Situational assessment of COVID-19 in Australia

Technical Report 15 March 2021 (released 28 May 2021)

Nick Golding[1], Freya M. Shearer[2], Robert Moss[2], Peter Dawson[3], Dennis Liu[4], Joshua V. Ross[4], Rob Hyndman[5], Pablo Montero-Manso[6], Gerry Ryan[1], Tobin South[4], Jodie McVernon[2,7,8], David J. Price[2,7], and James M. McCaw[2,7,9]

1. Telethon Kids Institute and Curtin University, Perth, Australia
2. Melbourne School of Population and Global Health, The University of Melbourne, Australia
3. Defence Science and Technology, Department of Defence, Australia
4. School of Mathematical Sciences, The University of Adelaide, Australia
5. Department of Econometrics and Business Statistics, Monash University, Australia
6. Discipline of Business Analytics, University of Sydney Business School, University of Sydney, Australia
7. Peter Doherty Institute for Infection and Immunity, The Royal Melbourne Hospital and The University of Melbourne, Australia
8. Murdoch Children's Research Institute, The Royal Children's Hospital, Australia
9. School of Mathematics and Statistics, The University of Melbourne, Australia

Preamble

This is the fifth technical report (released on 28 May 2021) in a series on COVID-19 situational assessment in Australia. All reports are available at the following link: `https://www.doherty.edu.au/about/reports-publications`.

The previous report was published on 29 July 2020. The focus of the current report is on COVID-19 situational assessment in Australia for the period from early August 2020 up to 16 February 2021. The report is divided into two sections:

- In Part I, we present time-series estimates of key situational awareness metrics, including state-wide transmission potential, the effective reproduction number (R_{eff}) of active cases and macro-/micro-physical distancing behaviour, for each Australian state/territory from 1 March 2020 up to 14 February 2021.

- In Part II, we report on five key epidemiological events that occurred in Australia during the period from August 2020 up to February 2021 and describe the situational analyses conducted at the time. We compare those real-time analyses with the retrospective assessment presented in Part I. In addition to the estimates of transmission potential and R_{eff}, we report and interpret forecasts of daily case incidence, where appropriate.

As of 15 March 2021 (the time of writing of this report), following a period of local near-elimination, the Australian National Notifiable Diseases Surveillance System reported a small number of new locally acquired cases in Queensland and New South Wales.

Part I: Key situational awareness metrics up to 14 February 2021

We use a novel semi-mechanistic model to estimate the ability of SARS-CoV-2 to spread in a population, informed by data on cases, population behaviours and health system effectiveness. Where the virus is present, the quantity we compute is the effective reproduction number (R_{eff}). In the absence of cases, it reflects the ability of the virus, if it were present, to spread in a population, which we define as the 'transmission potential'.

Applying this method provides an estimate of the transmissibility of SARS-CoV-2 in periods of high, low, and zero, case incidence, with a coherent transition in interpretation across changing epidemiological situations. A brief summary of the method is provided below (see also Figure 1) and full details are provided in the Appendix.

We provide time-series estimates of state-wide transmission potential (Figure 2), the effective reproduction number (R_{eff}) of active cases (Figure 3), and macro-/micro-distancing behaviour (Figures 5 and 6) for each Australian state/territory from 1 March 2020 up to of 14 February 2021, based on case data extracted from the Australian National Notifiable Diseases Surveillance System (NNDSS) on 15 February 2021. Time-series estimates of sub-components of these metrics are also provided, including the deviation in transmission from state-wide transmission potential (Figure 4), time-to-case detection (Figure 8), and population mobility (Figure 7). Finally, we include estimates of trends in state-wide transmission potential if we assume that only one of macro-distancing behaviour, micro-distancing behaviour, or the time-to-case detection had changed over time, *i.e.*, counterfactuals (Figures S1, S2 and S3).

Overview of method

We separately model transmission from locally acquired cases (local-to-local transmission) and from overseas acquired cases (import-to-local transmission). We model local-to-local transmission for each Australian state and territory using two components (Figure 1):

1. the average population-level trend in transmissibility driven by interventions that primarily target transmission from local cases, specifically changes in physical distancing behaviour and case targeted measures (Component 1); and

2. short-term fluctuations in R_{eff} to capture stochastic dynamics of transmission, such as clusters of cases and short periods of lower-than-expected transmission, and other factors factors influencing R_{eff} that are otherwise unaccounted for by the model (Component 2).

During times of disease activity, Components 1 and 2 are combined to provide an estimate of the R_{eff} as traditionally measured. In the absence of disease activity, Component 1 is interpreted as the potential for the virus, if it were present, to establish and maintain community transmission (> 1) or otherwise (< 1).

Sub-models for estimating local transmission potential (Component 1)

To estimate Component 1, we use three sub-models (Figure 1, labelled a, b and c). We distinguish between two types of distancing behaviour:

a. macro-distancing, defined as the reduction in the rate of non-household contacts and assessed through weekly nationwide surveys of the daily number of non-household contacts; and

b. micro-distancing, defined as the reduction in probability of transmission per non-household contact, and assessed through weekly nationwide surveys from which we estimate the proportion of the population who report keeping 1.5 m physical distance from non-household contacts at all times.

By synthesising data from these surveys and numerous population mobility data streams made available by technology company Google, we infer temporal trends in macro- and micro-distancing behaviour (sub-models a and b). Furthermore, using data from the NNDSS on the number of days from symptom onset to case notification for cases, we estimate the proportion of cases that are detected (and thus advised to isolate) by each day post-infection. By quantifying the temporal change in the probability density for the time-to-detection (sub-model c), the model estimates how earlier isolation of cases reduces the ability of SARS-CoV-2 to spread. Improvements in contact tracing, expanded access to testing, more inclusive case definitions, and other factors impacting case detection rates, all contribute to this estimated reduction in Component 1.

Estimating the relative transmissibility of SARS-CoV-2 VOC 202012/01

The recent emergence of SARS-CoV-2 variants of higher transmissibility, such as VOC 202012/01 in the UK, is accounted for in our model by an increase in the probability of infection per contact in the transmission potential. We performed an analysis of the relative transmissibility of SARS-CoV-2 VOC 202012/01 compared with non-VOCs in the UK, using:

- data from Public Health England on secondary attack rates among known contacts of cases;

- our model for estimating transmission potential based on macro- and micro-distancing data streams, which separately considers household and non-household rates of transmission;

- data on macro-distancing behaviour (from both the UK and Australia) and mobility and micro-distancing behaviour (from the UK).

This approach allows us to directly estimate the impact of VOC 202012/01 on the probability of transmission to a contact per unit of contact time. Because the relative number of household and non-household contacts and the time spent with these contacts changes under different levels of public health restrictions, our approach estimates increases in transmission potential/transmissibility of VOC 202012/01 relative to non-VOCs ranging from 40% [30, 50] under nationwide "stay-at-home" restrictions in Australia in March/April 2020 to 48% [35, 60] for a pre-pandemic baseline (R_0). See Appendix for details.

From 1 December 2020, we provide adjusted estimates of transmission potential for VOC 202012/01 based our time-series estimates of the relative increase in transmissibility of VOC 202012/01 compared with non-VOC lineages.

