Estimating temporal variation in transmission of COVID-19 and adherence to social distancing measures in Australia

Technical Report 15 May 2020

Nick Golding[1], Freya M. Shearer[2], Robert Moss[2], Peter Dawson[3], Lisa Gibbs[2], Eva Alisic[2], Jodie McVernon[2,4,5], David J. Price[2,4], and James M. McCaw[2,4,6]

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. Peter Doherty Institute for Infection and Immunity, The Royal Melbourne Hospital and The University of Melbourne, Australia
5. Murdoch Children's Research Institute, The Royal Children's Hospital, Australia
6. School of Mathematics and Statistics, The University of Melbourne, Australia

Key messages

Assessment of adherence to social distancing measures

- An analysis of trends in population mobility data streams up to 11 May was performed to assess adherence to social distancing policy.
- This analysis suggests that adherence to social distancing measures may have decreased in the past four weeks (Figure 2).
- Two waves of a national survey were conducted to assess how Australians are thinking, feeling and behaving in response to social distancing measures (Figures 3 and 4).
- We used a statistical model to analyse the survey data and estimate a 37% increase (from 2.78 to 3.80) in reported daily non-household contacts between 3 April and 6 May.

Estimates of current epidemic activity

- Due to very low case incidence, estimates of the effective reproduction number (R_{eff}) using existing methodologies are becoming increasingly unstable, *i.e.*, based primarily on model assumptions rather than actual case data.
- To overcome this limitation, we report preliminary estimates of local transmission potential (Figure 7) from a new method which estimates components of the effective reproduction number. This method uses both daily incident case counts and outputs from an analysis of population mobility (Figure 1).

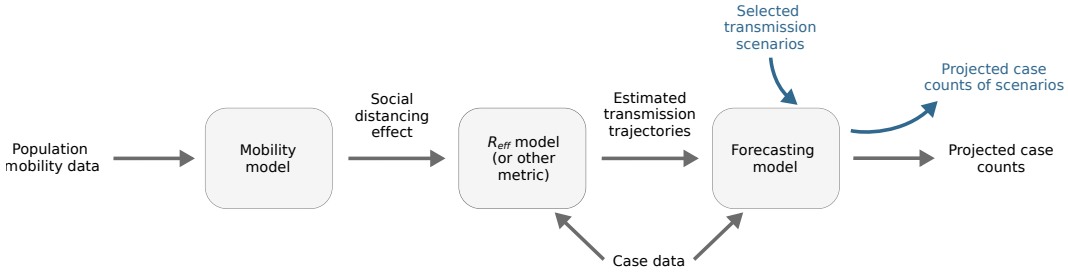
Forecasts of the daily number of new confirmed cases

- Estimates of local transmission potential were input into a mathematical model of disease dynamics which was projected forward to forecast the daily number of new confirmed cases.
- We report Australia-wide (Figure 8) and state-level (Figure 10) forecasts of the daily number of new confirmed cases up to 1 July, assuming that local transmission potential remains at its current estimated level.

Forecasts of the daily number of new confirmed cases under alternative future transmission scenarios

- A scenario analysis was performed to assess the potential impact of increased transmission following the relaxation of social distancing measures from 11 May.
- We project the daily number of new confirmed cases in Australia up to 1 July for three future scenarios: one where local transmission potential increases from 11 May to 1.1, one where it increases to 1.2, and another where it increases to 1.5 (Figures 9, S8, S9, and S10).

Figure 1: Epidemic forecasting workflow.



Assessment of adherence to social distancing measures through the analysis of trends in population mobility data streams

Summary

A number of data streams provide information on mobility before and in response to COVID-19 across Australian states/territories. Each of these data streams represents a different aspect of population mobility, but they show some common trends — reflecting underlying changes in behaviour. We use a latent variable statistical model to simultaneously analyse these data streams and quantify these underlying behavioural variables.

Data streams

We currently consider 10 different data streams, provided by three different technology companies: Apple and Citymapper provide regularly updated data on direction requests, while Google provide less regularly updated data on different measures of mobility from users' GPS data. Google provide GPS-derived indices of the amount of time spent (a combination of visits and lengths of stay) in locations of one of 6 types ('workplaces', 'residential', 'parks', 'grocery and pharmacy', 'retail and recreation', 'transit stations'). See Appendix for further details. Each data stream is encoded as a percentage change in the mobility metric, relative to a pre-COVID-19 baseline.

Access and privacy

The data streams provided by Apple, Citymapper and Google are all publicly available for the express purpose of supporting public heath bodies in their response to COVID-19. All of these datasets are fully anonymised and aggregated at the level of either states and territories or Australia's four largest cities, over each day. The large-scale aggregation of these datasets ensures the privacy of users — the smallest population is that of the Northern Territory: over 245,000.

Interpretation

The model identifies three COVID-19-related behavioural variables that explain the trends in all of these data streams. **The dominant behaviour in all data streams is a behavioural switch to increased social distancing occurring during the period when social distancing measures were implemented.** The switch to social distancing behaviour was most rapid around the date of the second social distancing measure we consider: closure of restaurants, bars, and cafes on 24 March. Also detected is a period of increased activity in some data streams in advance of this social distancing behaviour, apparently representing preparation for social distancing behaviour. This is most evident in Google's index of time spent at grocery stores and pharmacies (not shown in Figure 2). **The model also detects a decline in the social distancing variable over time *i.e.*, increasing mixing (Figure 2). Specifically, by 11 May, the impact of social distancing on time at parks is expected to have reduced by 30% on average across states (ranging from 7% in TAS to 65% in NT), the effect on requests for driving directions by 33% (22% in VIC to 50% in NT), and the effect on time at transit stations by 14% (7% in TAS to 30% in NT)**.

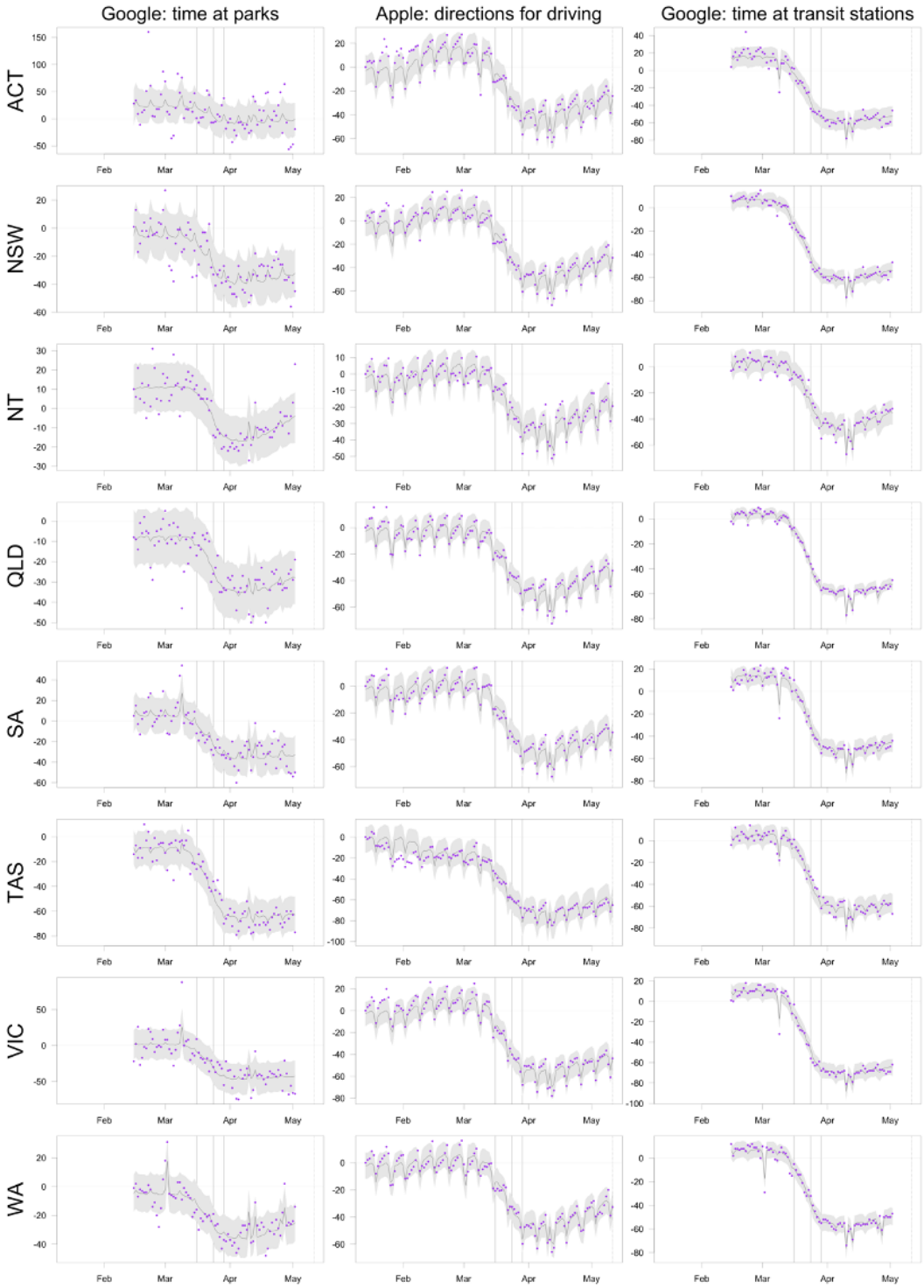
The largest reductions in the impacts of social distancing are evident in mobility data streams for lower transmission risk activities, including activities encouraged by public health authorities *e.g.*, exercising. There is a clear reduction in data streams representing higher-risk activities,

such as time at workplaces. However, these mobility data do not indicate whether the increase in lower transmission risk activities is mitigated by other behaviours that are not measured by these metrics – such as reducing contacts and adherence to the $4m^2$ rule.

Other behavioural variables driving the data streams but not related to COVID-19 are: a gradual increase in work- and school-related travel after the school holidays ended in February (evident in Apple's direction requests, and Google's time at transit stations and workplaces); reduced mobility on weekends (evident in weekly-cycles in most data streams); and reduced mobility on public holidays ('dips' evident in most data streams).

Plots of each data stream and our model fits for each state and territory are shown in the Appendix (Figures S12–S18), with annotation matching that in Figure 2.

Figure 2: Percentage change compared to a pre-COVID-19 baseline of three key mobility data streams in each Australian state and territory up to 11 May. Solid vertical lines give the dates of three social distancing measures: restriction of gatherings to 500 people or fewer; closure of bars, restaurants, and cafes; restriction of gatherings to 2 people or fewer. The dashed vertical line marks 11 May, the most recent date for which some mobility data are available. Blue 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 of the latent behaviours driving each data stream.



National survey to assess the public response to social distancing measures

Two waves of an online survey were conducted from 3 to 6 April and 30 April to 6 May to assess how people in Australia are thinking, feeling, and behaving in relation to the COVID-19 pandemic and the social distancing measures in place at the time. Survey design summary:

- First wave survey from 3–6 April 2020 (n=999)
- Recruitment was targeted to be representative of the adult population in Australia
- Second wave survey of the same individuals (where possible) from 30 April–6 May (n=1000)
- Preliminary results reported here correspond to the paired responses from 732 Australian residents aged 18 years and over who responded to both surveys
- Surveys were timed to occur in response to key changes in epidemic activity and public health policy/messaging

Estimating the change in number of reported non-household contacts

Respondents were asked to report the number of people they had contact with outside of their household in the past 24 hours. Contact was defined as "either a face to face conversation of at least three words or any form of physical contact, such as a handshake."

We used a statistical model to estimate the percentage change in the average daily number of non-household contacts between the first and second wave survey periods. The average daily number of non-household contacts reported in the first survey wave was 2.78 (95% CrI 2.44–3.17). For context, a contact survey conducted in the UK reported 10.8 pre-epidemic daily contacts [1] (note that these estimates also included contacts within the household).

For Australia, we estimate a 37% (95% CrI 16.8%–59.6%) increase in the average daily number of non-household contacts per respondent (from 2.78 to 3.80) in the second survey wave.

The timing of the first wave survey coincides with the approximate peak adherence to social distancing measures from our mobility analysis ($\approx$ 2 April). The estimated increase in numbers of daily non-household contacts between the two survey periods is consistent with estimated waning in social distancing over the same time period.

Assessing changes in risk perception and worry about COVID-19

Preliminary analysis of the survey data suggests that during the period from 30 April to 6 May, individuals were less worried about the COVID-19 outbreak in Australia and believed that they were less likely to become infected at some point in the future, compared to the period from 3 to 6 April. Note that these findings are likely to vary across diverse sub-populations within Australia.

Summary of findings:

- 83% reported being worried about the COVID-19 outbreak in Australia in the first survey period, compared to approximately 70% in the second survey period (Figure 3).
- 40% who had not tested positive for COVID-19 believed it was likely they would be infected at some point in the future in the first survey period, compared with 30% in the second survey period (Figure 4).

Figure 3: Self-reported concern about the COVID-19 outbreak in Australia for paired responses to the first wave (3–6 April) and second wave (30 April–6 May) national surveys.

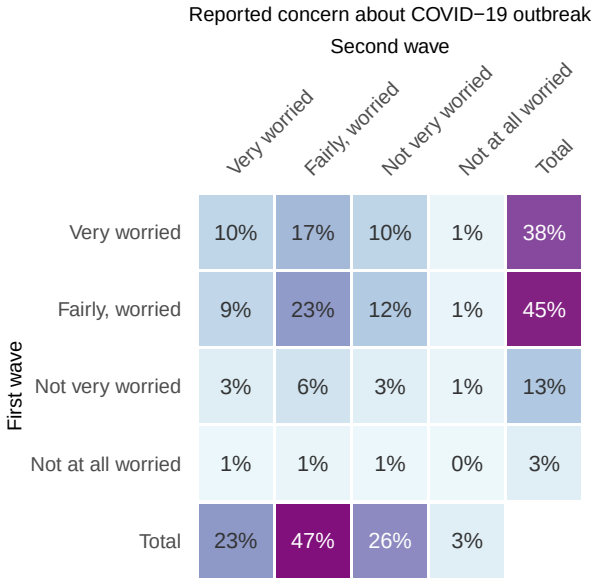
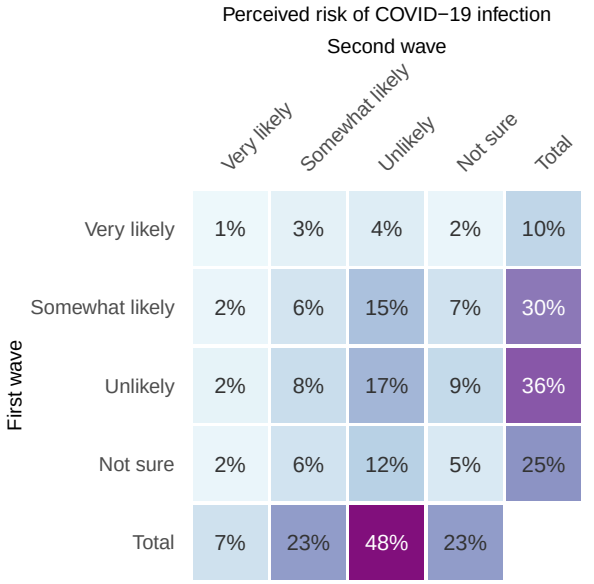


Figure 4: Perceived risk of COVID-19 infection for paired responses to the first wave (3–6 April) and second wave (30 April–6 May) national surveys.



