PHILOSOPHICAL TRANSACTIONS A

Phil. Trans. R. Soc. A.

doi:10.1098/not yet assigned

Heterogeneity in the onwards transmission risk between local and imported cases affects practical estimates of the time-dependent reproduction number

R. Creswell[1,†], D. Augustin[1,†], I. Bouros[1,†], H.J. Farm[1,†], S. Miao[2,†], A. Ahern[2,†], M. Robinson[1], A. Lemenuel-Diot[3], D.J. Gavaghan[1], B. Lambert[1], R.N. Thompson[4,5,*]

1. Department of Computer Science, University of Oxford, Oxford, OX1 3QD, United Kingdom.
2. Mathematical Institute, University of Oxford, Oxford, OX2 6GG, United Kingdom.
3. Roche Pharmaceutical Research and Early Development, Pharmaceutical Sciences, Roche Innovation Center Basel, Basel, CH-4070, Switzerland.
4. Mathematics Institute, University of Warwick, Coventry, CV4 7AL, United Kingdom. ORCID ID: 0000-0001-8545-5212.
5. Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research, University of Warwick, Coventry, CV4 7AL, United Kingdom.

Keywords: mathematical modelling; reproduction number; imported cases; branching processes; COVID-19; SARS-CoV-2

Summary

During infectious disease outbreaks, inference of summary statistics characterising transmission is essential for planning interventions. An important metric is the time-dependent reproduction number (R_t), which represents the expected number of secondary cases generated by each infected individual over the course of their infectious period. The value of R_t varies during an outbreak due to factors such as varying population immunity and changes to interventions, including those that affect individuals' contact networks. While it is possible to estimate a single population-wide R_t, this may belie differences in transmission between subgroups within the population. Here, we explore the effects of this heterogeneity on R_t estimates. Specifically, we consider two groups of infected hosts: those infected outside the local population (imported cases), and those infected locally (local cases). We use a Bayesian approach to estimate R_t, made available for others to use via an online tool, that accounts for differences in the onwards transmission risk from individuals in these groups. Using COVID-19 data from different regions worldwide, we show that different assumptions about the relative transmission risk between imported and local cases affect R_t estimates significantly, with implications for interventions. This

*Author for correspondence: robin.n.thompson@warwick.ac.uk

†These authors contributed equally to this research

emphasises the need to collect data during outbreaks describing heterogeneities in transmission between different infected hosts, and to account for these heterogeneities in methods used to estimate R_t.

Main Text

1. Introduction

Mathematical and computational models have been used during the COVID-19 pandemic to infer changes in transmissibility and to plan public health measures [1–7]. An important metric for assessing the effectiveness of current interventions during outbreaks is the time-dependent reproduction number (R_t – sometimes referred to informally as the "R number"), which represents the expected number of infections generated by someone infected at time t over the course of their infectious period [8–16]. This quantity varies during an outbreak in response to factors affecting transmission such as changes in public health measures, varying population immunity and pathogen evolution. If R_t remains below one, the number of cases each day will decrease; if instead R_t is persistently above one, the outbreak will grow. In the UK, the government has published estimates of R_t throughout the COVID-19 pandemic [17] alongside other values such as estimates of the epidemic growth rate and daily numbers of new cases, hospitalisations and deaths.

Different formal definitions of R_t have been proposed, most notably the instantaneous reproduction number and the case reproduction number [18]. The instantaneous reproduction number represents the expected number of infections generated (over the course of their infectious period) by someone who is infected at time t if transmission conditions do not change in the future (i.e., this quantity is a measure of instantaneous transmissibility). The case reproduction number, on the other hand, reflects the expected number of infections generated by someone who is infected at time t but accounts for changes in transmissibility that occur after time t (e.g. the subsequent introduction of public health measures). The instantaneous reproduction number has been proposed as the most appropriate definition to use for real-time inference, as this quantity reflects current transmissibility and does not require future changes in transmission conditions to be known [11]. For that reason, we use this definition of R_t for our analyses in this manuscript.

A range of methods have been developed for estimating R_t from outbreak data [11,12,19,20]. Two common approaches are the Cori method [8,9] and the Wallinga-Teunis method [21], which involve inferring the value of R_t from disease incidence time series (i.e. time series describing the number of new cases every day) and an estimate of the serial interval distribution (representing the time period between successive cases; specifically, the difference between the symptom onset times of infectors and infectees). Irrespective of the precise approach used to infer R_t, estimates can be updated and tracked as additional data become available during an outbreak.

Recent developments in the theory of R_t estimation include accounting for reporting delays [7] and considering the impacts of temporal changes in the serial interval [22]. Another consideration is the potential for heterogeneity in R_t between different subgroups in the population. The COVID-19 pandemic has highlighted that individuals in different settings (e.g. care homes as opposed to the wider population [23]) or with different characteristics (e.g. different ages [10,24–26] or vaccination statuses [27,28]) face different risks of both becoming infected and transmitting the virus. Shortly before the COVID-19 pandemic, the Cori method was extended to account for differences in the source locations of local and imported cases [9], but with an assumption that the expected numbers of onwards transmissions from local and imported cases are identical. With that assumption, that work illustrated that failing to differentiate between local and imported cases can lead to overestimation of the number of local infections and therefore overestimation of R_t [9].

Apart from their different origins, local and imported cases can differ in other ways. The risk of onwards transmission from an imported case may be different to the risk from a local case [29]. Imported cases may have visited regions with high case numbers and therefore respond more quickly to early signs of disease, isolating as soon as symptoms develop. This effect might be especially pronounced when a pathogen has first arrived in the local host population, when the infection risk may be higher outside the local population than within it. Imported cases may also be subject to increased testing for infection or pre-emptive home quarantine following travel, thereby lowering the risk of onwards transmission [30]. On the other hand, individuals who travel frequently may be likely to have more contacts with others than those who do not, potentially leading to a higher risk of onwards transmission for imported cases. For example, business travellers may participate in large numbers of meetings, thereby coming into contact with many other people. In either situation, an assumption that R_t is identical for both local cases and imported ones, as made previously [9], is not always appropriate.

In principle, the same disease incidence time series can occur with different divisions of the transmission risk between local and imported infectors (Fig 1). This has implications for pathogen control, since a scenario with substantial local transmission requires localised public health measures to disrupt chains of transmission and prevent spread. In contrast, a scenario with high transmission from imported cases may motivate travel restrictions to prevent importations. Here, we modify the Cori method to allow local and imported cases to have unequal risks of generating new infections. We analyse disease incidence time series recorded during the COVID-19 pandemic in different locations. Our main goal is not to provide a novel methodological approach for estimating R_t, but rather to explore as simply as possible the potential consequences on estimates of R_t of failing to account for differences in the onwards transmission risk from local and imported cases. To allow other researchers to repeat our analyses for similar data, we provide an open-source Python software library including a user-friendly web interface (https://sabs-r3-epidemiology.github.io/branchpro). Our research demonstrates the importance of accounting for differences in the transmission risk between imported and local cases. More widely, it indicates that careful consideration of heterogeneity in the transmission risk between population subgroups may be necessary to make robust public health policy decisions.

2. Methods

2.1 Inference of the time-dependent reproduction number

We modify the Cori method for estimating R_t [8,9] to account for differences in the onwards transmission risk between cases that arise locally compared to those originating elsewhere. In the underlying transmission model, new cases occur according to a time-varying branching process in which each local case is assumed to generate R_t new infections on average, and each imported case is expected to generate εR_t new infections on average, where $\varepsilon \geq 0$ indicates the relative transmission risk from an imported case compared to a local case. Here, we assume that R_t is the instantaneous reproduction number [8,9], representing the expected number of cases that an individual infected at time t is likely to generate over the course of their infection assuming that future pathogen transmissibility is fixed at the current level. Our focus is on estimating the extent of local transmission (i.e. the local time-dependent reproduction number [20]) characterised by R_t. As has been proposed previously [20], the value of R_t therefore reflects the potential for local transmission of the pathogen (rather than being an averaged quantity across both local and imported cases). Here, $\varepsilon < 1$ means that an

imported case is responsible for fewer infections (on average) than a local case, whereas $\varepsilon > 1$ indicates that an imported case generates more infections.