Figure 1: Depiction of the relationship between data sources and R_{eff} analysis components.
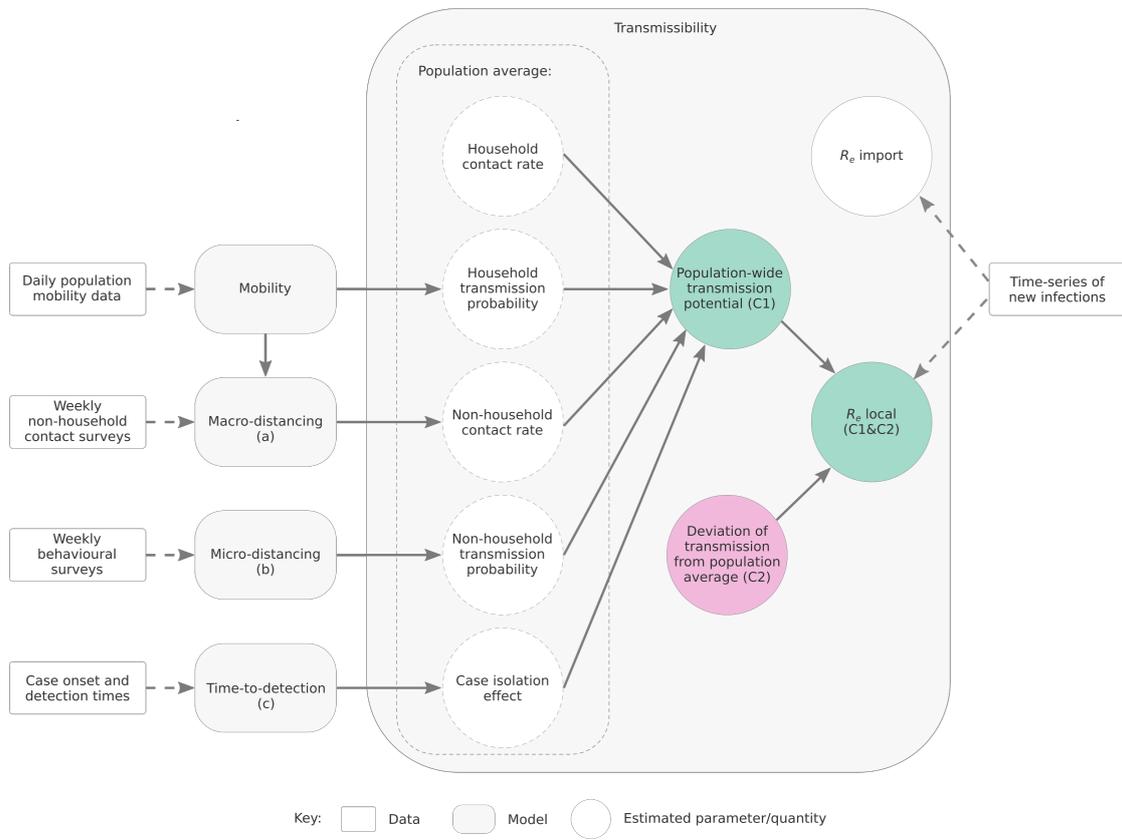
Figure 2: Estimates of state-wide transmission potential (model Component 1) by state/territory up to 14 February 2021 (light coloured ribbons = 90% credible intervals; dark coloured ribbons = 50% credible intervals). We provide two sets of estimates of transmission potential. In green: Estimates of transmission potential of non-VOC lineages. In grey: Estimates of transmission potential of VOC 202012/01, based on the estimated relative increase in transmissibility of VOC 202012/01 compared with non-VOCs, as described in the Appendix, from 1 December 2020. Note: the most recent estimate for VIC in this time-series is informed by behavioural survey and population mobility data collected prior to the activation of stay-at-home restrictions from 13–17 February 2021.
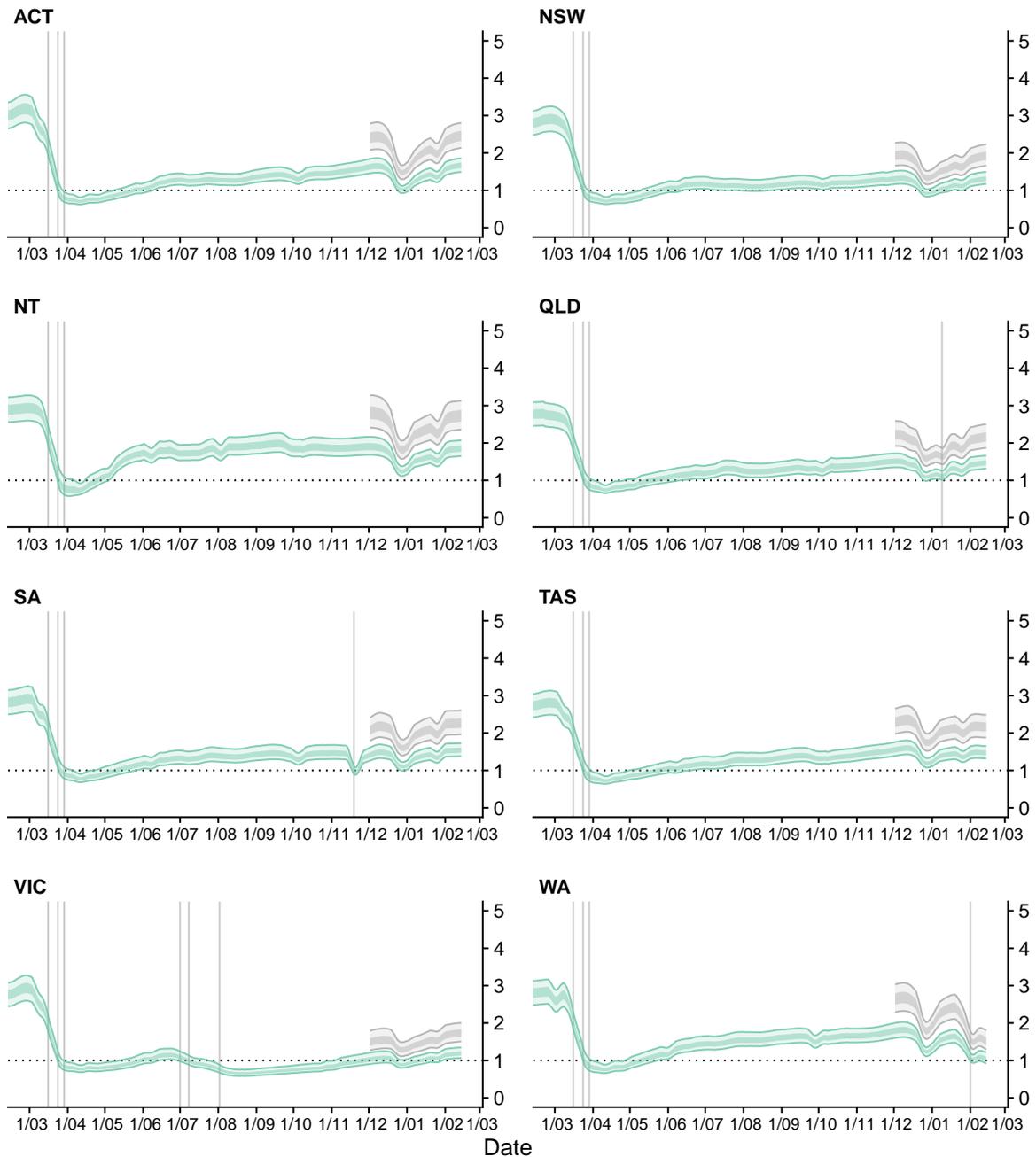
Figure 3: Estimates of R_{eff} for local active cases (model Component 1&2, see Appendix) up to 14 February 2021 for each state/territory (light green ribbon = 90% credible interval; dark green ribbon = 50% credible interval). Solid grey vertical lines indicate key dates of implementation of various physical distancing policies. Black dotted line indicates the target value of 1 for the effective reproduction number required for control. Local cases by inferred date of infection are indicated by grey ticks on the x-axis. For states/territories with very low numbers of local active cases, the estimates of R_{eff} for active cases is highly uncertain. The state-wide transmission potential should be referred to when assessing the risk of an epidemic becoming established given a seeding event.
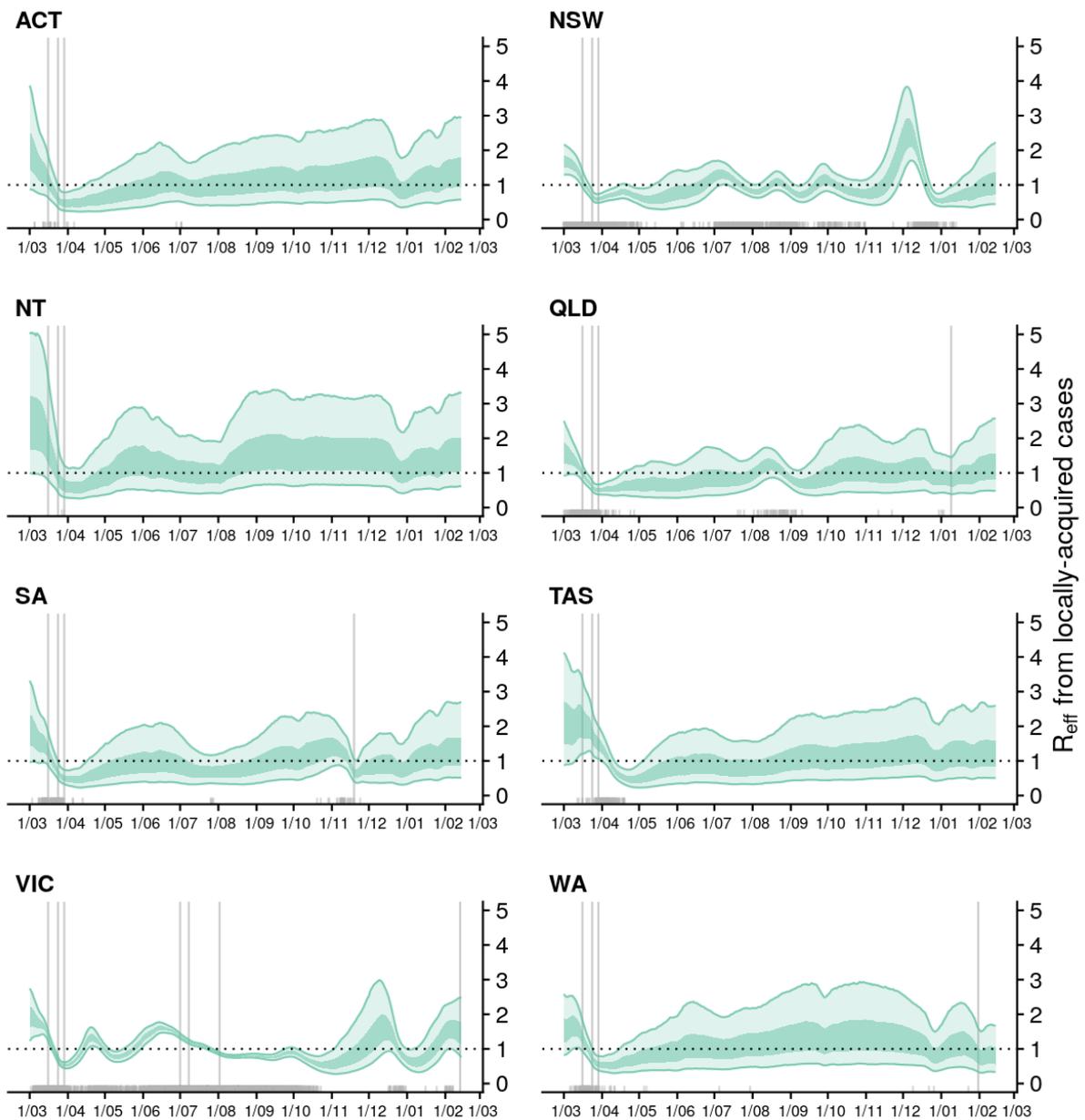
Figure 4: Deviation of transmission potential in local active cases (*e.g.*, clusters) from state-level local transmission potential of **non-VOC lineages** (model Component 2) for each state/territory up to 14 February 2021 (light pink ribbon = 90% credible interval; dark pink ribbon = 50% credible interval). Solid grey vertical lines indicate key dates of implementation of various physical distancing policies. Local cases by inferred date of infection are indicated by grey ticks on the x-axis. When Component 2 is positive, the virus is spreading faster than expected from the estimated transmission potential. Conversely, when Component 2 is negative, the virus is spreading slower than expected from the estimated transmission potential. Note: for the entire time-series deviations are from the local transmission potential of non-VOC lineages. Therefore, in VIC in early February 2021, where there were a small number of local active cases of VOC 202012/01, the negative deviation is an underestimate (*i.e.*, the negative deviation from the transmission potential of VOC 202012/01 would be larger).
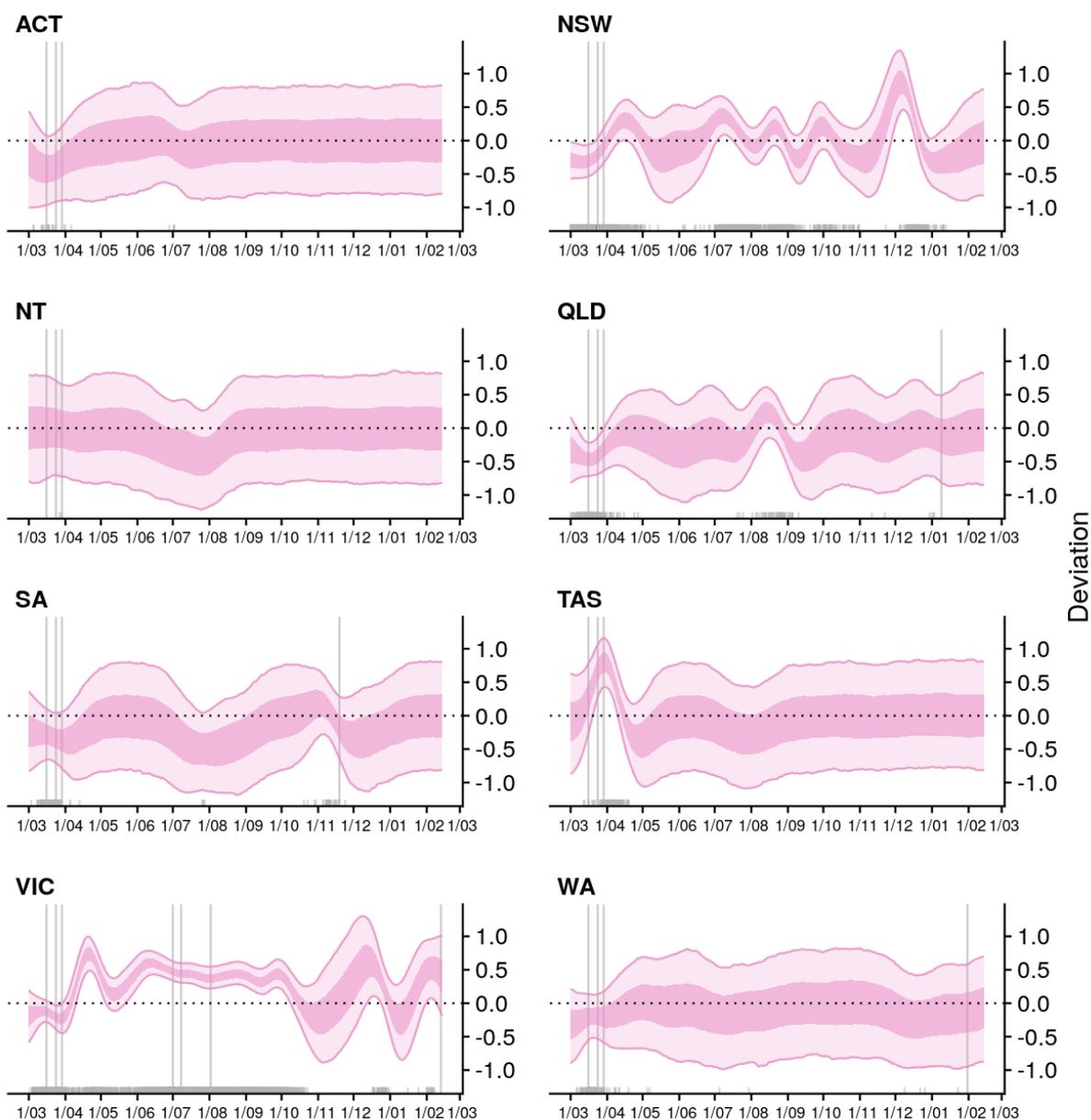
Figure 5: **Estimated trends in macro-distancing behaviour**, *i.e.*, reduction in the daily rate of non-household contacts, in each state/territory up to 14 February 2021 (light purple ribbons = 90% credible intervals; dark purple ribbons = 50% credible intervals). Estimates are informed by state-level data from nationwide surveys (indicated by the black lines and grey rectangles) and population mobility data. The width of the grey boxes corresponds to the duration of each survey (around 4 days) and the green ticks indicate the dates that public holidays coincided with surveys (when people tend to stay home, biasing down the number of non-household contacts reported on those days).
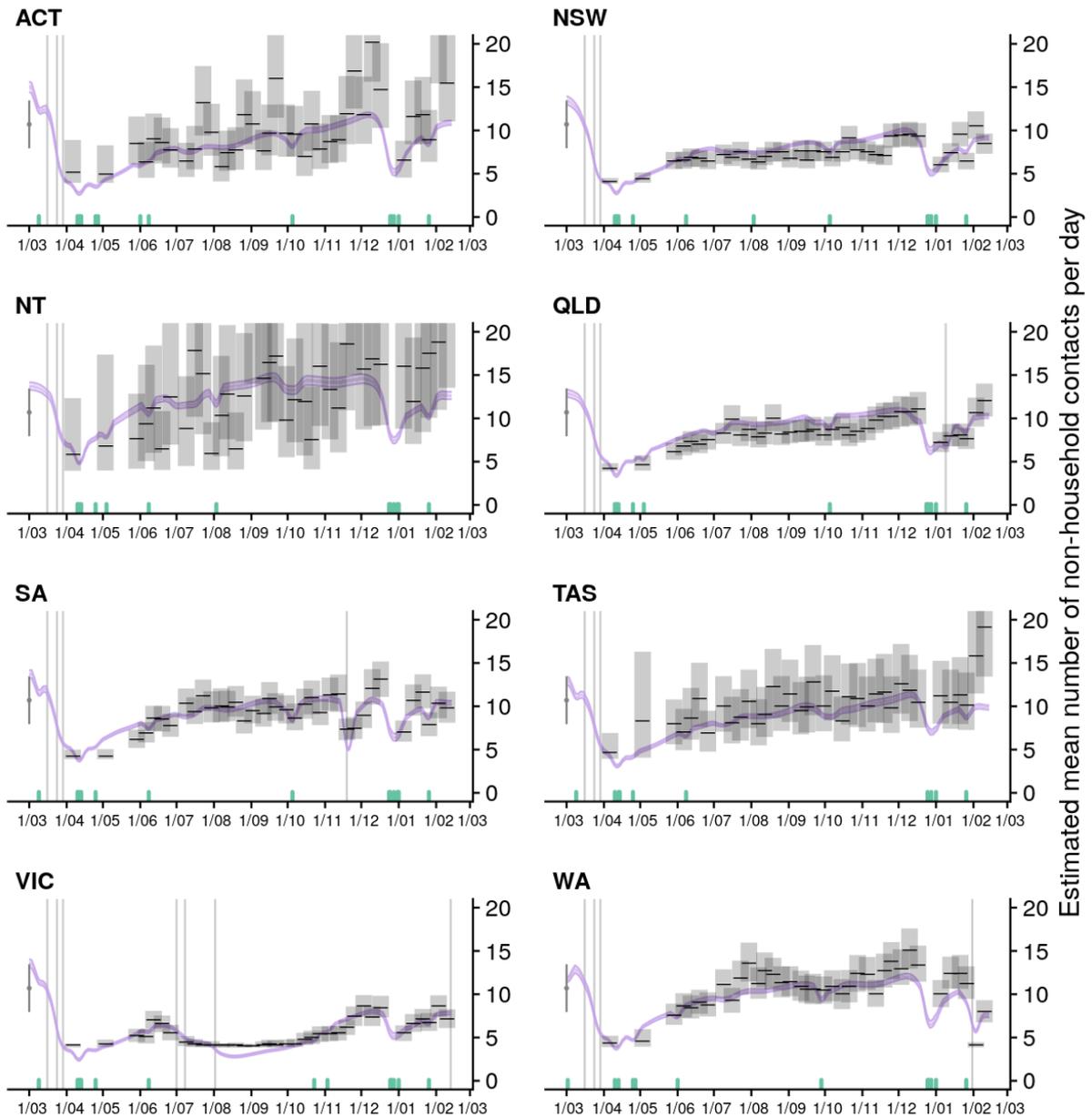
Figure 6: **Estimated trends in micro-distancing behaviour**, *i.e.* reduction in transmission probability per non-household contact, in each state/territory up to 14 February 2021 (light purple ribbons = 90% credible intervals; dark purple ribbons = 50% credible intervals). Estimates are informed by state-level data from nationwide weekly surveys since March 2020 (indicated by the black lines and grey boxes). The width of the grey boxes corresponds to the duration of each survey wave (around 4 days). Note: By February 2021, with the high volume of data included in the time-series, there were some issues fitting the existing model to the earliest parts of the time-series, notably in April, May and June 2020. However, these issues were not present in 2020 and the model performed as required, as shown in Figure S6. Estimates for 2021 remain reliable.
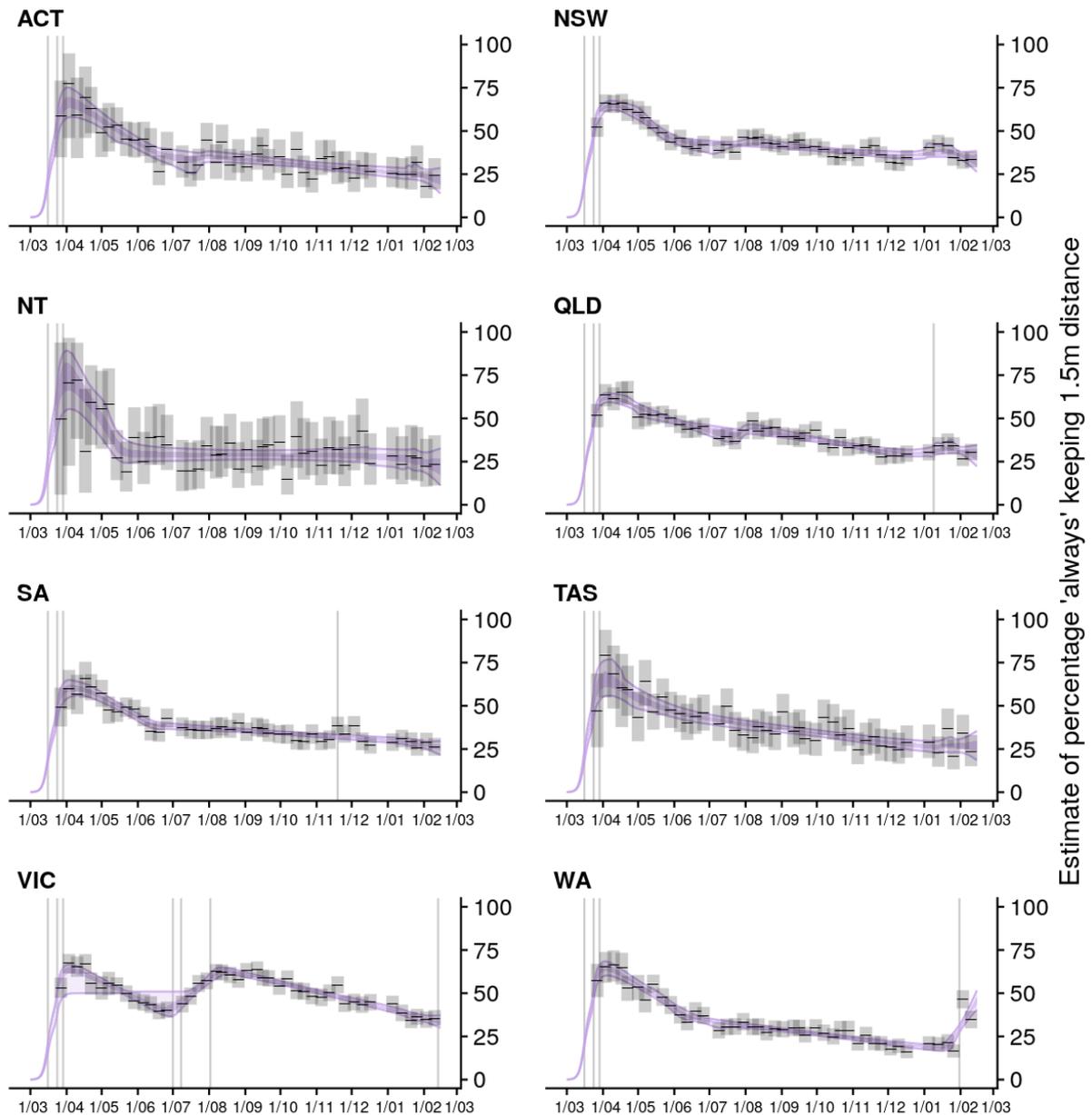
Figure 7: Percentage change compared to a pre-COVID-19 baseline of three key mobility data streams in each Australian state and territory up to 9 February. Solid vertical lines indicate dates of implementation of key physical distancing measures. The dashed vertical line marks 9 February, the most recent date for which some mobility data are available. Purple dots in each panel are data stream values (percentage change on baseline). Solid lines and grey shaded regions are the posterior mean and 95% credible interval estimated by our model.
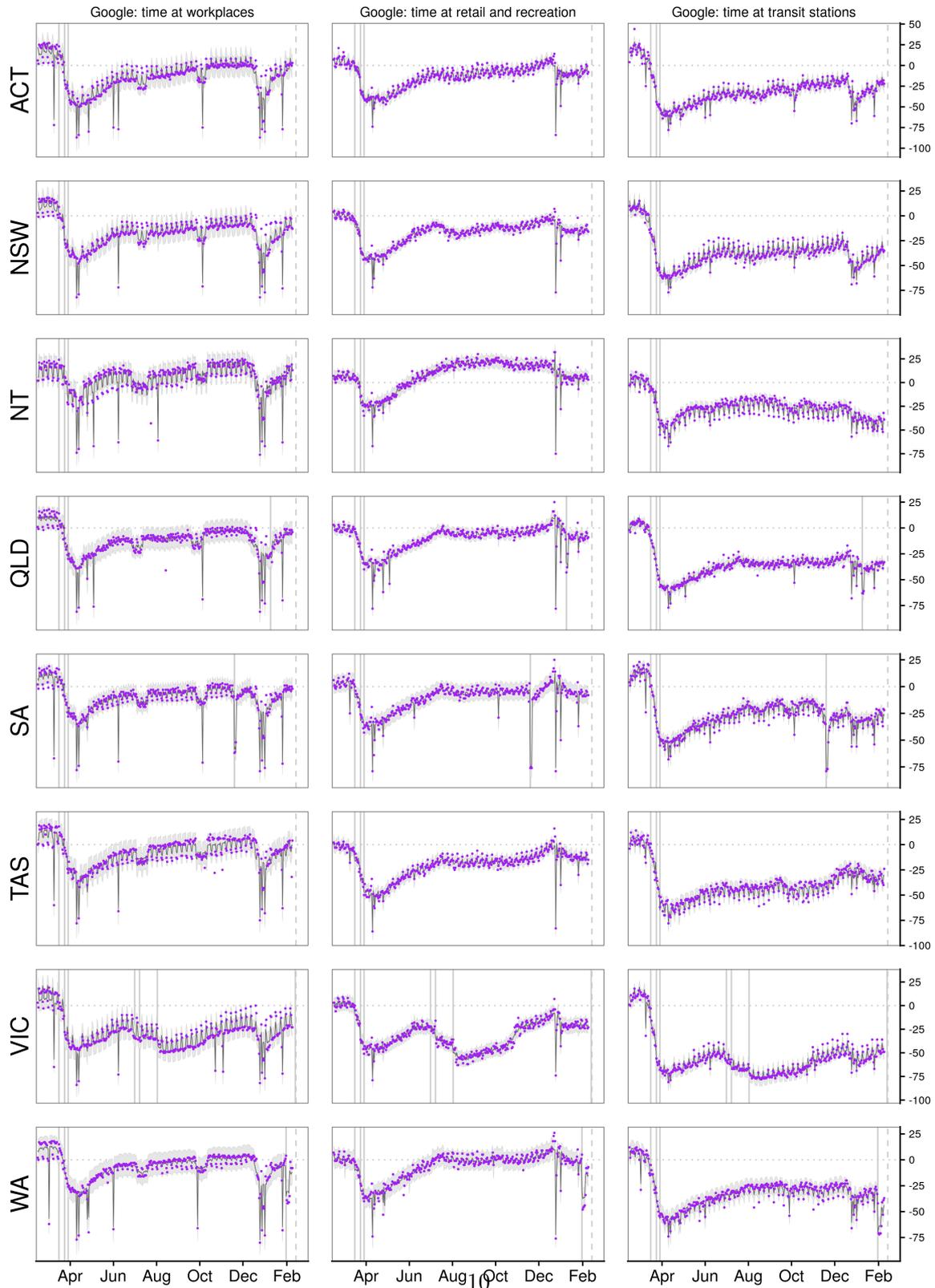
Figure 8: Estimated trend in distributions of time from symptom onset to notification for locally acquired cases for each state/territory up to 7 February 2021 (black line = median; yellow ribbons = 90% distribution quantiles; black dots = time-to-notification of each case). Faded regions indicate where a national trend is used due to low case counts.
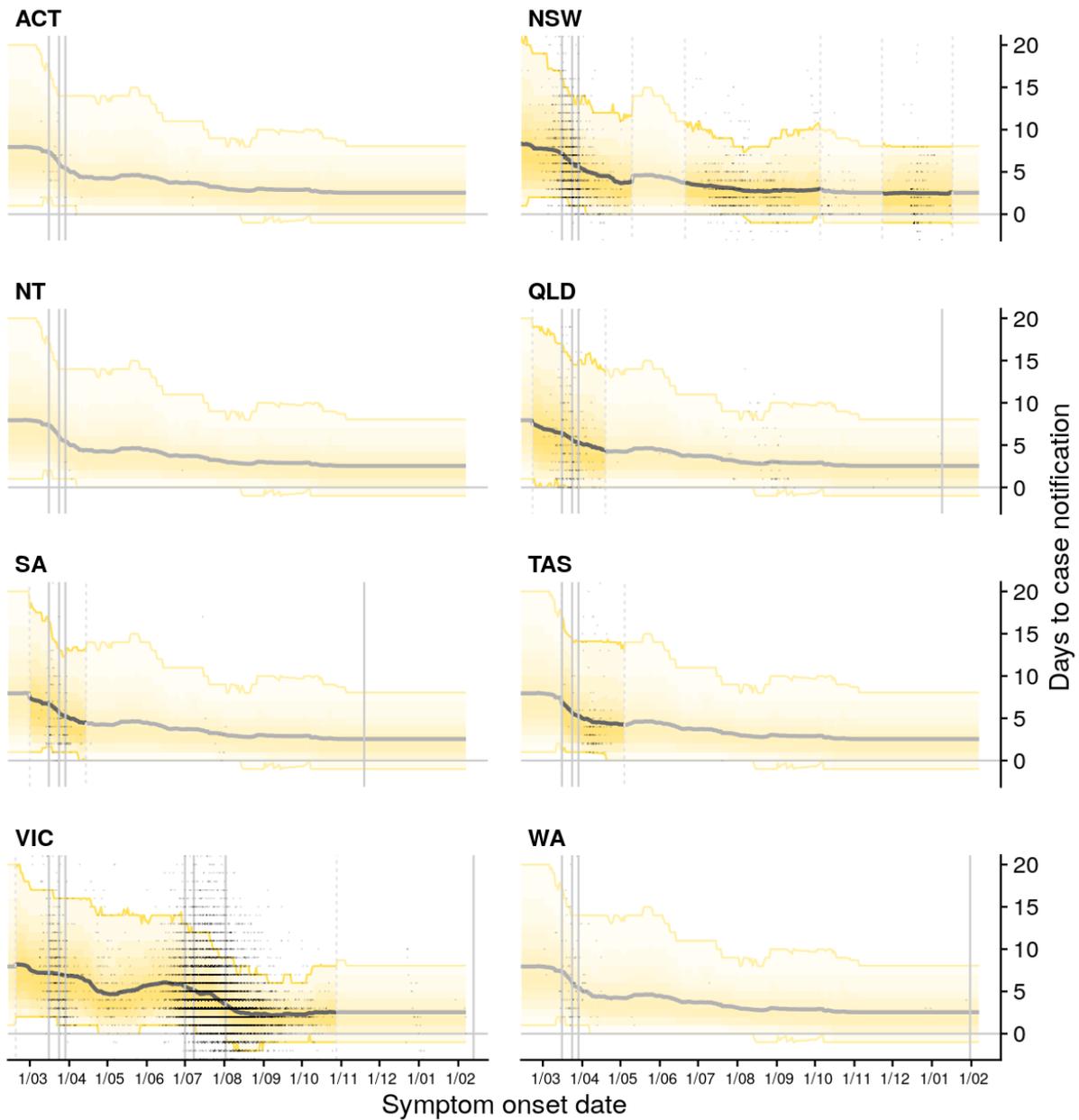
Part II: reporting on key epidemiological events from August 2020 up to February 2021

We report on five key epidemiological events that occurred in Australia during the period from August 2020 up to February 2021 and describe the situational analyses conducted at the time. We compare those real-time analyses with the retrospective assessment presented in Part I above. In addition to the estimates of transmission potential and R_{eff}, we report and interpret forecasts of daily case incidence. We focus on the following five events:

- Declining phase of the Victorian second wave epidemic in September 2020

- Localised outbreak in South Australia in November 2020

- Localised outbreak in New South Wales in December 2020

- Incursion of VOC 202012/01 in Western Australia in January 2021

- Localised outbreak of VOC 202012/01 in Victoria in February 2021

Overview of methodologies

The methods used for estimating transmission potential and R_{eff} have been briefly described at the beginning of Part I, with full details provided in the Appendix.

We report month-ahead state-level forecasts of the daily number of new confirmed cases from an 'ensemble forecast' of three independent models. Ensemble forecasts tend to produce improved estimates of both the central values, as well as improved estimates of the plausible, yet least likely forecasts (uncertainty). Our ensemble is generated by equally weighting the forecasts from each of the three models. A brief description of each method incorporated in the ensemble is given below (see Appendix for details):

- **SEEIIR Forecast**: A stochastic susceptible-exposed-infectious-recovered (SEEIIR) compartmental model that incorporates changes in local transmission potential via the estimated time-varying effective reproduction number (as shown in Figure 3).

- **Probabilistic Forecast**: A stochastic epidemic model that accounts for the number of imported-, symptomatic- and asymptomatic-cases over time. This model estimates the effective reproduction number corresponding to local and imported cases, and incorporates mobility data to infer the effect of macro-distancing behaviour. This model captures variation in the number and timing of new infections via probability distributions. The parameters that govern these distributions are inferred from the case and mobility data (*e.g.*, mean number of imported cases).

- **Time-Series Forecast**: A time-series model that does not account for disease transmission dynamics, but rather uses recent daily case counts to forecast cases into the future. Parameters of this 'autoregressive' model are estimated using global data accessible via the Johns Hopkins COVID-19 repository. Case counts from a specific time window prior to the forecasting date (the present) are used for model calibration. The number of days within this time window is chosen to optimise projections for Australian data.

Declining phase of the Victorian epidemic in September 2020

We report on situational analyses as of 12 September 2020, based on case data extracted from the NNDSS on 14 September 2020.

Context and situational assessment:

- At the time of analysis, the Victorian second wave epidemic had been in decline for almost seven weeks, with cases falling from a peak of 465 daily cases (by date of symptom onset) on 29 July 2020 to 26 daily cases on 12 September 2020. Stage 4 stay-at-home restrictions had been in place in metropolitan Melbourne for approximately six weeks, upgraded from Stage 3 stay-at-home restrictions on 2 August 2020.

- As of 12 September 2020, we estimated an R_{eff} of 0.75 [0.57, 0.96] (cf. 0.76 [0.68, 0.83] in the retrospective analysis) for active cases in VIC, with a 3% chance of exceeding 1 (Figure 3 and Table 1). The state-wide transmission potential estimated at the time was 0.59 [0.51, 0.69], indicating that levels of distancing behaviour were highly likely to prevent escalation of epidemic activity in the broader community (Figure 2). Note: some methodological updates made since September 2020 have resulted in a revised transmission potential of 0.69 [0.61, 0.81], with no material change in the interpretation.

- The strong positive deviation in actual transmission from state-wide transmission potential (evident in model Component 2, Figure 4) likely reflected heightened transmission in subsections of the population with higher-than-average rates of social contact. This positive deviation in Component 2 persisted throughout the course of the epidemic and was concordant with the demography and socio-economic circumstances of early affected areas, which had higher than average household sizes and a large proportion of essential and casualised workers who were unable to work from home.

- The forecast for VIC strongly suggested that case counts would continue to decline through September, as was ultimately the case (Figure 9). Unlike analyses that fit simple trends (such as linear regression, with or without accounting for over-dispersion), the forecast did admit a small possibility of sustained epidemic activity, or even an escalation of cases. Such uncertainty, which flows from our use of more realistic models of SARS-Cov-2 transmission, better reflects the uncertainty in future epidemic dynamics.

