Commentary on the use of the reproduction number R during the COVID-19 pandemic

Carolin Vegvari[1], Sam Abbott[2], Frank Ball[3], Ellen Brooks-Pollock[4], Robert Challen[5], Benjamin S Collyer[6], Ciara Dangerfield[7], Julia R Gog[8], Katelyn M Gostic[9], Jane M Heffernan[10], T Déirdre Hollingsworth[11], Valerie Isham[12], Eben Kenah[13], Denis Mollison[14], Jasmina Panovska-Griffiths[15], Lorenzo Pellis[16], Michael G Roberts[17], Gianpaolo Scalia Tomba[18], Robin N Thompson[19], and Pieter Trapman[20]

[1]Medical Research Council Centre for Global Infectious Disease Analysis, Department of Infectious Disease Epidemiology, School of Public Health, Imperial College London, London, United Kingdom
[2]Center for the Mathematical Modelling of Infectious Diseases, London School of Hygiene Tropical Medicine, London, UK
[3]School of Mathematical Sciences, University of Nottingham, Nottingham, UK, orcid.org/0000-0002-5599-2903 ristol Veterinary School, University of Bristol, Bristol, UK andNIHR Health Protection Research Unit in Behavioural Science and Evaluation at the University of Bristol
[5]EPSRC Centre for Predictive Modelling in Healthcare, University of Exeter, Exeter, Devon, UK and Somerset NHS Foundation Trust, Taunton, Somerset, UK
[6]Medical Research Council Centre for Global Infectious Disease Analysis, Department of Infectious Disease Epidemiology, School of Public Health, Imperial College London, London, United Kingdom
[7]Isaac Newton Institute for Mathematical Sciences
[8]Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge, UK
[9]Department of Ecology and Evolution, University of Chicago, Chicago IL USA
[10]Centre for Disease Modelling, Mathematics Statistics, York

University, Canada and COVID Modelling Task-Force, The Fields Institute, Canada
[11]Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford
[12]Department of Statistical Science, University College London, London, UK, orcid.org/0000-0001-5189-2844
[13]Division of Biostatistics, College of Public Health, The Ohio State University, Columbus, Ohio, USA
[14]Department of Actuarial Mathematics and Statistics, Heriot-Watt University Edinburgh, UK
[15]Department of Applied Health Research, University College London, London, UK and The Queen's College, Oxford University, Oxford, UK, orcid.org/0000-0002-7720-1121 and
[16]Department of Mathematics, The University of Manchester, Manchester, UK, prcid.org/0000-0002-3436-6487 and The Alan Turing Institute, London
[17]School of Natural and Computational Sciences and New Zealand Institute for Advanced Study, Massey University, Auckland, New Zealand, orcid.org/0000-0003-2693-5093
[18]Department of Mathematics, University of Rome Tor Vergata, Italy
[19]Mathematical Institute, University of Oxford, Oxford, UK, orcid.org/0000-0001-8545-5212 and Christ Church, University of Oxford, Oxford, UK and Center for the Mathematical Modelling of Infectious Diseases, London School of Hygiene Tropical Medicine, London, UK
[20]Department of Mathematics, Stockholm University, Stockholm, Sweden

6 October 2020

Abstract

Since the beginning of the COVID-19 pandemic, the reproduction number R has become a popular epidemiological metric used by policy makers and the media to communicate the state of the epidemic across countries. At its most basic, R is defined as the average number of secondary infections caused by

one primary infected individual. R seems convenient and easy to use, because the epidemic is expanding if $R > 1$ and contracting if $R < 1$. The magnitude of R indicates by how much transmission needs to be reduced to control the epidemic. However, using R in a naïve way can cause new problems. The reasons for this are threefold. 1) There is not just one definition of R but many, and the precise definition of R affects both its estimated value and how it should be interpreted. 2) Even with a particular clearly defined R, there may be different statistical methods used to estimate its value, and the choice of method will affect the estimate. 3) The availability and type of data used to estimate R vary, and it is not always clear what data should be included in the estimation. For example, should imported cases that are immediately quarantined count towards R, or should the data used to estimate R capture the potential of the local population to transmit the infection? In this review, we discuss when R is useful, when it may be of use but needs to be interpreted with care, and when it may be an inappropriate indicator of the progress of the epidemic. We also argue that careful definition of R, and the data and methods used to estimate it, can make R a more useful metric for future management of the epidemic.