The total number of new cases arising at timestep t can be split according to the sources of infection, $I_t = I_t^{loc} + I_t^{imp}$, where I_t^{loc} represents the number of new cases who were infected within the local population and I_t^{imp} represents the number of new cases who were infected elsewhere. The expected number of local cases at timestep t is then given by

$$\mathbf{E}(I_t^{loc}|\{I_k^{loc}\}_{k=0}^{t-1}, \{I_k^{imp}\}_{k=0}^{t-1}, \varepsilon, R_t, \boldsymbol{w}) = R_t \sum_{s=1}^{t}(I_{t-s}^{loc} + \varepsilon I_{t-s}^{imp})w_s.$$

In this expression, the vector $\boldsymbol{w}$ is the (discrete) serial interval distribution with entries w_s (which characterises the times between successive cases in a chain of transmission; w_1 is the probability that the serial interval is one day, w_2 is the probability that the serial interval is two days, and so on).

We define the transmission potential at timestep t to represent the expected number of local cases arising at timestep t if $R_t = 1$. Thus, the transmission potential at timestep t is given by $\Lambda_t(\boldsymbol{w}, \varepsilon) = \sum_{s=1}^{t}(I_{t-s}^{loc} + \varepsilon I_{t-s}^{imp})w_s$. We assume that the number of local cases in timestep t is drawn from a Poisson distribution with mean $R_t\Lambda_t(\boldsymbol{w}, \varepsilon)$. Hence, the probability of observing the local incidence $\{I_k^{loc}\}_{k=t-\tau}^{t}$ over a time window including $\tau + 1$ days (assuming that R_t is constant during that time window), conditional each day on all previous incidence data, is given by

$$\mathbf{P}(\{I_k^{loc}\}_{k=t-\tau}^{t}\big|\{I_k^{loc}\}_{k=0}^{t-\tau-1}, \{I_k^{imp}\}_{k=0}^{t-1}, \varepsilon, R_t, \boldsymbol{w}\big) = \prod_{k=t-\tau}^{t} \frac{\left(R_t\Lambda_k(\boldsymbol{w},\varepsilon)\right)^{I_k^{loc}} \exp\left(-R_t\Lambda_k(\boldsymbol{w},\varepsilon)\right)}{I_k^{loc}!}.$$

Data describing daily numbers of imported cases enter this expression through $\Lambda_t(\boldsymbol{w}, \varepsilon)$. The model therefore reflects how local cases arise using information about historical numbers of local and imported cases.

Assuming a gamma distributed prior for R_t, the posterior distribution for R_t over the time window $[t - \tau, t]$, conditional on $\boldsymbol{w}$, ε and the observed incidence data (denoted $p(R_t|\boldsymbol{w}, \varepsilon, \boldsymbol{I_{\leq t}})$ – we represent this by $p(.)$ rather than $\mathbf{P}(.)$ since the posterior is a continuous probability density function) is also a gamma distribution due to prior-likelihood conjugacy (see Cori *et al.* [8] and Thompson *et al.* [9] for further details). Specifically,

$$p(R_t|\boldsymbol{w}, \varepsilon, \boldsymbol{I_{\leq t}}) = \text{gamma}\big(R_t, \alpha + \textstyle\sum_{k=0}^{\tau} I_{t-k}^{loc}, \beta + \sum_{k=0}^{\tau}\Lambda_{t-k}(\boldsymbol{w}, \varepsilon)\big),$$

where, for notational convenience, here and above we have combined the disease incidence data into the variable $\boldsymbol{I_{\leq t}} = \{\{I_k^{loc}\}_{k=0}^{t}, \{I_k^{imp}\}_{k=0}^{t-1}\}$. In this expression, the parameters $\alpha > 0$ and $\beta > 0$ are the shape and rate

parameters of the gamma prior distribution for R_t. The function $\text{gamma}(x, a, b)$ corresponds to the probability density function of a gamma distribution with shape parameter a and rate parameter b, so that

$$\text{gamma}(x, a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx).$$

The inferred posterior, $p(R_t|\boldsymbol{w}, \varepsilon, \boldsymbol{I}_{\leq t})$, is based on local infectees appearing in the incidence data in the estimation window $[t-\tau, t]$, infected by local or imported infectors appearing in the incidence data at any time in $[0, t-1]$. Estimates of R_t at successive timesteps are generated by shifting the estimation window by one timestep and repeating the inference procedure. The purpose of this estimation window (rather than estimating R_t based on infectees appearing in the incidence time series on day t alone) is to increase the smoothness of successive R_t estimates, instead of inferring variations in R_t due to the inherent randomness in the epidemiological system (or any other factor affecting the numbers of cases observed each day; for example, daily fluctuations in the proportion of cases that are reported). This comes at the cost of missing changes in transmission occurring at a fine temporal resolution [8].

2.2 Accounting for uncertainty in the serial interval distribution

The approach described above involves estimating R_t using disease incidence time series and an estimate of the serial interval distribution, accounting for differences in both the source location of infection and onwards transmission risk between local and imported cases. However, there is often significant uncertainty in the serial interval distribution. To account for this, we consider a scenario in which there is a set of equally plausible serial interval distributions, $\{\boldsymbol{w}^{(i)}\}_{i=1}^n$. For a single value of i, the entries ${w_s}^{(i)}$ of the vector $\boldsymbol{w}^{(i)}$ correspond to the probability that the serial interval takes the value s days, conditional on $\boldsymbol{w}^{(i)}$ being the true serial interval distribution.

In our analyses of COVID-19 data, we use a set of equally plausible serial intervals, $\{\boldsymbol{w}^{(i)}\}_{i=1}^n$, obtained from a previous study (see below). To account for this uncertainty in the serial interval distribution when estimating R_t, we first estimate R_t separately for each plausible serial interval distribution, $\boldsymbol{w}^{(i)}$, giving the conditional posterior distribution $p(R_t|\boldsymbol{w}^{(i)}, \varepsilon, \boldsymbol{I}_{\leq t})$. We then combine these estimates to give a posterior distribution for R_t accounting for this uncertainty by calculating

$$p(R_t|\varepsilon, \boldsymbol{I}_{<t}) = \frac{1}{n}\sum_{i=1}^{n} p(R_t|\boldsymbol{w}^{(i)}, \varepsilon, \boldsymbol{I}_{\leq t}).$$

2.3. Data and parameterisation

In our main analyses, we consider five disease incidence time series datasets collected in different locations during the COVID-19 pandemic. The key feature of these datasets is that information is available which allows locally originating cases to be differentiated from those infected elsewhere. The datasets are:

1. Ontario, Canada (Fig 2a – left). Incidence data were obtained for the time period from 1 March – 20 April 2020 [31]. Cases were classified as imported if they reported travelling outside Ontario within 14 days of symptom onset. Cases with unknown recent travel status were assumed to have been infected locally.
2. New South Wales, Australia (Fig 2a – middle). Incidence data were obtained for the time period from 1 March – 13 April 2020. Cases were classified as imported if they were reported as "overseas acquired" in the Australian national COVID-19 database (see [30] for further details). Cases with unknown origin were assumed to have been infected locally.
3. Victoria, Australia (Fig 2a – right). Details as above for New South Wales.
4. Hong Kong (Fig 4a – left). Incidence data were obtained for the time period from 23 January – 24 March 2020 [32]. Cases were classified as imported if they were listed as "imported case, confirmed" in the Hong Kong Department of Health COVID-19 database (see [33] for further details). All other cases were classified as local cases.
5. Hainan Province, China (Fig 4a – right). Incidence data were obtained for the time period from 22 January – 20 February 2020 [34]. Cases were classified as imported if they either reported travel outside Hainan Province in the 14 days prior to symptom onset or reported any recent travel to a known COVID-19 outbreak area. All other cases were classified as local cases.