Table 1: Median estimates of state-wide transmission potential (model Component 1) and R_{eff} of current active cases (model Component 1&2) by state/territory as of 12 September 2020 [90% credible intervals]. The total number of cases (locally acquired or missing place of acquisition) with a symptom onset date recorded (or inferred) to be from 29 August–11 September inclusive is also shown, indicative of the number of local active cases at the time of analysis. For states/territories with very low numbers of local active cases, the estimates of R_{eff} for active cases is highly uncertain. The state-wide transmission potential should be referred to when assessing the risk of an epidemic becoming established given a seeding event. Note: estimates in this table were made at the time of analysis and may differ from those in the time-series as of 14 February 2020 as a result of updates to the case data and some technical details of the methods over time, as well as minor statistical variation and smoothing.

	State-wide transmission potential		R_{eff} of current active cases		*Cases
State	R_{eff} [90% CrI]	$P(R_{\text{eff}} > 1)$	R_{eff} [90% CrI]	$P(R_{\text{eff}} > 1)$	29 Aug – 11 Sept
ACT	1.05 [0.94, 1.17]	0.76	1.04 [0.63, 1.77]	0.60	0
NSW	0.88 [0.79, 1.02]	0.07	0.95 [0.69, 1.31]	0.39	79
NT	1.61 [1.43, 1.81]	1.00	1.59 [0.74, 2.98]	0.88	0
QLD	1.01 [0.89, 1.17]	0.53	0.96 [0.61, 1.41]	0.42	23
SA	1.08 [0.96, 1.23]	0.87	0.96 [0.40, 1.76]	0.45	1
TAS	1.22 [1.09, 1.39]	1.00	1.17 [0.39, 3.04]	0.63	0
VIC	0.59 [0.51, 0.69]	0.00	0.75 [0.57, 0.96]	0.03	758
WA	1.29 [1.16, 1.44]	1.00	1.25 [0.62, 2.15]	0.80	0

*Indication of the number of cases included in the R_{eff} analyses of local active cases. This includes cases coded as either locally acquired or missing place of acquisition within the NNDSS at the time of analysis. Our algorithms classify any cases that are missing place of acquisition as locally acquired.

Figure 9: Time series of new daily local cases of COVID-19 estimated in VIC from the forecasting ensemble model (50–90% confidence intervals coloured in progressively lighter blue shading) from 12 September to 10 October 2020. The observed daily counts of locally acquired cases are also plotted from 1 June by date of symptom onset (grey bars). Recent case counts are inferred to adjust for reporting delays (black dots).
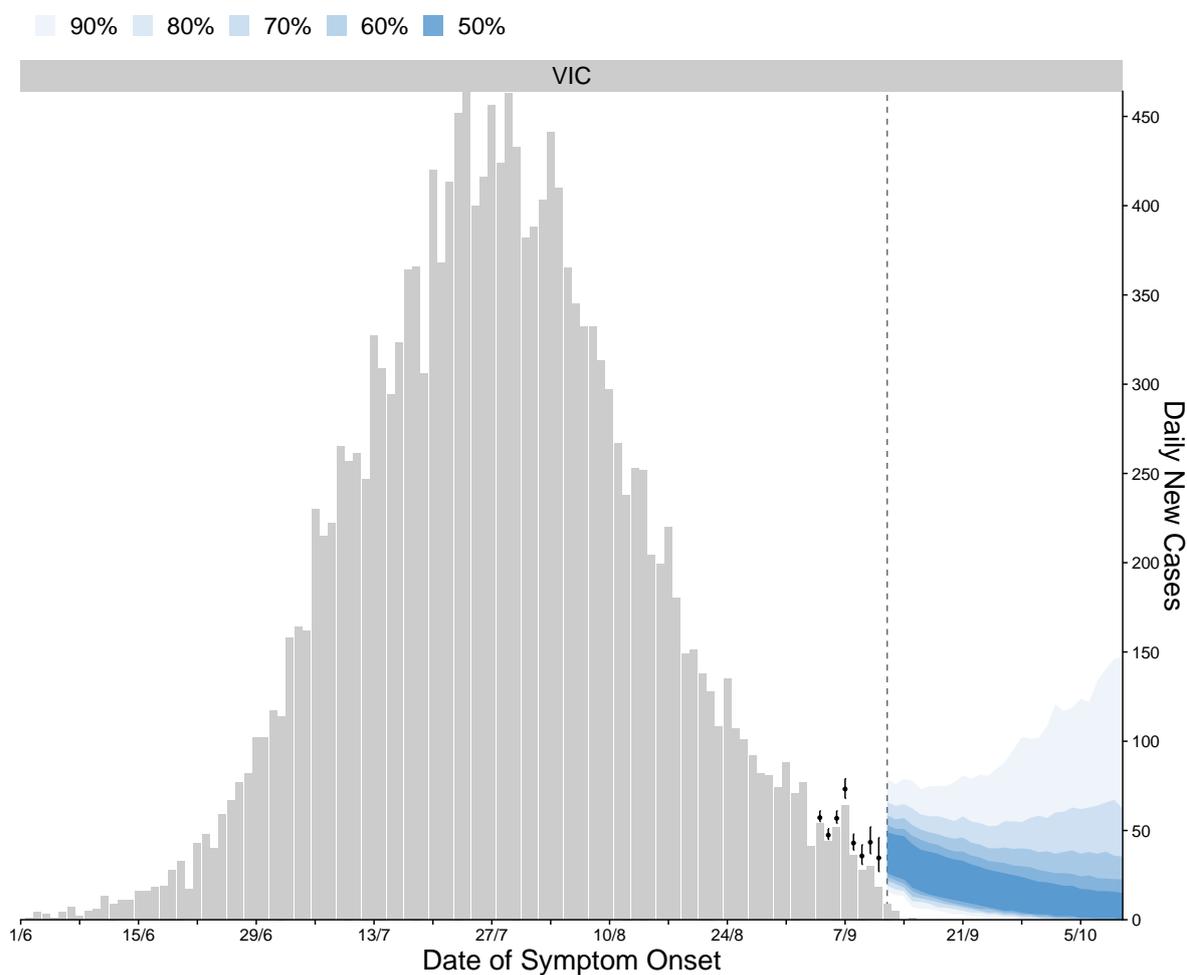
Localised outbreak in South Australia in November 2020

We report on situational analyses as of 22 November 2020, based on case data extracted from the NNDSS on 23 November 2020.

Context and situational assessment:

- In mid-November 2020, a sustained period of zero local case incidence in SA was disrupted by a breach of mandatory hotel quarantine which led to a cluster of more than 20 cases. At the time, SA was largely open/unrestricted and transmission potential was estimated to be 1.27 [1.14, 1.41] (cf. 1.39 [1.24, 1.60] in the retrospective analysis), suggesting that the risk of establishing an epidemic was reasonably high, and that if established, spread would be rapid. In the week prior to our illustrative analysis (14 November 2020), we estimated that the R_{eff} of active cases was above 1 — though highly uncertain due to the small number of cases — and the forecast at this time suggested an increase in epidemic activity through December.

- In response to the outbreak, South Australian authorities enacted a three-day period of stay-at-home restrictions across the state from 19 November 2020. This was in addition to an intensive public health response to trace and quarantine contacts. The outbreak was rapidly contained.

- Population mobility and rates of non-household contacts decreased substantially and rapidly around the time of activation of restrictions on 19 November 2020 (Figures 5 and 7). There was also some evidence that people changed their behaviour ahead of the announcement of restrictions (for example in Google's time at transit stations, Figure 7), likely in response to reported cases.

- As of 22 November 2020, we estimated that R_{eff} was well below 1. The small number of cases within the cluster meant that the future behaviour of the epidemic was difficult to predict at this time. The state-wide transmission potential of 0.73 [0.66, 0.79] estimated on that day indicated that levels of distancing behaviour were likely sufficient to prevent escalation of epidemic activity in the general population, if the current cluster was not definitively contained (Table 2). Due to the brief period of restrictions in SA, the approximate week long window over which macro- and micro-distancing data are collected, and the smoothing of transmission potential in our method, we are uncertain as to the lowest value of transmission potential obtained during the SA stay-at-home period. Our analysis at the time is presented in Figure S4. In the retrospective analysis (performed 14 February 2021, Figure 2), the depth of the trough in transmission potential is less pronounced, primarily due to the smoothing applied by the model. Based on our analyses at the time, which were made prior to the rebound in behaviour following the easing of restrictions, and knowledge of the response to the SA public health orders, we believe that the transmission potential was almost certainly under 1 at its lowest point.

- The forecast for SA as of 12 September 2020, suggested that case counts were highly likely to remain low or decline through December (Figure 10).

Table 2: Median estimates of state-wide transmission potential (model Component 1) and R_{eff} of current active cases (model Component 1&2) by state/territory as of 22 November 2020 [90% credible intervals]. The total number of cases (locally acquired or missing place of acquisition) with a symptom onset date recorded (or inferred) to be from 7 November–20 November 2020 inclusive is also shown, indicative of the number of local active cases at the time of analysis. For states/territories with very low numbers of local active cases, the estimates of R_{eff} for active cases is highly uncertain. The state-wide transmission potential should be referred to when assessing the risk of an epidemic becoming established given a seeding event. Note: estimates in this table were made at the time of analysis and may differ from those in the time-series as of 14 February 2020 as a result of updates to the case data and some technical details of the methods over time, as well as minor statistical variation and smoothing.

	State-wide transmission potential		R_{eff} of current active cases		*Cases
State	R_{eff} [90% CrI]	$P(R_{\text{eff}} > 1)$	R_{eff} [90% CrI]	$P(R_{\text{eff}} > 1)$	7 Nov–20 Nov
ACT	1.31 [1.19, 1.45]	1.00	1.30 [0.67, 2.51]	0.81	0
NSW	1.12 [1.02, 1.23]	0.98	1.03 [0.62, 1.60]	0.54	0
NT	1.51 [1.36, 1.68]	1.00	1.52 [0.75, 3.26]	0.89	0
QLD	1.21 [1.10, 1.34]	1.00	1.13 [0.52, 2.12]	0.63	1
SA	0.73 [0.66, 0.79]	0.00	0.70 [0.33, 1.28]	0.14	18
TAS	1.24 [1.13, 1.37]	1.00	1.26 [0.43, 4.08]	0.66	0
VIC	0.83 [0.75, 0.91]	0.00	0.80 [0.49, 1.27]	0.22	0
WA	1.40 [1.26, 1.54]	1.00	1.38 [0.86, 2.37]	0.90	0

*Indication of the number of cases included in the R_{eff} analyses of local active cases. This includes cases coded as either locally acquired or missing place of acquisition within the NNDSS at the time of analysis. Our algorithms classify any cases that are missing place of acquisition as locally acquired.

†One case in QLD was missing place of acquisition in the NNDSS at the time of analysis (23 November 2020). Our algorithms classify all such cases as local cases.

Figure 10: Time series of new daily local cases of COVID-19 estimated in SA from the forecasting ensemble model (50–90% confidence intervals coloured in progressively lighter blue shading) from 21 November to 19 December 2020. The observed daily counts of locally acquired cases are also plotted from 1 September by date of symptom onset (grey bars). Recent case counts are inferred to adjust for reporting delays (black dots).
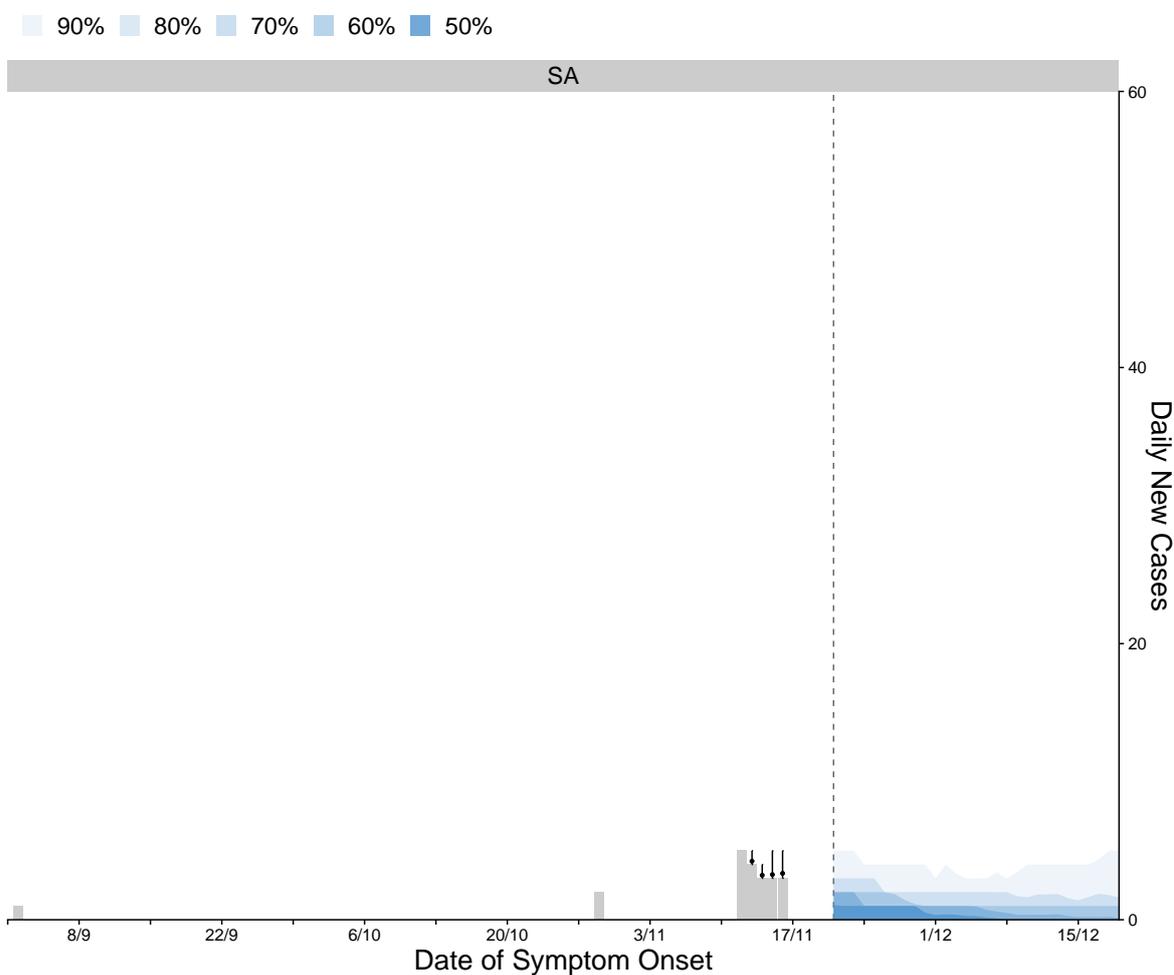
Localised outbreak in New South Wales in December 2020

We report on situational analyses as of 21 December 2020, based on case data provided by NSW Health on 22 December 2020.

Context and situational assessment:

- In mid-December 2020, two large super-spreading events occurred in the locality of Northern Beaches, leading to a substantial and rapidly developing outbreak in NSW. The origins of this outbreak are yet to be established. Concurrent with this outbreak, a breach of mandatory hotel quarantine resulted in a second, small cluster.

- In response to the Northern Beaches outbreak, NSW authorities enacted stay-at-home restrictions in the affected locality from 20 December 2020. This was in addition to an intensive public health response.

- Given that a significant proportion of cases arose from the two super-spreading events, an estimate of R_{eff} in which the date of symptom onset of cases are used to infer infection dates would not be reliable. Without actual dates of infection, the data would indicate to the model that these transmission events occurred over several days, implying a period of high transmission rather than a single major event followed by low transmission.

- NSW Health provided more detailed information on cases than is reported in the NNDSS, including the likely dates of infection based on epidemiological investigation. We used these data to compute an R_{eff} that more accurately captured the transmission behaviour (Table 3). As of 21 December 2020, we estimated an R_{eff} of 1.14 [0.58, 2.06] (cf. 0.92 [0.69, 1.20] in the retrospective analysis) for active cases in NSW, with a 63% chance of exceeding 1. For comparison, a naive estimate of R_{eff} using dates of symptom onset from these data would be around 3. The state-wide transmission potential of 1.38 [1.23, 1.59] (cf. 1.09 [0.99, 1.27] in the retrospective analysis) suggested that conditions were suitable for an epidemic to become established in the broader population if the outbreak was not definitely contained.

Table 3: Median estimates of state-wide transmission potential (model Component 1) and R_{eff} of current active cases (model Component 1&2) by state/territory as of as of 21 December 2020 [90% credible intervals]. The total number of cases (locally acquired or missing place of acquisition) with a symptom onset date recorded (or inferred) to be from 6 December–19 December inclusive is also shown, indicative of the number of local active cases at the time of analysis. For states/territories with very low numbers of local active cases, the estimates of R_{eff} for active cases is highly uncertain. The state-wide transmission potential should be referred to when assessing the risk of an epidemic becoming established given a seeding event. Note: estimates in this table were made at the time of analysis and may differ from those in the time-series as of 14 February 2020 as a result of updates to the case data and some technical details of the methods over time, as well as minor statistical variation and smoothing.

	State-wide transmission potential		R_{eff} of current active cases		*Cases
State	R_{eff} [90% CrI]	$P(R_{\text{eff}} > 1)$	R_{eff} [90% CrI]	$P(R_{\text{eff}} > 1)$	6 Dec–19 Dec
ACT	1.64 [1.47, 1.87]	1.00	1.29 [0.55, 2.82]	0.70	0
NSW	1.38 [1.23, 1.59]	1.00	1.14 [0.58, 2.06]	0.63	94
NT	1.75 [1.56, 2.02]	1.00	1.32 [0.57, 3.06]	0.71	0
QLD	1.43 [1.29, 1.62]	1.00	0.97 [0.39, 2.12]	0.47	†1
SA	1.43 [1.28, 1.65]	1.00	1.00 [0.42, 2.14]	0.50	0
TAS	1.57 [1.41, 1.79]	1.00	1.24 [0.55, 2.90]	0.67	0
VIC	1.08 [0.96, 1.29]	0.86	0.83 [0.34, 1.96]	0.35	0
WA	1.77 [1.59, 1.98]	1.00	1.26 [0.55, 2.66]	0.69	‡1

*Indication of the number of cases included in the R_{eff} analyses of local active cases. This includes cases coded as either locally acquired or missing place of acquisition within the NNDSS at the time of analysis. Our algorithms classify any cases that are missing place of acquisition as locally acquired.
†One recent case in QLD was missing place of acquisition in the NNDSS database at the time of analysis (21 December 2020). Our algorithms classify all such cases as local cases.
‡The recent local case in WA was acquired in hotel quarantine.

Incursion of VOC 202012/01 in Western Australia in January 2021

We report on situational analyses as of 7 February 2021, based on case data extracted from the NNDSS on 8 February 2020.

Context and situational assessment:

- In late January 2021, a breach of mandatory hotel quarantine resulted in one confirmed local case of SARS-CoV-2 Variant of Concern (VOC) 202012/01 in WA (reported on 31 January 2021). Around this time, the state-wide transmission potential of VOC 202012/01 was approximately 2.3 suggesting that conditions were highly suitable for an epidemic to become established in the general population if there were onward transmission from active cases.

- In response to this case, West Australian authorities enacted a five-day period of stay-at-home restrictions across the Perth, Peel and the South West regions from 31 January 2021. This was in addition to an intensive public health response.

- We estimated that substantial changes in macro- and micro-distancing behaviour occurred from around 31 January when stay-at-home restrictions were activated (Figures 5 and 6). This resulted in a substantial decrease in the estimated state-wide transmission potential.

- As of 7 February 2021, no further cases had been reported since the one on 31 January 2020. We estimated that the state-wide transmission potential of VOC 202012/01 was 1.22 [1.08, 1.45], which while much lower compared to the previous week, suggested that conditions remained suitable for an epidemic to become established in the general population if there were onward transmission from active cases (Table 4 and Figure 2). Our analysis at the time is presented in Figure S5. Our most recent estimate of transmission potential of VOC 202012/01 at 7 February 2021 is 1.54 [1.35, 1.85]. Like for SA, due to the short period of restrictions, the approximate week long window over which macro-/micro-distancing data are collected, and the smoothing of transmission potential in our method, we are uncertain as to the lowest value of transmission potential obtained during the WA stay-at-home period. While not visible in the time-series as of 14 February 2021, the transmission potential for non-VOCs was likely under 1 at its lowest point, based on estimates made prior to the rebound in behaviour that followed the easing of restrictions (see Figure S5).

- With only a single case, epidemic forecasts are not informative. They are dominated by stochasticity and so of minimal public health importance and not shown here.

Table 4: Median estimates of state-wide transmission potential (Component 1) and R_{eff} of current active cases (Component 1&2) by state/territory as of 7 February [90% credible intervals]. We provide two sets of estimates of transmission potential. In black: Estimates of transmission potential of non-VOC lineages. In blue: Estimates of transmission potential of VOC 202012/01, based on an estimated 47% [35, 58] increase in relative transmissibility of VOC 202012/01 compared with non-VOCs (see Appendix for details). The total number of cases (locally acquired or missing place of acquisition) with a symptom onset date recorded (or inferred) to be from 23 January–5 February inclusive is also shown, indicative of the number of local active cases at the time of analysis. For states/territories with very low numbers of local active cases, the estimates of R_{eff} for active cases is highly uncertain. The state-wide transmission potential should be referred to when assessing the risk of an epidemic becoming established given a seeding event. Note: estimates in this table were made at the time of analysis and may differ from those in the time-series as of 14 February 2020 as a result of updates to the case data and some technical details of the methods over time, as well as minor statistical variation and smoothing.

| | State-wide transmission potential | | | | R_{eff} of current active cases | | *Cases |
| | Non-VOCs | | VOC 202012/01 | | | | |
State	est. [90% CrI]	$P(>1)$	est. [90% CrI]	$P(>1)$	est. [90% CrI]	$(P>1)$	23 Jan – 5 Feb
ACT	1.57 [1.41, 1.76]	1.00	2.23 [1.97, 2.53]	1.00	1.24 [0.51, 2.83]	0.67	0
NSW	1.26 [1.13, 1.44]	1.00	1.78 [1.58, 2.06]	1.00	0.92 [0.37, 2.08]	0.43	0
NT	1.79 [1.60, 2.03]	1.00	2.55 [2.24, 2.93]	1.00	1.41 [0.59, 3.13]	0.75	0
QLD	1.44 [1.29, 1.63]	1.00	2.03 [1.81, 2.35]	1.00	1.07 [0.47, 2.22]	0.56	[†]1
SA	1.49 [1.34, 1.70]	1.00	2.11 [1.87, 2.44]	1.00	1.20 [0.51, 2.56]	0.63	0
TAS	1.46 [1.30, 1.69]	1.00	2.07 [1.82, 2.44]	1.00	1.16 [0.46, 2.67]	0.62	0
VIC	1.14 [1.02, 1.31]	0.98	1.61 [1.43, 1.88]	1.00	0.79 [0.33, 1.68]	0.31	1
WA	0.88 [0.78, 1.02]	0.08	1.22 [1.08, 1.45]	1.00	0.62 [0.27, 1.27]	0.15	1

*Indication of the number of cases included in the R_{eff} analyses of local active cases. This includes cases coded as either locally acquired or missing place of acquisition within the NNDSS at the time of analysis. Our algorithms classify any cases that are missing place of acquisition as locally acquired.