1 What is the reproduction number R?

Since the start of the novel coronavirus (SARS-CoV-2) pandemic, the reproduction number R has become a popular summary statistic, used by policy makers to assess the state of the epidemic and the efficacy of interventions and by the media to communicate the progress of the epidemic to the general public. The primary appeal of R is that it offers a single number that indicates whether the transmission of the pathogen is increasing or decreasing, depending on whether R is above or below one. Early R estimates for SARS-CoV-2 in different countries were in the range of 2.0 - 6.5 [35, 54]. However, the use of R can be problematic in terms of both its definition and its estimation. Its usefulness is precisely because it is a summary statistic rather than a basic parameter describing the dynamic processes of infection, transmission and recovery. To understand how it is calculated and how it can be affected by interventions, the epidemic process needs to be considered in more detail. When epidemic numbers are small or concentrated in possibly atypical parts of a population, it may be an unreliable descriptor of the state of the outbreak.

In this paper, we discuss these issues and determine the situations when the reproduction number R is most useful for assessing and communicating the state of an outbreak (see Figure 1).

1.1 The beginning of a pandemic - R_0

In the early stages of a new outbreak of an infectious disease we can define an initial R value, known as the *basic reproduction number* R_0, that is the average number of individuals infected by each infectious individual in a fully susceptible population [22, 31, 32]. An outbreak resulting from one infected individual may die out within a few infection generations by chance [20, 55]. Otherwise, if $R_0 > 1$, the incidence of cases will grow exponentially, with on average R_0^n cases in the n^{th} generation. Already, this simple description introduces a number of concepts and assumptions. An individual's *infection generation* specifies their position in the chain of infections, the $(n-1)^{\text{th}}$ generation infects the n^{th} generation, and so on. It also assumes an underlying scenario (model) in which the average number of susceptibles infected by each infective stays the same over successive infection generations, and ignores the depletion of susceptibles. (We refer to those members of the population who are uninfected and susceptible to infection as *susceptibles*, and those that are infected and infectious as *infectives*.) The potential importance of these assumptions depends on the contact structure of the population, to which we return below.

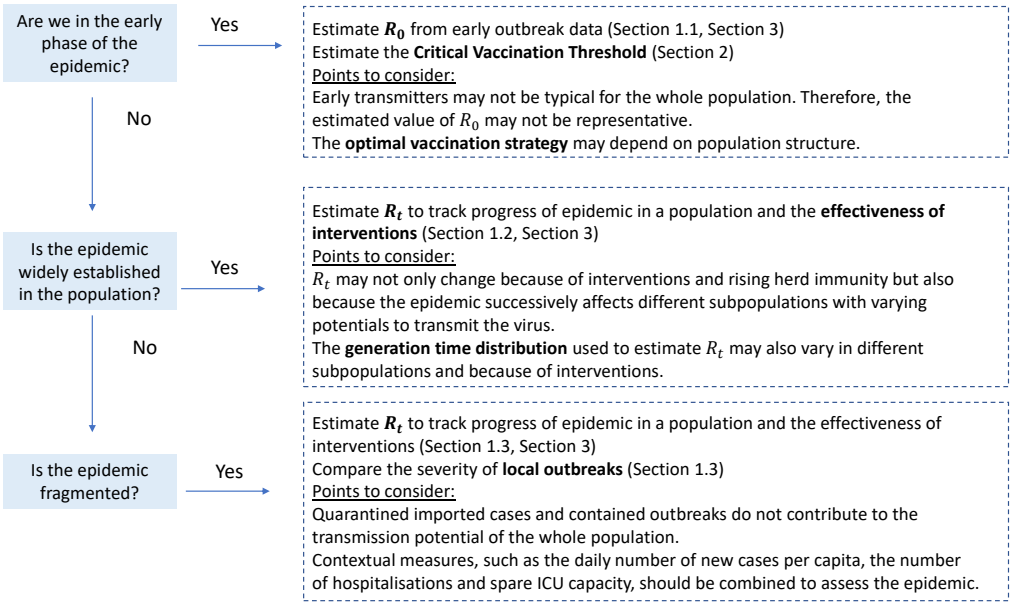

Figure 1: Flow chart summarising the main points explained in the main text depending on the state of the epidemic.