Estimates of current epidemic activity

We are currently developing new methods for assessing epidemic activity, which should be more informative when daily case incidence is very low. We report preliminary results from a new method which estimates components of the effective reproduction number using outputs from the population mobility analysis.

Overview

We developed a new model to estimate components of the effective reproduction number resulting from transmission from locally acquired cases and from overseas acquired cases. This model enables us to 1) estimate the relative temporal variation in transmission from local to local cases and from overseas-acquired to local cases; and 2) quantify the relative impacts of national-level interventions on transmission in Australia. Whilst both locally and overseas acquired cases contribute to Australia's case count, the transmission rates from each of these groups differs as they are each targeted by different interventions. Isolation and quarantine of overseas arrivals modifies the transmission rates of overseas acquired cases only, and social distancing measures modify transmission rates of locally acquired cases. By splitting R_{eff} between these two groups, the model enables us to estimate the relative impacts of various response policies on transmission in Australia, namely isolation and quarantine of overseas arrivals and social distancing of the general population.

We model local to local transmission and import to local transmission for each state/territory using three components:

1. the nationwide average trend in R_{eff} that is driven by interventions;
2. time-varying deviations from this national rate for each state, reflecting state to state differences in transmission; and
3. short-term fluctuations in R_{eff} in each state/territory to capture stochastic dynamics of transmission, such as clusters of cases and short periods of low transmission.

Whilst components 1 and 2 reflect the average local transmission potential at national and state levels, component 3 captures transmission within the sub-populations that have the most active cases at a given point in time. Component 3 is therefore useful for estimating the specific (heightened) transmission among clusters of cases — such as in healthcare workers in Tasmania and in meat processing workers in Victoria — but does not reflect changes in statewide transmission potential (Figure 6). In order to produce state and national level forecasts, we therefore combined components 1 and 2 to estimate the local transmission potential across the whole population of each state, *i.e.*, removing the contribution of short-term fluctuations that represent R_{eff} in clusters, rather than from the general population (Figure 7).

Interpretation

Where there is epidemic activity, the estimates of local transmission potential may be interpreted as the effective reproduction number, R_{eff}. In the absence of epidemic activity, this quantity reflects the ability for the virus, if it were present, to establish and maintain community transmission (> 1) or otherwise (< 1).

Previous methods used to estimate time-varying R_{eff} of SARS-CoV-2 in Australia (*i.e.*, including those presented in Technical Report dated 14 May 2020 [2, 3, 4]) showed an increase

in local transmission potential at a state-level as a result of localised outbreaks (*i.e.*, clusters), such as those seen in Tasmania and Victoria.

Figure 5: Depiction of how R_{eff} analysis components 1 and 2 feed into the forecasting model.

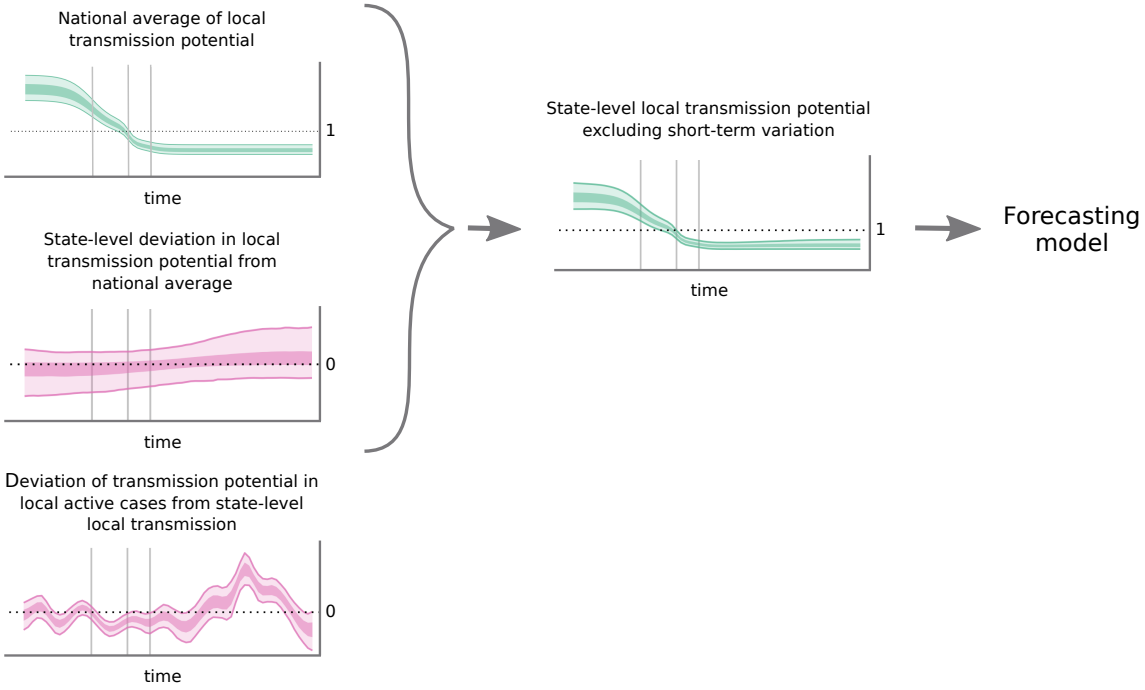


Figure 6: Deviation of transmission potential in local active cases (component 3) from state-level local transmission for each state/territory (light pink ribbon=90% credible interval; dark pink ribbon = 50% credible interval) up to 4 May, based on data up to and including 10 May. Solid grey vertical lines indicate key dates of implementation of various social distancing policies. Black dotted line indicates the target value of 1 for the effective reproduction number required for control.

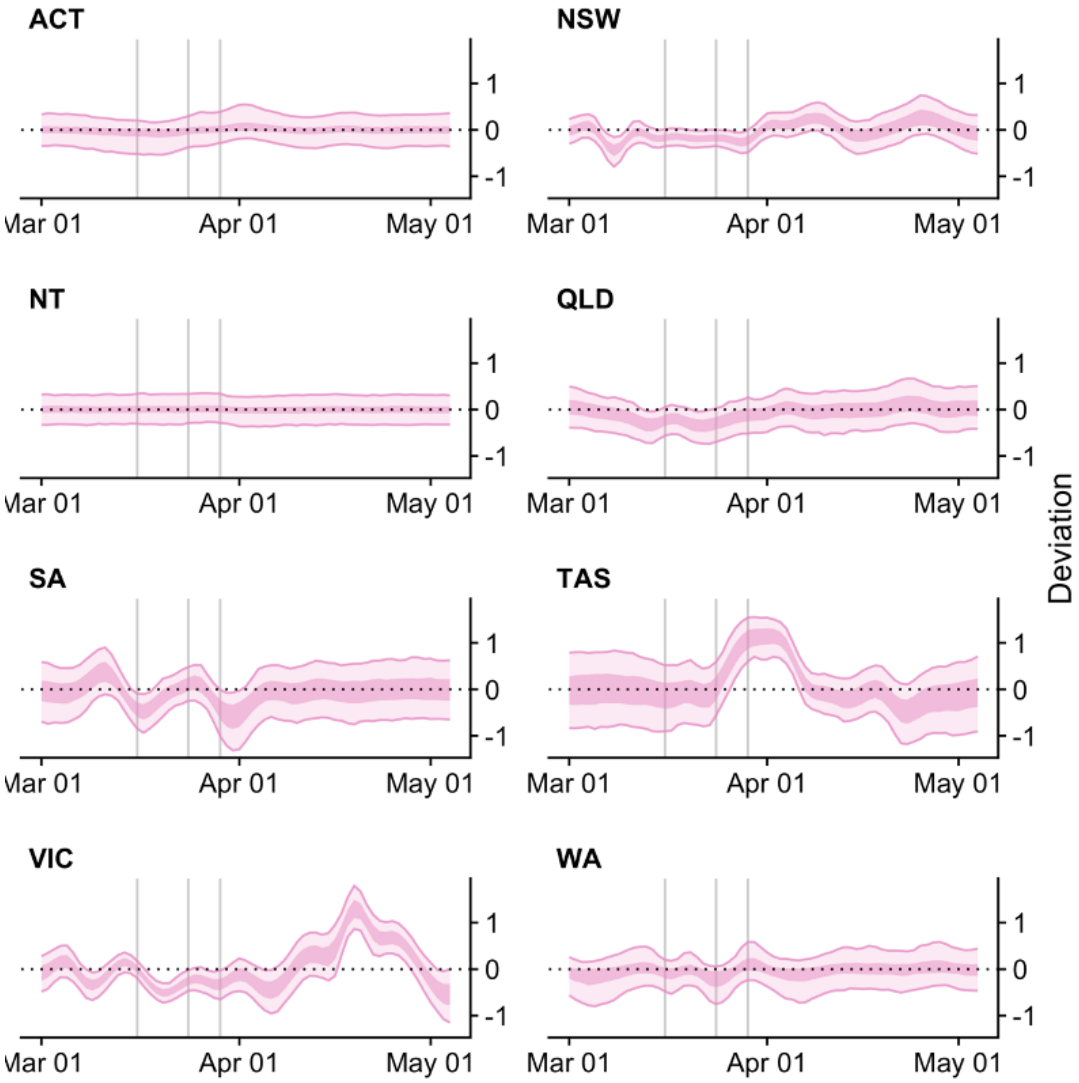
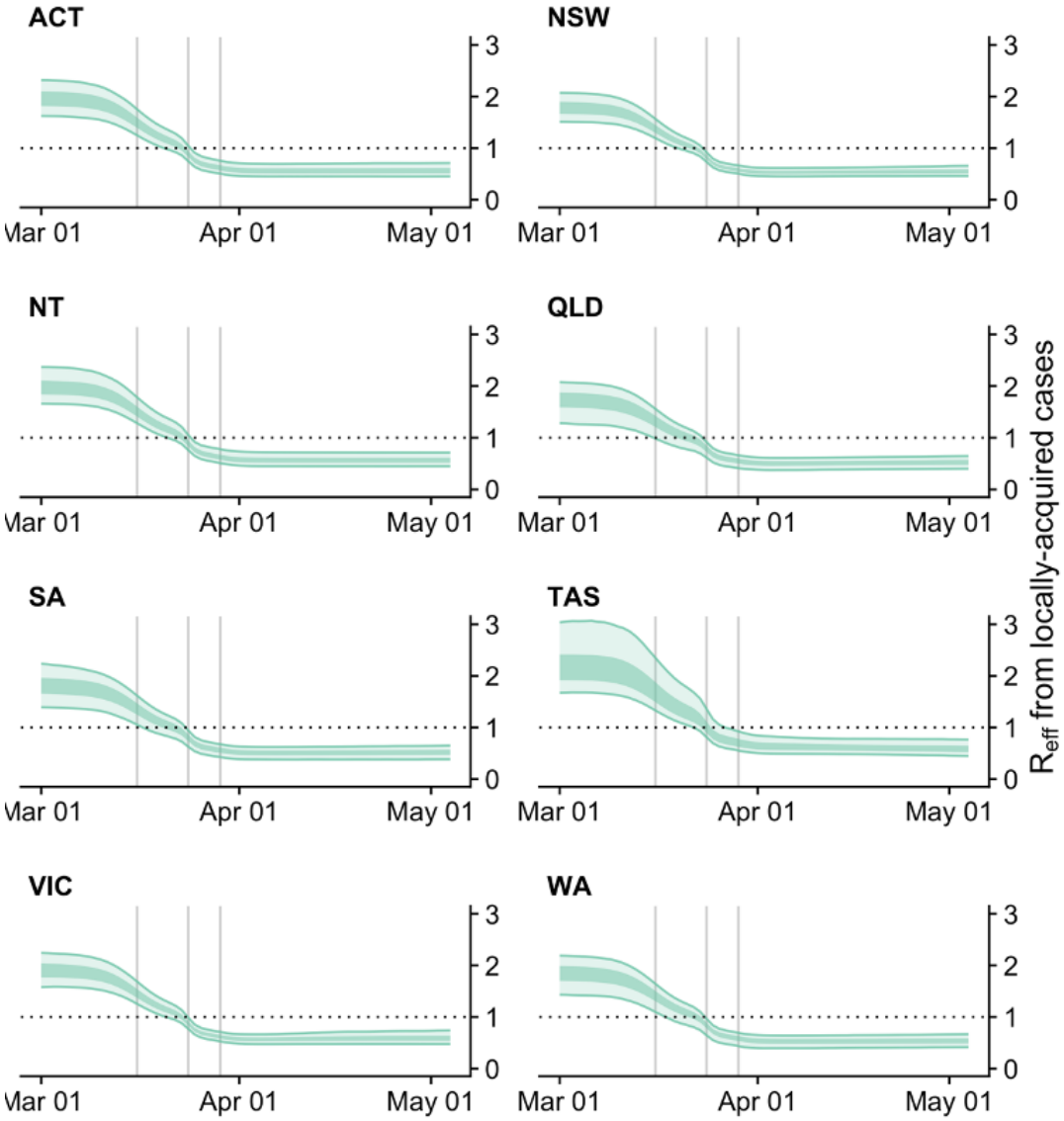


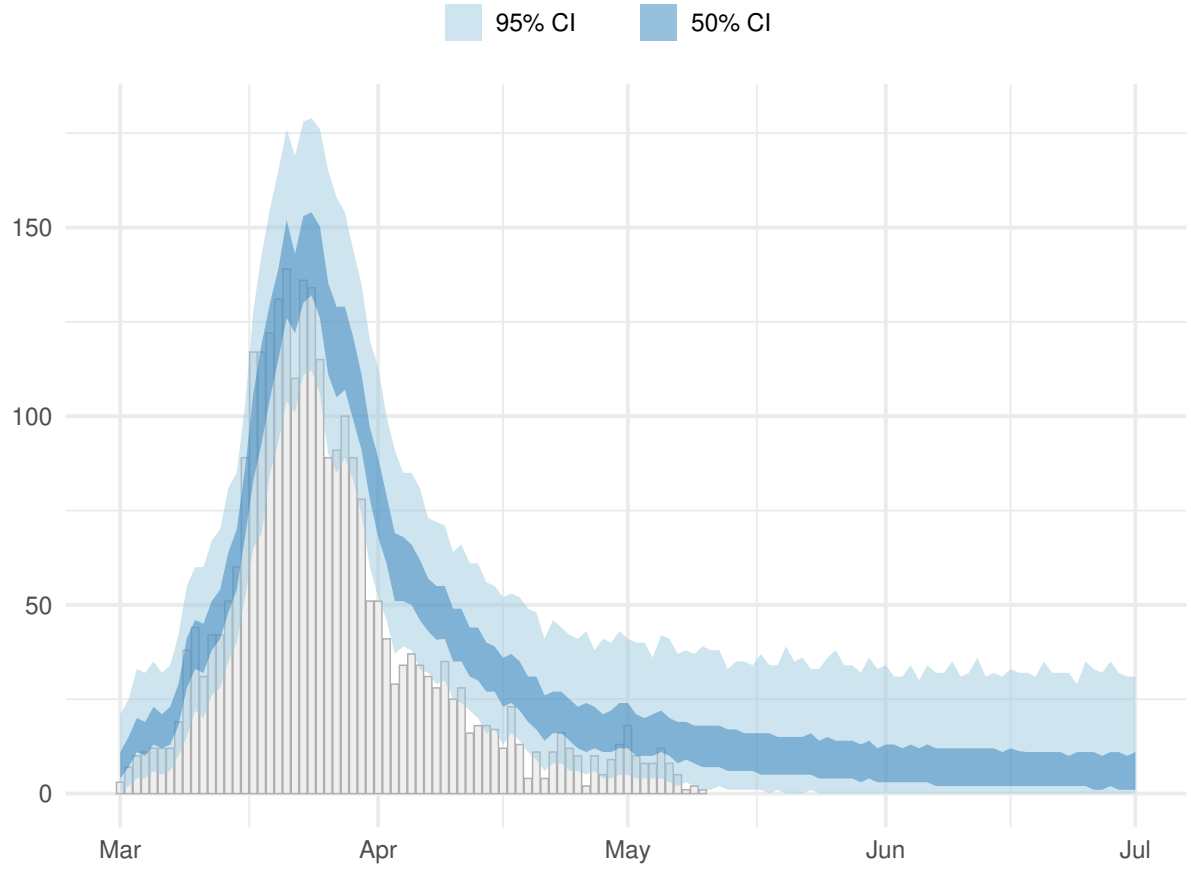
Figure 7: Time varying estimate of local transmission potential excluding the contribution of short-term fluctuations in transmission from clusters (combination of components 1 and 2) for each state/territory (light green ribbon=90% credible interval; dark green ribbon = 50% credible interval) up to 4 May, based on data up to and including 10 May. Solid grey vertical lines indicate key dates of implementation of various social distancing policies. Black dotted line indicates the target value of 1 for the effective reproduction number required for control. Where there is epidemic activity, this quantity may be interpreted as the effective reproduction number, R_{eff}. In the absence of epidemic activity, this quantity reflects the ability for the virus, if it were present, to establish and maintain community transmission (> 1) or otherwise (< 1).