†One recent case in QLD was missing place of acquisition in the NNDSS at the time of analysis (8 February 2020).

Localised outbreak of VOC 202012/01 in Victoria in February 2021

We report on situational analyses as of 14 February 2021, based on case data extracted from the NNDSS on 15 February 2021.

Context and situational assessment:

- In early February 2021, a breach of mandatory hotel quarantine led to a cluster of more than 20 cases of VOC 202012/01 in VIC.

- In response to the outbreak, Victorian authorities imposed a five-day period of Stage 4 stay-at-home restrictions across the state from 13 February 2021. This was in addition to an intensive public health response.

- As of 14 February 2021, we estimated an R_{eff} of 1.38 [0.76, 2.47] for active cases in VIC, with a 83% chance of R_{eff} exceeding 1. Our analysis suggested that R_{eff}, while still above 1, was in the early stages of decline (Table 5 and Figure 3)[1].

- The state-wide transmission potential of VOC 202012/01 of 1.73 [1.51, 2.01] in VIC suggested that conditions were suitable for an epidemic to become established if there were onward transmission from active cases (Table 5 and Figure 2). Note that this estimate of transmission potential was informed by survey and mobility data collected prior to the activation of Stage 4 restrictions on 13 February 2021.

- Given the small number of active cases, the future behaviour of the outbreak was highly uncertain. The forecast for VIC suggested that case counts would likely remain low through to early March, with the median estimate increasing from 2–5 cases per day (Figure 11). There was also substantial support for definitive control being achieved, particularly given the success of the public health response up to the time of analysis, with new cases confirmed to have been identified as contacts and placed in quarantine at least three days prior to symptom onset. However, the forecast did not exclude the possibility of increasing epidemic activity.

*Indication of the number of cases included in the R_{eff} analyses of local active cases. This includes cases coded as either locally acquired or missing place of acquisition within the NNDSS at the time of analysis. Our algorithms classify any cases that are missing place of acquisition as locally acquired.

---

[1]At the time of this report (1 March 2021), there had been very few additional cases associated with the cluster reported. The R_{eff} had continued its decline and was under 1 by late February 2021.

Table 5: Median estimates of state-wide transmission potential (Component 1) and R_{eff} of current active cases (Component 1&2) by state/territory as of 14 February [90% credible intervals]. We provide two sets of estimates of transmission potential. In black: Estimates of transmission potential of non-VOC lineages. In blue: Estimates of transmission potential of VOC 202012/01, based on an estimated 47% [35, 58] increase in relative transmissibility of VOC 202012/01 compared with non-VOCs (see Appendix for details). The total number of cases (locally acquired or missing place of acquisition) with a symptom onset date recorded (or inferred) to be from 30 January – 12 February inclusive is also shown, indicative of the number of local active cases at the time of analysis. For states/territories with very low numbers of local active cases, the estimates of R_{eff} for active cases is highly uncertain. The state-wide transmission potential should be referred to when assessing the risk of an epidemic becoming established given a seeding event.

| | State-wide transmission potential | | | | | R_{eff} of current active cases | | *Cases |
| | Non-VOCs | | VOC 202012/01 | | | | | |
State	est. [90% CrI]	$P(>1)$	est. [90% CrI]	$P(>1)$		est. [90% CrI]	$(P>1)$	30 Jan –12 Feb
ACT	1.67 [1.49, 1.86]	1.00	2.46 [2.13, 2.81]	1.00		1.29 [0.55, 2.92]	0.71	0
NSW	1.31 [1.17, 1.50]	1.00	1.92 [1.66, 2.23]	1.00		0.99 [0.44, 2.10]	0.49	0
NT	1.85 [1.65, 2.07]	1.00	2.73 [2.36, 3.13]	1.00		1.43 [0.61, 3.27]	0.75	0
QLD	1.46 [1.31, 1.66]	1.00	2.16 [1.87, 2.50]	1.00		1.14 [0.49, 2.52]	0.61	0
SA	1.54 [1.38, 1.73]	1.00	2.27 [1.97, 2.61]	1.00		1.20 [0.51, 2.74]	0.66	0
TAS	1.47 [1.32, 1.65]	1.00	2.16 [1.88, 2.49]	1.00		1.17 [0.53, 2.54]	0.64	0
VIC	1.18 [1.06, 1.35]	1.00	1.73 [1.51, 2.01]	1.00		1.38 [0.76, 2.47]	0.83	16
WA	1.03 [0.92, 1.22]	0.65	1.50 [1.29, 1.80]	1.00		0.77 [0.34, 1.67]	0.30	0

Figure 11: Time series of new daily local cases of COVID-19 estimated in VIC from the forecasting ensemble model (50–90% confidence intervals coloured in progressively lighter blue shading) from 13 February to 13 March 2021. The observed daily counts of locally acquired cases are also plotted from 1 September 2020 by date of symptom onset (grey bars). Recent case counts are inferred to adjust for reporting delays (black dots).

Acknowledgements

Appendix

For full methodological details on the population mobility analysis, please refer to our previous Technical Report (dated 15 May 2020) available at the following link:

`https://www.doherty.edu.au/about/reports-publications`

Supplementary figures

Figure S1: Estimate of average state-level trend in local transmission potential, if we assume that only macro-distancing behaviour had changed and not micro-distancing behaviour or the time-to-detection, for each state/territory up to 7 February 2021 (light blue ribbon = 90% credible interval; dark blue ribbon = 50% credible interval). Solid grey vertical lines indicate key dates of implementation of various physical distancing policies. Black dotted line indicates the target value of 1 for the effective reproduction number required for control.
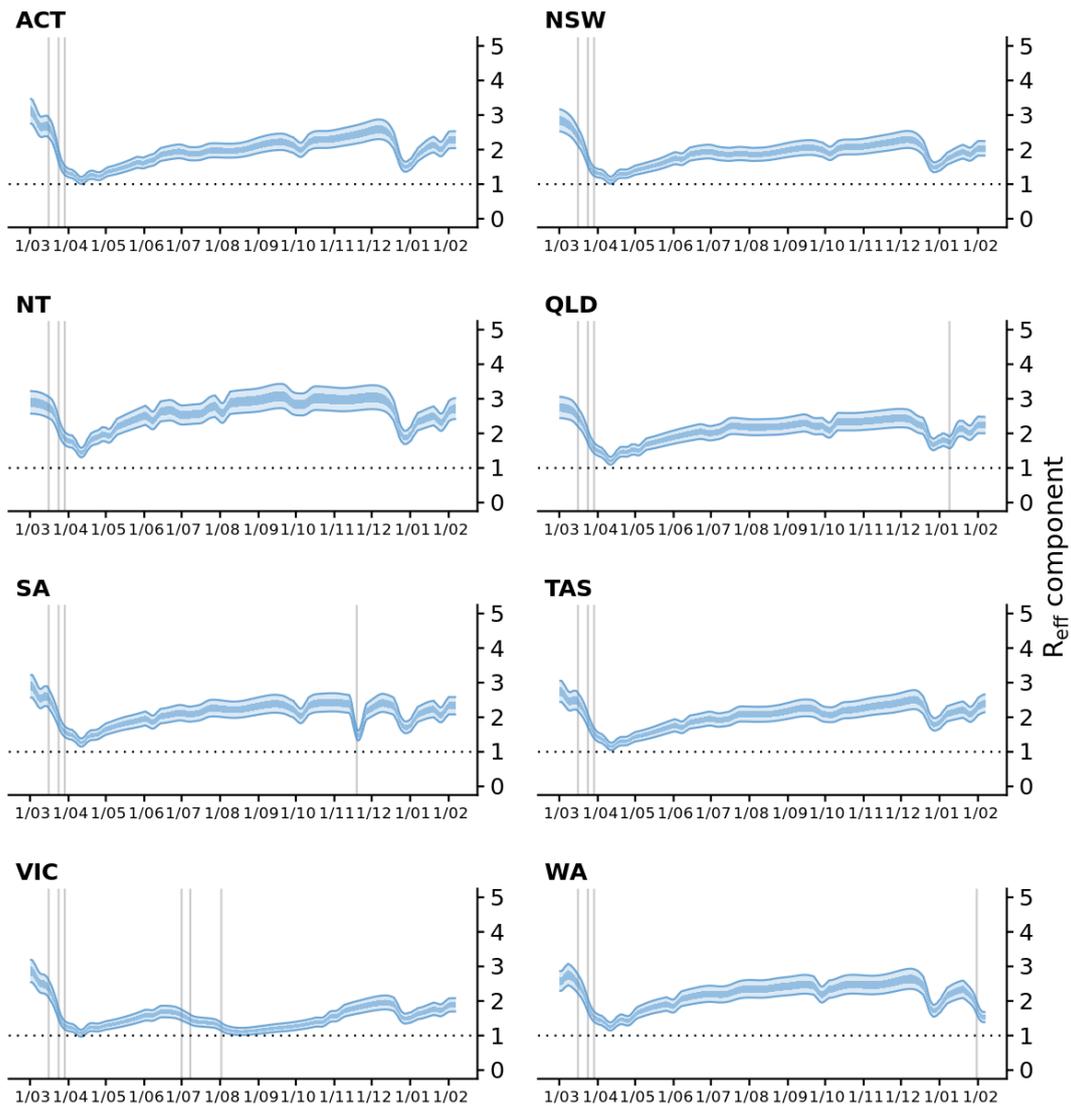
Figure S2: Estimate of average state-level trend in local transmission potential, if we assume that only micro-distancing behaviour had changed and not macro-distancing behaviour or the time-to-detection, for each state/territory up to 7 February 2021 (light purple ribbon = 90% credible interval; dark purple ribbon = 50% credible interval). Solid grey vertical lines indicate key dates of implementation of various physical distancing policies. Black dotted line indicates the target value of 1 for the effective reproduction number required for control.
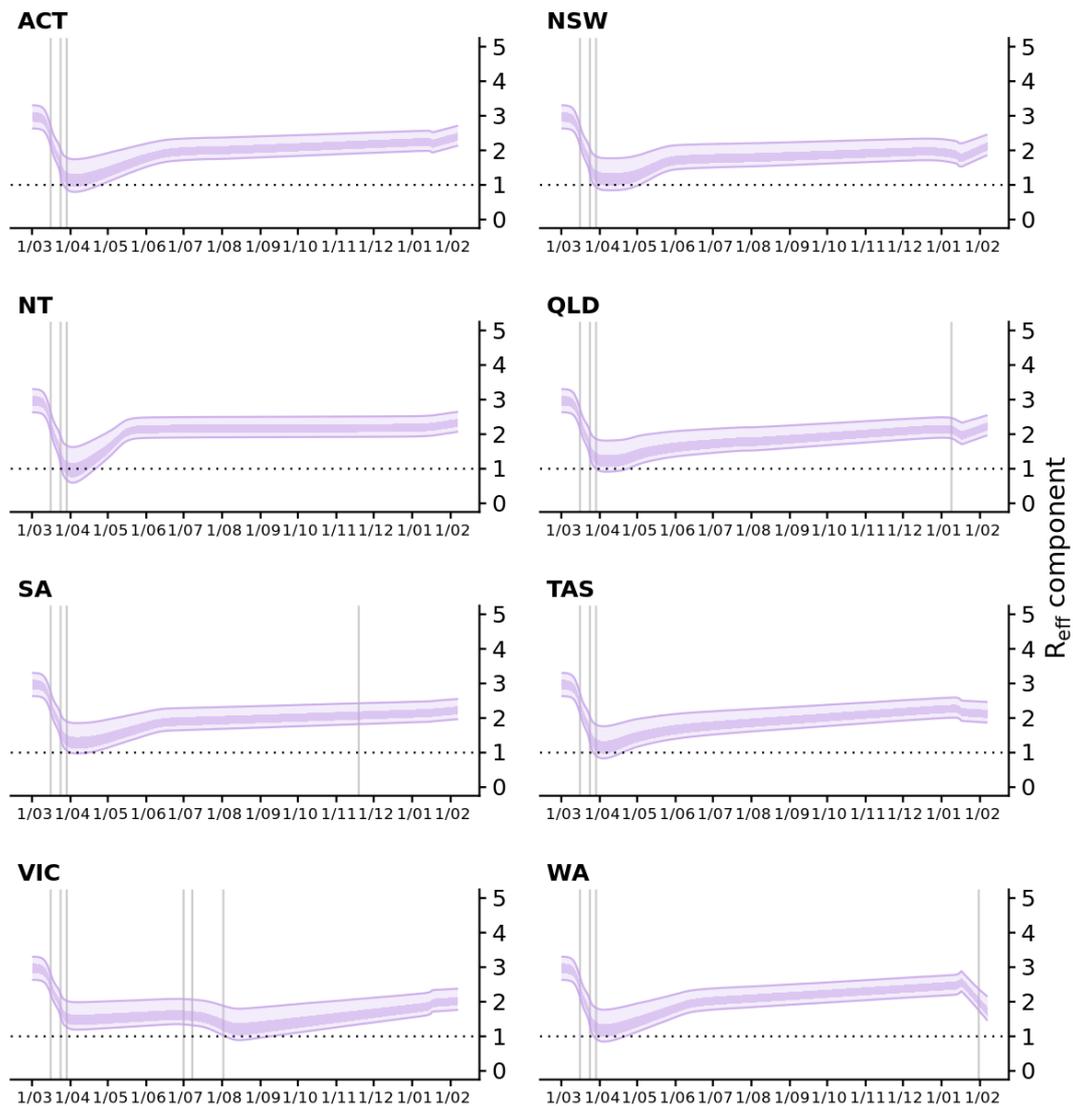
Figure S3: Estimate of average state-level trend in local transmission potential, if we assume that only the time-to-detection had changed and not macro-distancing or micro-distancing behaviour, for each state/territory up to 7 February 2021 (light yellow ribbon = 90% credible interval; dark yellow ribbon = 50% credible interval). Solid grey vertical lines indicate key dates of implementation of various physical distancing policies. Black dotted line indicates the target value of 1 for the effective reproduction number required for control.
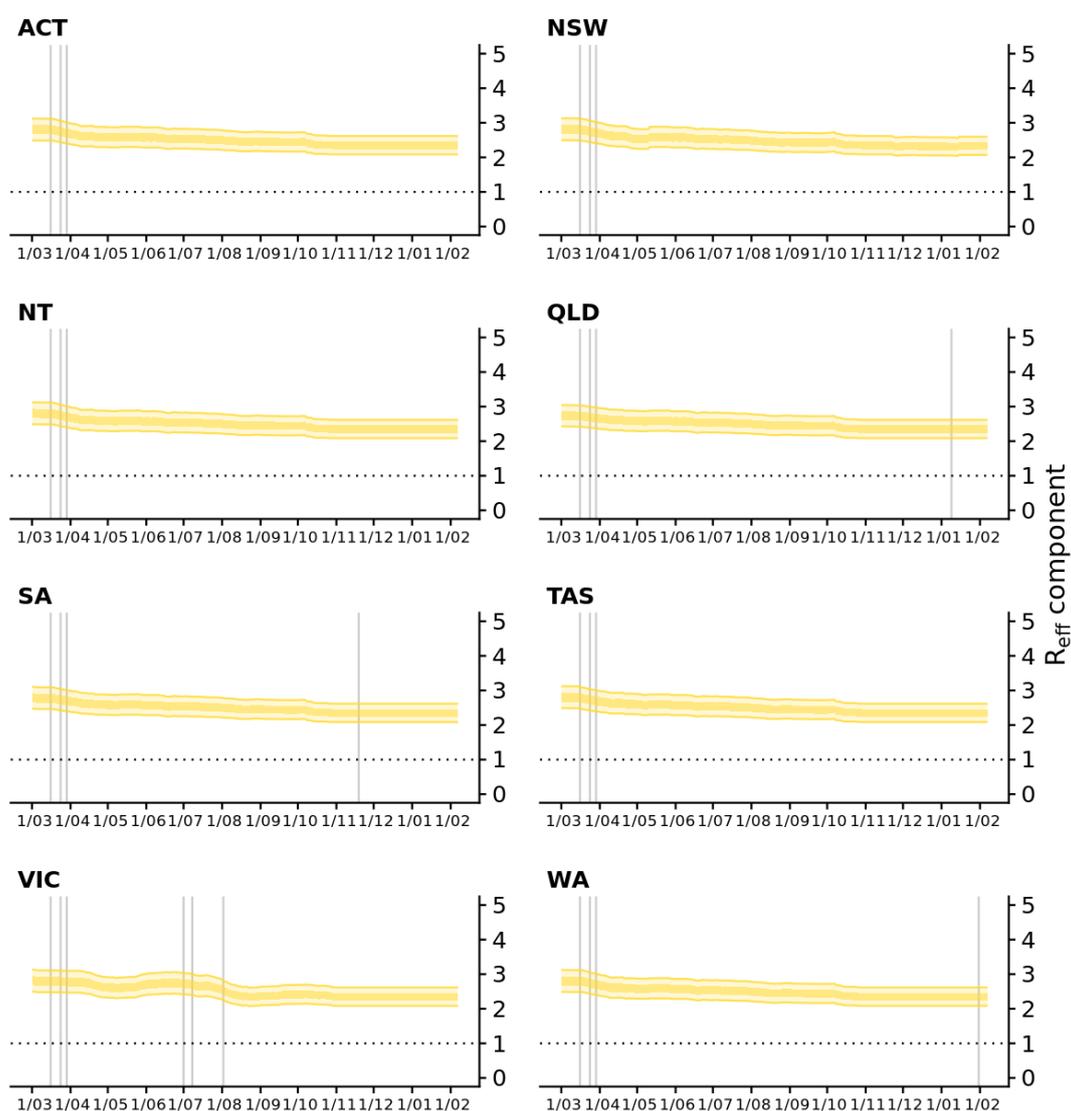
Figure S4: Estimate of state-wide transmission potential (model Component 1) by state/territory **up to 22 November 2020**. Light green ribbon=90% credible interval; dark green ribbon = 50% credible interval. Solid grey vertical lines indicate key dates of implementation of various physical distancing policies.
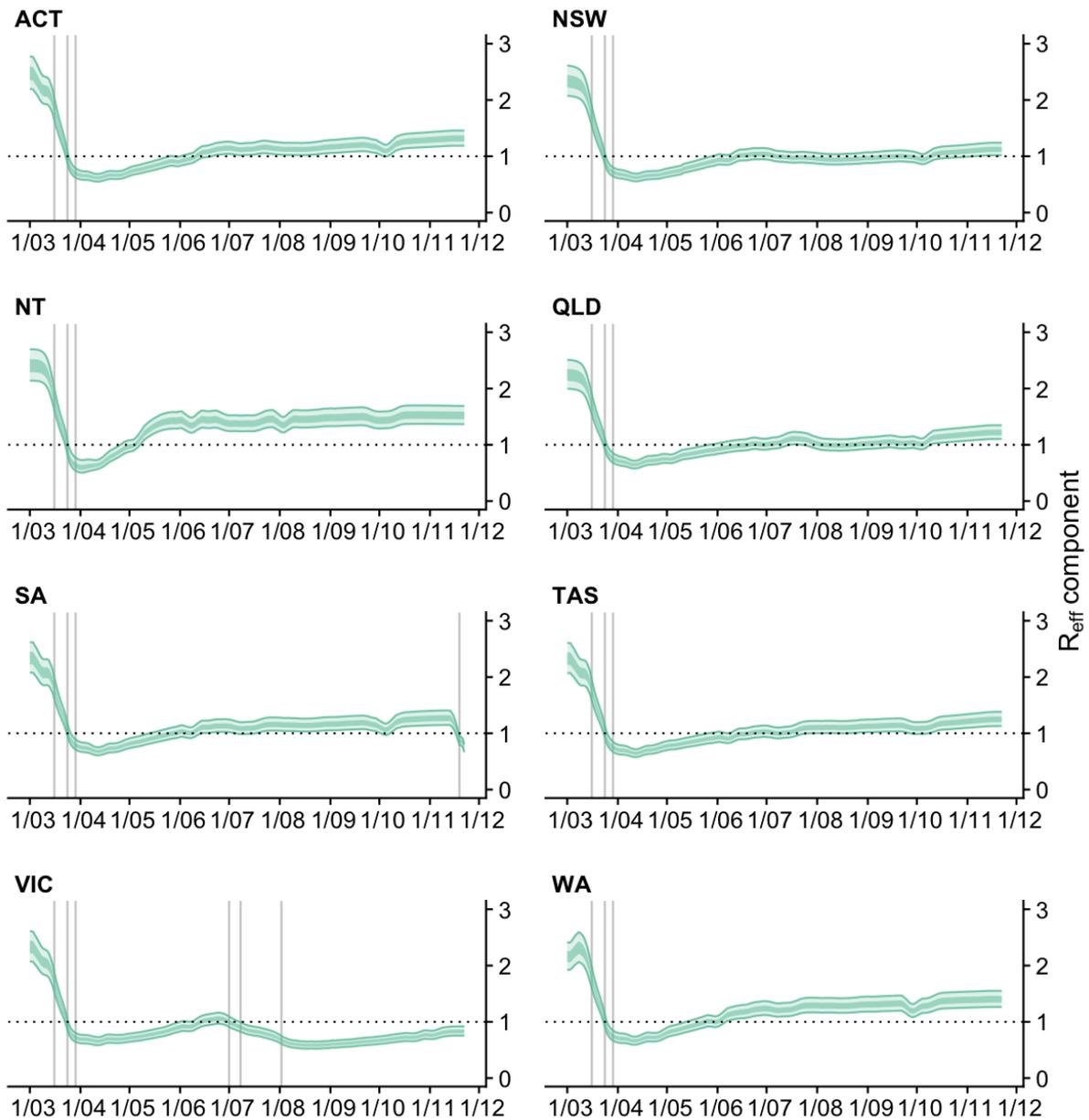
Figure S5: Estimate of state-wide transmission potential (model Component 1) by state/territory **up to 7 February 2021** (light coloured ribbons = 90% credible intervals; dark coloured ribbons = 50% credible intervals). We provide two sets of estimates of transmission potential. In green: Estimates of transmission potential of non-VOC lineages. In grey: Estimates of transmission potential of VOC 202012/01, based on the estimated relative increase in transmissibility of VOC 202012/01 compared with non-VOCs, as described in the Appendix, from 1 December 2020.
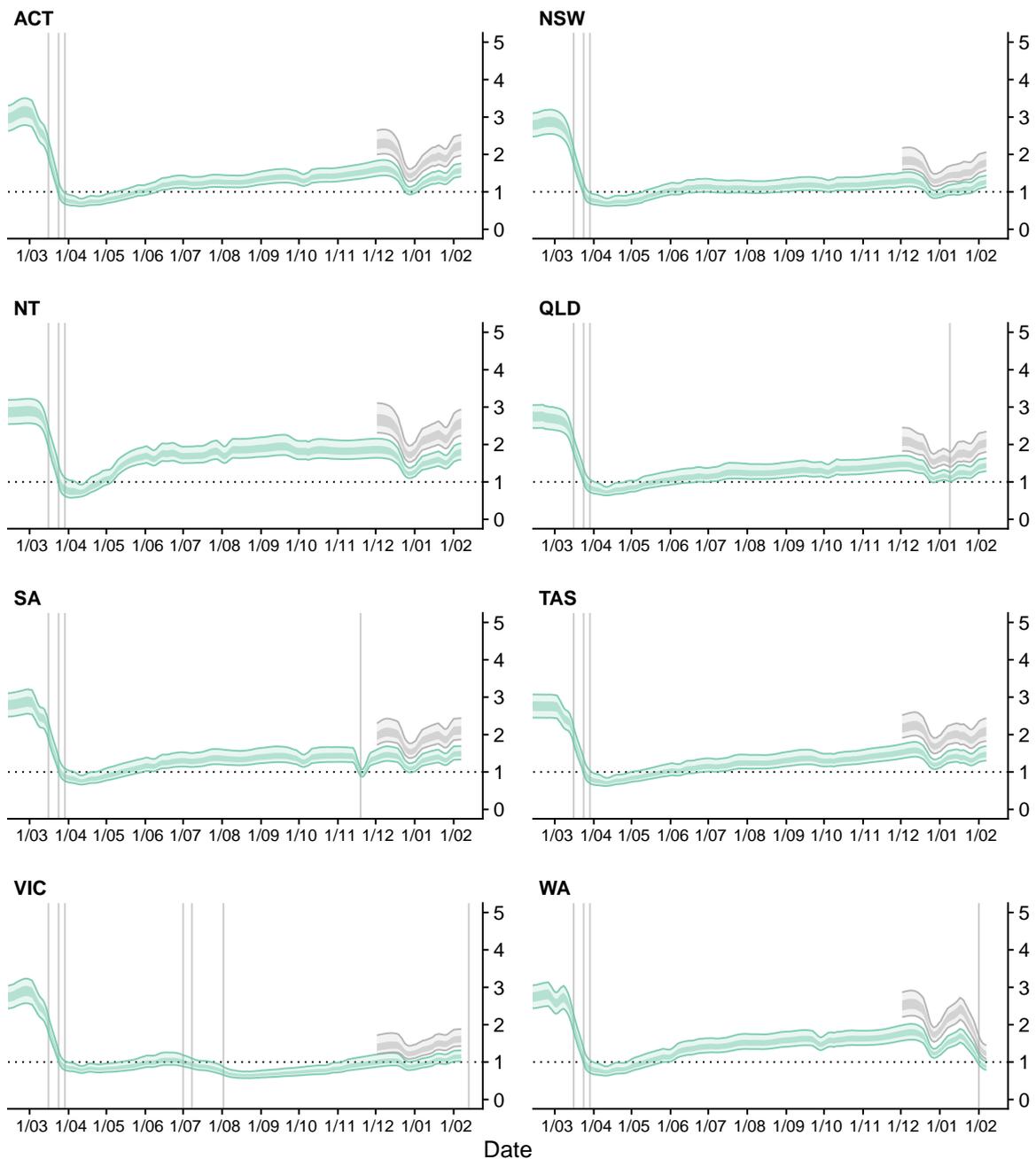
Figure S6: Estimated trends in micro-distancing behaviour, *i.e.* reduction in transmission probability per non-household contact, in each state/territory **up to 13 December 2020** (light purple ribbons = 90% credible intervals; dark purple ribbons = 50% credible intervals). Estimates are informed by state-level data from nationwide weekly surveys since March 2020 (indicated by the black lines and grey boxes). The width of the grey boxes corresponds to the duration of each survey wave (around 4 days).
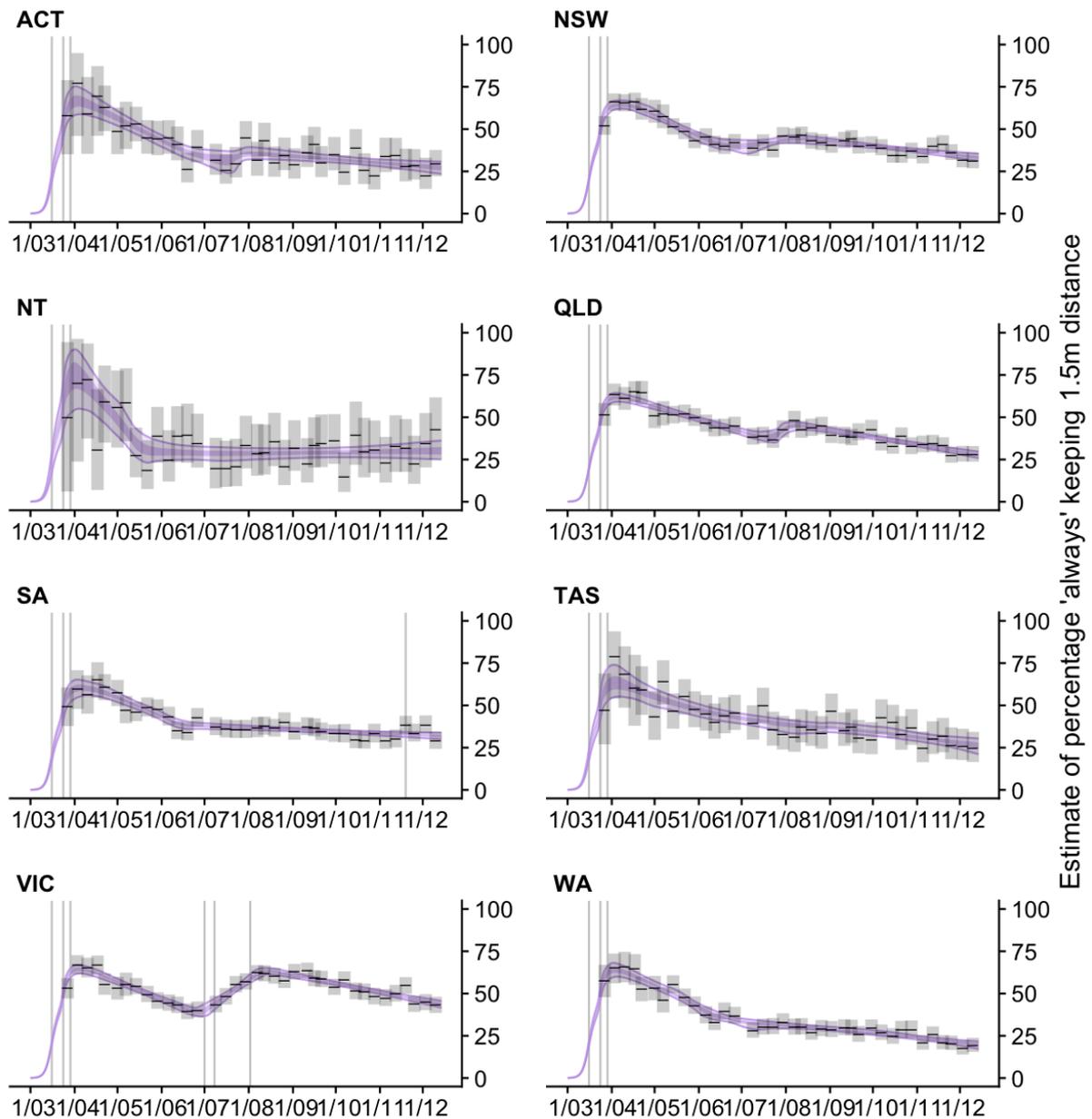
Supplement: model of SARS-CoV-2 transmissibility