We chose to analyse the first three datasets in the main text due to their differing outbreak trajectories in the time periods considered. Specifically, the Ontario dataset represents a growing outbreak, the New South Wales dataset represents a full outbreak wave with a large number of imported cases compared to local cases, and the Victoria dataset represents a full outbreak wave with more local cases than imported ones. We chose to analyse the fourth and fifth datasets because further information was available from those locations with which

it was possible to approximate the value of ε. This allowed us to demonstrate inference of R_t in scenarios in which the relative transmission risk from imported and local cases is known.

In addition to our analyses in the main text, we considered datasets from other locations and display similar analyses in the Supplementary Information (Supplementary Figures 1-6); specifically, we considered COVID-19 disease incidence time series datasets from five other Australian states, New Zealand and Hawaii, and we considered a disease incidence time series dataset for MERS in Saudi Arabia in 2014-15. The key feature of all these datasets is that information was available with which to classify cases as either local or imported. In the analysis of MERS in Saudi Arabia, imported cases were not those who had arrived from a geographically distinct location. Instead, in that analysis, imported cases were those who were likely to have been infected directly from the animal reservoir.

For the serial interval in all our analyses of COVID-19 incidence datasets, we considered an estimate for SARS-CoV-2 obtained by Nishiura *et al.* [35]. Specifically, those authors fitted a lognormal distribution to data from known infector-infectee transmission pairs using Markov chain Monte Carlo (MCMC), thereby obtaining a set of equally plausible possible serial interval distributions. We considered the set of serial interval distributions obtained by Nishiura *et al.* [35] using both certain and probable infector-infectee pairs while accounting for right-truncation (i.e. the possibility that a dataset detailing infector-infectee pairs observed when the outbreak is growing excludes some transmissions with longer serial intervals that have not yet occurred). For our inference procedure, we used $n = 1{,}000$ randomly selected MCMC draws from their analysis, where each draw characterises a continuous distribution. Since our approach considers the number of new cases each day, we require a discrete serial interval distribution. We therefore "discretised" the continuous distributions into daily timesteps using the method described by Cori *et al.* [8] (see web appendix 11 of that article). The set of $n = 1{,}000$ serial interval distributions used in our analysis (i.e. $\{\boldsymbol{w}^{(i)}\}_{i=1}^{n}$) is shown in Supplementary Figure 7.

We fixed the parameters of the gamma distributed prior for R_t so that both the mean and standard deviation were equal to five (to do this, we chose $\alpha = 1$ and $\beta = 0.2$). The rationale for this choice is that a large standard deviation ensures that the prior is relatively uninformative, while a high mean ensures that the outbreak is unlikely to be determined as under control ($R_t < 1$) unless there is substantial evidence from the data supporting this conclusion. In all of our analyses of COVID-19 incidence data, R_t was estimated using a weekly sliding window, so that $\tau = 6$ days. In the figures, the posterior distribution for R_t shown on day t is based on a sliding window that ends on day t (i.e. the sliding window $[t - \tau, t]$).

2.4 Correctness and reproducibility of results

We followed a range of software development practices to guard against coding errors and to ensure code reusability: these included collaborative coding using Github to manage merging of code via pull requests, unit testing of functions and classes (with 100% test coverage) and continuous integration testing. To ensure reproducibility of results, all analyses for this paper can be rerun by cloning our Github repository (https://github.com/SABS-R3-Epidemiology/transmission-heterogeneity-results) and executed via a single command from the terminal.

3. Results

3.1 Effect of the relative transmission risk on estimates of R_t

To explore how different assumptions about the relative transmission risk from imported and local cases affect R_t estimates, we initially applied our method to data from the first three locations described in Methods (Fig 2). We considered three different assumptions about the relative transmission risk. First, we assumed that imported cases were each expected to generate fewer infections than local cases ($\varepsilon = 0.25$; Fig 2b – blue). Second, we assumed instead that imported cases were each expected to generate more infections than local cases ($\varepsilon = 2$; Fig 2b – red). Third, we made the standard assumption [9] that the transmission risk from each local case was identical to the transmission risk from each imported case ($\varepsilon = 1$; Fig 2b – black). These analyses highlight that different assumed values of ε lead to different inferred R_t values. As might be expected, assuming a larger value of ε leads to smaller estimated values of R_t, since more transmission is then attributed to imported cases rather than local cases.

We then went on to consider the implications for public health policy of differences in the relative transmissibility of imported and local cases. For the dataset from Ontario (Fig 2a – left), the numbers of local cases broadly increased throughout the time period considered. A key question in that setting is "Is $R_t > 1$?", since this determines whether sustained local transmission will occur. If so, fast detection that $R_t > 1$ is crucial

to allow interventions to be introduced quickly to prevent further exponential growth of the outbreak. In the left panel of Fig 3a, posterior mean estimates of R_t each day from 8 March – 20 April 2020 are shown for a range of values of ε. The first date on which the mean estimate of R_t is above one and remains above one thereafter is shown for different values of ε in the left panel of Fig 3b (grey). This indicates that a smaller assumed value of ε leads to an earlier conclusion that R_t is greater than one for this dataset. The proportion of the period considered for which the mean R_t estimate is above one also depends on the assumed value of ε (Fig 3c – left).

While a policy-maker may choose to strengthen control measures when the mean estimate of R_t increases above one, a more risk averse choice could be to conclude that the outbreak is not under control if an upper percentile of the posterior distribution of R_t exceeds one. For example, for the Ontario dataset, when $\varepsilon = 1.2$, the mean estimate of R_t is (and remains) above one from 11 April 2020 onwards (Fig 3b – left, grey), whereas the 97.5th percentile estimate of R_t remains above one from the earlier date of 23 March 2020 onwards (Fig 3b – left, green dashed). By using an approach like the one described here, policy-makers can adjust their decision making according to their chosen level of risk aversion. This simply involves specifying the percentile value of R_t to track to guide decision making regarding strengthening and relaxing public health measures.

During the COVID-19 pandemic, public health measures have been relaxed in many regions and countries when the outbreak has been assessed as being under control. We therefore considered the incidence dataset from New South Wales and estimated when policy-makers could conclude that R_t had fallen below one (Fig 3b – middle). In this scenario, a larger assumed value of ε led to an earlier date on which R_t was assessed to be below one (and remained below one thereafter). In order for policy-makers to be more certain that R_t is below one when relaxing restrictions, one possibility is to conclude that R_t is below one when a high percentile value of the posterior for R_t has fallen below one. For example, for this dataset, if the mean estimate of R_t is considered and the value $\varepsilon = 1.2$ is assumed, then R_t is inferred to fall and remain below one on 15 March 2020 (Fig 3b – middle, grey), whereas if instead the 97.5th percentile estimate of R_t is considered, then R_t is inferred to fall below one on the later date of 19 March 2020 (Fig 3b – middle, green dashed).

As the final component of these analyses, we considered the disease incidence time series dataset from Victoria and repeated the analysis that we conducted for the dataset from New South Wales. We found that, if a high value of ε is assumed, then the outbreak is inferred to be under control ($R_t < 1$) for the majority of the time period under consideration (Fig 3c, right). However, if instead the value of ε is lower, then R_t may be

estimated to be greater than one early in the outbreak. For small values of ε, so that initial estimated values of R_t are high, the most policy-relevant question may again be to determine when R_t has fallen below one (Fig 3b – right).