Forecasts of the daily number of new confirmed cases nationally

We used our estimates of local transmission potential and observed cases to generate preliminary forecasts of the daily number of new confirmed cases nationally (Figure 8). Estimates of local transmission potential were input into a mathematical model of disease dynamics. A sequential Monte Carlo method was used to infer the model parameters and appropriately capture the uncertainty, conditional on each of a number of sampled trajectories of the estimated local transmission potential up to 10 May, from which point they were assumed to be constant. The model was subsequently projected forward from up to 1 July, to forecast the number of reported cases, assuming a case detection probability of 80%.

Figure 8: Time series of new daily confirmed cases of COVID-19 estimated from the forecasting model up to 10 May and projected forward up to 1 July (light blue shading = 95% confidence intervals, dark blue shading = 50% confidence intervals), assuming that local transmission potential remains at its current estimated level (see Figure S4). The observed case counts are also plotted (grey bars).



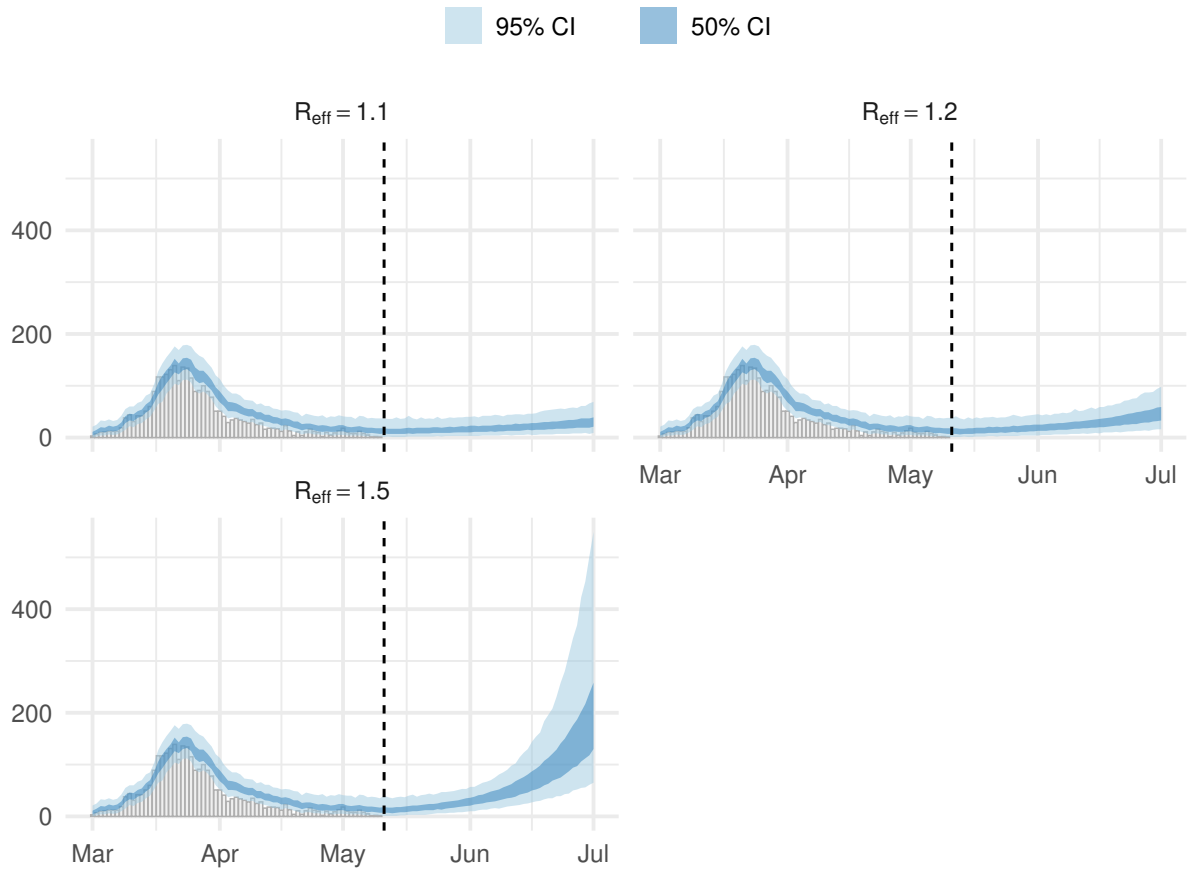
Forecasts of the daily number of new confirmed cases nationally under alternative future transmission scenarios

We performed a scenario analysis to assess the potential impact of increased transmission following the relaxation of social distancing measures from 11 May.

As per the above analysis, we used estimates of local transmission potential up to 4 May, from which point they were assumed to be constant, as inputs into a mathematical model of transmission dynamics. We subsequently projected the transmission potential forward from 5 May to 10 May to forecast the number of reported cases. Then from 11 May, we explore three future scenarios: one where local transmission potential increases from 11 May to 1.1, one where it increases to 1.2, and another where it increases to 1.5 (Figure 8). For each scenario, the model was subsequently projected forward from 11 May to 1 July.

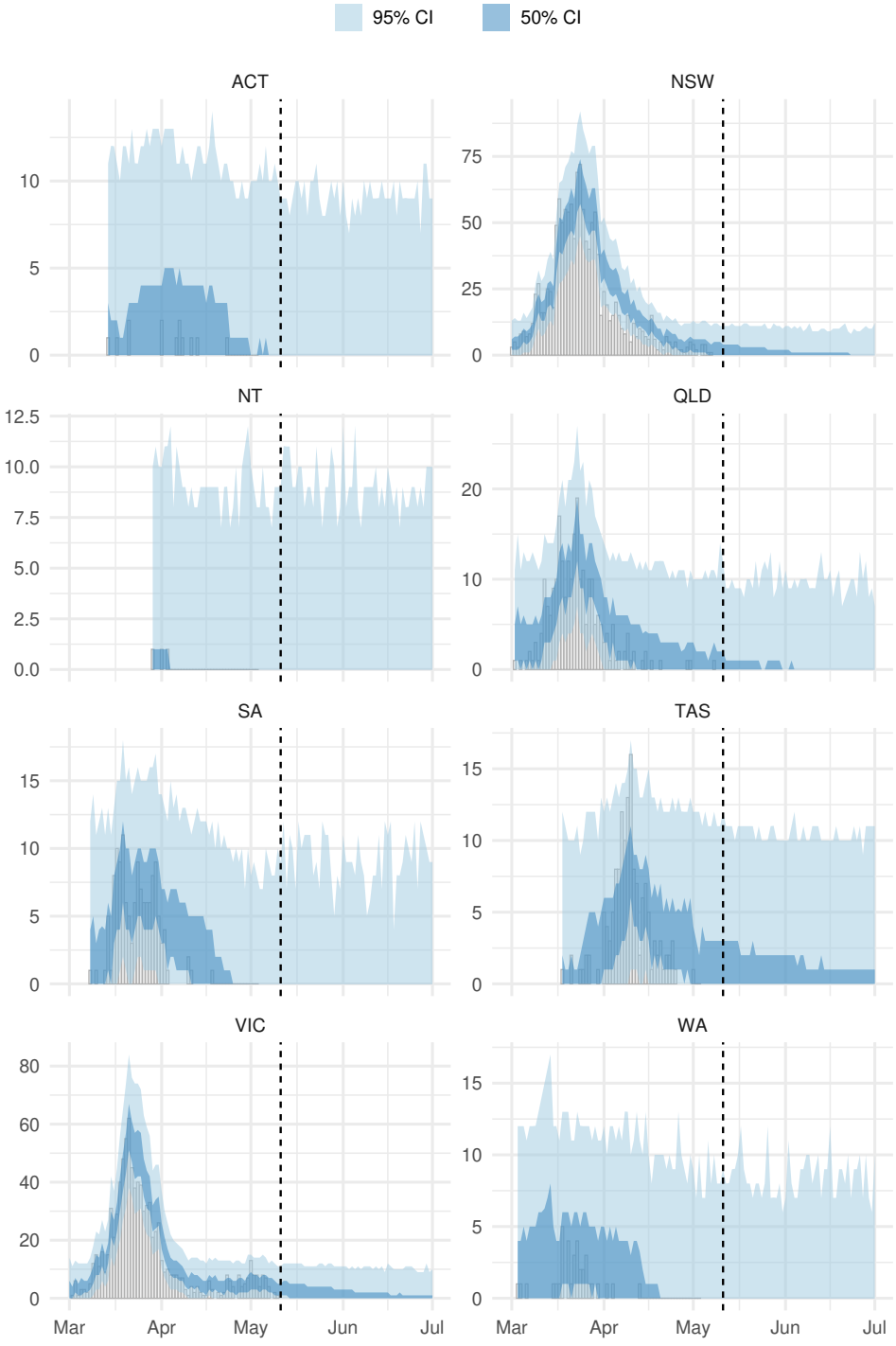
Scenario analyses for each state/territory are provided in the Supplementary Appendix (Figures S6–S8).

Figure 9: Time series of new daily confirmed cases of COVID-19 estimated from the forecasting model up to 10 May and projected forward up to 1 July (light blue shading = 95% confidence intervals, dark blue shading = 50% confidence intervals), assuming that local transmission potential increases to 1.1 (top left), 1.2 (top right), and 1.5 (bottom left), respectively from 11 May (see Figures S5, S6, and S7). The observed case counts are also plotted (grey bars).



Forecasts of the daily number of new confirmed cases for each state/territory

Figure 10: Time series of new daily confirmed cases of COVID-19 estimated from the forecasting model up to 10 May and projected forward up to 1 July (light blue shading = 95% confidence intervals, dark blue shading = 50% confidence intervals), assuming that local transmission potential remains at its current estimated level (see Figure S4). The observed case counts are also plotted (grey bars).



Acknowledgements

This report represents surveillance data reported through the Communicable Diseases Network Australia (CDNA) as part of the nationally coordinated response to COVID-19. We thank public health state from incident emergency operations centres in state and territory health departments, and the Australian Government Department of Health, along with state and territory public health laboratories. We thank members of CDNA for their feedback and perspectives on the results.

Supplementary Appendix

Assessment of adherence to social distancing measures through the analysis of trends in population mobility data streams

Summary

A number of data streams provide information on mobility before and in response to COVID-19 across Australian states/territories. Each of these data streams represents a different aspect of population mobility, but they show some common trends — reflecting underlying changes in behaviour. We use a latent variable statistical model to simultaneously analyse these data streams and quantify these underlying behaviours.

Data streams

We currently consider 10 different data streams, provided by three different technology companies: Apple and Citymapper provide regularly updated data on direction requests, whilst Google provides less regularly updated data on different measures of mobility from users' GPS data. Each data stream is encoded as a percentage change in the mobility metric, relative to a pre-COVID-19 baseline.

Apple:
Apple provide three data streams from their Apple Maps app of the total number of user requests for directions, one each for driving, walking, and public transport. No further details are provided on the nature of these journeys. Separate versions of the directions for driving data stream are provided for each Australian state/territory, whilst the directions for walking and public transport are provided only for Australia's four largest cities. We assume direction requests in these cities are representative of the states to which they belong. Apple's proportional change data is relative to a baseline of the metric in each location on 13 January 2020. Daily counts are calculated using Pacific Standard Time (*i.e.*, the time in California), which largely but not entirely overlaps with the subsequent day in Australia. We therefore assign mobility data to the subsequent date. These data are updated daily and can be accessed at: https://www.apple.com/covid19/mobility.

Citymapper:
Citymapper provide a composite index based on the numbers of requests for directions using the Citymapper app. This is a composite metric of requests for different transport mode, though Citymapper does not provide driving directions and is primarily used for public transport directions, with some use for walking, cycling, and cab directions. No further details are provided on the nature of these journeys. Separate versions of these data streams are provided for Sydney, Melbourne, and for all of Australia. We use the city-level data as representative of mobility within that state, and do not currently use the national composite data. Citymapper's proportional change data are relative to a baseline of the metric between 6 January and 2 February 2020. Daily counts are calculated using UTC (*i.e.*, the time in London), which largely but not entirely overlaps with the same day in Australia. We therefore assign mobility data to the date provided. These data are updated daily and can be accessed at: https://citymapper.com/CMI.

Google:
Google provide GPS-derived indices of the amount of time spent (a combination of visits and lengths of stay) in locations of one of 6 types. The following are Google's descriptions of these places:

- retail and recreation: "places like restaurants, cafes, shopping centers, theme parks, museums, libraries, and movie theaters"
- grocery & pharmacy: "places like grocery markets, food warehouses, farmers markets, specialty food shops, drug stores, and pharmacies"
- parks: "places like national parks, public beaches, marinas, dog parks, plazas, and public gardens"
- transit stations: "places like public transport hubs such as subway, bus, and train stations"
- workplaces: "places of work"
- residential: "places of residence"

This visitation information is derived from the aggregated GPS tracks of users with the Location History setting enabled (off by default). No further details are provided on the users or the nature of these visits. Separate versions of these data streams are provided for each state and territory, and for all of Australia. We do not currently use the national composite data. Google's proportional change data are relative to a baseline of the metric between 3 January and 6 February 2020. Daily counts are calculated using UTC (i.e., the time in London), which largely but not entirely overlaps with the same day in Australia. We therefore assign mobility data to the date provided. These data are updated intermittently. Recently the data have been released every week but providing data only up to about a week before the release date so that data are typically between one and two weeks out of date. The data can be accessed at: https://www.google.com/covid19/mobility/.

Representativeness

All of these data streams are derived from apps that are primarily smartphone-based. As such they overrepresent demographics that are more likely to own smartphones and have account with these technology companies. Data on the demographics of the users that contributed to these particular data streams are not available. Given that school-age children are underrepresented in COVID-19 cases in Australia, and that those over 65 are likely to be less mobile, these data may be unintentionally biased towards those contributing most strongly to transmission.

Apple's and Citymapper's direction requests likely represent deliberate visits to a specific location, and a location that the user either does not usually visit (and therefore requires directions) or would benefit from live traffic updates for. Whilst Google's index of the time spent in residential places is likely to be a good proxy for people staying at home, it does not exclude the possibility that this time is spent in another person's household.