Overview

We use a novel semi-mechanistic model to estimate the ability of SARS-CoV-2 to spread in a population, informed by data on cases, population behaviours and health system effectiveness. Where the virus is present, the quantity we compute is the effective reproduction number (R_{eff}). In the absence of cases, it reflects the ability of the virus, if it were present, to spread in a population, which we define as the 'transmission potential'.

Applying this method provides an estimate of the transmissibility of SARS-CoV-2 in periods of high, low, and zero, case incidence, with a coherent transition in interpretation across the changing epidemiological situations.

We separately model transmission from locally acquired cases (local-to-local transmission) and from overseas acquired cases (import-to-local transmission). We model local-to-local transmission for each Australian state and territory using two components (Figure 1):

1. the average population-level trend in transmissibility driven by interventions that primarily target transmission from local cases, specifically changes in physical distancing behaviour and case targeted measures (Component 1); and

2. short-term fluctuations in R_{eff} to capture stochastic dynamics of transmission, such as clusters of cases and short periods of lower-than-expected transmission, and other factors factors influencing R_{eff} that are otherwise unaccounted for by the model (Component 2).

During times of disease activity, Components 1 and 2 are combined to provide an estimate of the local R_{eff} as traditionally measured. In the absence of disease activity, Component 1 is interpreted as the potential for the virus, if it were present, to establish and maintain community transmission (> 1) or otherwise (< 1).

Modelling the impact of physical distancing

Overview

To investigate the impact of distancing measures on SARS-CoV-2 transmission, we distinguish between two types of distancing behaviour: 1) macro-distancing *i.e.*, reduction in the rate of non-household contacts; and 2) micro-distancing *i.e.*, reduction in transmission probability per non-household contact.

We used data from nationwide surveys to estimate trends in specific macro-distancing (average daily number of non-household contacts) and micro-distancing (proportion of the population always keeping 1.5m physical distance from non-household contacts) behaviours over time. We used these survey data to infer state-level trends in macro- and micro-distancing behaviour over time, with additional information drawn from trends in mobility data.

Estimating changes in macro-distancing behaviour

To estimate trends in macro-distancing behaviour, we used data from: two waves of a national survey conducted in early April and early May by the University of Melbourne; and weekly waves of a national survey conducted by the Behavioural Economics Team of the Australian Government (BETA)/Department of Health from late May. Respondents were asked to report the number of individuals that they had contact with outside of their household in the previous 24 hours. Note that the first wave of the University of Melbourne survey was fielded four days

after Australia's most intensive physical distancing measures were recommended nationally on 29 March 2020.

Given these data, we used a statistical model to infer a continuous trend in macro-distancing behaviour over time. This model assumed that the daily number of non-household contacts is proportional to a weighted average of time spent at different types of location, as measured by Google mobility data. The five types of places are: parks and public spaces; residential properties; retail and recreation; public transport stations; and workplaces. We fit a statistical model that infers the proportion of non-household contacts occurring in each of these types of places from:

- a survey of location-specific contact rates pre-COVID-19 Rolls *et al.* (2015); and

- a separate statistical model fit to the national average numbers of non-household contacts from a pre-COVID-19 contact survey and contact surveys fielded post-implementation of COVID-19 restrictions.

Waning in macro-distancing behaviour is therefore driven by Google mobility data on increasing time spent in each of the different types of locations since the peak of macro-distancing behaviour.

Estimating changes in micro-distancing behaviour

To estimate trends in micro-distancing behaviour, we used data from weekly national surveys (first wave from 27–30 March) to assess changes in behaviour in response to COVID-19 public health measures. Respondents were asked to respond to the question: 'Are you staying 1.5m away from people who are not members of your household' on a five point scale with response options "No", "Rarely", "Sometimes", "Often" and "Always".

These behavioural survey data were used in a statistical model to infer the trend in micro-distancing behaviour over time. Micro-distancing behaviour was assumed to be non-existent prior to the first epidemic wave of COVID-19, and the increase in micro-distancing behaviour to its peak was assumed to follow the same trend as macro-distancing behaviour — implying that the population simultaneously adopted both macro- and micro-distancing behaviours around the times that restrictions were implemented. The behavioural survey data was then used to infer the date of peak micro-distancing behaviour (assumed to be the same in all states), the proportion of the population adopting micro-distancing behaviour, and the rate at which micro-distancing behaviour is waning from that peak in each state.

Incorporating estimated changes in distancing behaviour in the model of R_{eff}

These state-level macro-distancing and micro-distancing trends were then used in the model of R_{eff} to inform the reduction in non-household transmission rates. Since the macro-distancing trend is calibrated against the number of non-household contacts, the rate of non-household transmission scales directly with this inferred trend. The probability of transmission per non-household contact is assumed to be proportional to the fraction of survey participants who report that they always maintain 1.5m physical distance from non-household contacts. The constant of proportionality is estimated in the R_{eff} model.

The estimated rate of waning of micro-distancing is sensitive to the metric used. If a different metric of micro-distancing (*e.g.*, the fraction of respondents practicing good hand hygiene) were used, this might affect the inferred rate of waning of micro-distancing behaviour, and therefore increasing R_{eff}.

Modelling the impact of quarantine of overseas arrivals

We model the impact of quarantine of overseas arrivals via a 'step function' reflecting three different quarantine policies: self-quarantine of overseas arrivals from specific countries prior to March 15; self-quarantine of all overseas arrivals from March 15 up to March 27; and mandatory quarantine of all overseas arrivals after March 27 (Figure S7). We make no prior assumptions about the effectiveness of quarantine at reducing R_{eff} import, except that each successive change in policy increased that effectiveness.

Figure S7: Nationwide average reduction in R_{eff} that is due to quarantine of overseas arrivals estimated from the R_{eff} model (light orange ribbon=90% credible interval; dark orange ribbon = 50% credible interval). Note that this trend does not capture time-varying fluctuations in R_{eff} in each state/territory. Solid grey vertical lines indicate key dates of implementation of key response policies. Black dotted line indicates the target value of 1 for the effective reproduction number required for control. Note: A simple but naïve upper bound on R_{eff} import can be computed by assuming that all locally acquired cases arose from imported cases, and therefore computing the ratio of the numbers of local and imported cases. This results in a maximum possible value of the average R_{eff} import of 0.57.

Impact of quarantine of overseas arrivals

Component of R_{eff} due to quarantine of overseas arrivals
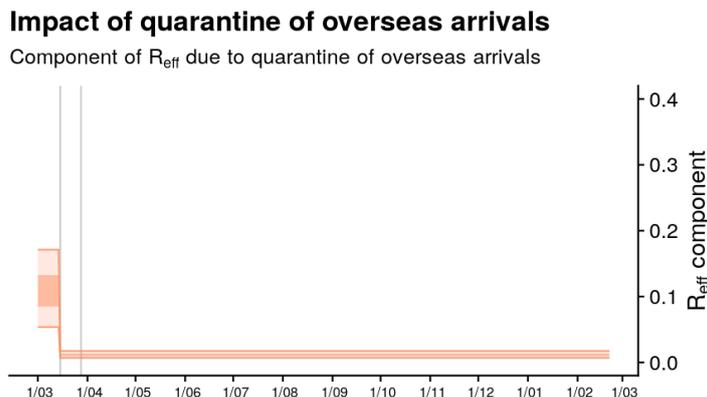
Model limitations

Note that while we have data on whether cases are locally acquired or overseas acquired, no data are currently available on whether each of the locally acquired cases were infected by an imported case or by another locally acquired case. This data would allow us to disentangle the two transmission rates. Without this data, we can separate the denominators (number of infectious cases), but not the numerators (number of newly infected cases) in each group at each point in time. The model we have developed enables us to estimate these effects from the currently available data but missing data reduces the precision of these estimates. For example, we currently cannot account for state-level variation in the impacts of quarantine of overseas arrivals or connect them to specific policies.

Should these data become available, this method will enable us to provide more precise estimates of R_{eff}.

Model description

We developed a semi-mechanistic Bayesian statistical model to estimate R_{eff}, or $R(t)$ hereafter, the effective rate of transmission of of SARS-CoV-2 over time, whilst simultaneously quantifying the impacts on $R(t)$ of a range of policy measures introduced at national and regional levels in Australia.

Observation model

A straightforward observation model to relate case counts to the rate of transmission is to assume that the number of new locally-acquired cases $N_i^L(t)$ at time t in region i is (conditional on its expectation) Poisson-distributed with mean $\lambda_i(t)$ given by the product of the total infectiousness of infected individuals $I_i(t)$ and the time-varying reproduction rate $R_i(t)$:

$$N_i^L(t) \sim \text{Poisson}(\lambda_i(t)) \tag{1}$$

$$\lambda_i(t) = I_i(t)R_i(t) \tag{2}$$

$$I_i(t) = \sum_{t'=0}^{t} g(t')N_i(t') \tag{3}$$

$$N_i(t') = N_i^L(t) + N_i^O(t) \tag{4}$$

where the total infectiousness, $I_i(t)$, is the sum of all active infections $N_i(t')$ — both locally-acquired $N_i^L(t')$ and overseas-acquired $N_i^O(t')$ — initiated at times t' prior to t, each weighted by an infectivity function $g(t')$ giving the proportion of new infections that occur t' days post-infection. The function $g(t')$ is the probability of an infector-infectee pair occurring t' days after the infector's exposure, *i.e.*, a discretisation of the probability distribution function corresponding to the generation interval.

This observation model forms the basis of the maximum-likelihood method proposed by White and Pagano (2007) [1] and the variations of that method by Cori et al. (2013) [2], Thompson et al. (2019) [3] and Abbott et al. (2020) [4] that have previously been used to estimate time-varying SARS-CoV-2 reproduction numbers in Australia.

We extend this model to consider separate reproduction rates for two groups of infectious cases, in order to model the effects of different interventions targeted at each group: those with locally-acquired cases $I_i^L(t)$, and those with overseas acquired cases $I_i^O(t)$, with corresponding reproduction rates $R_i^L(t)$ and $R_i^O(t)$. These respectively are the rates of transmission from imported cases to locals, and from locally-acquired cases to locals. We also model daily case counts as arising from a Negative Binomial distribution rather than a Poisson distribution to account for potential clustering of new infections on the same day, and use a state- and time-varying generation interval distribution $g_i(t',t)$ (detailed in *Surveillance effect model*):

$$N_i^L(t) \sim \text{NegBinomial}(\mu_i(t), r) \tag{5}$$

$$\mu_i(t) = I_i^L(t)R_i^L(t) + I_i^O(t)R_i^O(t) \tag{6}$$

$$I_i^L(t) = \sum_{t'=0}^{t} g_i(t,t')N_i^L(t) \tag{7}$$

$$I_i^O(t) = \sum_{t'=0}^{t} g_i(t,'t)N_i^O(t) \tag{8}$$

where the negative binomial distribution is parameterised in terms of its mean $\mu_i(t)$ and dispersion parameter r. In the commonly used probability and dispersion parameterisation with probability ψ the mean is given by $\mu = \psi r/(1 - \psi)$.

Note that if data were available on the whether the source of infection for each locally-acquired case was another locally-acquired case or an overseas-acquired cases, we could split this into two separate analyses using the observation model above; one for each transmission source. In the absence of such data, the fractions of all transmission attributed to sources of each type is implicitly inferred by the model, with an associated increase in parameter uncertainty.

We provide the model with additional information on the rate of import-to-local transmission by adding a further likelihood term to the model for known events of import-to-local transmission since the implementation of mandatory hotel quarantine:

$$K \sim \text{Poisson}\left(\sum_{i=1}^{8}\sum_{t=\tau_2}^{\tau_3} R_i^O(t)N_i^0(t)\right) \tag{9}$$

where K is the total number of known events of transmission *from* overseas-acquired cases occurring within Australia from $\tau_2 = $ 2020-03-28 to $\tau_3 = $ 2020-12-31. These events are largely transmission events within hotel quarantine facilities, some of which led to outbreaks of local-to-local transmission. Prior to this period, import-to-local transmission events cannot be reliably distinguished from local-to-local transmission events.

When estimating R_{eff} from recent case count data, care must be taken to account for under-reporting of recent cases (those which have yet to be detected), because failing to account for this under-reporting can lead to estimates of R_{eff} that are biased downwards. We correct for this right-truncation effect by first estimating the fraction of locally-acquired cases on each date that we would expect to have detected by the time the model is run (detection probability), and correcting both the infectiousness terms $I_i^L(t)$, and the observed number of new cases $N_i^L(t)$. We calculate the detection probability for each day in the past from the empirical cumulative distribution function of delays from assumed date of infection to date of detection over a recent period (see *Surveillance effect model*). We correct the infectiousness estimates $I_i^L(t)$ by dividing the number of newly infected cases on each day $N_i^L(t)$ by this detection probability — to obtain the expected number of new infections per day — before summing across infectiousness. We correct the observed number of new infections by a modification to the negative binomial likelihood; multiplying the expected number of cases by the detection probability to obtain the expected number of cases observed in the (uncorrected) time series of locally-acquired cases.