Thus, R_0 (and other R values to be defined later) is not just a property of the infectious agent (*pathogen*). It depends on demography, and whatever human behaviour is associated with the possibility of infectious contact (an *effective contact* is one that results in transmission if made with a susceptible, while a contact in the common sense of the word has a certain probability of transmission). For the simplest models, $R_0 > 1$ implies that an introduction of infection will result in an

epidemic. Furthermore, if there were no interventions or changes in behaviour, then the proportion of the population infected during the entire course of an epidemic would be the non-zero solution of the equation $P = 1 - e^{-R_0 P}$ (for example, if $R_0 = 2$, then P is approximately 0.8). This result is referred to as the *final size equation*, and underscores the fact that during an epidemic it is not generally true that everybody will be infected at some point.

Individuals may vary considerably in their susceptibility to infection and in their propensity to pass it on through their biology or behaviour. Age is often an important determinant. If the population is grouped in some way, so that for instance some groups have higher R values than others, then the overall outbreak is expected to grow as described by an R_0 that depends on all of these values, and also depends on how each group infects the others, i.e. on the R values between groups as well as within them (R_0 is then the dominant eigenvalue of the matrix of R values [21,22,31]). The first few stages of the outbreak may be atypical, depending on which group is first infected.

For the simplest mathematical model of the beginning of an outbreak, it is assumed that because only a small fraction of the population has been infected, all potential contacts are with susceptibles. This may be an unrealistic assumption because human interaction networks tend to be clustered (for example, through households, workplaces or schools). Growth through successive generations of infection, which is the basis for defining R_0, does not translate simply into time, because the *generation interval* of an infection (the time interval back from the instant when a susceptible is infected, to that when their infector was infected) is variable, and infection generations may overlap temporally. Typically, growth in the early stages is faster than the simple assumption of a fixed average generation time would suggest and this is a major problem in estimating R_0 from early outbreak data. In addition, the implicit assumption is that all infectives are identifiable as such. If there is a significant proportion of asymptomatic cases, an estimate of R_0 may be affected by the time from when an asymptomatic infective has become infected to when he/she is expected to infect susceptibles. If this timing is the same for asymptomatic and symptomatic cases, then the estimate for R_0 will be unaffected.

1.2 The second simplest case: where an outbreak is widespread - R_t

When the pandemic is well-established in a country (or region), with large numbers of cases most of which are internal to the country, an 'effective reproduction number' at time t, R_t (sometimes denoted R_e or R_{eff}), is a useful descriptor of the progress of the outbreak (Figure 1). Again, the concept is of an average of how many new cases each infectious case causes. The value of R_t may be affected by interventions: typically the aim is to reduce R_t below one and to as small a value as possible. For

models including detailed, and therefore complex, contact networks there may be more than one way of defining R_t; however, definitions should always agree that the value of R_t is 1 when the expected number of new infections is constant.

The relevance of the assumptions here (large numbers of cases, mostly internal to the region) is that in such circumstances we expect R_t to have a fairly stable value that changes substantially over time only when interventions are introduced or cease. The definition of R_t here is in terms of actual new infectious cases, i.e. excluding potentially infectious contacts with individuals who have had the disease and are immune to reinfection. As the number of immune individuals grows large compared to the entire population, the spread of infections will gradually slow, because many contacts will be with immune individuals, and hence the value of R_t will be reduced. The level of immunity at which $R_t = 1$ is the *herd immunity threshold* (see Section 2 on vaccination and herd immunity below).

1.3 When the outbreak is at a low level or fragmented – the concept of R may be less useful

If the outbreak is at a low level either because it has run its course or because of successful interventions, the definition and the use of an R value are problematic (Figure 1). At low levels of prevalence there will (as in the early stages of the outbreak) be greater statistical variability. Additionally, there are likely to be heterogeneities associated with the infection being unevenly spread among different subgroups of the population (possibly dependent on age, behaviour or geographical location [56]), with some parts of the population having had more exposure than others. There may also be local variability in interventions, and it may not be easy to allow for the effect of some cases being introductions from outside the population under consideration. If the outbreak is fragmented, particularly when close to elimination, it will make more sense to think of it as composed of separate local outbreaks, which can be modelled separately, rather than trying to specify an average R value overall.