Model description

We simultaneously analyse all of these data streams using a statistical latent variable model [5]. A latent variable model is akin to fitting a linear regression model to each data stream using the same set of covariates, but where the covariates are learned from the data streams rather than being specified in advance. These inferred covariates are termed latent variables. Latent variables are unitless quantities that summarise the patterns that are shared by some or all of the data streams. Because all data streams share the same latent variables but have their own regression coefficients ('loadings'), the same latent variable can imply a percentage increase on one data stream, but a different percentage decrease in another.

In this analysis, we consider each latent variable to be an index of some underlying population-level behavioural pattern. This enables us to distil these multiple data streams down into a smaller set of behavioural patterns we are most interested in. If a latent variable has a large influence on, and good statistical fit to multiple data streams of different types of mobility, that gives us confidence that the pattern is reflective of general behaviours, rather than being a quirk of a single mobility index.

Traditional latent variable models are purely statistical – they make no initial assumptions about the structure of the underlying latent variables. We developed a semi-mechanistic Bayesian latent variable analysis that uses prior knowledge to define a parametric function (with unknown parameters) for each latent variable, so that the latent variables represent known or hypothesised patterns of population-level behaviour. Three of these latent variables are related to COVID-19, and three describe non-COVID-19-related trends. The three COVID-19 related latent variables considered are: 'preparation' – a surge of activity in preparation of social distancing, 'social distancing' – a switch to more socially-distant behaviour, and 'waning distancing' – a partial reversion to a non-social-distancing behavioural state. The three non-COVID-19-related latent variables are: 'back to work' – a behavioural switch associated with the start of school term one (and return to work for many parents), 'day of the week' – different behaviours on each day of the week, including weekday vs weekend behaviours, and 'public holidays' – different behaviours associated with public holidays in each state. The model fits for each data stream incorporate the impacts of all of these latent variables, though we are most interested in the COVID-19 related variables

Because each data stream weights the different latent variables differently, these latent variables are considered only as relative effects they have no absolute magnitude or sign. We therefore define each of them as functions with values constrained between 0 and 1. The following describes the parametric functions chosen for each latent variable to reflect prior knowledge of the behaviours they represent:

- Preparation – a smooth, symmetric unimodal peak representing behaviours in advance of anticipated social distancing behaviour. The date and width of the peak are estimated from the data.

- Social Distancing – a smooth monotone increasing function representing the population switching to a behavioural state of increased social distance. We assume that the switching happens in bursts around each of three major social distancing measures introduced at the national level: restriction of gatherings to 500 or fewer people on 16 March; closure of restaurants, bars, and cafes on 24 March; restriction of gatherings to 2 or fewer people (except South Australia) on 29 March. The relative magnitudes of the behavioural switches associated with each of these social distancing measures, as well as the rate of change around each, are estimated from the data.

- Waning Distancing – a linear increasing function representing a waning of social distancing behaviour from a peak as the population switches back to a socially non-distant behavioural state. The overall date of the peak is estimated from the data, as is the fraction of the peak that has been reverted (I.e. how far it has reverted back to the baseline) in each data stream.

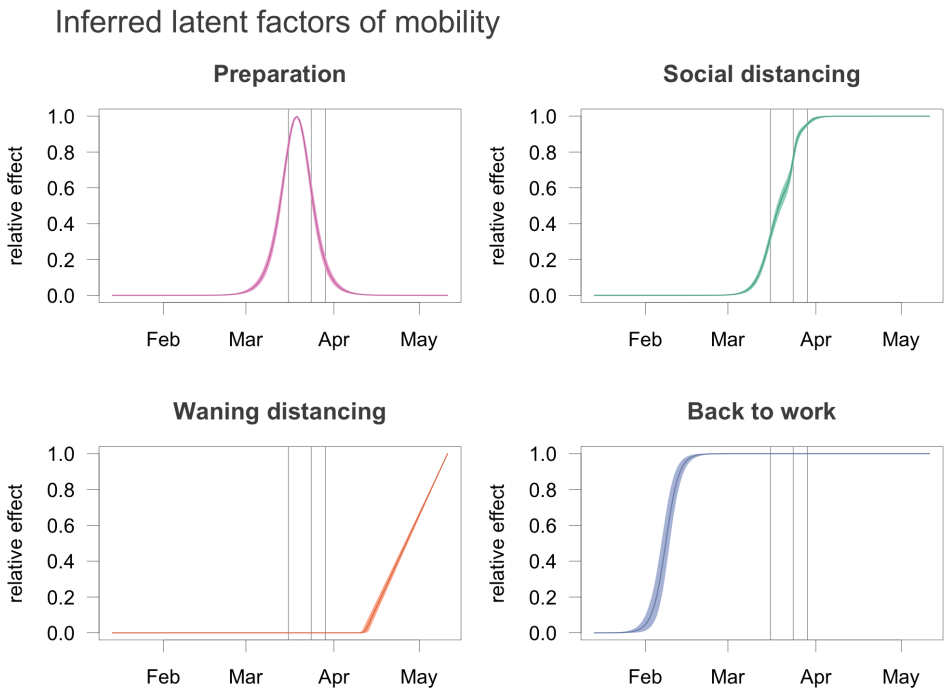
- Back to Work – an increasing sigmoid function representing a switch from one behavioural state during the school holidays to another during term one. The location (date by which 50% of people have switched) and slope of this sigmoid are estimated from the data.

- Day of the Week – a flexible curve over the day of the week, representing a gradient of behaviour between the peak of activity in the work week, and the trough at the weekend.

The curve is constrained between 0 and 1, with a value of 1 constrained to fall on a Sunday. The three parameters describing the shape of the curve are estimated from the data.

- Public Holidays – a series of independent offsets applied to each public holiday in each state to represent different behavioural states on these days. The value of the offset for each holiday is estimated from the data and scaled to between 0 and 1.

Figure S1: Plots of three COVID related latent factors ('Preparation', 'Social Distancing', and 'Waning Distancing') and one non-COVID related latent factor, ('Back to work'), inferred by the latent variable statistical model from multiple mobility data streams. Note that the height of these latent factors is not interpretable - they are all scaled between 0 and 1.



National survey to assess public response to social distancing measures

Estimating the change in number of reported non-household contacts

We estimated the population mean numbers of household contacts in each survey period, and the change in that mean number of contacts for individuals in that population using a statistical model. For each survey respondent, the number of contacts outside of the household was assumed to follow Negative Binomial distribution. The variance of that distribution — day-to-day variability in each respondent's daily number of contacts — was assumed to be the same for all respondents, but each respondent was assumed to have a different mean number of contacts per day. Each respondent's mean number of contacts per day in the first survey period was drawn from a Lognormal distribution. The mean of this distribution represents the population-wide average number of contacts per day in the first period, and the variance represents variability between members of the population in their average daily numbers of contacts. The statistical model can be represented, and was fitted, as a generalised linear mixed effects model with Negative Binomial sampling distribution, random effect on participant ID, and fixed effect of the survey wave.

Each respondent's mean number of contacts during the second survey period was assumed to be the mean during the first period multiplied by a population-wide parameter for the change in mean numbers of contacts, *i.e.*, the rate of change was assumed to be the same for all members of the population, but reflects change in individuals' mean number of contacts, rather than change in the population mean number of contacts.

The survey respondents were selected to be representative of the wider Australian population. To further increase the representativeness of the survey respondents, we incorporated survey weights into the model likelihood so that our parameter estimates represents the Australian population, rather than the population of survey respondents.

Further details on the survey methodology can be found in [6].

Estimates of current epidemic activity

We report preliminary results from a new method which estimates components of the effective reproduction number using outputs from the population mobility analysis.

Overview

We developed a new model to estimate components of the effective reproduction number resulting from transmission from locally acquired cases and from overseas acquired cases. This model enables us to 1) estimate the relative temporal variation in transmission from local to local cases and from overseas-acquired to local cases and 2) quantify the relative impacts of national-level interventions on transmission in Australia. Whilst both locally and overseas acquired cases contribute to Australia's case count, the transmission rates from each of these groups differs as they are each targeted by different interventions. Quarantine of overseas arrivals modifies the transmission rates of overseas acquired cases only, and social distancing measures modify transmission rates of locally acquired cases. By splitting R_{eff} between these two groups, the model enables us to estimate the relative impacts of various response policies on transmission in Australia, namely quarantine of overseas arrivals and social distancing of the general population.

We model local to local transmission and import to local transmission for each state/territory using three components:

1. the nationwide average reduction in R_{eff} that is due to interventions;
2. time-varying deviations from this national rate for each state, reflecting state to state differences in transmission; and
3. short-term fluctuations in R_{eff} in each state/territory to capture stochastic dynamics of transmission, such as clusters of cases and short periods of low transmission.

Modelling the impact of social distancing

We model the impact of social distancing by assuming that any reduction in R_{eff} local is proportional to the trend of reduced mobility seen in the population mobility analysis. The population mobility analysis identifies a common trend in all available data streams, whereby population mobility was reduced around the dates that three social distancing restrictions were implemented. The reduction accelerated around the time of the second national restriction — closure of bars restaurants and cafes — resulting in a noticeable kink in this curve. The model estimates a constant of proportionality between this index and the reduction in R_{eff} local. The overall effect of social distancing restrictions on transmission is therefore estimated from case data, but the shape and timing of the impact are estimated from mobility data, which are much more informative (Figure 6). **This proportionality assumption is under review and may be relaxed in a future iteration — enabling us to more confidently assign reductions in transmission to specific policies.**

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 S3). 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 S2: Nationwide average reduction in R_{eff} that is due to social distancing estimated from the R_{eff} model (light green ribbon=90% credible interval; dark green ribbon = 50% credible interval). Note that this trend does not capture time-varying fluctuations in R_{eff} in each state/territory due to stochastic dynamics of transmission, such as the healthcare-associated outbreak in Tasmania. Solid grey vertical lines indicate key dates of implementation of social distancing policies. Black dotted line indicates the target value of 1 for the effective reproduction number required for control.

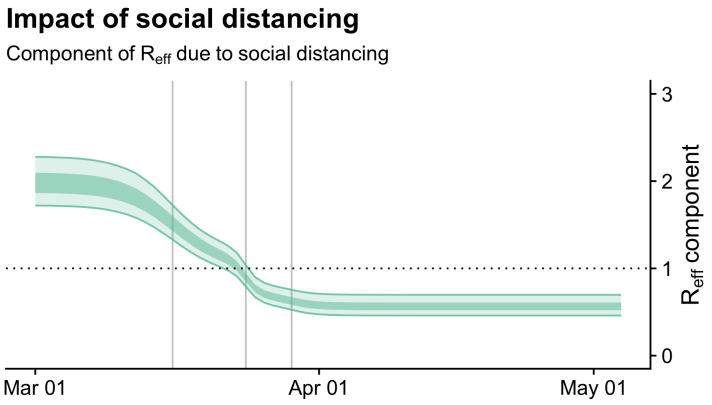
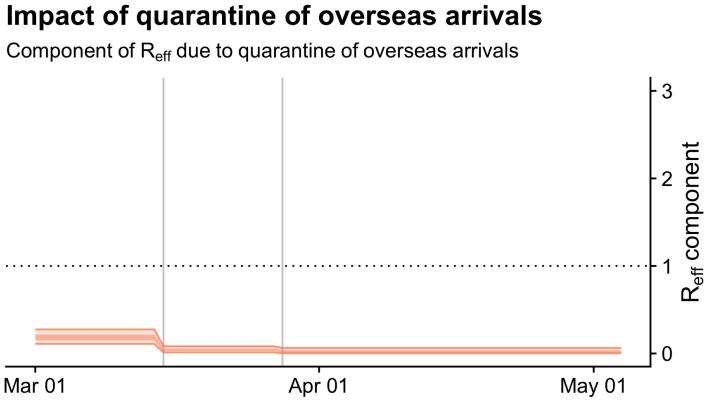


Figure S3: Nationwide average reduction in R_{eff} that is due to quarantine and isolation of overseas arrivals estimated from the R_{eff} model (light orange ribbon=90% credible interval; dark orange ribbon = 50% credible interval). Note that this trend does not capture time-varying fluctuations in 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.



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 accounting for the missing data reduces the precision of these estimates. For example, we currently cannot account for state-level variation in the impacts of interventions 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 $R_i(t)$:

$$N_i^L(t) \sim Poisson(\lambda_i(t)) \tag{1}$$
$$\lambda_i(t) = I_i(t)R_i(t) \tag{2}$$
$$I_i(t) = \sum_{t'=1}^{t} g(t')N_i(t') \tag{3}$$
$$N_i(t') = N_i^L(t) + N_i^O(t) \tag{4}$$

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) [2] and the variations of that method by Cori et al. (2013) [3], Thompson et al. (2019) [7] 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:

$$N_i^L(t) \sim Poisson(\lambda_i(t)) \tag{5}$$

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

$$I_i^L(t) = \sum_{t'=1}^{t} g(t')N_i^L(t) \tag{7}$$

$$I_i^O(t) = \sum_{t'=1}^{t} g(t')N_i^O(t) \tag{8}$$

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.

Reproduction rate models
We model the reproduction rates of the two sub-populations in a semi-mechanistic way:

$$R_i^L(t) = R_0 D(t) e^{\epsilon_i^L(t)} \tag{9}$$

$$R_i^O(t) = R_0 Q(t) e^{\epsilon_i^O(t)} \tag{10}$$

with each reproduction rate modelled as a product of: the reproduction rate under initial conditions and no interventions R_0; deterministic functions $D(t)$ and $Q(t)$ that modify R_0 over time to respectively represent the impacts of social distancing measures and quarantine (and isolation) of overseas arrivals at a national level, and correlated time series of random effects $\epsilon_i^L(t)$ and $\epsilon_i^O(t)$ to represent stochastic fluctuations in the reporting rate in each state. These error terms are themselves decomposed into population-wide changes in $R_i(t)$ not captured by the impacts of the policy measures, and the more stochastic fluctuations in the reproduction rate among cases at each point in time — for example due to the clusters in sub-populations with higher or lower reproduction rates than the general population.

Intervention models
We model the effect of $D(t)$ as being proportional (on the log scale) to an index of the proportional change in population mobility in response to physical distancing measures $d(t)$ – estimated from multiple data streams of population mobility as described in the next section — which has initial value 0 before distancing measures were implemented and value 1 at its maximum extent:

$$D(t) = 1 - \beta d(t) \tag{11}$$

We model $Q(t)$ 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:

$$Q(t) = \begin{cases} q_1 & t < \tau_1 \\ q_2 & \tau_1 \leq t < \tau_2 \\ q_3 & \tau_2 \leq t \end{cases} \tag{12}$$

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 each group $\epsilon_i^L(t)$ and $\epsilon_i^O(t)$ are each modelled as a zero-mean Gaussian process (GP) with covariance structure reflecting temporal correlation in errors within each state (but independent between states). The covariance function (kernel) for each Gaussian process is constructed as the sum of covariance functions representing different components of error in the observed reproduction rates. Below we follow the notation of additive GP kernels that is more common in the field of probabilistic machine learning than in statistics when discussing correlated error terms. This corresponds to the summation of the covariance matrices returned by these functions and equally to the summation of GPs, each following these functions.