Reproduction rate models

We model the onward reproduction rates for overseas-acquired and locally-acquired cases in a semi-mechanistic way. Reproduction rates for local-to-local transmission are modelled as a combination of a deterministic model of the population-wide transmission potential for that type of case, and a correlated time series of random effects to represent stochastic fluctuations in the reporting rate in each state over time. Import-to-local transmission is modelled in a mechanistic way:

$$R_i^L(t) = \exp(\log(R_i^*(t)) - \sigma^2 + \epsilon_i(t)) \tag{10}$$
$$R_i^O(t) = R_i^*(0)Q(t) \tag{11}$$

For locally-acquired cases, the state-wide average transmission rate at time t, $R_i^*(t)$, is given by a deterministic epidemiological model of population-wide transmission potential that considers the effects of distancing behaviours. The correlated time series of random effects $\epsilon_i(t)$ represents stochastic fluctuations in these local-local reproduction rates in each state over time — for example due to clusters of transmission in sub-populations with higher or lower

reproduction rates than the general population. We consider that the transmission potential $R_i^*(t)$ is the average of individual reproduction rates over the entire state population, whereas the effective reproduction number $R_i^L(t)$ is the average of individual reproduction rates among a (non-random) sample of individuals – those that make up the active cases at that point in time. We therefore expect that the long-term average of $R_i^L(t)$ will equate to $R_i^*(t)$. The relationship between these two is therefore defined such that the hierarchical distribution over $R_i^L(t)$ is marginally (with respect to time) a log-normal distribution with mean $R_i^*(t)$. The parameter σ^2 is the marginal variance of the ϵ_i, as defined in the kernel function of the Gaussian process.

For overseas-acquired cases the population-wide transmission rate at time t, $R_i^*(0)Q(t)$, is the baseline rate of transmission ($R_i^*(0) = R_0$; local-to-local transmission potential in the absence of distancing behaviour or other mitigation) multiplied by a quarantine effect model, $Q(t)$, that encodes the efficacy of the three different overseas quarantine policies implemented in Australia (described below).

We model $R_i^*(t)$, the population-wide rate of local-to-local transmission at time t, as the sum of two components: the rate of transmission to members of the same household, and to members of other households. Each of these components is computed as the product of the number of contacts, and the probability of transmission per contact. The transmission probability is in turn modelled as a binomial process considering the duration of contact with each person and the probability of transmission per unit time of contact. This mechanistic consideration of the contact process enables us to separately quantify how macro- and micro-distancing behaviours impact on transmission, and to make use of various ancillary measures of both forms of distancing:

$$R_i^*(t) = s_i(t)(HC_0(1 - (1-p)^{HD_0 h_i(t)d}) + NC_0\delta_i(t)d(1 - (1-p)^{ND_0})\gamma_i(t)) \tag{12}$$

where: $s(t)$ is the effect of surveillance on transmission, due to the detection and isolation of cases (detailed below); HC_0 and NC_0 are the baseline (*i.e.*, before adoption of distancing behaviours) daily rates of contact with, respectively, people who are, and are not, members of the same household; HD_0 and ND_0 are the baseline average total daily duration of contacts with household and non-household members (measured in hours); d is the average duration of infectiousness in days; p is the probability of transmitting the disease per hour of contact, and; $h_i(t)$, $\delta_i(t)$, $\gamma_i(t)$ are time-varying indices of change relative to baseline of the duration of household contacts, the number of non-household contacts, and the transmission probability per non-household contact, respectively (modifying both the duration and transmission probability per unit time for non-household contacts).

The first component in equation (12) is the rate of household transmission, and the second is the rate of non-household transmission. Note that the duration of infectiousness d is considered differently in each of these components. For household members, the daily number of household contacts is typically close to the total number of household members, hence the expected number of household transmissions saturates at the household size; so the number of days of infectiousness contributes to the probability of transmission to each of those household members. This is unlikely to be the case for non-household members, where each day's non-household contacts may overlap, but are unlikely to be from a small finite pool. This assumption would be unnecessary if contact data were collected on a similar timescale to the duration of infectiousness, though issues with participant recall in contact surveys mean that such data are unavailable.

The parameters HC_0, HD_0, and ND_0 are all estimated from a contact survey conducted in Melbourne in 2015 [5]. NC_0 is computed from an estimate of the total number of contacts per day for adults from [6], minus the estimated rate of household contacts. Whilst [5] also provides

an estimate of the rate of non-household contacts, the method of data collection (a combination of 'individual' and 'group' contacts) makes it less comparable with contemporary survey data than the estimate of [6].

The expected duration of infectiousness d is computed as the mean of the non-time-varying discrete generation interval distribution:

$$d = \sum_{t'=0}^{\infty} t' g * (t') \tag{13}$$

and change in the duration of household contacts over time $h_i(t)$ is assumed to be equivalent to change in time spent in residential locations in region i, as estimated by the mobility model for the data stream *Google: time at residential*. In other words, the total duration of time in contact with household members is assumed to be directly proportional to the amount of time spent at home. Unlike the effect on non-household transmission, an increase in macro-distancing is expected to slightly increase household transmission due to this increased contact duration.

The time-varying parameters $\delta_i(t)$ and $\gamma_i(t)$ respectively represent macro- and micro-distancing; behavioural changes that reduce mixing with non-household members, and the probability of transmission for each of non-household member contact. We model each of these components, informed by population mobility estimates from the mobility model and calibrated against data from nationwide surveys of contact behaviour.

Surveillance effect model

Disease surveillance — both screening of people with COVID-like symptoms and performing contact tracing — can improve COVID-19 control by placing cases in isolation so that they are less likely to transmit the pathogen to other people. Improvements in disease surveillance can therefore lead to a reduction in transmission potential by isolating cases more quickly, and reducing the time they are infectious but not isolated. Such an improvement changes two quantities: the population average transmission potential $R^*(t)$ is reduced by a factor $s_i(t)$; and the generation interval distribution $g(t, t')$ is shortened, as any transmission events are more likely to occur prior to isolation.

We model both of these functions using a region- and time-varying estimate of the discrete probability distribution over times from infection to detection $f_i(t, t')$:

$$g_i(t, t') = \frac{f_i(t, t') g^*(t')}{s_i(t)} \tag{14}$$

$$s_i(t) = \sum_{t'=0}^{\infty} f_i(t, t') g^*(t') \tag{15}$$

where $g^*(t')$ is the baseline generation interval distribution, representing times to infection in the absence of detection and isolation of cases, $s_i(t)$ is a normalising factor — and also the effect of surveillance on transmission — and $f_i(t, t')$ is a region- and time-varying probability density over periods from infection to isolation t'. In states/territories and at times when cases are rapidly found and placed in isolation, the distribution encoded by $f_i(t, t')$ has most of its mass on small delays, average generation intervals are shortened, and the surveillance effect $s_i(t)$ tends toward 0 (a reduction in transmission). At times when cases are not found and isolated until after most of their infectious period has passed, $f_i(t, t')$ has most of its mass on large delays, generation intervals are longer on average, and $s_i(t)$ tends toward 1 (no effect of reduced transmission).

We model the region- and time-varying distributions $f_i(t, t')$ empirically via a time-series of empirical distribution functions computed from all observed infection-to-isolation periods

observed within an adaptive moving window around each time t. Since dates of infection and isolation are not routinely recorded in the dataset analysed, we use 5 days prior to the date of symptom onset to be the assumed date of infection, and the date of case notification to be the assumed date of isolation. This will overestimate the time to isolation and therefore underestimate the effect of surveillance when a significant proportion of cases are placed into isolation prior to testing positive – e.g. during the tail of an outbreak being successfully controlled by contact tracing.

For a given date and state/territory, the empirical distribution of delays from symptom onset to notification is computed from cases with symptom onset falling within a time window around that date, with the window selected to be the smallest that will yield at least 500 observations; but constrained to between 1 and 8 weeks.

Where a state/territory does not have sufficient cases to reliably estimate this distribution in an 8 week period, a national estimate is used instead. Specifically, if fewer than 100 cases, the national estimate is used, if more than 500 the state estimate is used, and if between 100 and 500 the distribution is a weighted average of state and national estimates. The national estimate is obtained via the same method but with no upper limit on the window size and excluding data from Victoria since 14 June, since the situation during the Victorian outbreak after this time is not likely to be representative of surveillance in states with few cases.

Macro-distancing model
The population-wide average daily number of non-household contacts at a given time can be directly estimated using a contact survey. We therefore used data from a series of contact surveys commencing immediately after the introduction of distancing restrictions to estimate $\delta_i(t)$ independently of case data. To infer a continuous trend of $\delta_i(t)$, we model the numbers of non-household contacts at a given time as a function of mobility metrics considered in the mobility model. We model the log of the average number of contacts on each day as a linear model of the log of the ratio on baseline of five Google metrics of time spent at different types of location: residential, transit stations, parks, workplaces, and retail and recreation:

$$\log(\delta_i(t)) = (\omega \odot \mathbf{m}) \log(\mathbf{M}_i(t)). \tag{16}$$

where ω is the the vector of 5 coefficients, $\mathbf{m}$ is an vector of length 5 containing of ones, except for the element corresponding to time at residential locations, which has value 1, and $\odot$ indicates the elementwise product. This constrains the direction of the effect of increasing time spent at each of these locations to be positive (more contacts), except for time at residential, which we constrain to be negative. The intercept of the linear model (average daily contacts at baseline) is given an prior formed from the daily number of non-household contacts in a pre-COVID-19 contact survey [5]. Since our aim is to capture general trends in mobility rather than daily effects, we model the weekly average of the daily number of contacts, by using smoothed estimates of the Google mobility metrics.

Whilst we aim to model weekly rather than daily variation in contact rates, when fitting the model to survey data we account for variation among responses by day of the week by modelling the fraction of the weekly number of contacts falling on each day of the week (the length-seven vector in each state and time $\mathbf{D}_i(t)$) and using this to adjust the expected number of contacts for each respondent based on the day of the week they completed the survey. To account for how the weekly distribution of contacts has changed over time as a function of mixing restrictions (*e.g.*, a lower proportion of contacts on weekdays during periods when stay-at-home orders were in place) we model the weekly distribution of contacts itself as a function of deviation in the

weekly average of the daily number of contacts, with length-seven vector parameters α and θ. We use the softmax (normalised exponential) function to transform this distribution to sum to one, then multiply the resulting proportion by 7 to reweight the weekly average daily contact rate to the relevant day of the week.

Combining the baseline average daily contact rate NC_0, mobility-driven modelled change in contact rates over time $\delta_i(t)$, and time-varying day of the week effects $\mathbf{D}_i(t)$ we obtain an expected number of daily contacts for each survey response NC_k:

$$\log(NC_k) = \log(NC_0) + \log(\delta_{i[k]}(t[k])) + \log(\mathbf{D}_{i[k]}(t[k]) * 7)_{d[k]} \tag{17}$$
$$\mathbf{D}_i(t) = \text{softmax}(\alpha + \theta \log(\delta_i(t))) \tag{18}$$

where $i[k]$, $t[k]$, and $d[k]$ respectively indicate the state, time, and day of the week on which respondent k filled in the survey.

We model the number of contacts from each survey respondent as a draw from an interval-censored discrete lognormal distribution. This choice of distribution enables us to account for the *ad-hoc* rounding of reported numbers of contacts (responses larger than 10 tend to be 'heaped' on multiples of 10 and 100), whilst also accounting for heavy upper tail in numbers of reported contacts. The support of this distribution is the integers from 0 to 10 inclusive, and the intervals 11-20, 21-50, and 50-999. Reported daily contact rates ≥ 1000 are excluded as these are considered implausible for our definition of a contact. The probability mass function of this distribution is the integral across these ranges of a lognormal distribution with parameters μ_k and τ, parameterised such that the mean of the distribution is NC_k:

$$\mu_k = \log(NC_k) - \tau^2/2 \tag{19}$$

Micro-distancing model
Unlike with macro-distancing behaviour and contact rates, there is no simple mathematical framework linking change in micro-distancing behaviours to changes in non-household transmission probabilities. We must therefore estimate the effect of micro-distancing behaviour on transmission via case data. We implicitly assume that any reduction in local-to-local transmission potential that is not explained by changes to the numbers of non-household contacts, the duration of household contacts, or improved disease surveillance is explained by the effect of micro-distancing on non-household transmission probabilities.

Whilst it is not necessary to use ancillary data to estimate the effect that micro-distancing has at its peak, we use behavioural survey data to estimate the temporal trend in micro-distancing behaviour, in order to estimate to what extent adoption of that behaviour has waned and how that has affected transmission potential.

We therefore model γ_t (see equation (12)) as a function of the proportion of the population adhering to micro-distancing behaviours. We consider adherence to the '1.5m rule' as indicative of this broader suite of behaviours due to the availability of data on this behaviour in a series of weekly behavioural surveys beginning prior to the last distancing restriction being implemented [7]. We consider the number $m_{i,t}^+$ of respondents in region i on survey wave commencing at time t replying that they 'always' keep 1.5m distance from non-household members, as a binomial sample with sample size $m_{i,t}$. We model $c_i(t)$, the proportion of the population in region i responding that they always comply as a function of time, composed of an initial adoption phase, a date of peak compliance, and a subsequent piecewise linear trend. We assume that the temporal pattern in the initial rate of adoption of the behaviour is the same as for macro-distancing behaviours — the adoption curve estimated from the mobility model. In other words,

we assume that all macro- and micro-distancing behaviours were adopted simultaneously around the time the first population-wide restrictions were put in place in March and April 2020. However we do not assume that these behaviours peaked at the same time or subsequently followed the same temporal trend. The model for the proportion complying with this behaviour is therefore:

$$m_{i,t}^+ = \text{Binomial}(m_{i,t}, c_i(t)) \tag{20}$$

$$c_i(t) = \mathbf{w}(t, \kappa_i)\mathbf{h}_i - \kappa_{i,1}(1 - d_i(t)) \tag{21}$$

$$logit(\kappa_{i,0}/T) \sim N(\mu_{\kappa_0}, \sigma_{\kappa_0}^2) \tag{22}$$

$$logit(\kappa_{i,l-1}/\kappa_{i,l}) \sim N(\mu_{\kappa_l}, \sigma_{\kappa_l}^2) \tag{23}$$

$$logit(h_{i,l}) \sim N(\mu_{h_l}, \sigma_{h_l}^2) \tag{24}$$

where $d_i(t)$ is the latent function for adoption of distancing behaviour, estimated from the mobility model (scaled from 0 at baseline to 1 at maximum), the vector of weights against the components of the piecewise linear model $\mathbf{w}(t, \kappa_i)$ are computed as a function of time t and the vector of inflection times for state i; κ_i ($\kappa_{i,0}$ being the date of peak compliance). For each inflection l, the timing $\kappa_{i,l}$) and height $h_{i,l}$ of the inflection in each state are drawn from a hierarchical distribution to share information between states on behavioural changes. The prior on timings within each state are is structured to ensure increasing dates of inflection, to ensure identifiability of the model.

Given $c_i(t)$, we model $\gamma_i(t)$ as a functiosn of the degree of micro-distancing relative to the peak:

$$\gamma_i(t) = 1 - \beta(c_i(t)/\kappa_{i,0}) \tag{25}$$

with β inferred from case data in the main R_{eff} model.

Overseas quarantine model
We model the effect of overseas quarantine $Q(t)$ via a monotone decreasing step function with values constrained to the unit interval, and with steps at the known dates τ_1 and τ_2 of changes in quarantine policy:

$$Q(t) = \begin{cases} q_1 & t < \tau_1 \\ q_2 & \tau_1 \leq t < \tau_2 \\ q_3 & \tau_2 \leq t \end{cases} \tag{26}$$

where $q_1 > q_2 > q_3$ and all parameters are constrained to the unit interval.

Error models
The correlated time-series of errors in the log of the effective reproduction rate for local-to-local transmission in each region $\epsilon_i(t)$ is each modelled as a zero-mean Gaussian process (GP) with covariance structure reflecting temporal correlation in errors within each region, but independent between regions. We use a rational quadratic covariance function k_{RQ}, enabling periods of comparatively smooth variations, with occasional more rapid fluctuations, to represent the sudden rapid growth of cases that can occur with a high-transmission cluster. Kernel parameters are assumed to be the same across regions:

$$\epsilon_i \sim GP(\mathbf{0}, k_{RQ}(t, t')) \tag{27}$$

$$k_{RQ}(t, t') = \sigma^2 \exp\left(1 + \frac{(t - t')^2}{2\alpha l^2}\right)^{-\alpha} \tag{28}$$

Components of local transmission potential

We model the rate of transmission from locally acquired cases as the product of the time-varying mechanistic model of transmission rates $R_i^*(t)$, and a temporally-correlated error term $e^{\epsilon_i(t)}$. This structure enables inference of mechanistically interpretable parameters whilst also ensuring that statistical properties of the observed data are represented by the model. Moreover, these two parts of the model can also be interpreted in epidemiological terms as two different components of transmission rates:

1. **Component 1** – transmission rates averaged over the whole state population, representing how macro- and micro-distancing affect the potential for widespread community transmission $(R_i^*(t))$, and

2. **Component 2** – the degree to which the transmission rates of the population of current active cases deviates from the average statewide transmission rate $(e^{\epsilon_i(t)})$.

Component 2 reflects the fact that the population of current active cases in each state at a given time will not be representative of the the state-wide population, and may be either higher (*e.g.*, when cases arise from a cluster in a high-transmission environment) or lower (*e.g.*, when clusters are brought under control and cases placed in isolation).

Component 1 can therefore be interpreted as the expected rate of transmission if cases were widespread in the community. The product of Components 1 and 2 can be interpreted as the rate of transmission in the sub-population making up active cases at a given time.

Where a state has active cases in one or more clusters, the product of these components gives the apparent rate of transmission in those clusters. Where a state has no active cases, the product of Components 1 and 2 gives the rate of spread expected if an index case were to occur in a random sub-population. Because the amplitude of this error term is learned from the data, this is informative as to the range of plausible rates of spread that might be expected from a case being introduced into a random sub-population.

Parameter values and prior distributions

Tables S1 and S3 give the prior distributions of parameters in the semi-mechanistic and time-series (ϵ) parts of the model respectively. Table S2 gives fixed parameter values used in the semi-mechanistic part of the model.

The parameters of the generation interval distribution are the posterior mean parameter estimates corresponding to a lognormal distribution over the serial interval estimated by [8]. The shape of the generation interval distribution for SARS-CoV-2 in comparable populations is not well understood, and a number of alternative distributions have been suggested by other analyses. A sensitivity analysis performed by running the model with alternative generation interval distributions (not presented here) showed that parameter estimates were fairly consistent between these scenarios, and the main findings were unaffected. A full, formal analysis of sensitivity to this and other assumptions will be presented in a future publication.

No ancillary data are available to inform p, the probability of transmission per hour of contact in the absence of distancing behaviour. However, at $t = 0$, holding HC_0, NC_0 HD_0, and ND_0 constant, there is a deterministic relationship between p and $R_i^*(0)$ (the basic reproduction rate, which is the same for all states). The parameter p is therefore identifiable from transmission rates at the beginning of the first epidemic wave in Australia. We define a prior on p that corresponds to a prior over $R_i^*(0)$ matching the averages of the posterior means and 95% credible intervals for 11 European countries as estimated by [9] in a sensitivity analysis where the mean generation interval was 5 days — similar to the serial interval distribution assumed here. This corresponds to a prior mean of 2.79, and a standard deviation of 1.70 for $R_i^*(0)$.

This prior distribution over p was determined by a Monte-Carlo moment-matching algorithm, integrating over the prior values for HC_0, NC_0 HD_0, and ND_0.

Model fitting
We fitted (separate) models of $c_i(t)$ and $NC_0\delta_i(t)$ to survey data alone in order to infer trends in those parameters as informed by survey data. These are shown in Figures 5–6. We used the posterior means of each of these model outputs as inputs into the R_{eff} model. The posterior variance of each of these quantities is largely consistent over time and between states, and the absolute effect of each is scaled by other parameters (e.g. β), meaning that uncertainty in these quantities is largely not identifiable from uncertainty in other scaling parameters. As a consequence, propagation of uncertainty in these parameters into the R_{eff} model (as was performed in a previous iteration of the model) has little impact on estimates of R_{eff} and transmission potential, so is avoided for computational brevity.

Inference was performed by Hamiltonian Monte Carlo using the R packages `greta` and `greta.gp` [10, 11]. Posterior samples of model parameters were generated by 10 independent chains of a Hamiltonian Monte Carlo sampler, each run for 1000 iterations after an initial, discarded, 'warm-up' period (1000 iterations per chain) during which the sampler step size and diagonal mass matrix was tuned, and the regions of highest density located. Convergence was assessed by visual assessment of chains, ensuring that the potential scale reduction factor for all parameters had values less than 1.1, and that there were at least 1000 effective samples for each parameter.

Visual posterior predictive checks were performed to ensure that the observed data were consistent with the posterior predictive density over all cases (and survey results), and over time-varying case predictions within each state.

Table S1: Parameters in the semi-mechanistic part of the time-varying model of R_{eff}. Prior on weights for ω correspond to Google mobility metrics in the following order: parks, residential, retail and recreation, transit stations, workplaces.

Prior distribution	Parameter description
$r^{-1/2} \sim \text{N}^+(0, 0.5)$	Overdispersion of observed daily new infections
$\text{logit}(p) \sim \text{N}(2.57, 0.08^2)$	Transmission probability per hour contact time
$HC_0 \sim \text{N}^+(2.09, 0.06^2)$	Baseline average daily household contacts
$NC_0 \sim \text{N}^+(10.70, 0.28^2)$	Baseline average daily non-household contacts
$HD_0 \sim \text{N}^+(1.05, 1.68^2)$	Baseline daily duration per household contact (hours)
$ND_0 \sim \text{N}^+(0.687, 0.05^2)$	Baseline daily duration per non-household contact (hours)
$\omega \sim \text{N}^+(0, 1^2)$	Mobility-metric weights for non-household contact rates
$\alpha \sim \text{N}(0, 1)$	Effect of day of the week on non-household contact rates
$\theta \sim \text{N}(0, 1)$	Effect of day-of-week/mobility interaction on contact rate responses
$r_{NC}^{-1/2} \sim \text{N}^+(0, 0.5)$	Overdispersion of daily non-household contacts
$\mu_{\kappa_{i,0}} \sim \text{N}(0, 10^2)$	Hierarchical mean for peak microdistancing timing
$\mu_{\kappa_{i,l}} \sim \text{N}(0, 10^2)$	Hierarchical mean for microdistancing inflection timing
$\mu_{h_{i,l}} \sim \text{N}(0, 10^2)$	Hierarchical mean for microdistancing inflection height
$\sigma_{\kappa_{i,0}} \sim \text{N}^+(0, 0.5^2)$	Hierarchical s.d. for peak microdistancing timing
$\sigma_{\kappa_{i,l}} \sim \text{N}^+(0, 0.5^2)$	Hierarchical s.d. for microdistancing inflection timing
$\sigma_{h_{i,l}} \sim \text{N}^+(0, 0.5^2)$	Hierarchical s.d. for microdistancing inflection height
$\beta \sim \text{U}(0, 1)$	Microdistancing effect on transmission
$q_1 \sim \text{U}(0, 1)$	Effect of quarantine of overseas arrivals (phase 1)
$q_2 \times q_1 \sim \text{U}(0, 1)$	Relative effect of quarantine (phase 2 vs 1)
$q_3 \times q_2 \sim \text{U}(0, 1)$	Relative effect of quarantine (phase 3 vs 2)

Table S2: Fixed parameters in the semi-mechanistic part of the time-varying model of R_{eff}.