3.2 Realistic values of the relative transmission risk

In the analyses presented in Section 3.1, we demonstrated clearly that the assumed relative transmission risk between imported and local cases affects R_t estimates, impacting policy-relevant conclusions drawn from disease incidence time series data. The relative transmission risk may differ between settings. In some scenarios, it may be possible to inform estimates of ε with real-world data. Here we provide two examples, in the context of SARS-CoV-2 transmission in Hong Kong and Hainan Province (the fourth and fifth disease incidence time series datasets described in Methods). Additional possible approaches for estimating the value of ε are described in the Discussion.

First, we considered the dataset from Hong Kong (Fig 4a – left). A previous study [33] reconstructed the transmission network of cases in that region (between 23 January 2020 and 8 January 2021; although in principle a similar analysis could be conducted at a smaller spatial scale for shorter time periods, as would likely be most useful for early real-time estimation of R_t), inferring the "outdegree" of imported and local cases. Based on the aggregated data shown in Table 1 of that study, the mean outdegree was 0.74 for imported cases and 3.68 for local cases, which corresponds to a value of $\varepsilon = 0.2$. We therefore compared estimated values of R_t for $\varepsilon = 0.2$ (Fig 4b – left, green) with analogous estimates under the standard assumption that $\varepsilon = 1$ (Fig 4b – left, grey). Since a value of $\varepsilon = 0.2$ leads to less transmission being attributed to imported infections than when $\varepsilon = 1$, estimated values of R_t are higher when $\varepsilon = 0.2$. In terms of decision making during an ongoing outbreak, time periods when the mean estimated value of R_t is greater than one for $\varepsilon = 0.2$ and less than one for $\varepsilon = 1$ may be particularly concerning. In these periods, the outbreak might erroneously be inferred as being under control if the incorrect assumption that $\varepsilon = 1$ is made. In the analysis shown in the left panel of Fig 4b, this is the case for 20.8% of the time period considered. Of course, similarly to Section 3.1, analogous analyses could be performed based on different percentile estimates of R_t rather than the mean estimated value.

Second, we considered the dataset from Hainan Province, China. A previous study [34] compared the epidemiological features of imported and local cases in that province, and found that imported cases tended to belong to older age groups than local cases. We applied a contact matrix for China [36] to the age distributions of imported and local cases, and thus estimated the expected number of contacts per day for imported cases (10.5) and local cases (13.4). To approximate the value of ε, we divided the expected number of contacts per day for imported cases by the analogous value for local cases, giving $\varepsilon = 0.785$. We then compared estimates of R_t for that more realistic value of $\varepsilon = 0.785$ (Fig 4b – right, green) to estimates of R_t under the standard assumption that $\varepsilon = 1$ (Fig 4b – right, grey). Since a value of $\varepsilon = 1$ is only slightly larger than $\varepsilon = 0.785$, and the data from Hainan Province suggest only limited local transmission, we found that incorrectly assuming that $\varepsilon = 1$ did not have a substantial effect on inferred R_t values for this dataset.

4. Discussion

Summary statistics for tracking pathogen transmissibility are increasingly used during infectious disease outbreaks to guide decision making. Throughout the COVID-19 pandemic, R_t has been estimated in regions and countries worldwide (see e.g. [7]). This metric is useful and straightforward to interpret, corresponding to the number of individuals that one infected host is expected (on average) to go on to infect. As well as providing information about whether an outbreak is growing or declining, the value of R_t can be used to determine the proportion of transmissions that must be prevented for a growing outbreak to decline.

In this article, we have presented a modified version of the commonly used Cori method for inferring R_t [8,9]. We have accounted for different transmission risks from local and imported cases, rather than assuming that the transmission risk is identical for individuals in these groups. We provide an accompanying online software tool for estimating R_t (https://sabs-r3-epidemiology.github.io/branchpro) where users can upload their own data (disease incidence time series and an estimate of the serial interval distribution – or multiple equally plausible serial interval distributions as described in Methods). We have conducted a systematic analysis of the dependence of inferred R_t values on the assumed relative transmission risk from an imported case compared to a local case (ε; see Figs 2 and 3). We also considered examples in which it was possible to approximate the value of ε from other data sources (Fig 4). In general, larger assumed values of ε lead to smaller R_t estimates. This is important, since assuming an unrealistically high value of ε may lead to the outbreak being falsely determined as under control. When an outbreak is ongoing, we have shown that the speed at which local

transmission can be inferred as being either under control or not depends on the assumed value of ε. This dependence on ε demonstrates clearly that whether or not an outbreak is under control cannot always be inferred accurately from summary statistics that do not account for differences in the transmission risk between imported and local cases (e.g., the growth rate of overall cases). We have also shown how different percentile estimates of R_t can be used to guide decision making, according to the policy-maker's level of acceptable risk.

A previous approach for estimating R_t allows infectees to have been infected either within or outside the local population [9]. However, in that framework, an assumption is made that the transmission risk from a local case is identical to the analogous risk from an imported case. The potential for different transmission risks from imported and local cases has implications for optimising interventions, since if the risk of transmission is predominantly from imported cases, then travel restrictions and interventions that prevent transmissions from imported cases (e.g. quarantine of incoming travellers) may be the optimal interventions. If instead the transmission risk is highest from local cases, then interventions such as social distancing and face coverings that reduce transmission from all infected individuals in the population may be necessary. In scenarios in which a novel pathogen variant is being imported into a new location from somewhere it is already widespread, the composition of variants causing local and imported cases might affect the relative transmission risk [37], although we note that in our modelling framework it is only the imported cases themselves that are assumed to represent a different transmission risk (rather than all infected individuals in a chain of transmission starting with an imported case).

A recent, closely related study by Tsang *et al.* [29] involved estimating independent R_t values for local and imported cases throughout an outbreak. A benefit of that approach is that it does not require an assumption to be made about the relative transmission risk from imported compared to local cases. However, there are substantial logistical challenges to estimating independent R_t values for local and imported cases: this requires local cases who were infected by other local cases to be distinguished from those who were infected by imported cases. This may be possible either on a small scale or in locations with extensive contact tracing [29,38], but, in many situations, it is infeasible. In the absence of data with which to estimate R_t for local and imported cases independently, and without known changes in the relative transmission risk from imported compared to local cases, then assuming a constant relative transmission risk between the two types of case as we have done seems reasonable. To obtain an idea about whether the relative transmission risk (i.e., the

parameter ε in our model) is likely to be less than or greater than one, we considered examples in which we approximated ε using a reconstructed transmission network and based on the age characteristics of local and imported cases (Fig 4). In both examples that we considered, the estimated value of ε was less than one, suggesting a lower transmission risk from imported cases than from local cases. Other approaches for inferring ε are also possible. One way to estimate ε is to analyse data containing both local and imported cases in small-scale settings in which infector-infectee transmission pairs can be identified or estimated, such as household or contact tracing studies. Another option might be to perform forwards contact tracing on imported cases at a single stage of the outbreak. If the value of εR_t can be estimated from the contact tracing data at that stage, ε could then be estimated from the population-level incidence data. The contribution of imported cases to transmission is likely to vary by the time in the outbreak and by location [30,39]. In principle, estimates of ε could be updated based on the latest available contact tracing data.