The structure for error in locally-acquired cases is as follows:

$$\epsilon_i^L \sim GP(\mathbf{0}, k^L(t,t')) \tag{13}$$
$$k^L(t,t') = k_{SE}(t,t') + k_{RQ}(t,t') + k_{IID}^L(t,t') \tag{14}$$
$$k_{SE}(t,t') = \sigma_1^2 \exp\left(-\frac{(t-t')^2}{2l_1^2}\right) \tag{15}$$
$$k_{RQ}(t,t') = \sigma_2^2 \exp\left(1 + \frac{(t-t')^2}{2\alpha l_2^2}\right)^{-\alpha} \tag{16}$$
$$k_{IID}^L(t,t') = \begin{cases} \sigma_3^2 & t = t' \\ 0 & t \neq t' \end{cases} \tag{17}$$

where k_{SE} is a squared-exponential kernel encoding smooth, long-term deviations from the mean, k_{RQ} is a rational-quadratic kernel encoding shorter-term deviations from the mean with varying smoothness, and k_{IID} is an IID or white-noise kernel encoding independent deviations from the mean on each day. The three kernels are used to separately encode three different types of variation. The squared exponential kernel represents long-term trends in the population-wide transmission rate $R_i^L(t)$ that are not captured by the effects of social distancing — such as improvements in surveillance, or behavioural changes not captured by changes in mobility data. This kernel also represents time-independent region-level differences in transmission rates. The rational quadratic kernel represents the short-term fluctuations in the reproduction rate that are due to large clusters of cases, with infections occurring over a period of days to weeks. The IID kernel represents clustering of cases on a single day — overdispersion in the observed numbers of new local infections per day relative to the assumption of independent events implicit in the Poisson observation model. This is equivalent to using a Poisson-Lognormal observation distribution conditional on the time series component of the model; functionally equivalent to a Negative Binomial observation distribution.

The structure for error in overseas-acquired cases is somewhat simpler:

$$\epsilon_i^O \sim GP(\mathbf{0}, k^O(t,t')) \tag{18}$$
$$k^O(t,t') = k_{bias}(t,t') + k_{IID}^O(t,t') \tag{19}$$
$$k_{bias}(t,t') = \sigma_4^2 \tag{20}$$
$$k_{IID}^O(t,t') = \begin{cases} \sigma_5^2 & t = t' \\ 0 & t \neq t' \end{cases} \tag{21}$$

which consists of a bias kernel, k_{bias}, representing time-independent region-level differences in transmission rates, and an IID kernel, k^O_{IID}, representing daily overdispersion due to clustering in new cases acquired from individuals with overseas-acquired infections. Note that this GP formulation is equivalent to the following IID Gaussian random effects model:

$$\epsilon^O_i(t) \sim N(\mu_i, \sigma_5^2) \tag{22}$$

$$\mu_i \sim N(0, \sigma_4^2) \tag{23}$$

Components of local transmission potential
We model the rate of transmission from locally acquired cases using the basic reproduction rate, the impact of social distancing, and an error term. However it is easier to consider these epidemiologically by decomposing the expression for $R^L_i(t)$ into three different components:

1. the national trend in $R(t)$ due to social distancing,
2. the state-level trends in population-wide $R_i(t)$, and
3. the deviation in this statewide average transmission rate due to case clusters.

Component 1 of the transmission rate for locally-acquired cases is simply given by $R_0 D(t)$. The latter two correspond to different sub-components of the correlated error term $\epsilon^L_i(t)$: $\epsilon^{L,1}_i$ which comprises long-term variation in the region-wide transmission rate, characterised by k_{SE}; and $\epsilon^{L,2}_i(t)$, characterised by $k_{RQ} + k^L_{IID}$, which comprises short-term variation in the transmission rate among the infectious cases due to clusters of cases in sub-populations with transmission rates that are higher (or possibly lower) than the population average. The additive structure of the kernel enables us to perform an additive decomposition of $\epsilon^L_i(t)$ into these two terms.

Component 2 is therefore given by $R_0 D(t) e^{\epsilon^{L,1}_i}$, and Component 3 (which we consider on the log-scale for ease of visualisation) is $\epsilon^{L,2}_i(t)$.

Parameter values and priors
Tables S1 and S3 give the prior distributions of parameters in the semi-mechanistic and time-series (ϵ^L and ϵ^O) 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 means of the Log-normal distribution over the serial interval estimated by [8].

The parameters of the Lognormal prior over R_0 were computed such that the distribution matched 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_0.

We assumed slightly informative priors over the lengthscales of components 2 and 3 of the time series over the effective reproduction number for local-local transmission (parameters l_1 and l_2), such that longer lengthscales were more likely for component 2 (time series of the state-wide average transmission rate) than for component 3 (time series of the average transmission rate over active cases).

Model fitting
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 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.

Table S1: Parameters in the semi-mechanistic part of the time-varying model of R_{eff}

Prior distribution	Parameter description
$R_0 \sim \text{lognormal}(1.02, 0.10^2)$	Basic reproduction number
$\beta \sim \text{U}(0, 1)$	Effect of social distancing behaviours
$q_1 \sim \text{U}(0, 1)$	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
$g(t) = \int_{t-1}^{t} \text{lognormal}(\tau \mid 1.377, 0.567^2)\, \text{d}\tau$	Generation interval function

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

Prior distribution	Parameter description
$\sigma_1 \sim \text{N}^+(0, 0.25^2)$	Amplitude of deviation; local R_{eff} statewide
$l_1 \sim \text{lognormal}(4, 0.5^2)$	Temporal correlation; local R_{eff} statewide
$\sigma_2 \sim \text{N}^+(0, 0.25^2)$	Amplitude of deviation; local R_{eff} active cases
$l_2 \sim \text{lognormal}(2, 0.5^2)$	Temporal correlation; loca R_{eff} active cases
$\alpha \sim \text{lognormal}(2, 0.5^2)$	Correlation mixture weights; local R_{eff} active cases
$\sigma_3 \sim \text{N}^+(0, 0.25^2)$	Overdispersion; local R_{eff} active cases
$\sigma_4 \sim \text{N}^+(0, 0.5^2)$	Amplitude of deviation; import R_{eff} statewide
$\sigma_5 \sim \text{N}^+(0, 0.5^2)$	Overdispersion; import R_{eff} active cases

Forecasts of the daily number of new confirmed cases

Compartmental model of infection

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:

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

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

$$
\begin{aligned}
S^{\text{Pr}}(t) &= 1 - \exp\left(-\beta(t) \cdot \left[I_1(t) + I_2(t)\right]/N\right) & S^{\text{out}}(t) &\sim Bin(S(t), S^{\text{Pr}}(t)) \\
E_1^{\text{Pr}}(t) &= 1 - \exp\left(2 \cdot \sigma(t)\right) & E_1^{\text{out}}(t) &\sim Bin(E_1(t), E_1^{\text{Pr}}(t)) \\
E_2^{\text{Pr}}(t) &= 1 - \exp\left(2 \cdot \sigma(t)\right) & E_2^{\text{out}}(t) &\sim Bin(E_2(t), E_2^{\text{Pr}}(t)) \\
I_1^{\text{Pr}}(t) &= 1 - \exp\left(2 \cdot \gamma(t)\right) & I_1^{\text{out}}(t) &\sim Bin(I_1(t), I_1^{\text{Pr}}(t)) \\
I_2^{\text{Pr}}(t) &= 1 - \exp\left(2 \cdot \gamma(t)\right) & I_2^{\text{out}}(t) &\sim Bin(I_2(t), I_2^{\text{Pr}}(t)) \\
S(t+1) &= S(t) - S^{\text{out}}(t) & E_1(t+1) &= E_1(t) + S^{\text{out}}(t) - E_1^{\text{out}}(t) \\
E_2(t+1) &= E_2(t) + E_1^{\text{out}}(t) - E_2^{\text{out}}(t) & I_1(t+1) &= I_1(t) + E_2^{\text{out}}(t) - I_1^{\text{out}}(t) \\
I_2(t+1) &= I_2(t) + I_1^{\text{out}}(t) - I_2^{\text{out}}(t) & R(t+1) &= R(t) + I_2^{\text{out}}(t)
\end{aligned}
$$

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 $\mathrm{X}(t)$ 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:

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

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
	$R_{\text{eff}}(t)$	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	0.8
	p_{obs}	The observation probability	0.8
	k	The dispersion parameter	100
(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 $R_{\text{eff}}(t)$, 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 $R_{\text{eff}}(t)$ trajectories for each particle.

References

[1] Joël Mossong, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Massari, Stefania Salmaso, Gianpaolo Scalia Tomba, Jacco Wallinga, Janneke Heijne, Malgorzata Sadkowska-Todys, Magdalena Rosinska, and W. John Edmunds. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLOS Medicine*, 5(3):1–1, 2008.

[2] 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.

[3] 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.

[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 Bartholomew, Martin Knott, and Irini Moustaki. *Latent Variable Models and Factor Analysis: A Unified Approach*. Wiley, London, United Kingdom, 2011.

[6] Freya M. Shearer, Lisa Gibbs, Eva Alisic, Katitza Marinkovic Chavez, Niamh Meagher, Lauren Carpenter Phoebe Quinn, Colin MacDougall, and David J. Price. Distancing measures in the face of COVID-19 in Australia. Available from: `https://www.doherty.edu.au/uploads/content_doc/social_distancing_survey_wave1_report_May142.pdf`, 2020.

[7] 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.

[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.

Supplementary figures to forecasting and alternative transmission scenarios

Figure S4: Time varying estimate of local to local transmission potential (light green ribbon = 90% credible interval; dark green ribbon = 50% credible interval) for each Australian state/territory used as an input to generate the forecasts shown in Figures 8 and 10. Estimates were made up to 4 May and projected forward from 11 May to 1 July (indicated by grey box), **assuming that mean local transmission potential remains at current estimate levels**. Solid grey vertical lines indicate key dates of implementation of various social distancing policies. Black dotted line indicates the target value of 1 for the effective reproduction number required for control. Where there is epidemic activity, this quantity may be interpreted as the effective reproduction number, R_{eff}. In the absence of epidemic activity, this quantity reflects the ability for the virus, if it were present, to establish and maintain community transmission (> 1) or otherwise (< 1).

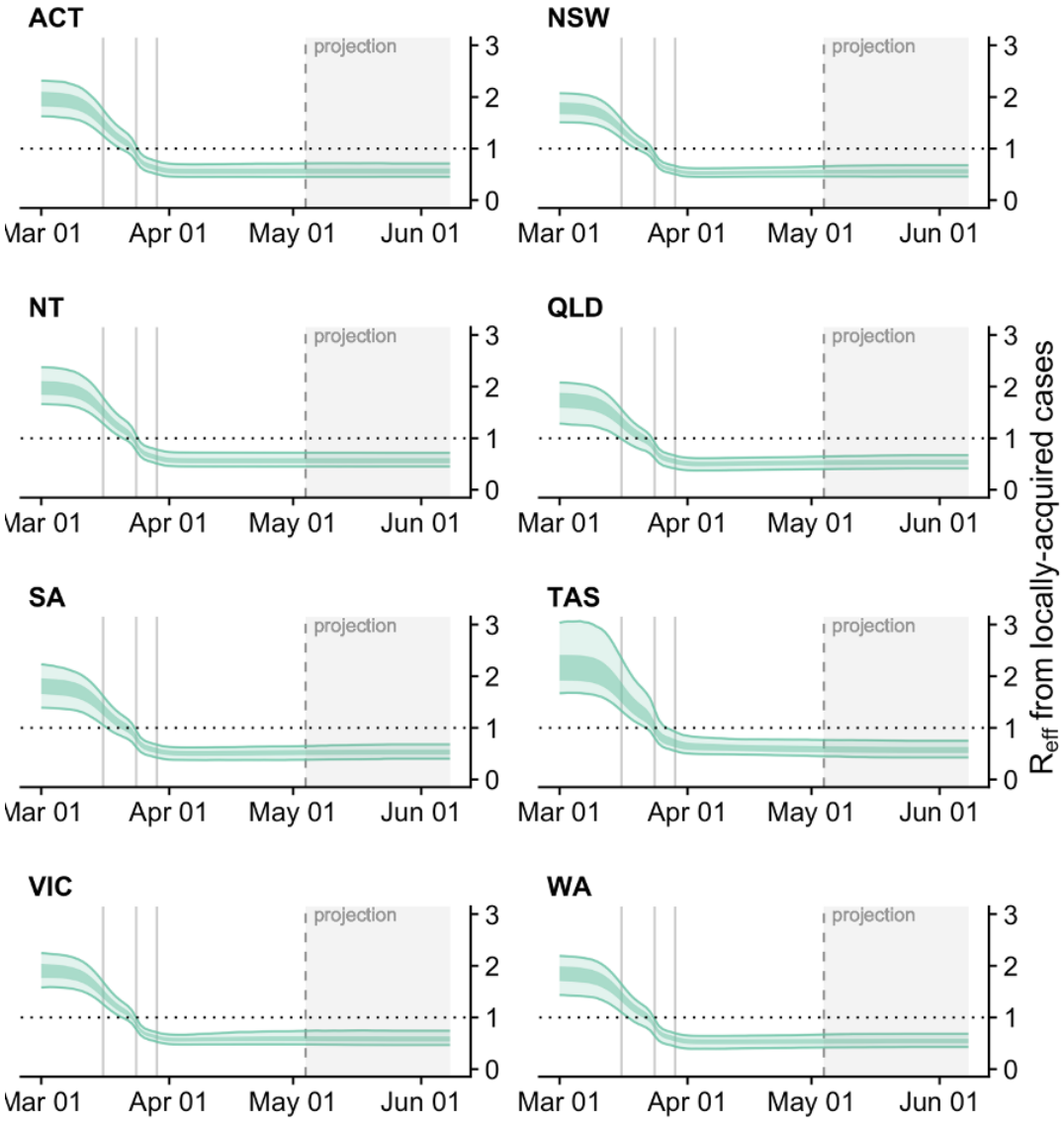


Figure S5: Time varying estimate of local to local transmission potential (light green ribbon = 90% credible interval; dark green ribbon = 50% credible interval) for each Australian state/territory used as an input to generate the alternate transmission scenario shown in 9. Estimates were made up to 4 May and projected forward from 11 May to 1 July (indicated by grey box), **assuming that mean local transmission potential increases to 1.1 from 11 May.** Solid grey vertical lines indicate key dates of implementation of various social distancing policies. Black dotted line indicates the target value of 1 for the effective reproduction number required for control. Where there is epidemic activity, this quantity may be interpreted as the effective reproduction number, R_{eff}. In the absence of epidemic activity, this quantity reflects the ability for the virus, if it were present, to establish and maintain community transmission (> 1) or otherwise (< 1).

