COVID-19 response illustrates that traditional academic reward structures and metrics do not reflect crucial contributions to modern science. *PLOS Biology*, *18*(10), e3000913. https://doi.org/10.1371/journal.pbio.3000913