In our main analyses, we have considered scenarios in which imported cases are individuals who have been infected in other geographical locations. However, an imported case may be defined as any case with an infection source outside the local host population. In the Supplementary Information, we consider an analysis of MERS cases in Saudi Arabia in 2014-15 (Supplementary Figures 5-6), where cases are likely to have arisen both via human-to-human transmission and from an animal reservoir (specifically, from dromedary camels [40]). In that analysis, imported cases are assumed to be those reporting regular contacts with camels. It is possible that those individuals typically live in less densely populated areas than individuals who do not have regular contacts with camels, meaning that the relative risk of an imported case transmitting the virus is lower on average than the analogous risk from a local case. Like our analyses of COVID-19 datasets, our analysis of the MERS incidence data illustrates that assumptions about the relative transmission risk between local and imported cases can affect estimates of R_t and conclusions about whether or not local human-to-human transmission is under control.

We also conducted an additional supplementary analysis in which we generated synthetic epidemic datasets and investigated further the conditions under which mischaracterising the relative transmissibility of imported and local cases affects estimates of R_t substantially. Specifically, we generated synthetic data for different values of ε and different strengths of local transmission. We calculated the error in estimates of R_t if the standard assumption that $\varepsilon = 1$ is made (Supplementary Figure 8). This suggests that the largest errors occur when the relative transmissibility of imported (compared to local) cases differs substantially from one, and when imported cases represent a high proportion of the overall cases observed in the population.

In the research that we have presented, we sought to explore the relationship between heterogeneities in the onwards transmission risk between different groups of infectious individuals and inferred values of R_t. Practical applications of this approach should consider incorporating additional features into the modelling framework. An important consideration when assessing pathogen transmissibility during outbreaks is that R_t represents the average number of onwards infections over multiple infected individuals and transmission events. However, different infected individuals may generate very different numbers of infections [10,41–43]. The potential for super-spreading events at which large numbers of infections occur could be built into the underlying transmission model and into the resulting R_t estimates, although it may then be impossible to generate an analytic expression for the posterior for R_t. We sought to demonstrate the general principle that population heterogeneity can affect estimates of R_t. To do this as simply as possible, we used a model with only two different groups of infected hosts (i.e., local and imported cases). However, many different sources of heterogeneity exist within host populations. For example, there may be substantial differences in the transmission risk between other subgroups of the population: for example, risk may vary by age [10,24,25] and vaccination status [27]. Geographically distinct populations could be linked in a transmission model, so that spatial heterogeneity in R_t can be explored. In principle, compartmental models can be developed in which a range of different sources of heterogeneity are included, and R_t may be estimated using those compartmental models. It might also be possible to include further sources of heterogeneity in a renewal equation framework as studied here. These possibilities represent interesting avenues for future research.

Here, we assumed that the data represent disease incidence time series, and that the serial interval (the time between successive symptomatic cases in a transmission chain) is always positive. In reality, pre-symptomatic infections occur, and serial intervals may take negative values [44–46] with infectors developing symptoms after some of the individuals who they infect. While the assumption of a positive valued serial interval distribution has been made in many previous studies in which R_t has been estimated for different pathogens, this issue can be avoided by using the incidence of infections and the generation time distribution [44,46,47] rather than the incidence of cases and the serial interval distribution [11]. The subtle difference here is that incidence time series of cases do not reflect the times at which individuals were first infected, but instead reflect the times at which individuals were recorded as infected (which occurs after infection, for example when individuals display symptoms). Use of the incidence of infections and the generation time distribution may require the incidence of infections to be inferred from the incidence of cases, for example using an assumed

incubation period distribution and the Richardson-Lucy deconvolution technique [48]. We note that the serial interval distribution may be different to the generation time distribution (specifically, pre-symptomatic transmission can lead to shorter serial intervals than generation times [46]). Another potential extension to our research is incorporation of different serial interval (or generation time) distributions for local and imported cases [38], particularly given that part of an imported case's infectious period may occur before they enter the local population. Reconstructed transmission networks might provide insights into these distributions.

More broadly, we note that R_t is only one summary statistic for tracking changes in transmission during an infectious disease outbreak. This metric does not provide information about the speed of the outbreak, which is better measured by the growth rate of cases [49,50]. Furthermore, current incidence of cases, hospitalisations and deaths are also key inputs to policy decisions. For example, an outbreak with R_t close to one is likely to have more detrimental impacts if case numbers are high compared to if case numbers are low. Nonetheless, R_t has been useful for guiding interventions during the COVID-19 pandemic, in combination with these other statistics. We therefore contend that studies that improve understanding of the impacts of factors affecting R_t estimates, such as heterogeneity in the onwards transmission risk between different infectious hosts, are valuable and an important component of preparedness for future outbreaks.

Additional Information

Data Accessibility

The user-friendly web interface for estimating R_t while accounting for different transmission risks from local and imported cases can be found at https://sabs-r3-epidemiology.github.io/branchpro. All data and computing scripts required to reproduce the results presented here are available at https://github.com/SABS-R3-Epidemiology/transmission-heterogeneity-results. The source code of the branchpro Python package, which we developed to perform the inference presented in this article, is available at https://github.com/SABS-R3-Epidemiology/branchpro. No restrictions exist on data availability.

Authors' Contributions

RNT conceived of and designed the study; RC, DA, IB, HJF, SM and AA performed the research; MR, AL-D, DJG, BL and RNT supervised the research; RNT wrote the original draft; All authors revised the manuscript; All authors read and approved the manuscript.

Competing Interests

The authors declare that they have no competing interests.

Funding Statement

This research was funded by the EPSRC via a doctoral training partnership studentship in the Department of Computer Science at the University of Oxford (RC), by the EPSRC via the CDT in Sustainable Approaches to

Biomedical Science: Responsible and Reproducible Research – SABS:R^3 grant EP/S024093/1 (IB, HJF, SM, AA and MR), by the Clarendon Fund (DA) and by the UKRI via grant EP/V053507/1 (RNT). AL-D was funded as a Roche Pharmaceutical Research employee. The funders had no role in study design, data analysis, preparation of the manuscript or the decision to publish.

References

1. Brooks-Pollock E, Danon L, Jombart T, Pellis L. Modelling that shaped the early COVID-19 pandemic response in the UK. Philos Trans R Soc B. 2021;376: 20210001.

2. Thompson RN. Epidemiological models are important tools for guiding COVID-19 interventions. BMC Med. 2020;18: 152.

3. Davies NG, Kucharski AJ, Eggo RM, Gimma A, CMMID COVID-19 working group, Edmunds WJ. The effect of non-pharmaceutical interventions on COVID-19 cases, deaths and demand for hospital services in the UK: a modelling study. Lancet Public Health. 2020;5: e375–e385.

4. Moore S, Hill EM, Dyson L, Tildesley MJ, Keeling MJ. Modelling optimal vaccination strategy for SARS-CoV-2 in the UK. PLOS Comput Biol. 2020;17: e1008849.

5. Dehning J, Zierenberg J, Spitzner EP, Wibral M, Neto JP, Wilczek M, et al. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. Science. 2020; eabb9789.

6. Birrell P, Blake J, Van Leeuwen E, Gent N, De Angelis D. Real-time nowcasting and forecasting of COVID-19 dynamics in England: the first wave. Philos Trans R Soc B. 2020;376: 20200279.

7. Abbott S, Hellewell J, Thompson RN, Sherratt K, Gibbs HP, Bosse NI, et al. Estimating the time-varying reproduction number of SARS-CoV-2 using national and subnational case counts. Wellcome Open Res. 2020;5: 112.

8. Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. Am J Epidemiol. 2013;178: 1505–12.

9. Thompson RN, Stockwin JE, van Gaalen RD, Polonsky JA, Kamvar ZN, Demarsh PA, et al. Improved inference of time-varying reproduction numbers during infectious disease outbreaks. Epidemics. 2019;29: 100356.

10. Thompson RN, Hollingsworth TD, Isham V, Arribas-Bel D, Ashby B, Britton T, et al. Key questions for modelling COVID-19 exit strategies. Proc R Soc B Biol Sci. 2020;287: 20201405.