1.4 Relating R to details of the infection process

If the population is heterogeneous or structured, defining a reproduction number needs care, as the number of new cases an infective is expected to cause will depend on both their infectiousness and how well connected they are. It has been shown that in the early stages of an epidemic, when the relevant contact structures of a population are not known and interventions are not targeted, assuming a homogeneous contact structure results in conservative estimates of R_0 and the required control effort. However, designing targeted intervention strategies requires reliable information on infectious contact structures [59]. There are several basic ways to use structured population models to capture departures from the simplest epidemic models. The

four most common are (i) household models, (ii) multi-type models, (iii) network models and (iv) spatial models.

In a household model, every person in the population is assumed to be part of a single household, which is typically small, and may even be of size one. Those in the same household have a higher probability of infecting each other than is the case for two people chosen randomly from the population. In this model, reproduction numbers can still be defined [6, 27]. The most commonly used is the household reproduction number R_*, which is the expected number of members of other households that are infected by people from a primary infected household. It is still possible to consider the average number of susceptibles infected by a single infectious person. However, in order for this to be useful, the average has to be computed in a sophisticated way, because the number of people a person can infect will depend on how many members of the same household are still susceptible when s/he becomes infectious [47].

A second way of modelling heterogeneity in the population is to assume that the population can be subdivided into groups. The groups may be defined through age bands, social activity levels, health status, type of job, place of residence and so on. Characteristics such as susceptibility, infectivity and frequency of contact may depend on an individual's group, but all those in a single group have the same characteristics. It is often assumed that all these groups are large. If there are regular inter-group contacts then the largest eigenvalue of the so-called next generation matrix [21, 22] has many similar properties to those of R_0 for an epidemic spreading in a homogeneously mixing population, although the final size equation is generally not satisfied.

A third way of introducing heterogeneity is to represent the population by a network, where transmission is only possible between people sharing a link in the network. For many network models it is still possible to define a reproduction number [37]. It is important to note that the person initially infected in a population is often atypical and should be ignored in computing or estimating the reproduction number. A useful extension is a mixture of a network model and a homogeneous mixing model, in which both regular and casual contacts are captured. In this extension, a reproduction number with the desired threshold properties can be defined [5].

Sometimes most transmission is restricted to people living close to each other, and spatial models are useful when physical location should be incorporated. For these, it is often difficult to define a reproduction number because there is no phase in which the number of infecteds is growing exponentially [19, 48]. If standard estimation methods are used where there is a considerable spatial component then the estimates will be close to one, even when the spread is highly supercritical and transmission needs to be much reduced in order to control the epidemic.

2 R, vaccination and herd immunity

As immunity builds up in a population through infection during the course of an epidemic, even when the contact rate between individuals remains the same (assuming no change in interventions), both the chance that a contact is susceptible to infection, and the effective reproduction number, R_t, will decrease. Herd immunity is achieved when enough individuals have become immune so that R_t falls below the value 1 without the need to reduce contacts among individuals by non-pharmaceutical interventions.

Vaccination provides another means of building up immunity in a population. Depending on the coverage, it can slow or halt the spread of an epidemic, preventing individual infection or limiting experiences of disease. All vaccination programs aim to achieve sufficient immunity in the population that $R_t < 1$ without modifying contact patterns among individuals. In this situation, there are insufficient susceptibles in the population for sustained transmission. The susceptible proportion of a population for which $R_t = 1$ is known as the *critical vaccination threshold (CVT)*. When the susceptible proportion is below this threshold, there is herd immunity, which means that the population is protected from a major outbreak even though not everyone is vaccinated or otherwise immune.

In simple mathematical models (e.g. models in which the population is only subdivided into susceptible, infected and recovered individuals), the CVT is determined by the basic reproduction number R_0. Specifically, vaccination of a uniform randomly chosen proportion $1 - \frac{1}{R_0}$ of the population is sufficient to create herd immunity and prevent an epidemic, as long as the vaccine-induced immunity is sufficiently long-lasting [52]. As a simple example, if $R_0 = 2$ then 50% of a population would need to be vaccinated or otherwise immune to prevent outbreaks. If $R_0 = 3$, as is approximately the case for COVID-19, then 67% of a population would need to be vaccinated or immune. When setting such vaccination targets, waning immunity needs to be taken into account. The implementation and impact of a vaccination programme depends on whether vaccination is performed before or during an outbreak [13, 33].