Parameter value	Parameter description
$\tau_1 = 2020\text{-}03\text{-}15$	Date of change from arrivals policy phase 1 to 2
$\tau_2 = 2020\text{-}03\text{-}28$	Date of change from arrivals policy phase 2 to 3
$\tau_3 = 2020\text{-}12\text{-}31$	Date of end of observed quarantine spillover period
$T = 2021\text{-}02\text{-}09$	Date of most recent mobility data
$g^*(t) = \int_{t-1}^{t} \text{lognormal}(\tau \mid 1.377, 0.567^2)\, d\tau$	Baseline generation interval function

Table S3: Parameters used in the time-series part of the time-varying model of R_{eff}.

Prior distribution	Parameter description
$\sigma \sim \text{N}^+(0, 0.5^2)$	State-level component of amplitude of deviation R_{eff}
$l \sim \text{lognormal}(3, 1)$	Temporal correlation R_{eff}
$\alpha \sim \text{lognormal}(3, 1)$	Correlation mixture weights R_{eff}

Estimating the relative transmissibility of SARS-CoV-2 VOC 202012/01

Overview

Rapid estimates of the relative transmissibility of VOC 202012/01 compared with non-VOC lineages were made by two groups in mid-January (not peer-reviewed at the time): the London School of Hygiene and Tropical Medicine (LSHTM) and Imperial College. LSHTM estimated a 31% [27%, 34%] increase in the reproduction number based on their regional time-varying model assuming an average generation interval of 3.6 days[2] and Imperial estimated a 50–75% increase in the reproduction number assuming an average generation interval of 6.5 days[3].

These rapidly generated estimates assume that the increased transmissibility of VOC 202012/01 can be represented as a multiplicative increase in R_{eff}. However, the impact of VOC 202012/01 on transmissibility is likely to depend on the level of restrictions and population behaviour. We therefore performed an independent analysis of the relative transmissibility of SARS-CoV-2 VOC 202012/01 compared with non-VOCs in the UK, using:

- data from Public Health England on secondary attack rates among known contacts of cases;

- our model for estimating transmission potential in Australia (the outputs of which are routinely reported to AHPPC) which separately considers household and non-household rates of transmission;

- data on macro-distancing behaviour (from both the UK and Australia) and mobility and micro-distancing behaviour (from the UK).

This approach allow us to directly estimate the impact of VOC 202012/01 on the probability of transmission to a contact per unit of contact time and therefore account for variability in relative transmissibility between high-restriction and low-restriction scenarios (see below for details).

We estimate increases in transmission potential/transmissibility of VOC 202012/01 relative to non-VOCs ranging from 40% [30, 50] under nationwide "stay-at-home" restrictions in Australia in March/April 2020 to 48% [35, 60] for a pre-pandemic baseline (R_0).

Updates to our method may be possible as more information becomes available on VOC 202012/01 and other emerging variants of higher transmissibility.

Methodology

We carried out an analysis to independently estimate the relative transmissibility of SARS-CoV-2 VOC 202012/01 compared to non-VOCs, and to account for variability in relative transmissibility between high-restriction and low-restriction scenarios. We sought to directly estimate the impact of VOC 202012/01 on the probability of transmission to a contact per unit of contact time. A change to this parameter is consistent with the hypothesis that the increased growth rates in cases associated with VOC 202012/01 are due to increased viral shedding during infection. With an estimate of this parameter, we can modify our estimates of transmission potential in Australia, whilst accounting for estimated changes to the fraction of contacts that are made with household members, the duration of time spent in the household, and changes to micro-distancing behaviours. We estimated this key parameter by adapting the mathematical

---

[2] https://cmmid.github.io/topics/covid19/local-r-sgtf.html
[3] https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-42-sars-cov-2-variant/

model of household and non-household transmission that forms part of our routine methodology for estimating transmission potential in Australia, and fitting it to data from Public Health England on secondary attack rates among contacts for VOC 202012/01 and non-VOCs in nine English regions.

Model for household and non-household attack rates

Our existing transmission potential model explicitly considers secondary attack rates among household members and non-household members, modelled as a function of: the probability of transmission per unit contact time; the average duration of contacts with household and non-household members; and modification of the non-household attack rate. The latter is a combined effect of reductions in the per-unit-contact-time transmission probability and in the average duration of non-household contacts.

We explicitly model the household secondary attack rate at time/location i as:

$$HSAR_i = 1 - (1 - p)^{HD_i} \tag{29}$$

where p is the probability of transmission per unit of contact time, and HD_i is the average duration of household contacts at time and place i, summed over the full course of infection. We model the secondary attack rate for non-household members as:

$$OSAR_i = \gamma_i * (1 - (1 - p)^{OD_0}) \tag{30}$$

where OD_0 is the average duration of non-household contacts per 24 hours at baseline (prior to the pandemic and restrictions), and γ_i is the reduction in non-household secondary attack rates as a function of micro-distancing behaviour. We infer the parameters HD_i and γ_i from data on mobility and behavioural change as:

$$HD_i = HD_0 * h_i \tag{31}$$

$$\gamma_i = 1 - \beta * d_i \tag{32}$$

where HD_0 is the average duration of household contacts over the full infectious period at baseline, h_i is proportional change in the amount of time spent in the household, inferred from the Google mobility metric 'Time at Residential', d_i is the degree of adherence of micro-distancing behaviour, scaled to range from 0 at baseline to 1 at the peak of micro-distancing, and β is a free parameter controlling the impact of micro-distancing on reducing non-household transmission that is fitted to Australian case data.

Effect of VOC

We model the effect of VOC 202012/01 on per-unit-contact time probability of transmission via a parameter for the power of the probability of **not transmitting** per unit of contact time:

$$p^* = 1 - (1 - p)^{\phi} \tag{33}$$

where p^* and p are the per-unit-contact time probabilities of transmission for VOC 202012/01 and non-VOCs, respectively, and ϕ is a free parameter constrained to be positive that controls the relative infectiousness. $\phi = 1$ would imply the two variants have the same transmissibility. The aim of this analysis is to infer ϕ from UK attack rate data.

Fitting to UK attack rate data

Public Health England's Technical Report 3 on the VOC 202012/01 Table 6 reports numbers of contacts of cases with VOC 202012/01 and non-VOCs in nine English regions, and the number of those contacts that became cases, between 2020/09/20 and 2021/01/04: `https://www.gov.uk/government/publications/investigation-of-novel-sars-cov-2-variant-variant-of-concern-20201201`. We fit a model that separately considers attack rates in each of these regions, using region-specific estimates of mobility, micro-distancing, and macro-distancing. By considering all nine regions as independent observations (rather than aggregating the data for all of England), we increase statistical power and consider the effect of the variant at different levels of restrictions. The fact that we see similar estimates of attack rates across all nine regions gives us confidence that higher attack rates are due to biological differences between variants rather than founder effects or confounding with outbreaks in specific settings.

Unfortunately, these data are not provided dis-aggregated at a finer temporal resolution. Nor are the attack rate estimates dis-aggregated by whether or not the contacts were household members. We must therefore adapt our model to estimate an overall attack rate over contacts, and adjust it for non-random ascertainment of contacts in the PHE dataset.

We can estimate the overall secondary attack rate for each region SAR_i as a combination of household and non-household secondary attack rates weighted by w_i, the fraction of contacts that are household members:

$$SAR_i = w_i * HSAR_i + (1 - w_i) * OSAR_i \tag{34}$$
$$w_i = HC/(HC + OC_i * ID) \tag{35}$$

where HC is the average number of household contacts (assumed the same for each region), OC_i is the average number of non-household contacts per 24 hours, and ID is the average duration of infectiousness in days. Our model assumes that household contacts stay the same throughout the course of infection, but that there is a different set of non-household contacts on each day.

The overall secondary attack rates estimated by this model correspond to the average number of contacts specified as HC and OC_i. Whilst the number of household contacts is likely to be consistent between analyses, the operational contact definition used by the contact tracing teams that provided the PHE data is likely to yield a smaller number of contacts than the contact surveys used to estimate OC_i. Moreover, the number of contacts will not be a random sample of the larger number of contacts, since operational contact tracing will target those individuals with a greater risk of transmission. The consequence of this is that observed attack rates are biased upwards. This will also affect estimates of the relative transmissibility of VOC 202012/01 from these raw data — reducing the apparent transmissibility. We account for these issues by introducing a free parameter ψ to relate the 'true' and observed attack rates: SAR_i^{ψ}.

Full model

We specified a Bayesian statistical model to estimate ϕ and the other parameters from UK attack rate data as follows:

$$C_i \sim Binomial(N_i, SAR_i^{\psi}) \tag{36}$$

$$C_i^* \sim Binomial(N_i^*, SAR_i^{*\psi}) \tag{37}$$

$$SAR_i = w_i * HSAR_i + (1 - w_i) * OSAR \tag{38}$$

$$SAR_i^* = w_i * HSAR_i^* + (1 - w_i) * OSAR^* \tag{39}$$

$$w_i = HC/(HC + OC_i * ID) \tag{40}$$

$$HSAR_i = 1 - (1 - p)^{HD_i} \tag{41}$$

$$HSAR_i^* = 1 - (1 - p^*)^{HD_i} \tag{42}$$

$$OSAR_i = \gamma_i * \left(1 - (1 - p)^{OD_0}\right) \tag{43}$$

$$OSAR_i^* = \gamma_i * \left(1 - (1 - p^*)^{OD_0}\right) \tag{44}$$

$$HD_i = HD_0 * h_i \tag{45}$$

$$\gamma_i = 1 - \beta * d_i \tag{46}$$

where N_i and C_i are the number of contacts, and the number of those contacts that became cases in each English region i, and all variables with superscript $*$ correspond to infection with the VOC 202012/01, and those without correspond to non-VOCs. The model was fitted by MCMC using the same algorithm and software as the model for R_{eff}. The model was run until there were at least 1000 effective samples of each parameter. Convergence was assessed visually and by the potential scale reduction factor (less than 1.01 for all parameters). Calibration of the model was assessed by posterior predictive checks over each of C_i, C_i^*, and the empirical estimate of the ratio of attack rates between variants for each region: $\frac{C_i^*/N_i^*}{C_i/N_i}$, and indicated good fit.

Parameter values and prior distributions

When fitting the R_{eff} model for Australia, the parameters OC_i (non-household contacts per 24 hours), h_i (relative time spent at home), and d_i (relative micro-distancing effect) are all informed by bespoke statistical models tailored to the Australian situation and surveys carried out only in Australia. We developed equivalent estimates of these parameters for the UK from a range of other sources.

To estimate OC_i we used the macro-distancing model fitted to Australian contact survey data to predict the number of non-household contacts per days in each English region, based on the values of Google mobility metrics for those regions. Google mobility data were downloaded for each English county, aggregated up to compute the average value over each region, and then averaged for each region over the period over which attack rate data were collected. Predictions of the Australian contact model were visually compared with summary statistics of non-household contact rates from April to August 2020 as estimated by the UK's CoMix survey series and found to have good calibration.

We used the aggregated estimate of change in Google's time at residential to inform h_i.

To estimate d_i, we analysed data on adherence to the UK's 2m rule using data for each English region from regular YouGov behavioural surveys conducted in partnership with Imperial College London. We calculated the number of people responding that they had not broken the 2m rule ("come into physical contact with (within 2 meters / 6 feet)") in the past seven days. This is analogous to the 1.5m rule question used to define our micro-distancing metric

in Australia. This time series was analysed using a Binomial Generalised Additive Model to estimate a time-series of the metric for each region over the course of the pandemic. This time-series was re-scaled to have maximum value 1 and then averaged over the time period over which attack rate data were collected.

The model comprised six parameters; four for which we have existing estimates (p, HC, HD_0, and OD_0) and two for which we do not (ϕ and ψ). We defined an informative prior for p based on a normal approximation to the posterior for this parameter from the Australian R_{eff} model. This assumes *a priori* that the non-VOCs in the UK have equivalent infectiousness to the variants that have circulated in Australia to date, though the parameter can be amended by the UK attack rate model fitting procedure if this is inconsistent with the data.

For HC, HD_0, and OD_0, we used the same priors as we use in fitting the Australian model of R_{eff}– based on surveys of contact behaviour in Australian prior to the pandemic. The average number of household contacts in each English region as reported in the YouGov surveys agreed closely with this Australian prior for HC. We chose to use the Australian estimate rather than the UK estimates since the posterior estimate of p was estimated contingent on this distribution.

Both ϕ and ψ must be positive and a value of 1 indicates no effect (of the variant or of bias in contact acquisition, respectively). We therefore specified minimally informative positive-truncated normal prior distributions for both parameters, with mode (μ parameter of the normal distribution) of 1. For ψ we set the standard deviation of the normal prior distribution, σ, to 1 to allow a large range of values, and for ϕ we set it to 1. Prior predictive checks on the ratio of attack rates between VOC 202012/01 and non-VOCs (SAR_i^*/SAR_i) with this prior on ϕ confirmed that the prior was vague with respect to the relative transmissibility of VOC 202012/01 versus non-VOCs. In other words, multiplicative increases in transmissibility of VOC 202012/01 estimated from other studies were within the bulk of the prior distribution, as were larger increases and decreases in transmissibility.

β was fixed at the posterior mean as estimated from the Australian model. In the absence of a time series of attack rate data, it is not possible to estimate this parameter independently for the UK, and the value of the parameter is poorly statistically identified in this model due to potential confounding with other parameters – especially ψ. For this reason, uncertainty in β was not considered in this analysis.

Supplement: ensemble forecasts of the daily number of new local cases

We report month-ahead state-level forecasts of the daily number of new confirmed cases cases — *synthesised* from three independent models (known as an 'ensemble forecast'). Ensemble forecasts are more accurate than any individual forecast alone — biases and variances in each model that result from different modelling choices balance against each other to improve predictions. Hence, ensemble forecasts tend to produce improved estimates of both the central values, as well as improved estimates of the plausible yet unlikely forecasts (uncertainty). Here, the ensemble has been generated by equally weighting the forecasts from each model. A brief description of each method incorporated in the ensemble is given below (and full methodological details are provided in subsequent sections):

- **SEEIIR Forecast**: A stochastic susceptible-exposed-infectious-recovered (SEEIIR) compartmental model that incorporates changes in local transmission potential via the estimated time-varying effective reproduction number (as shown in Figure 3).

- **Probabilistic Forecast**: A stochastic epidemic model that accounts for the number of imported-, symptomatic- and asymptomatic-cases over time. This model estimates the effective reproduction number corresponding to local and imported cases, and incorporates mobility data to infer the effect of macro-distancing behaviour. This model captures variation in the number and timing of new infections via probability distributions. The parameters that govern these distributions are inferred from the case and mobility data (*e.g.*, mean number of imported cases).

- **Time-Series Forecast**: A time-series model that does not account for disease transmission dynamics, but rather uses recent daily case counts to forecast cases into the future. Parameters of this 'autoregressive' model are estimated using global data accessible via the Johns Hopkins COVID-19 repository. Case counts from a specific time window prior to the forecasting date (the present) are used for model calibration. The number of days within this time window is chosen to optimise projections for Australia data.

The SEEIIR and Probabilistic Forecasts explicitly incorporate dynamics of disease transmission and the impact of public health measures on transmission over time via R_{eff}. The Time-Series Forecast does not explicitly incorporate either of these factors. The Time-Series Forecast is expected to accurately forecast new daily case numbers over a shorter time period, whereas disease-specific models are anticipated to provide more accurate forecasts several weeks into the future.

1. SEEIIR Forecast

Model Description

We used a discrete-time stochastic SEEIIR model to characterise infection in each Australian jurisdiction. Let $S(t)$ represent the number of *susceptible* individuals, $E_1(t) + E_2(t)$ represent the number of *exposed* individuals, $I_1(t) + I_2(t)$ represent the number of *infectious* individuals, and $R(t)$ the number of *removed* individuals, at time t. Symptom onset is assumed to coincide with the transition from I_1 to I_2. Note that the two exposed and infectious classes are specified in order to obtain a Gamma distribution (with shape parameter 2) on the duration of time in the exposed and infectious classes, respectively. It is assumed that 10 exposures were introduced

into the E_1 compartment at time τ, to be inferred, giving initial conditions:

$$S(0) = N - E_1(0) \qquad\qquad E_1(0) = 10$$
$$E_2(0) = 0 \qquad\qquad I_1(0) = 0$$
$$I_2(0) = 0 \qquad\qquad R(0) = 0$$
$$\sigma(t) = \begin{cases} 0 & \text{if } t < \tau \\ \sigma & \text{if } t \geq \tau \end{cases} \qquad\qquad \gamma(t) = \begin{cases} 0 & \text{if } t < \tau \\ \gamma & \text{if } t \geq \tau \end{cases}$$
$$\beta(t) = R_{\text{eff}}(t) \cdot \gamma(t)$$

The number of individuals leaving each compartment on each *daily* time-step follows a Binomial distribution, as follows:

$$S = 1 - \exp\left(-\beta(t) \cdot [I_1(t) + I_2(t)]/N\right) \qquad S \sim Bin(S(t), S)$$
$$E_1 = 1 - \exp\left(2 \cdot \sigma(t)\right) \qquad E_1 \sim Bin(E_1(t), E_1)$$
$$E_2 = 1 - \exp\left(2 \cdot \sigma(t)\right) \qquad E_2 \sim Bin(E_2(t), E_2)$$
$$I_1 = 1 - \exp\left(2 \cdot \gamma(t)\right) \qquad I_1 \sim Bin(I_1(t), I_1)$$
$$I_2 = 1 - \exp\left(2 \cdot \gamma(t)\right) \qquad I_2 \sim Bin(I_2(t), I_2)$$
$$S(t+1) = S(t) - S \qquad E_1(t+1) = E_1(t) + S - E_1$$
$$E_2(t+1) = E_2(t) + E_1 - E_2 \qquad I_1(t+1) = I_1(t) + E_2 - I_1$$
$$I_2(t+1) = I_2(t) + I_1 - I_2 \qquad R(t+1) = R(t) + I_2$$

We modelled the relationship between model incidence and the observed daily COVID-19 case counts (y_t) using a Negative Binomial distribution with dispersion parameter k, since the data are non-negative integer counts and are over-dispersed when compared to a Poisson distribution. Let $X(t)$ represent the state of the dynamic process *and* particle filter particles at time t, and x_t represent a realisation, i.e., $x_t = (s_t, e_{1t}, e_{2t}, i_{1t}, i_{2t}, r_t, \sigma_t, \gamma_t, \beta_t)$. The probability of being observed (i.e., of being reported as a notifiable case) is the product of two probabilities: that of entering the I_2 compartment, $p_{\text{inc}}(t)$, and the observation probability p_{obs}. In order to improve the stability of the particle filter for very low (or zero) incidence, we also allowed for the possibility of a very small number of observed cases that are *not* directly a result of the community-level epidemic dynamics (bg_{obs}). The observation process is thus defined as:

$$\mathcal{L}(y_t \mid x_t) \sim (\mathbb{E}[y_t], k)$$
$$\mathbb{E}[y_t] = (1 - p_{\text{inc}}(t)) \cdot bg_{\text{obs}} + p_{\text{inc}}(t) \cdot p_{\text{obs}} \cdot N$$
$$p_{\text{inc}}(t) = \frac{I_2(t) + R(t) - I_2(t-1) - R(t-1)}{N}$$

We used a bootstrap particle filter, as previously described in the context of our Australian seasonal influenza forecasts [12, 13, 14, 15, 16], to generate forecasts at each day.

Parameters and model prior distributions

Model and inference parameters are described in Table S4. Note that the transmission model assumes that the population mixes homogeneously. Since Australia is one of the most urbanised countries in the world, for each jurisdiction we used capital city residential populations (including the entire metropolitan region, as listed in Table S5) in lieu of the residential population of each jurisdiction as a whole.

		Description	Value
(i)	N	The population size	Table S5
	$R_{\mathrm{eff}}(t)$	The time-varying effective reproduction number	See text
	σ	The inverse of the latent period (days^{-1})	See text
	γ	The inverse of the infectious period (days^{-1})	See text
	τ	The time of the initial exposures (days)	$\sim \mathcal{U}(0, 50)$
(ii)	bg_{obs}	The background observation rate	0.05
	p_{obs}	The observation probability	0.8
	k	The dispersion parameter	10
(iii)	N_{px}	The number of particles	2000
	$N_{\min}$	The minimum number of effective particles	$0.25 \cdot N_{px}$

Table S4: Parameter values for (i) the transmission model; (ii) the observation model; and (iii) the bootstrap particle filter.

Jurisdiction	N
Australian Capital Territory	410,199
New South Wales	5,730,000
Queensland	2,560,000
South Australia	1,408,000
Northern Territory	154,280
Tasmania	240,342
Victoria	5,191,000
Western Australia	2,385,000

Table S5: The population sizes used for each forecast.

The prior distributions for $R_{\mathrm{eff}}(t)$, σ, and γ were constructed in a separate analysis, not described here. Parameters σ and γ were sampled from a multivariate log-normal distribution that was defined to be consistent with a generation interval with mean=4.7 and SD=2.9, and sampled independent $R_{\mathrm{eff}}(t)$ trajectories for each particle.

2. Probabilistic Forecast

We provide a generative model of the dynamics of SARS-CoV-2 in Australia. This allows us to forecast COVID-19 cases by state/territory, and nationally. The model links distancing measures — captured via Google Mobility Indices, and a "Micro-distancing" parameter estimated via weekly national survey data — to the effective reproduction number of local infectious individuals, allowing us to produce forecasts under scenarios of change in Government-imposed distancing measures.