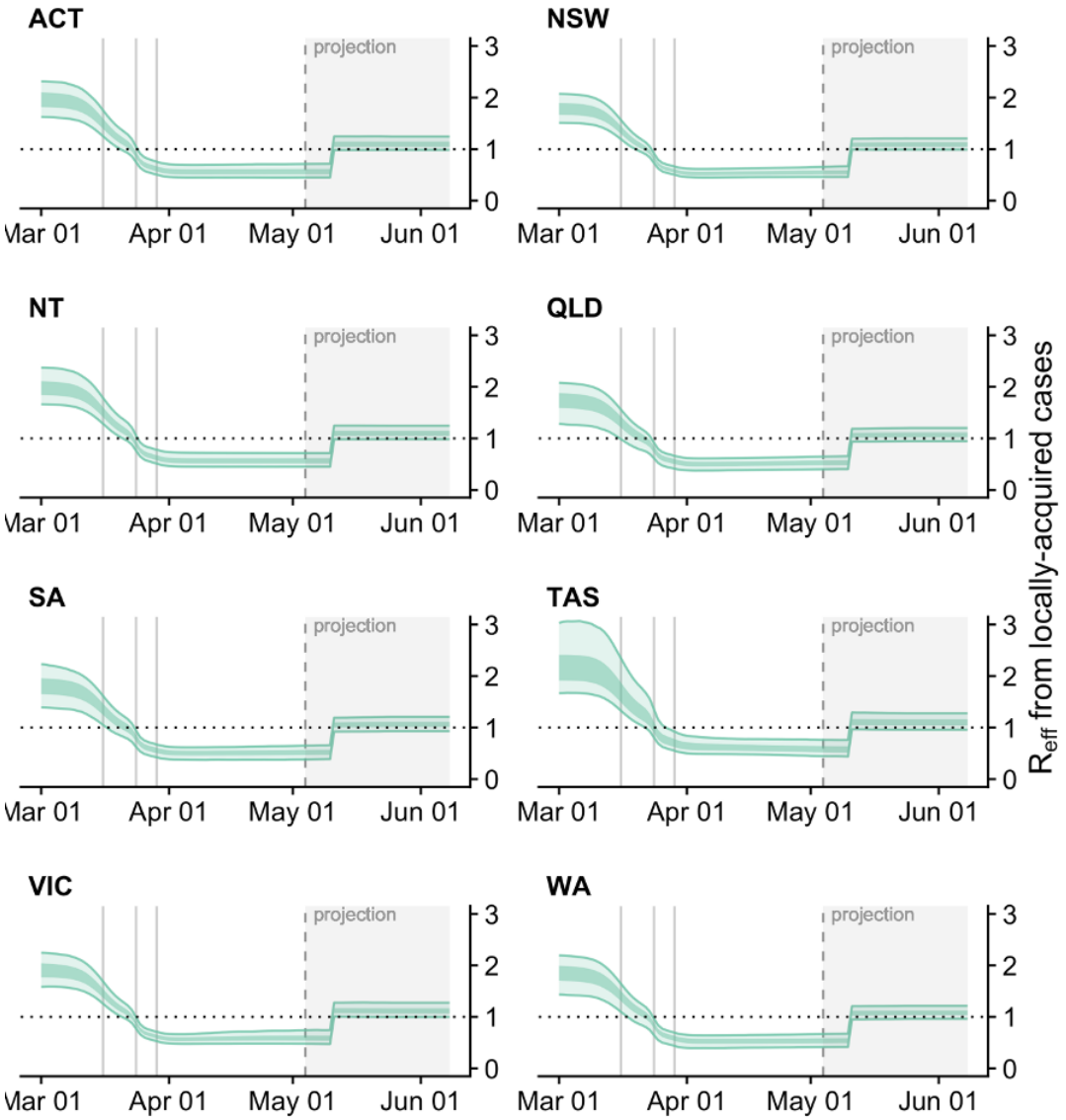


Figure S6: Time varying estimate of local to local transmission potential (light green ribbon = 90% credible interval; dark green ribbon = 50% credible interval) for each Australian state/territory used as an input to generate the alternate transmission scenario shown in 9. Estimates were made up to 4 May and projected forward from 11 May to 1 July (indicated by grey box), **assuming that mean local transmission potential increases to 1.2 from 11 May.** Solid grey vertical lines indicate key dates of implementation of various social distancing policies. Black dotted line indicates the target value of 1 for the effective reproduction number required for control. Where there is epidemic activity, this quantity may be interpreted as the effective reproduction number, R_{eff}. In the absence of epidemic activity, this quantity reflects the ability for the virus, if it were present, to establish and maintain community transmission (> 1) or otherwise (< 1).

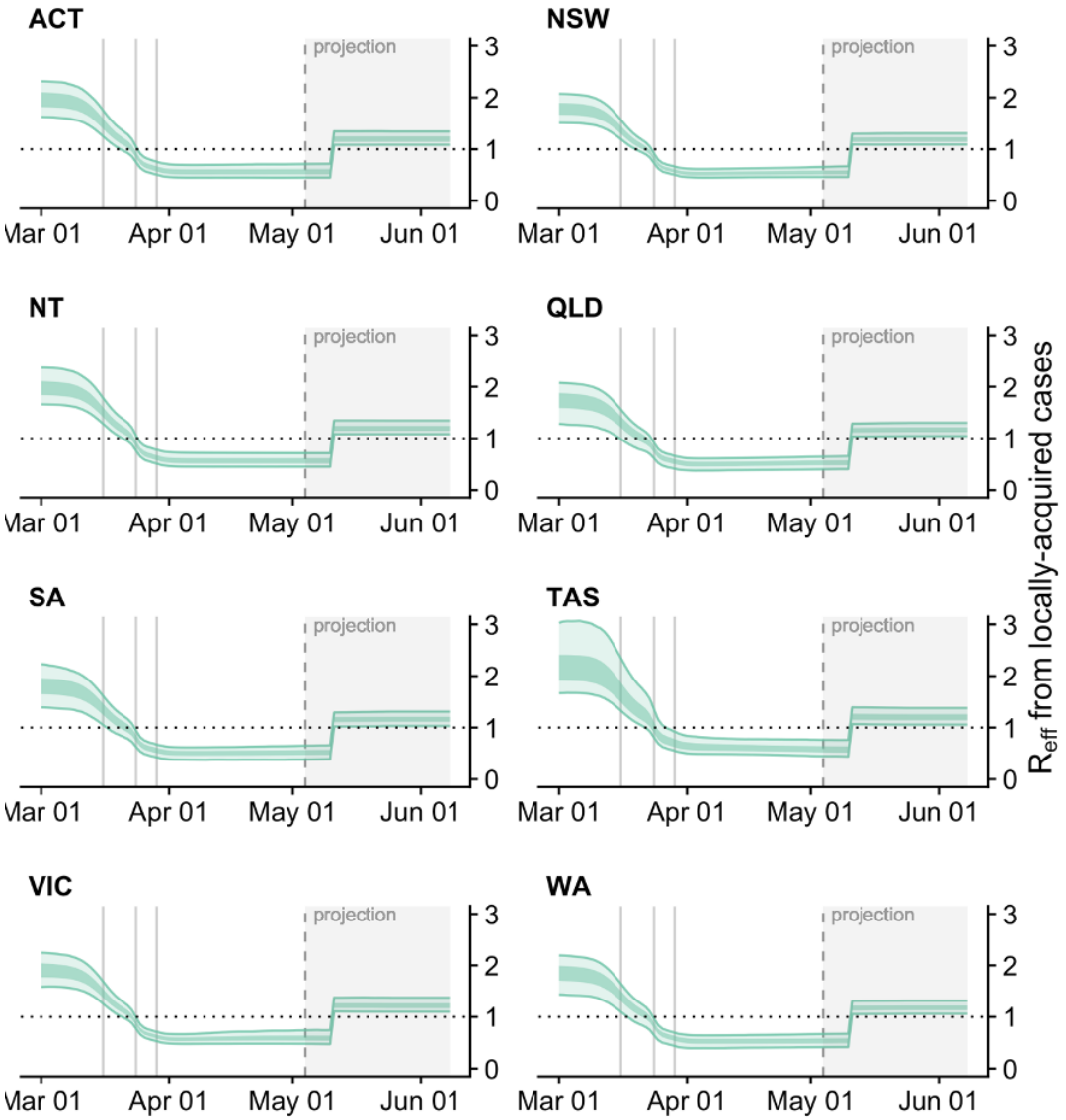


Figure S7: Time varying estimate of local to local transmission potential (light green ribbon = 90% credible interval; dark green ribbon = 50% credible interval) for each Australian state/territory used as an input to generate the alternate transmission scenario shown in 9. Estimates were made up to 4 May and projected forward from 11 May to 1 July (indicated by grey box), **assuming that mean local transmission potential increases to 1.5 from 11 May**. Solid grey vertical lines indicate key dates of implementation of various social distancing policies. Black dotted line indicates the target value of 1 for the effective reproduction number required for control. Where there is epidemic activity, this quantity may be interpreted as the effective reproduction number, R_{eff}. In the absence of epidemic activity, this quantity reflects the ability for the virus, if it were present, to establish and maintain community transmission (> 1) or otherwise (< 1).

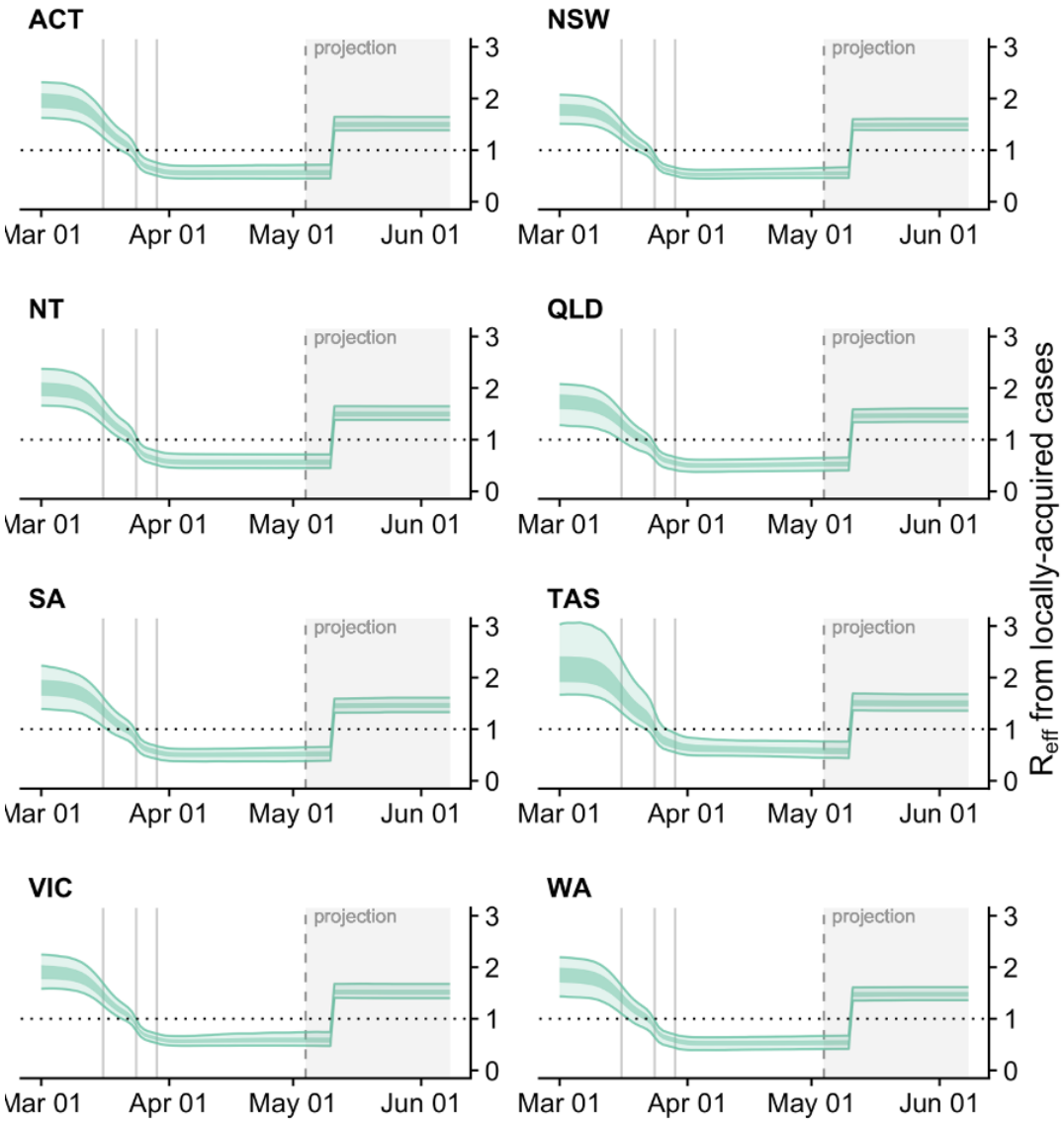


Figure S8: Time series of new daily confirmed cases of COVID-19 estimated from the forecasting model up to 10 May and projected forward up to 1 July (light blue shading = 95% confidence intervals, dark blue shading = 50% confidence intervals), **assuming that local transmission potential increases to 1.1 from 11 May** (indicated by the dashed black line). The observed case counts are also plotted (grey bars).

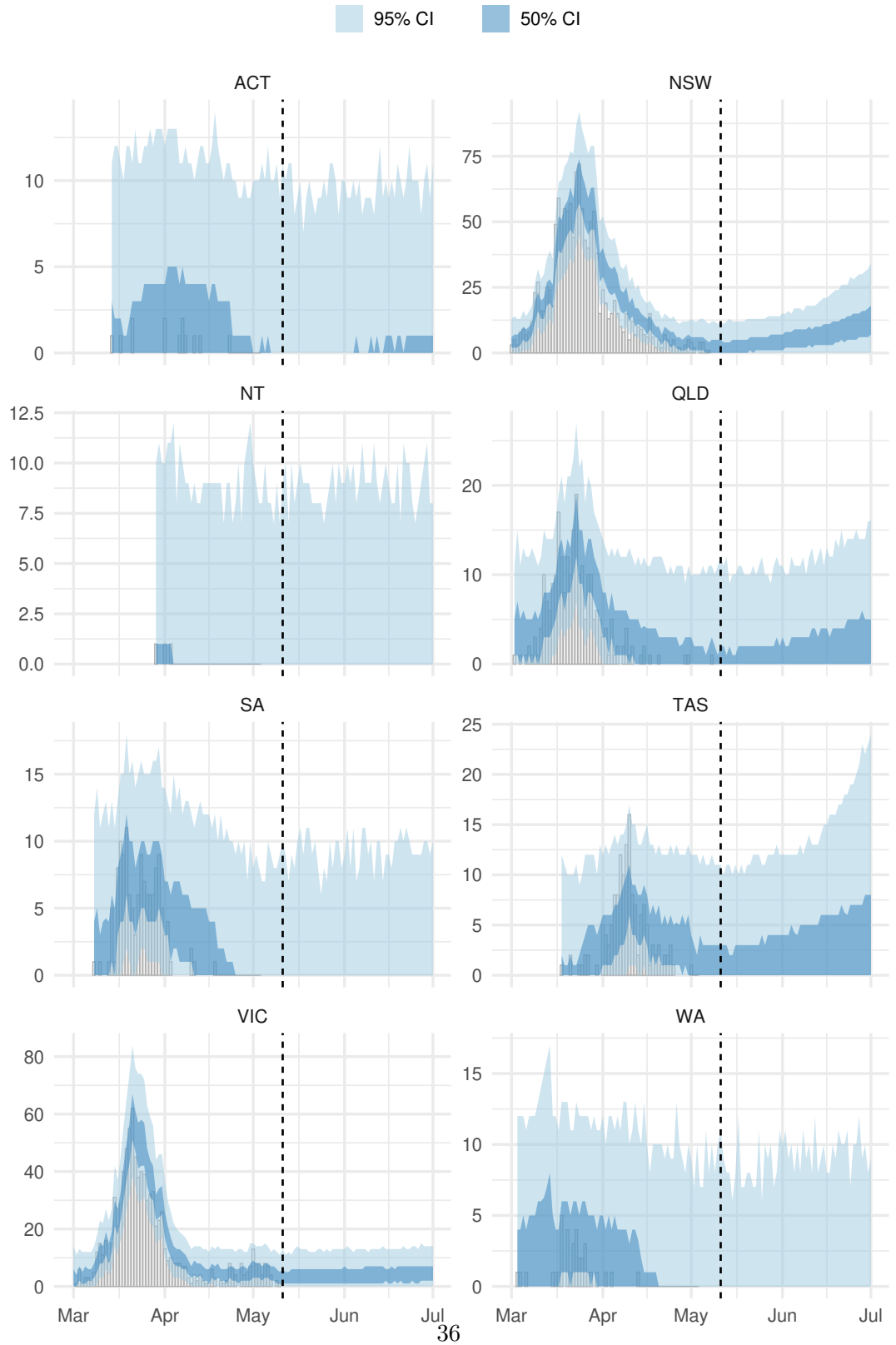


Figure S9: Time series of new daily confirmed cases of COVID-19 estimated from the forecasting model up to 10 May and projected forward up to 1 July (light blue shading = 95% confidence intervals, dark blue shading = 50% confidence intervals), **assuming that local transmission potential increases to 1.2 from 11 May** (indicated by the dashed black line). The observed case counts are also plotted (grey bars).