As outlined above, population structure affects the reproduction numbers R_0 and R_t as well as the probability that an epidemic will spread. Therefore, it has important effects on the threshold for herd immunity and the optimal vaccination strategy. For models with small mixing groups such as households, the basic reproduction number R_0, as defined in Section 1.1, does not provide a good indicator of whether or not an epidemic can take off because repeated contacts within households are likely even in the early stages of an outbreak. However, in the early stages of an epidemic, between-household contacts are likely to be with individuals in otherwise fully susceptible households, so the reproduction number R_* which is given by the average number of between-household contacts that emanate from a typical within-

household epidemic [4,7] can be used instead. For household models, herd immunity is achieved if a uniform randomly chosen proportion $1 - \frac{1}{R_*}$ of all *households* in a population is fully vaccinated.

For COVID-19, a toy model has been used to illustrate the effect of population heterogeneity on herd immunity. It showed [11] that age structure and variation in social contacts among individuals could reduce the herd immunity threshold to 43%, almost a third less than that for a homogeneous population. Assuming a more extreme variation in social contact rates and that the most exposed individuals become infected first, another study estimates that the herd immunity threshold in some populations could be as low as 20% [29]. In addition, there is some indication that immunity gained from infection by some common cold coronavirus strains may provide cross immunity to SARS-COV-2 [50, 62]. There have also been reports that immunity gained from COVID-19 infection may wane, reducing individual and population levels of immunity over time. If these observations are indeed applicable here, the herd immunity threshold could be further modified [50].

One important difference between immunisation by vaccination and by infection is that, during an epidemic, individuals with higher susceptibilities and/or larger numbers of contacts are likely to be infected earlier. If herd immunity is to be achieved by vaccination, optimal planning can reduce the coverage required to achieve herd immunity. For example, in an illustrative households model for variola minor infections in Brazil, it is shown that under the optimal vaccination strategy the proportion of the population that needs to be vaccinated is a third less than under a strategy that fully vaccinates randomly chosen households [3]. If a COVID-19 vaccine is developed, demand will surely exceed supply initially. Designing optimal vaccination strategies for different settings that take into account population structure alongside other public health concerns, e.g. protecting the vulnerable, could greatly enhance the chances of achieving herd immunity and the cost effectiveness of vaccination as an intervention.

3 How can R be estimated?

Before estimating R, the purpose of the estimation needs to be clarified. Is it intended simply to track the changes in the trajectory of case numbers over time? Or is it intended to assess the potential of a population to transmit a pathogen perhaps in the context of considering interventions? If the latter, the relevant population needs to be defined. Depending on the purpose, different data sets and statistical methods can be used.

There are several approaches to estimating R_t from epidemiological data. In the most direct method, high-quality contact tracing data can be used, in theory at least, to estimate both R_t and the generation time interval, and this has been

attempted for COVID-19 [23]. However, contact tracing of SARS-CoV-2 infections is notoriously difficult because of the high proportion of asymptomatic infections. Moreover, effective contact tracing reduces the number of contacts of traced individuals so that the corresponding estimates will be biased.

More commonly, R_t can be estimated by inferring the rate of infection transmission within a dynamical model fitted to observed cases, hospitalisations, deaths or a combination of those [49, 60]. Dynamical models have been used widely to forecast the spread of COVID-19 and the effect of interventions. These models allow the impact of assumed changes in specific interventions on R_t to be explored, so estimating R_t in this way can be convenient. Dynamical models can be described by systems of differential equations and assume very large to infinite population sizes. In completely deterministic dynamical models, the uncertainty in estimated R_t values depends only on data and parameter uncertainty, and not on stochastic uncertainty. However, if the number of new infections is small, the value of R_t is strongly affected by chance events, which increases the uncertainty in the estimate. This situation can be addressed by use of stochastic models or incorporating stochastic assumptions in otherwise deterministic model frameworks.