Inferring Effective Reproduction Numbers and Social Isolation Measures

We use the framework developed by Cori et al. (2013) [2] and Thompson et al. (2019) [3] to estimate the effective reproduction number, R_{eff}. The number of local and imported cases observed each day — available from the National Notifiable Disease Surveillance System (NNDSS) — is used to infer the distribution of the overall R_{eff} for an assumed generation interval and

incubation period distribution (see Time distributions). The model described below relates population mobility measures to this R_{eff} estimate and is then used to project the reproduction number forward in order to generate forecasts of cases.

These estimates of the effective reproduction number are assumed to be dependent upon the proportion of observed imported cases out of all observed cases (ρ), the impact of "macro" social isolation measures — captured via Google Mobility Indices, ($\boldsymbol{\omega}(t)$) — and "micro-distancing" (M_d) — captured via a weekly national survey ($\gamma(t)$).

We link these previous estimates of the effective reproduction number with the distancing measures via the model:

$$\hat{\mu}(t) = \rho(t)R_I + (1 - \rho(t))\,R_L(t), \tag{47}$$

$$R_L(t) = R_{Li}M_d(t) \times 2 \times \text{logistic}\left(\boldsymbol{\beta^T}\boldsymbol{\omega}(t)\right) \tag{48}$$

in which:

- $\rho(t)$ (inferred; jurisdiction level) is the proportion of imported cases (of all cases) on day t;

- R_I (inferred; national level) is the effective reproduction number of imported cases;

- R_{Li} (inferred; jurisdiction level) is the effective reproduction number of local symptomatic cases at base levels of mobility for jurisdiction i;

- $M_d(t)$ (inferred; jurisdiction level) is a micro-distancing factor on day t, which allows for the effective reproduction number to vary in accordance with compliance to micro-distancing policies assessed from the weekly national behavioural surveys conducted by the Department of Health, and modelled through the following relationship:

$$M_d = (1 + \theta)^{-\gamma(t)}, \quad \theta, \gamma(t) > 0, \tag{49}$$

  where γ is an estimate of the proportion of individuals always complying to the recommended health policy of maintaining 1.5 metres of distance from non-household members (θ is inferred; national);

- $\boldsymbol{\beta}$ (inferred; national level) is a vector of parameters that link Google Mobility Indices to the effective reproduction number of local cases via the logistic function, which ranges between 0 and 1; and,

- $\boldsymbol{\omega}(t)$ (state level) are the 7-day future moving average of Google Mobility Indices on day t.

We assume that (the likelihood)

$$\hat{\mu}(t) \sim \text{Gamma}(k(t), \theta(t)),$$

with $k(t) = \frac{R_{\text{eff}}(t)^2}{\sigma^2(t)}$ and $\theta(t) = \frac{\sigma^2(t)}{R_{\text{eff}}(t)}$, to match the distribution of R_{eff} estimates from earlier work [17].

We perform inference in a Bayesian framework, using Hamiltonian Monte Carlo through the software package `pystan` [18]. We use a hierarchical model to infer the R_{eff} of local cases at base levels of mobility and micro-distancing for jurisdiction i, R_{Li} as follows:

$$R_{Li} \sim \text{Gamma}(R_{L0}^2/\sigma^2, \sigma^2/R_{L0}),$$
$$R_{L0} \sim \text{Gamma}(1.8^2/0.01, 0.01/1.8),$$
$$\sigma^2 \sim \text{Exponential}(20).$$

Other prior distributions are:

$$\rho(t) \sim \text{Beta}\left(1 + I(t), 1 + L(t)\right);$$
$$\boldsymbol{\beta} \sim \text{Normal}(0, 1);$$
$$R_I \sim \text{Gamma}(1.25, 2.5);$$
$$\theta \sim \text{Log Normal}(0, 0.5),$$

where $I(t)$ and $L(t)$ are the number of imported and local cases on day t respectively.

We use case data for every Australian state through the month of March, 2020, and additionally in Victoria from 01/06/2020 to 28/10/2020 and in New South Wales from 01/06/2020 to 19/01/2021 (approximately corresponding to the second wave), and generate the posterior predictive distribution of the reproduction number in each of these jurisdictions over time.

Forecasts of mobility indices
Our forecasts are produced by first forecasting the Google Mobility Indices (using a random walk with drift).

For all Google mobility indices in each State, the differences in each successive day were assumed normally distributed and the mean and covariance matrix of the distribution were estimated through a maximum likelihood estimator. This was applied to the previous 28 days from the last Google mobility index entry. The indices were then forecast for the next six weeks plus any additional days yet to be provided by Google by successively adding a sample from the estimated multi-variate normal distribution each day. Each index is capped at a maximum of 0% or the historical maximum of baseline, whichever is higher, and -50% of baseline or the historical minimum, whichever is lower, to maintain reasonable estimates of the trend.

An independent normal distribution is used in a similar manner to forecast micro-distancing behaviour.

These forecasts for each state/territory are then used to create a posterior prediction of the local effective reproduction number.

Generative model
We simulate the number of cases using a branching process based on the estimated reproduction number described above. The generative model contains three types of infectious individuals: Imported (I_I); Asymptomatic (I_A), and; Symptomatic (I_S).

Secondary cases
Each case is assumed to generate a number of cases drawn from a Negative Binomial distribution, with parameters k and, $R_I/(R_I + k)$, $\alpha_A R_L/(\alpha_A R_L + k)$, $\alpha_S R_L/(\alpha_S R_L + k)$ for imported, asymptomatic and symptomatic individuals, respectively.

The parameters R_I and R_L (the effective reproduction numbers for import-to-local and local-to-local cases, respectively) are sampled from their posterior distributions.

The posterior predictive samples of the local reproduction number, which varies each day due to changes in social mobility and adherence to micro-distancing, are each generated by a single sample of parameters from the posterior distribution (described in the social mobility model

Table S6: Detection probabilities of Symptomatic, Asymptomatic and Imported cases for each jurisdiction.

Jurisdiction	q_S	q_A	q_I
NSW	0.90	0.05	0.95
QLD	0.90	0.05	0.95
SA	0.70	0.05	0.95
TAS	0.40	0.05	0.95
VIC	0.35	0.05	0.95
WA	0.70	0.05	0.95
ACT	0.95	0.70	0.95
NT	0.95	0.70	0.95

calibration). This generates a single posterior predictive set of local reproduction numbers over time, varying by the changes in mobility and micro-distancing for the jurisdiction. Hence, every infected case on the same infection day will have the same reproduction number.

The parameter k is fixed at 0.1 in our analysis, according to existing estimates [19]. This value allows for heterogeneity in the transmissibility of cases — so-called *super spreading* — in that the mean is realised with high variance. The parameter α_S corresponds to the contribution of transmissibility of symptomatic local cases and the parameter α_A corresponds to the contribution of transmissibility of asymptomatic local cases.

The R_{eff} estimate generated via [4] using the NNDSS case data does not readily distinguish between symptomatic and asymptomatic cases, and cases observed in this initial outbreak are all assumed to be symptomatic. The effective reproduction number is the average number of secondary infections caused by an infected individual, and can be characterised as

$$R_{\text{eff}} = \frac{s_{t+1}}{s_t}, \tag{50}$$

where s_t is the number of detected symptomatic cases in generation t.

In order to correctly attribute the contributions of symptomatic and asymptomatic cases to secondary cases, we require

$$s_{t+1} = (S_t \alpha_S R_L + A_t \alpha_A R_L) p_S q_S \tag{51}$$

where S_t is the true number of local symptomatic cases (*i.e.*, consisting of both observed and unobserved cases), A_t is the true number of local asymptomatic cases, p_S is the probability of being symptomatic and q_S is the probability of detecting a local symptomatic case.

Using Equations 50 and 51, and for local cases where $R_{\text{eff}} = R_L$, we have

$$\alpha_S p_S + \alpha_A (1 - p_S) = 1 \tag{52}$$

In this forecast we assume that p_S is 0.2, and the relative infectiousness of asymptomatic cases is 50% of symptomatic cases. It follows from Equation 52 that $\alpha_S = 1.67$ and $\alpha_A = 0.833$.

The generative model must also consider probabilities of observing infectious cases. Infectious individuals are detected, and hence become a case, with probabilities q_I, q_A and q_S respectively. Table S6 contains the values used in this forecast.

Time distributions
We use an incubation period – the delay between infection and symptom onset – that is Gamma distributed with shape parameter 5.807 and scale parameter 0.948 (Lauer et al. (2020) [20]).

Based on this incubation period, we follow (Abbott et al. (2020) [21]) and use the generation interval defined by a Gamma distribution with mean 3.635 days and standard deviation 3.075 days, as described in Ganyani et al. (2020) [22]. We estimate a reporting delay distribution from the case data by finding the maximum likelihood estimates of the parameters of a Gamma distribution of the time between symptom onset, where known, to the notification date; this gave a shape parameter of 1.82, and a scale parameter of 2.88.

Imports

We additionally assume a Poisson number of new imported infectious individuals on day t, where mean parameter λ_t is inferred from data. We use six time periods ($i = 1, \ldots, 6$) corresponding to:

- 01/03/2020 to 06/03/2020;

- 07/03/2020 to 13/03/2020;

- 14/03/2020 to 18/03/2020;

- 19/03/2020 to 23/03/2020;

- 24/03/2020 to 14/04/2020;

- 15/04/2020 to 15/01/2021; and,

- 16/01/2021 onwards.

Cases were classified as imported or locally-acquired according to their reported place of acquisition. If place of acquisition is unknown or missing, the cases were assumed to be locally acquired. In the event that symptom onset date is missing, the date is inferred as follows: 'notification date - 5' if notification date is recorded, otherwise, 'notification receive date - 6'.

To assign the imported cases to the period in which they are likely to have arrived (as infectious), we have subtracted 4 days from their symptom onset date. Within each state/territory ($j = 1, \ldots, 8$) and in each period, i, we assume that a Poisson number of imports are subsequently detected, $N_{i,j}$. That is, $N_{i,j} \sim \text{Poisson}(\lambda_{i,j})$, independently on each day.

Assuming *a priori* $\lambda_{i,j} \sim \text{Gamma}(\alpha, \beta)$, we have *a posteriori* that $\lambda_{i,j} \sim \text{Gamma}(a_{i,j}, b_i)$ where

$$
a_{i,j} = \begin{cases} \alpha + n_{i,j} & \text{if } i \neq 4, \\ \alpha + 1.3 n_{i,j}, & \text{if } i = 4 \end{cases}
$$
$$
b_i = \beta + m_i,
$$

in which $n_{i,j}$ is the total number of detected imported cases in period i in state/territory j, and m_i is the number of days in the period i ($m = (m_1, m_2, m_3, m_4, m_5, m_6) = (6, 8, 4, 5, 22, 276)$; the final period m_7 includes the remaining days in the forecast). The number of imported infectious individuals in period i in state/territory j, $D_{i,j}$, that are to be subsequently detected are simulated each day from its *posterior predictive distribution*,

$$
D_{i,j} \sim \text{NegBin}(a_{i,j}, 1/(b_i + 1)).
$$

We specified $\alpha = 1$ and $\beta = 1/5$ for the prior distribution, though this choice has little impact on the posterior distribution. Having generated $d_{i,j}$ for each day in period i in state/territory

j, we sum the total number of such cases, $s_{i,j}$. Subsequently, we simulate the total number of undetected imported cases,

$$U_{i,j} \sim \text{NegBin}(s_{i,j}, q_I),$$

where q_I is the detection probability of imported cases. The undetected imported cases are allocated to the days in the period i with equal probability.

This process is simulated first, before seeding with any local cases and simulating forward.

Model Initialisation

We use the number observed cases on a chosen start date (well before the forecast period) to initialise the model. Specifically:

- Given n_S symptomatic local cases, and n_I imported cases on 01/09/2020, we generate:

  - Undetected symptomatic individuals, $U_S \sim \text{NegBin}(n_S, q_S)$,
  - Undetected imports $U_I \sim \text{NegBin}(n_I, q_I)$; and
  - Asymptomatic individuals, $I_A \sim \text{NegBin}(I_S, p_S)$.

- Assign an infection time to the U_S, U_I and I_A individuals from the Generation Time distribution.

- For any infection time which is after the period being considered, sample those detected with probabilities q_S, q_I and q_A, respectively. For detected and symptomatic cases, sample the time to symptom onset from the Incubation Period distribution.

Model Re-initialisation

We expect events that are difficult to forecast precisely to occur, such as large cluster outbreaks. When such outbreaks occur, we add to our model state additional cases determined by performing the initialisation step on the day the threshold is exceeded. These events are detected via a threshold on the cumulative cases (over a moving 3-day period) – i.e., when the moving average exceeds the forecast case counts by a factor of 10. The additional n cases are distributed across the 3-day period by adding $n/6$, $n/3$ and $n/2$ to each corresponding day respectively. Simulations are only permitted to re-initialise a maximum of 10 times. If this is exceeded, the simulation is excluded from the forecast.

The model allows for the re-introduction of cases when there has been an extended period of zero reported local cases. If a case is reported in the NNDSS data and the simulation does not have any infections occurring in the previous seven days, then the missing cases are added to the simulation state as described above.

Conditioning on data

Where the number of cases in a simulation period exceeds a threshold for the jurisdiction, the simulation is excluded from the forecast. There are three thresholds used to exclude a forecast: in the 14 days preceding the forecast date, the threshold is the maximum of 10 or 1.5 times the number of local cases during that period; in the 46 days before the start of the previous period (60 days prior to the forecast date), the threshold is 100 or 4 times the number local cases during the period; and, after the date of the forecast, the threshold is 10,000 cases or the total number of local and imported cases in the data.

When a simulation is rejected for not matching the data of observed cases, this will then reject the corresponding sample from the posterior distribution, thereby only including in the forecast samples from the posterior distribution that were able to provide a reasonable explanation of the observed data.

Code Availability

The code used to generate the simulated cases can be accessed at the repository `https://github.com/tdennisliu/covid19-forecasting-aus`.

3. Time Series Forecast

We estimated a simple autoregressive model using available case data from many countries, obtained from the Johns Hopkins COVID19 repository. For each state, the model uses data from the previous 11 days to estimate the possible trajectory of cases over the next few weeks.

Model Description

We fit a global autoregressive model to all available data from the Johns Hopkins COVID19 repository.

Some data quality issues were detected and the resulting data cleaned or omitted. In particular:

- data by region in the UK, Spain and Italy were added.

- series with fewer than 500 cumulative cases were removed;

- series with fewer than 15 days of data were removed;

- series with anomalous data were removed (*e.g.*, negative case numbers, or very large step changes).

Let $n_{t,i}$ = the number of daily cases on day t in country (or region) i, and let $y_{t,i} = \log(n_{t,i} + 0.5)$. We fit an autoregressive model of order p:

$$y_{t,i} = \phi_1 y_{t-1,i} + \cdots + \phi_p y_{t-p,i} + \varepsilon_{t,i},$$

where $\varepsilon_{t,i}$ are independent $N(0, \sigma_i^2)$ errors. The model is estimated using ordinary least squares estimation, with no stationarity constraints. The parameters are scale free other than the error variance σ_i^2. Consequently the model is estimated by first scaling all data to have the same mean and variance, to avoid any one country dominating in the estimation. Then the model is applied to the raw data from each country or region when forecasting.

The value of $p = 24$ is chosen to minimize the average 7-day-ahead mean absolute error on recent Australian data (so 24 model parameters, estimated from approximately 10^4 daily observations). We can afford to have a large value of p due to the large data set used to estimate the model.

It is not intended to be a model of the disease development, and contains no terms that describe public health measures or related policies. However, the model is highly adaptive to different stages of the pandemic including rapid increases, periods of containment, and periods where there are few cases. Time series models of this kind tend to produce relatively accurate short-term forecasts, but are probably less accurate than epidemiological models in the long-term.

The model code is available at `https://github.com/pmontman/covid19forec`.

References

[1] Laura F. White and Marcello Pagano. A likelihood-based method for real-time estimation of the serial interval and reproductive number of an epidemic. *Stat Med*, 27(16):2999–3016, 2008.

[2] Anne Cori, Neil M. Ferguson, Christophe Fraser, and Simon Cauchemez. A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am J Epidemiol*, 178(9):1505–1512, 2013.

[3] Robin Thompson, Jake Stockwin, Rolina D. van Gaalen, Jonathan Polonsky, Zhian Kamvar, Alex Demarsh, Elisabeth Dahlqwist, Siyang Li, Eve Miguel, Thibaut Jombart, Justin Lessler, Simone Cauchemez, and Anne Cori. Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics*, 29:100356, 2019.

[4] Sam Abbott, Joel Hellewell, James Munday, Robin Thompson, and Sebastian Funk. *EpiNow: Estimate realtime case counts and time-varying epidemiological parameters*, 2020. R package version 0.1.0.

[5] David A. Rolls, Nicholas L. Geard, Deborah J. Warr, Paula M. Nathan, Garry L. Robins, Philippa E. Pattison, James M. McCaw, and Jodie McVernon. Social encounter profiles of greater Melbourne residents, by location – a telephone survey. *BMC Infectious Diseases*, 15(1):494, December 2015.

[6] Kiesha Prem, Alex R. Cook, and Mark Jit. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLOS Computational Biology*, 13(9):1–21, 09 2017.

[7] Department of the Prime Minister and Cabinet. BETA COVID-19 Barometer Survey, 2020.

[8] Hiroshi Nishiura, Natalie M Linton, and Andrei R Akhmetzhanov. Serial interval of novel coronavirus (COVID-19) infections. *Int J Infect Dis*, 93:284–6, 2020.

[9] Seth Flaxman, Swapnil Mishra, Axel Gandy, Juliette T Unwin, Helen Coupland, Thomas A Mellan, Harrison Zhu, Tresnia Berah, Jeffrey W Eaton, Pablo NP Guzman, Nora Schmit, Lucia Cilloni, Kylie EC Ainslie, Marc Baguelin, Isobel Blake, Adhiratha Boonyasiri, Olivia Boyd, Lorenzo Cattarino, Constanze Ciavarella, Laura Cooper, Zulma Cucunubá, Gina Cuomo-Dannenburg, Amy Dighe, Bimandra Djaafara, Ilaria Dorigatti, Sabine van Elsland, Rich FitzJohn, Han Fu, Katy Gaythorpe, Lily Geidelberg, Nicholas Grassly, Will Green, Timothy Hallett, Arran Hamlet, Wes Hinsley, Ben Jeffrey, David Jorgensen, Edward Knock, Daniel Laydon, Gemma Nedjati-Gilani, Pierre Nouvellet, Kris Parag, Igor Siveroni, Hayley Thompson, Robert Verity, Erik Volz, Caroline Walters, Haowei Wang, Yuanrong Wang, Oliver Watson, Xiaoyue Xi Peter Winskill, Charles Whittaker, Patrick GT Walker, Azra Ghani, Christl A Donnelly, Steven Riley, Lucy C Okell, Michaela AC Vollmer, Neil M. Ferguson, and Samir Bhatt. Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries. 2020.

[10] Nick Golding. greta: simple and scalable statistical modelling in R. *Journal of Open Source Software*, 4(40):1601, 2019.

[11] Nick Golding. *greta.gp: Gaussian Process Modelling in greta*, 2020. R package version 0.1.5.9001.

[12] Robert Moss, Alex Zarebski, Peter Dawson, and James M. McCaw. Forecasting influenza outbreak dynamics in Melbourne from Internet search query surveillance data. *Influenza and Other Respiratory Viruses*, 10(4):314–323, July 2016.

[13] Robert Moss, Alex Zarebski, Peter Dawson, and James M. McCaw. Retrospective forecasting of the 2010–14 Melbourne influenza seasons using multiple surveillance systems. *Epidemiology and Infection*, 145(1):156–169, January 2017.

[14] Robert Moss, James E. Fielding, Lucinda J. Franklin, Nicola Stephens, Jodie McVernon, Peter Dawson, and James M. McCaw. Epidemic forecasts as a tool for public health: interpretation and (re)calibration. *Australian and New Zealand Journal of Public Health*, 42(1):69–76, February 2018.

[15] Robert Moss, Alexander E. Zarebski, Sandra J. Carlson, and James M. McCaw. Accounting for healthcare-seeking behaviours and testing practices in real-time influenza forecasts. *Tropical Medicine and Infectious Disease*, 4:12, January 2019.

[16] Robert Moss, Alexander E. Zarebski, Peter Dawson, Lucinda J. Franklin, Frances A. Birrell, and James M. McCaw. Anatomy of a seasonal influenza epidemic forecast. *Communicable Diseases Intelligence*, 43:1–14, March 2019.

[17] David J. Price, Freya M. Shearer, Michael T. Meehan, Emma McBryde, Robert Moss, Nick Golding, Eamon J. Conway, Peter Dawson, Deborah Cromer, James Wood, Sam Abbott, Jodie McVernon, and James M. McCaw. Early analysis of the Australian COVID-19 epidemic. *eLife*, 9:e58785, 2020.

[18] Stan Development Team. Stan Modeling Language: User's Guide and Reference Manual. *Version 2.17.0*, 2017.

[19] Akira Endo, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Sam Abbott, Adam J. Kucharski, and Sebastian Funk. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. *Wellcome Open Research*, 5:67, 2020.

[20] Stephen A. Lauer, Kyra H. Grantz, Qifang Bi, Forrest K. Jones, Qulu Zheng, Hannah R. Meredith, Andrew S. Azman, Nicholas G. Reich, and Justin Lessler. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Annals of Internal Medicine*, 172(9):577–582, 2020. PMID: 32150748.

[21] Sam Abbott, Joel Hellewell, Robin N. Thompson, Katharine Sherratt, Hamish P. Gibbs, Nikos I. Bosse, James D. Munday, Sophie Meakin, Emma L. Doughty, June Young Chun, Yung-Wai Desmond Chan, Flavio Finger, Paul Campbell, Akira Endo, Carl A.B. Pearson, Amy Gimma, Tim Russell, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Stefan Flasche, Adam J. Kucharski, Rosaling M. Eggo, and Sebastian Funk. Estimating the time-varying reproduction number of SARS-CoV-2 using national and subnational case counts [version 2; peer review: 1 approved with reservations]. *Wellcome Open Research*, 5(112), 2020.

[22] Tapiwa Ganyani, Cécile Kremer, Dongxuan Chen, Andrea Torneri, Christel Faes, Jacco Wallinga, and Niel Hens. Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020. *Eurosurveillance*, 25(17), 2020.