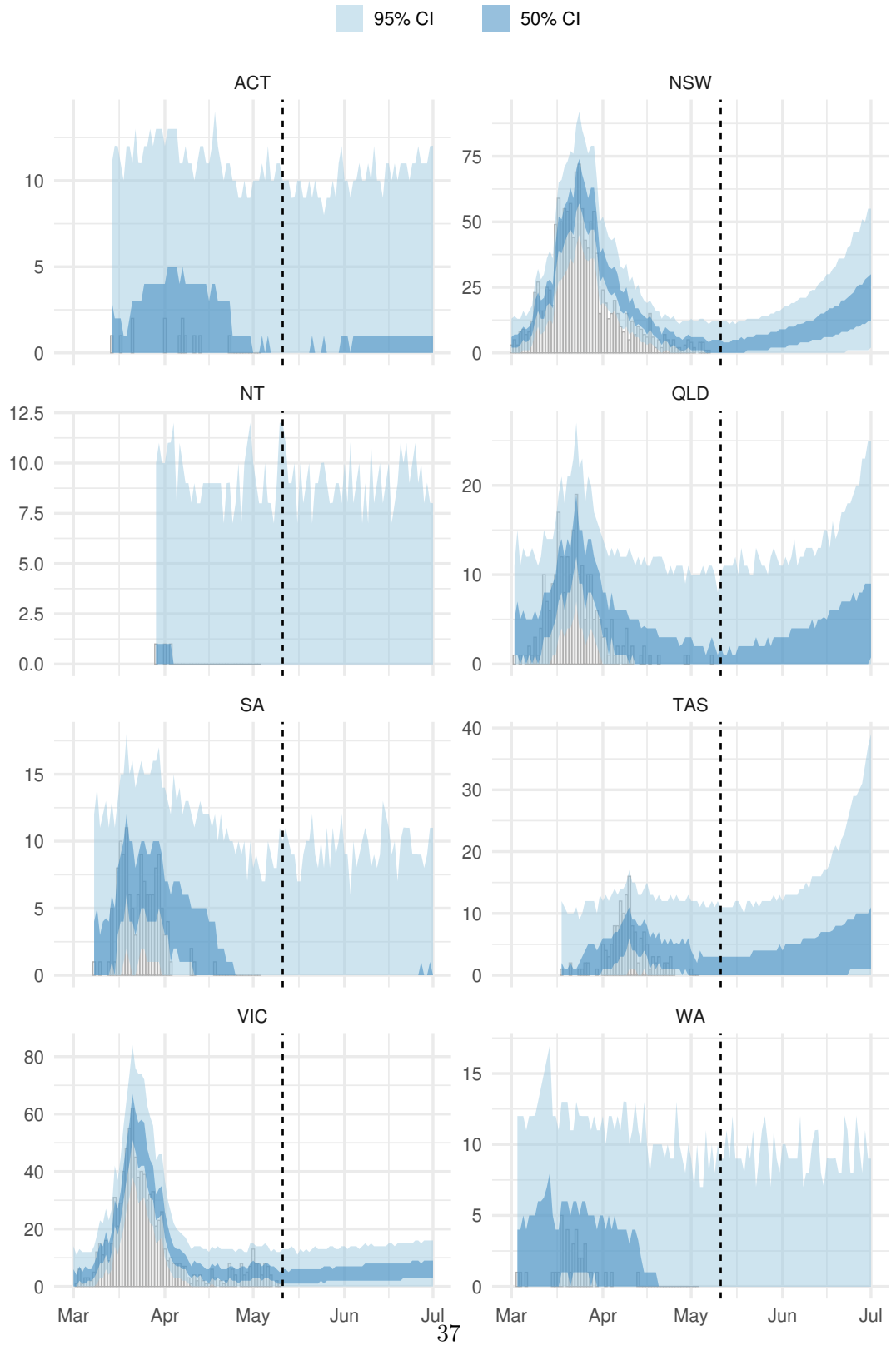
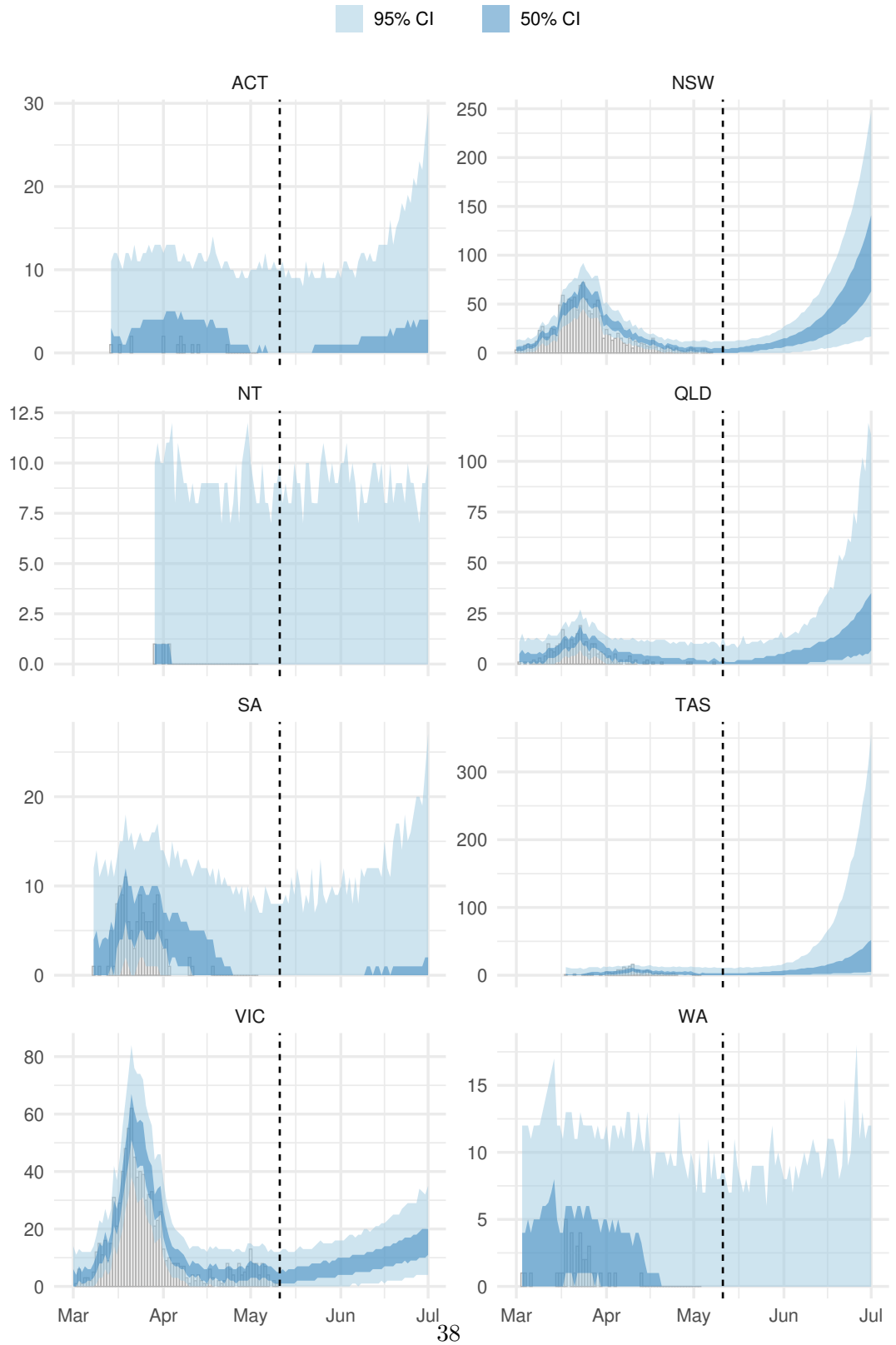


Figure S10: Time series of new daily confirmed cases of COVID-19 estimated from the forecasting model up to 10 May and projected forward up to 1 July (light blue shading = 95% confidence intervals, dark blue shading = 50% confidence intervals), **assuming that local transmission potential increases to 1.5 from 11 May** (indicated by the dashed black line). The observed case counts are also plotted (grey bars).



Supplementary figures to population mobility analysis

Figure S11: Percentage change compared to a pre-COVID-19 baseline of a number of key mobility data streams in the Australian Capital Territory. Solid vertical lines give the dates of three social distancing measures: restriction of gatherings to 500 people or fewer; closure of bars, restaurants, and cafes; restriction of gatherings to 2 people or fewer. The dashed vertical line marks the most recent date for which some mobility data are available. Blue 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 of the latent behavioural factors driving each data stream.

Figure S12: Percentage change compared to a pre-COVID-19 baseline of a number of key mobility data streams in New South Wales. Solid vertical lines give the dates of three social distancing measures: restriction of gatherings to 500 people or fewer; closure of bars, restaurants, and cafes; restriction of gatherings to 2 people or fewer. The dashed vertical line marks the most recent date for which some mobility data are available. Blue 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 of the latent behavioural factors driving each data stream.

Figure S13: Percentage change compared to a pre-COVID-19 baseline of a number of key mobility data streams in Northern Territory. Solid vertical lines give the dates of three social distancing measures: restriction of gatherings to 500 people or fewer; closure of bars, restaurants, and cafes; restriction of gatherings to 2 people or fewer. The dashed vertical line marks the most recent date for which some mobility data are available. Blue 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 of the latent behavioural factors driving each data stream.

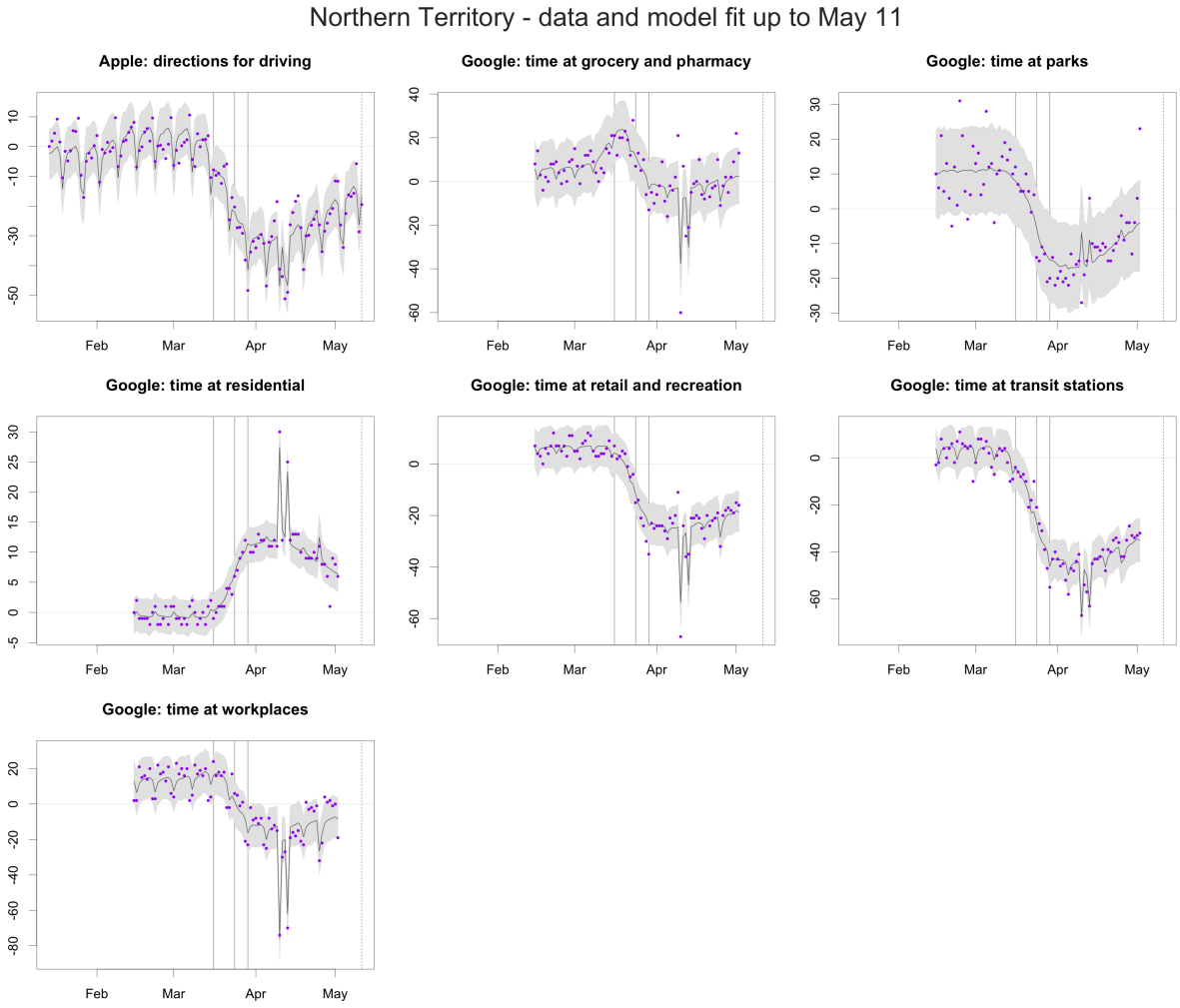


Figure S14: Percentage change compared to a pre-COVID-19 baseline of a number of key mobility data streams in Queensland. Solid vertical lines give the dates of three social distancing measures: restriction of gatherings to 500 people or fewer; closure of bars, restaurants, and cafes; restriction of gatherings to 2 people or fewer. The dashed vertical line marks the most recent date for which some mobility data are available. Blue 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 of the latent behavioural factors driving each data stream.