11. Gostic KM, McGough L, Baskerville E, Abbott S, Joshi K, Tedijanto C, et al. Practical considerations for measuring the effective reproductive number, Rt. PLoS Comput Biol. 2020.

12. White LF, Moser CB, Thompson RN, Pagano M. Statistical estimation of the reproductive number from case notification data. Am J Epidemiol. 2020; kwaa211.

13. Nishiura H, Chowell G. The effective reproduction number as a prelude to statistical estimation of time-dependent epidemic trends. Mathematical and Statistical Estimation Approaches in Epidemiology. 2009. pp. 103–121.

14. Cowling BJ, Lau MSY, Ho LM, Chuang SK, Tsang T, Liu SH, et al. The effective reproduction number of pandemic influenza: Prospective estimation. Epidemiology. 2010;21: 842–846.

15. Cauchemez S, Boëlle PY, Donnelly CA, Ferguson NM, Thomas G, Leung GM, et al. Real-time estimates in early detection of SARS. Emerg Infect Dis. 2006;12: 110–113.

16. Cheng Q, Liu Z, Cheng G, Huang J. Heterogeneity and effectiveness analysis of COVID-19 prevention and control in major cities in China through time-varying reproduction number estimation. Sci Rep. 2020;10: 21953.

17. UK Government. The R value and growth rate. 2021.

18. Fraser C. Estimating individual and household reproduction numbers in an emerging epidemic. PLoS ONE. 2007;2: e758.

19. Obadia T, Haneef R, Boëlle PY. The R0 package: A toolbox to estimate reproduction numbers for epidemic outbreaks. BMC Med Inform Decis Mak. 2012;12: 147.

20. Vegvari C, Abbott S, Ball F, Brooks-Pollock E, Challen R, Collyer BS, et al. Commentary on the use of the reproduction number R during the COVID-19 pandemic. Stat Methods Med Res. 2021.

21. Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. Am J Epidemiol. 2004;160: 509–516.

22. Ali ST, Wang L, Lau EHY, Xu XK, Du Z, Wu Y, et al. Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions. Science. 2020;369: 1106–1109.

23. Ladhani SN, Chow JY, Janarthanan R, Fok J, Crawley-Boevey E, Vusirikala A, et al. Increased risk of SARS-CoV-2 infection in staff working across different care homes: enhanced COVID-19 outbreak investigations in London care homes. J Infect. 2020;81: 621–624.

24. Davies NG, Klepac P, Liu Y, Prem K, Jit M, CMMID COVID-19 Working Group, et al. Age-dependent effects in the transmission and control of COVID-19 epidemics. Nat Med. 2020.

25. Keeling MJ, Hill EM, Gorsich EE, Penman B, Guyver-Fletcher G, Holmes A, et al. Predictions of COVID-19 dynamics in the UK: Short-term forecasting and analysis of potential exit strategies. PLoS Comput Biol. 2021;17.

26. Lovell-Read FA, Shen S, Thompson RN. Estimating local outbreak risks and the effects of non-pharmaceutical interventions in age-structured populations: SARS-CoV-2 as a case study. J Theor Biol. 2022;535: 110983.

27. Keeling MJ, Dyson L, Hill E, Moore S, Tildesley MJ. Road map scenarios and sensitivity: Step 4. 2021 [cited 4 Aug 2021]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/993358/s1288_Warwick_RoadMap_Step_4.pdf

28. Sachak-Patwa R, Byrne HM, Dyson L, Thompson RN. The risk of SARS-CoV-2 outbreaks in low prevalence settings following the removal of travel restrictions. Commun Med. 2021;1: 39.

29. Tsang TK, Wu P, Lau EHY, Cowling BJ. Accounting for imported cases in estimating the time-varying reproductive number of COVID-19 in Hong Kong. J Infect Dis. 2021.

30. Price DJ, Shearer FM, Meehan MT, McBryde E, Moss R, Golding N, et al. Early analysis of the Australian COVID-19 epidemic. eLife. 2020;9: e58785.

554 31. Government of Ontario. COVID-19 data: Likely source of infection. 2021.

555 32. Hong Kong Department of Health. Latest local situation of COVID-19. Available: https://data.gov.hk/en-
556 data/dataset/hk-dh-chpsebcddr-novel-infectious-agent/resource/ec4b49af-83e0-4c71-a3ba-14120e453b9d

557 33. Liu Y, Gu Z, Liu J. Uncovering transmission patterns of COVID-19 outbreaks: A region-wide comprehensive
558 retrospective study in Hong Kong. EClinicalMedicine. 2021;36: 100929.

559 34. Wu B, Lei Z-Y, Wu K-L, He J-R, Cao H-J, Fu J, et al. Compare the epidemiological and clinical features of imported and
560 local COVID-19 cases in Hainan, China. Infect Dis Poverty. 2020;9: 143.

561 35. Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. Int J Infect Dis.
562 2020;93: 284–286.

563 36. Prem K, Cook AR, Jit M. Projecting social contact matrices in 152 countries using contact surveys and demographic
564 data. PLoS Comput Biol. 2017;13: e1005697.

565 37. Challen R, Dyson L, Overton CE, Guzman-Rincon LM, Hill EM, Stage HB, et al. Early epidemiological signatures of novel
566 SARS-CoV-2 variants: establishment of B.1.617.2 in England. medRxiv. 2021.

567 38. Li M, Liu K, Song Y, Wang M, Wu J. Serial interval and generation interval for imported and local infectors, respectively,
568 estimated using reported contact-tracing data of COVID-19 in China. Front Public Health. 2021;8: 577431.

569 39. Russell TW, Wu JT, Clifford S, Edmunds WJ, Kucharski AJ, Jit M, et al. Effect of internationally imported cases on
570 internal spread of COVID-19: a mathematical modelling study. Lancet Public Health. 2021;6: e12-20.

571 40. Haagmans BL, Al Dhahiry SHS, Reusken CBEM, Raj VS, Galiano M, Myers R, et al. Middle East respiratory syndrome
572 coronavirus in dromedary camels: An outbreak investigation. Lancet Infect Dis. 2014;14: 140–145.

573 41. Endo A, CMMID COVID-19 working group, Abbott S, Kucharski AJ, Funk S. Estimating the overdispersion in COVID-19
574 transmission using outbreak sizes outside China. Wellcome Open Res. 2020;5: 67.

575 42. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease
576 emergence. Nature. 2005;438: 355–359.

577 43. Akhmetzhanov AR, Jung S-M, Cheng H-Y, Thompson RN. A hospital-related outbreak of SARS-CoV-2 associated with
578 variant Epsilon (B.1.429) in Taiwan: transmission potential and outbreak containment under intensified contact
579 tracing, January–February 2021. Int J Infect Dis. 2021;110: 15–20.

580 44. Hart WS, Maini PK, Thompson RN. High infectiousness immediately before COVID-19 symptom onset highlights the
581 importance of continued contact tracing. eLife. 2021;10: e65534.

582 45. Du Z, Xu X, Wu Y, Wang L, Cowling BJ, Meyers LA. Serial interval of COVID-19 among publicly reported confirmed
583 cases. Emerg Infect Dis. 2020;26: 1341–1343.

584 46. Hart WS, Miller E, Andrews NJ, Waight P, Maini PK, Funk S, et al. Generation time of the alpha and delta SARS-CoV-2
585 variants: an epidemiological analysis. Lancet Infect Dis. 2022; S1473309922000019.

586 47. Hart WS, Abbott S, Endo A, Hellewell J, Miller E, Andrews N, et al. Inference of the SARS-CoV-2 generation time using
587 UK household data. eLife. 2022;11: e70767.

588 48. Goldstein E, Dushoff J, Ma J, Plotkin JB, Earn DJD, Lipsitch M. Reconstructing influenza incidence by deconvolution of
589 daily mortality time series. PNAS. 2009;106: 21829.

590 49. Dushoff J, Park SW. Speed and strength of an epidemic intervention. Proc R Soc B Biol Sci. 2021;288: 20201556.

591 50. Pellis L, Scarabel F, Stage HB, Overton CE, Chappell LH, Fearon E, et al. Challenges in control of Covid-19: short
592 doubling time and long delay to effect of interventions. Philos Trans R Soc B. 2021;376: 20200264.

593
594

Figures

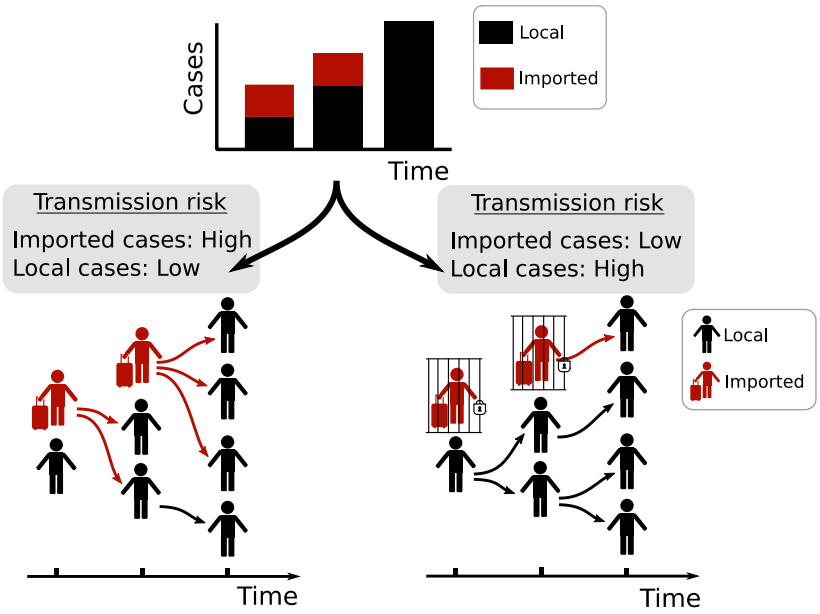


Figure 1. A disease incidence time series dataset can be generated by different combinations of transmission risks from imported and local cases. In the first scenario (bottom left), observed cases are mostly due to infections by imported cases, whereas in the second scenario (bottom right) observed cases are mostly due to infections by local cases. In the bottom panels, red arrows represent infections generated by imported cases and black arrows represent infections generated by local cases. An individual who is infected by an imported case is classified as a local case, since they have themselves been infected locally. Despite the same overall incidence, the two scenarios shown correspond to different risks of sustained local transmission (the risk of sustained local transmission is higher in the second scenario – bottom right), with implications for public health measures.

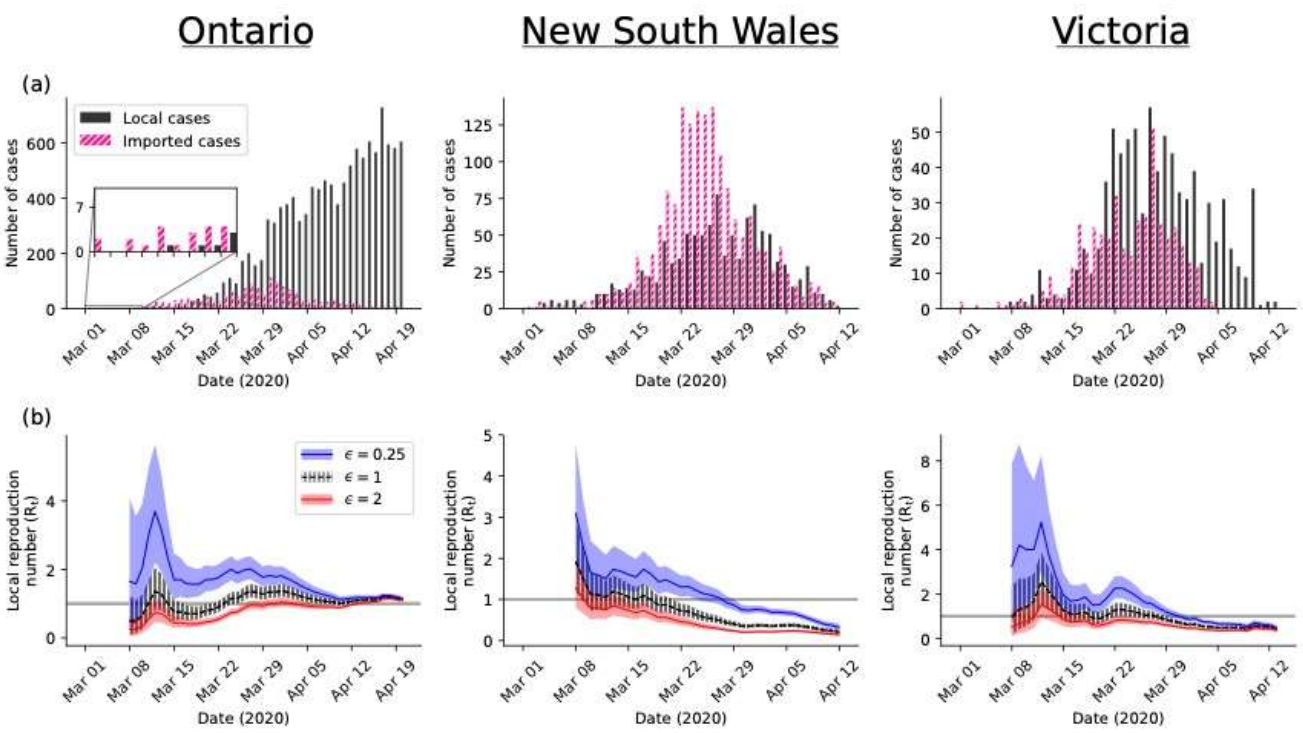
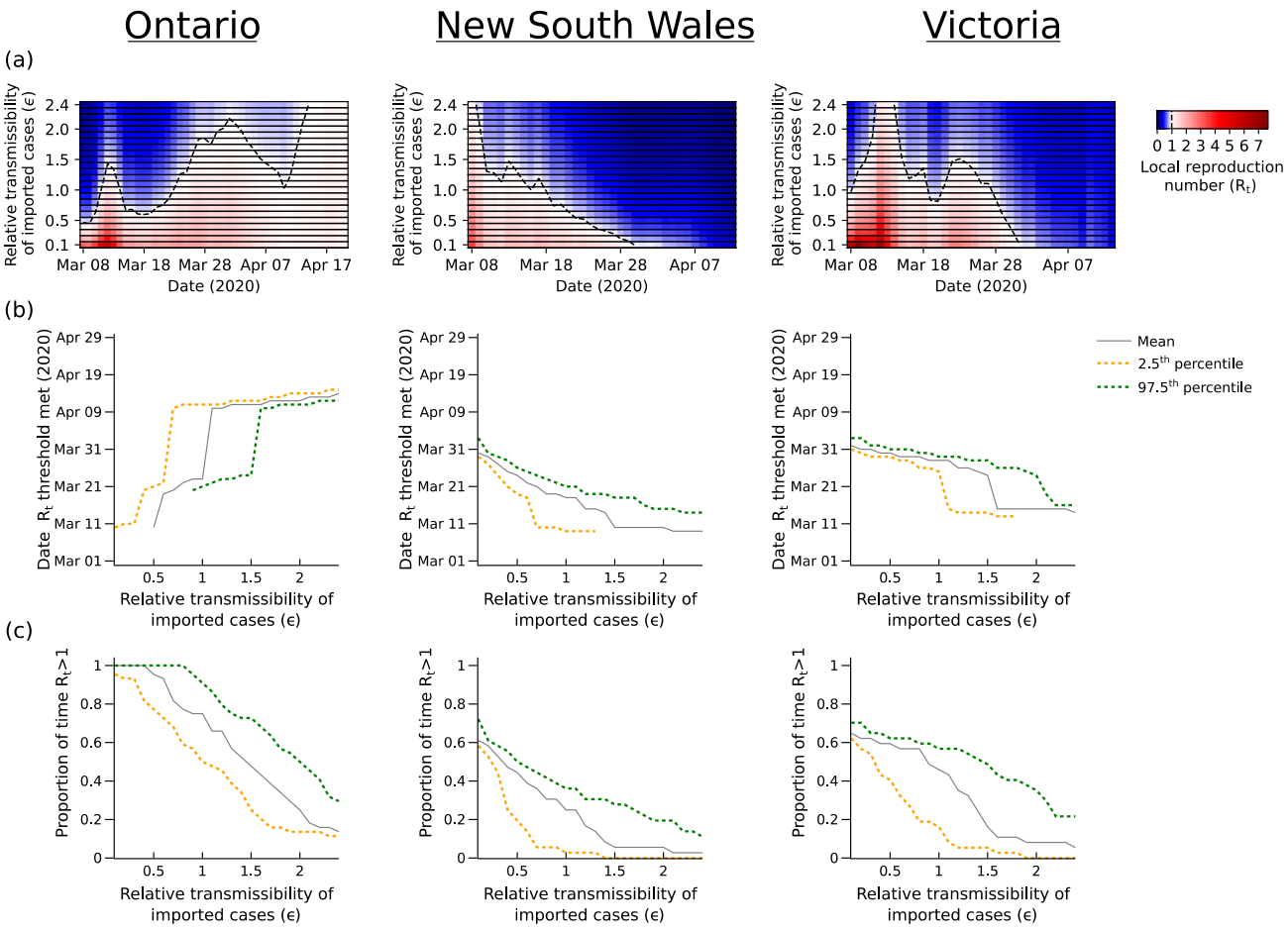