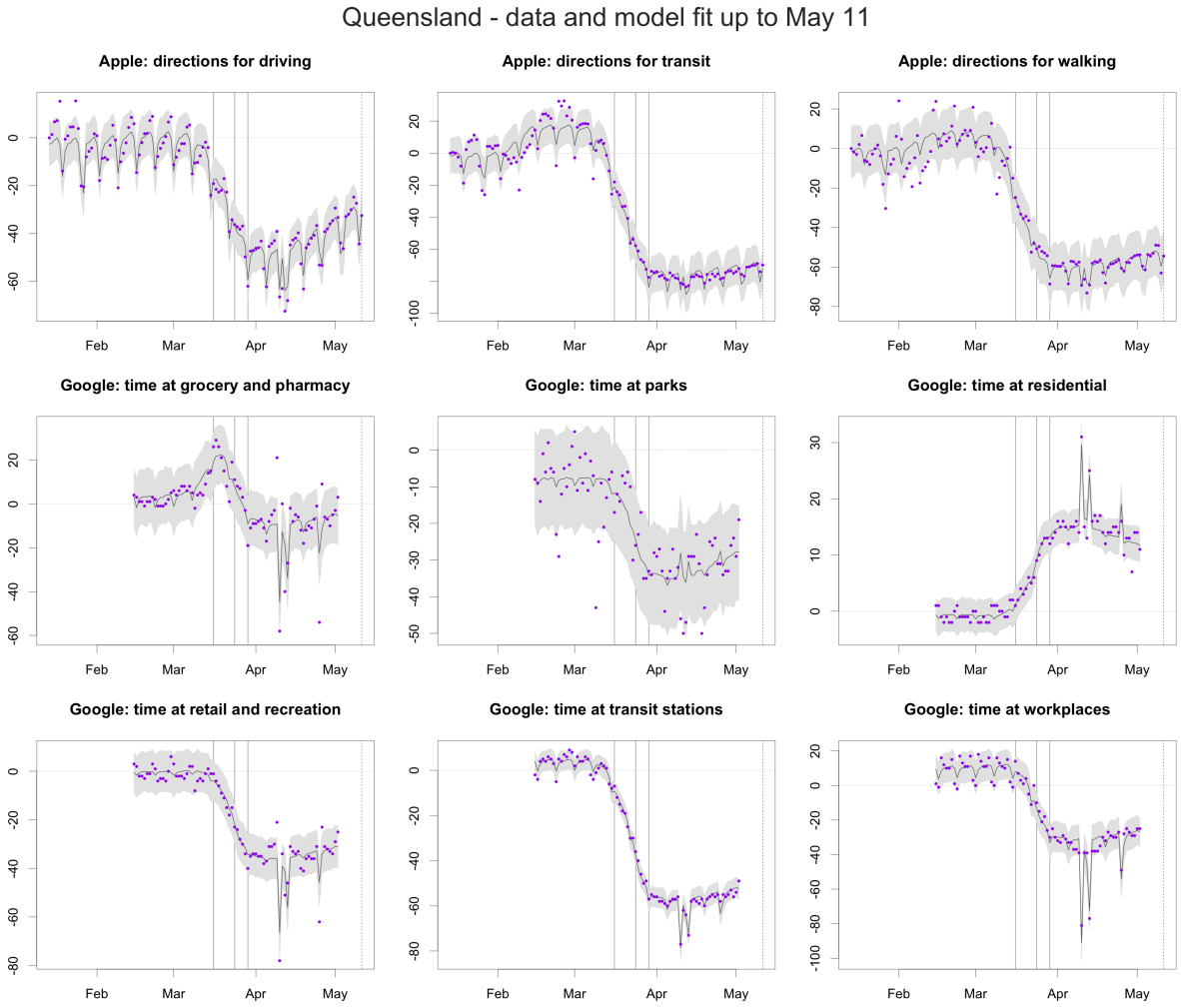


Figure S15: Percentage change compared to a pre-COVID-19 baseline of a number of key mobility data streams in South Australia. Solid vertical lines give the dates of three social distancing measures: restriction of gatherings to 500 people or fewer; closure of bars, restaurants, and cafes; restriction of gatherings to 2 people or fewer. The dashed vertical line marks the most recent date for which some mobility data are available. Blue 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 of the latent behavioural factors driving each data stream.

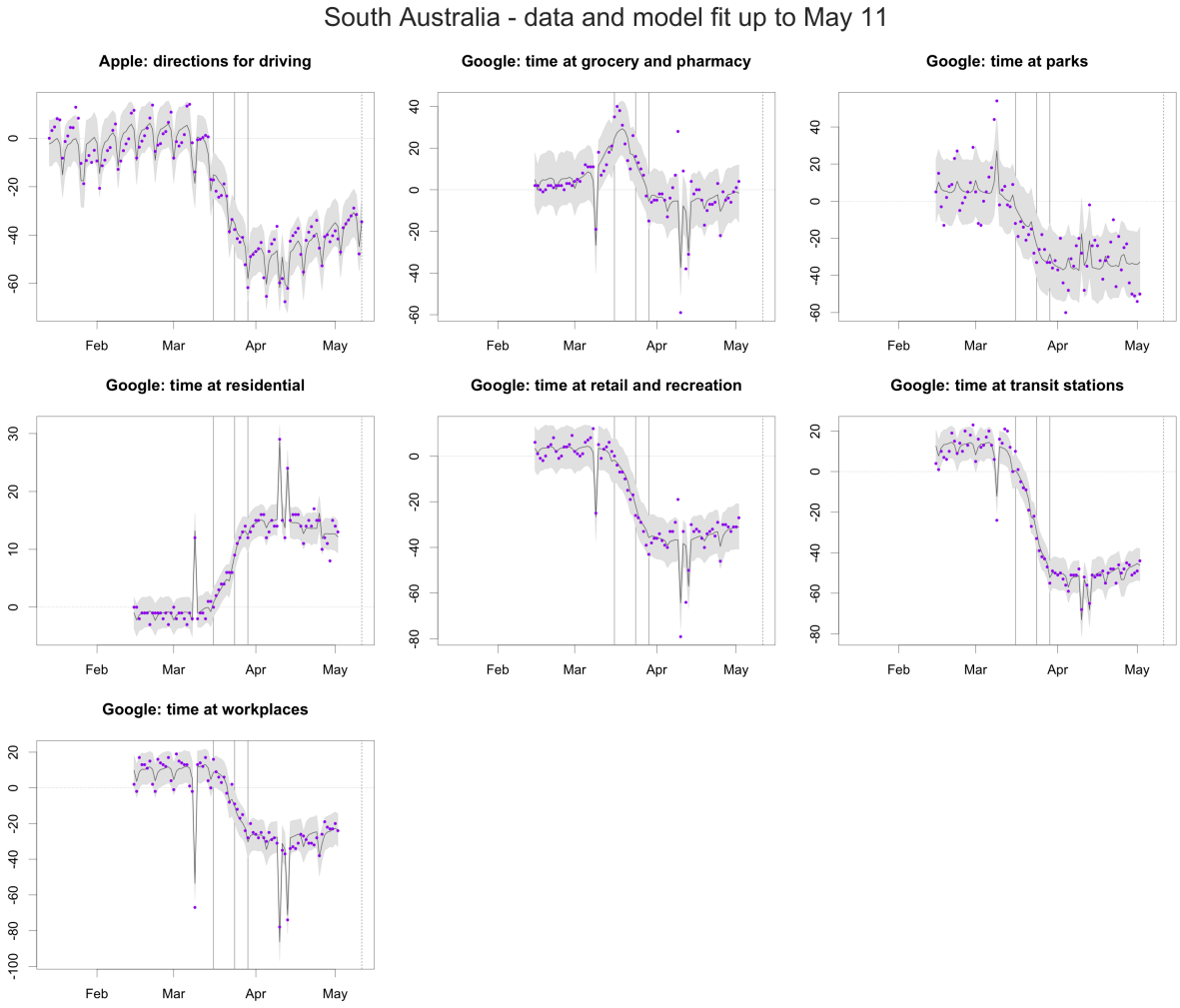


Figure S16: Percentage change compared to a pre-COVID-19 baseline of a number of key mobility data streams in Tasmania. Solid vertical lines give the dates of three social distancing measures: restriction of gatherings to 500 people or fewer; closure of bars, restaurants, and cafes; restriction of gatherings to 2 people or fewer. The dashed vertical line marks the most recent date for which some mobility data are available. Blue 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 of the latent behavioural factors driving each data stream.

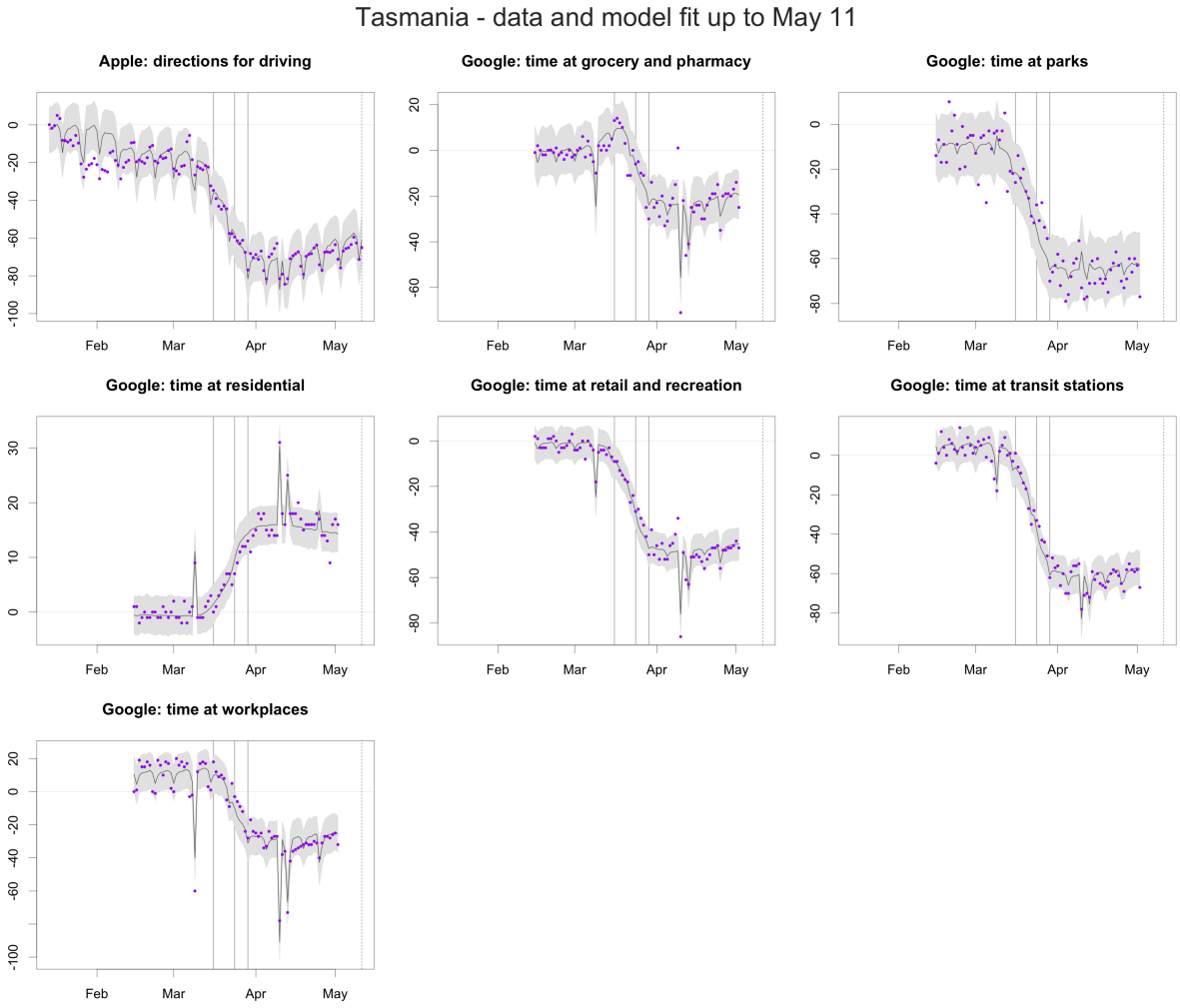


Figure S17: Percentage change compared to a pre-COVID-19 baseline of a number of key mobility data streams in Victoria. Solid vertical lines give the dates of three social distancing measures: restriction of gatherings to 500 people or fewer; closure of bars, restaurants, and cafes; restriction of gatherings to 2 people or fewer. The dashed vertical line marks the most recent date for which some mobility data are available. Blue 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 of the latent behavioural factors driving each data stream.

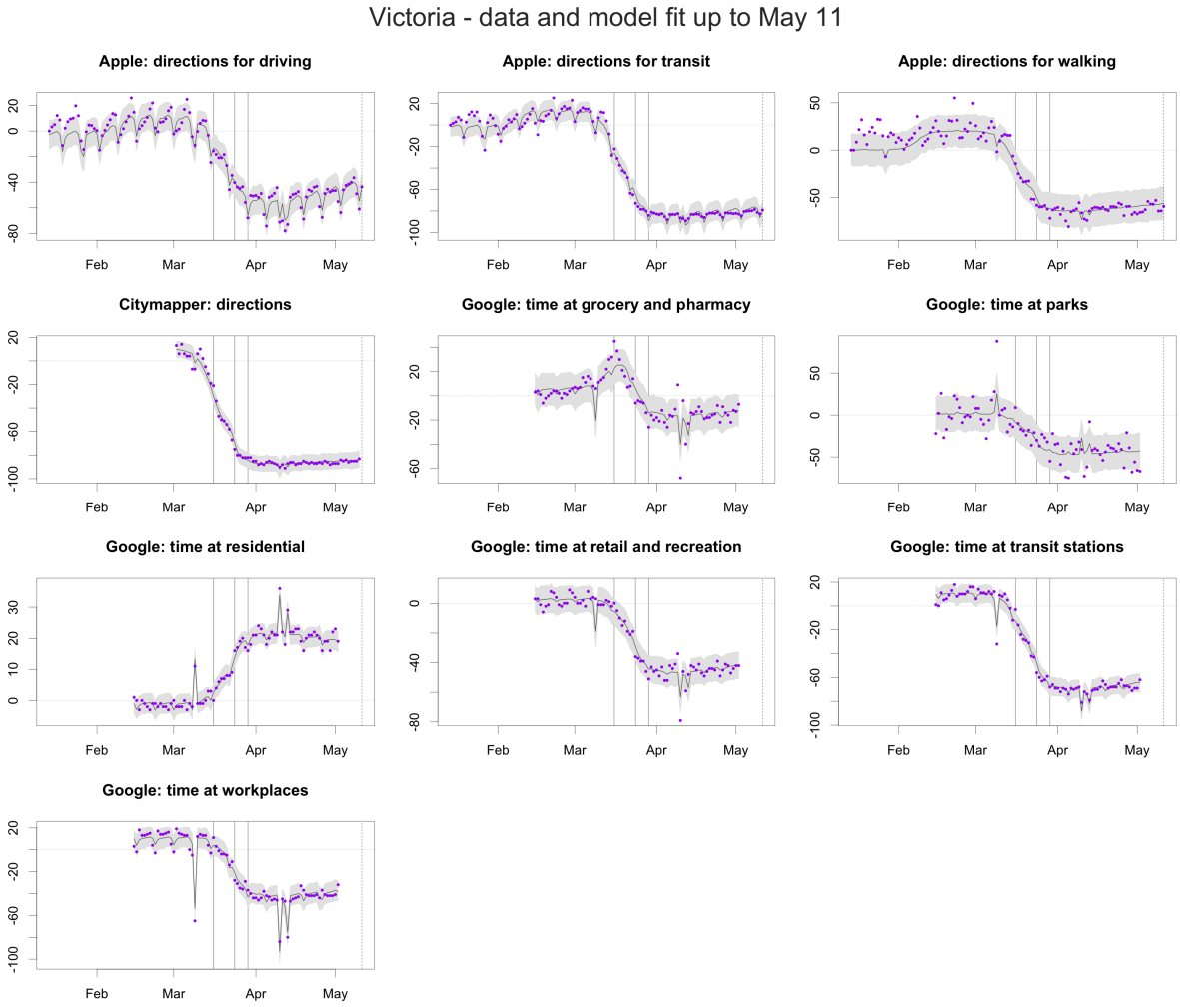


Figure S18: Percentage change compared to a pre-COVID-19 baseline of a number of key mobility data streams in Western Australia. Solid vertical lines give the dates of three social distancing measures: restriction of gatherings to 500 people or fewer; closure of bars, restaurants, and cafes; restriction of gatherings to 2 people or fewer. The dashed vertical line marks the most recent date for which some mobility data are available. Blue 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 of the latent behavioural factors driving each data stream.