Figure 2. Inference of the local reproduction number (R_t) under different assumptions about the relative transmission risk from imported and local cases. (a) The COVID-19 incidence time series datasets used in our main analyses, for Ontario (left), New South Wales (centre) and Victoria (right). Black bars represent the daily numbers of local cases, and pink bars represent the daily numbers of imported cases. (b) Inferred R_t values for different assumed values of the relative transmission risk from an imported cases compared to a local case (ε). The grey horizontal line represents the threshold $R_t = 1$, and shaded regions represent the 95% central credible interval of the R_t estimates.

614



615

616 Figure 3. Implications of differences in the assumed relative transmission risk from imported and local cases on policy-
617 making. (a) Inferred mean R_t values for different values of the relative transmissibility of imported cases compared to local
618 cases (ε). (b) Dates on which the estimated values of R_t cross policy-relevant thresholds (in scenarios where the thresholds
619 are crossed at some stage in the outbreak; otherwise dates are not plotted). For Ontario (left), the date shown represents
620 the first date when the estimated R_t value is above one and remains above one for the remainder of the time period
621 considered (until 20 April 2020). This represents the first date when the outbreak is not inferred to be under control for the
622 remainder of the time period. For New South Wales (centre) and Victoria (right), the date shown represents the first date
623 on which the estimated R_t value is below one and remains so for the remainder of the time period considered (until 13
624 April 2020). This represents the first date on which the outbreak could be concluded as being under control for the
625 remainder of the time period. (c) The proportion of the time periods considered for which the inferred R_t values are above
626 one (so the outbreak is not inferred to be under control). In b and c, results are shown for the mean values of the posterior
627 for R_t (grey), and well as for the 2.5th (yellow dotted) and 97.5th (green dotted) percentile values of the posterior for R_t
628 (which span the 95% central credible interval).
629

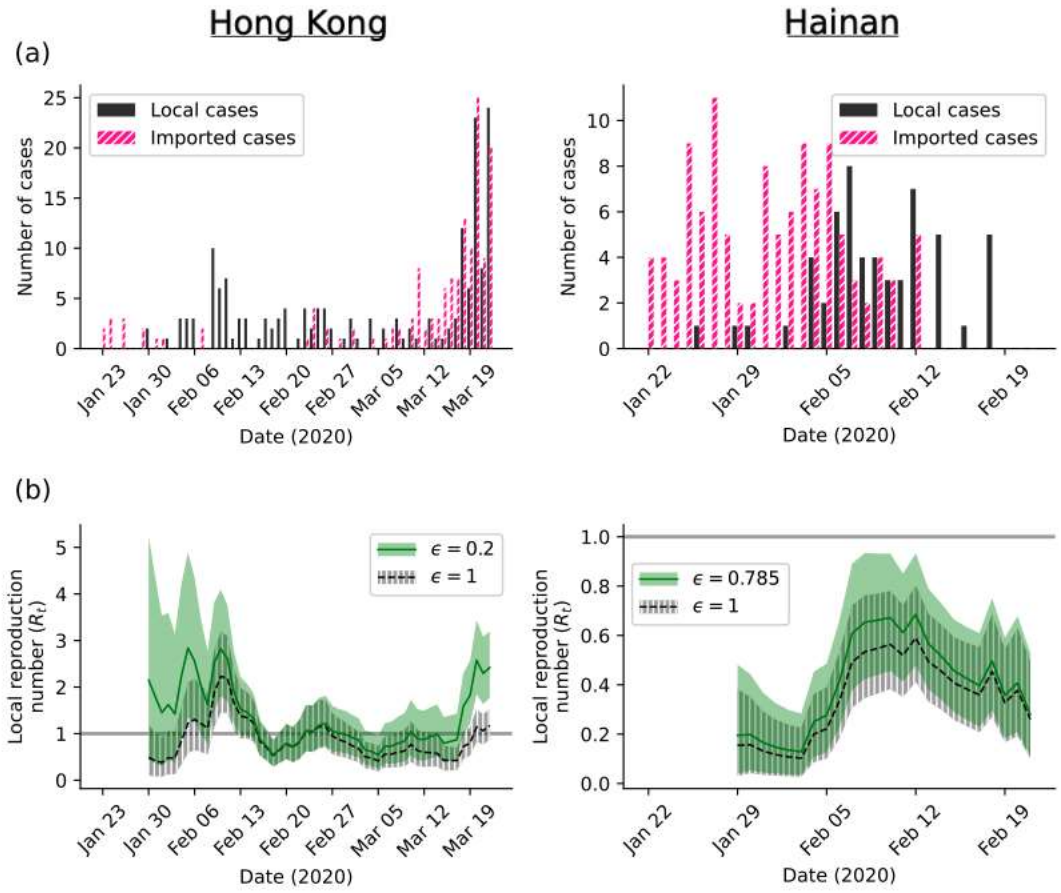


Figure 4. Inference of the local reproduction number (R_t) for estimated values of the relative transmission risk from imported and local cases. (a) The COVID-19 incidence time series datasets used in our main analyses, for Hong Kong (left) and Hainan Province, China (right). Black bars represent the daily numbers of local cases, and pink bars represent the daily numbers of imported cases. (b) Inferred R_t values for different assumed values of the relative transmission risk from an imported cases compared to a local case (ε), for Hong Kong (left) and Hainan Province (right). The grey horizontal line represents the threshold $R_t = 1$, and shaded regions represent the 95% central credible interval of the R_t estimates. The values $\varepsilon = 0.2$ for Hong Kong and $\varepsilon = 0.785$ for Hainan were estimated from alternative data sources, as described in the text.