But this approach is not without drawbacks. Not least, R_t estimates from dynamical models depend critically on assumptions (e.g. model structure and which parameter values are estimated), and on data quality. Another potential drawback is that many parameters of dynamical models are often assumed to be fixed over time. These approaches are therefore less suited to capture the effects of gradual, continuous changes in behaviour, mobility or social network structure. However, gradual changes in dynamic models can be incorporated by assuming that transmission parameters change over given intervals, while at the same time the possible amount of change is constrained to avoid big jumps caused by a small number of noisy data points [10]. In this way, dynamical models that include change-points in the rate of infection near specific interventions can infer the impact of control policies, as well as the effect of susceptible depletion.

There is also a difference in how R_t is estimated between compartmental and agent- or individual-based models. In an agent-based model, it is possible simply to count exactly how many secondary infections are caused by each primary infection. Thus, all details of the epidemic – including time-varying viral loads, population-level and localised immunity, interventions, network factors, and other effects – are automatically incorporated, and do not need to be considered separately [45]. As agent-based models explicitly include stochastic effects, the uncertainty in R_t estimates can be greater than for those derived from deterministic dynamical models. Because of the greater number of parameters included in dynamical and particularly agent-based models, they require more data and more different types of data than the simpler statistical models below to identify estimates for all parameters.

A third approach uses statistical models to estimate R_t, and continuous

changes in it, empirically from case notification data. These methods make minimal structural assumptions about epidemic dynamics, and only require users to specify the distribution of the generation interval. They are agnostic to population susceptibility or epidemic phase, but as we discuss below, care must still be taken to avoid quantitative and temporal biases. The most common empirical methods are the Cori method [18, 57] and the Wallinga-Teunis method [61]. Drawbacks of some statistical models include that they cannot be used to combine different data streams into a coherent picture.

Where genome sequences from viral samples taken from infected patients are available and the date of sampling is known, R_t can also be estimated using phylogenetic methods. An evolutionary model is fitted that best explains the patterns of nucleotide substitution in the dated samples. The fitted model parameters include the nucleotide substitution rate and the population size of the virus at a given time in the past. Using a metapopulation analogy, the effective population size of a pathogen has been shown to be proportional to the number of infected individuals and inversely proportional to the transmission rate from which the reproduction number can be determined [39].

3.1 Statistical methods to estimate R

In this paragraph we discuss two frequently used simple statistical methods to estimate R and common issues associated with them. The Cori and Wallinga-Teunis methods estimate subtly different versions of R_t; the Cori method generates estimates of the instantaneous reproduction number and the Wallinga-Teunis method generates estimates of the case reproduction number [18, 25]. The key difference is that the instantaneous reproduction number gives an average R_t for a homogeneous population at a single point in time, whereas the case reproduction number can accommodate individual heterogeneity, but blurs over several dates of transmission. Furthermore, the case reproduction number is a leading estimator of the instantaneous reproduction number, i.e. it depends on data from after the time for which the reproduction number is to be estimated, and must be adjusted accurately to infer the impact of time-specific interventions [30].

The instantaneous reproduction number represents the expected number of infections generated at time t by currently infectious individuals [18]. For real-time analysis, one of the benefits of estimating the instantaneous reproduction number is that it does not require information about future changes in transmissibility, and it reflects the effectiveness of control measures in place at time t. But as an aggregate measure of transmission by all individuals infected in the past (who may now be shedding virus), it does not easily consider heterogeneity in transmission. In contrast, the case reproduction number represents the expected number of infections generated by an individual who is first infected at time t, and has yet to progress through the full

course of viral shedding. This leads to "right censoring" when the case reproduction number is estimated in real-time; if all infections generated by individuals who were infected at time t have not yet been observed, then the data must be adjusted [14, 15, 44] or the case reproduction number will be underestimated.

The Cori method and the Wallinga-Teunis method involve inferring the values of R_t that are most consistent with observed incidence data (for a review, see [30]). In the Cori method, typically this inference is carried out by assuming that R_t is constant over fixed time windows. Smoothing windows are used to avoid spurious fluctuations in estimates of R_t. These can occur if imperfect observation and reporting effects, rather than actual bursts in transmission, are the main source of noise in the data. Cross-validation and proper scoring rules can be used to avoid under- or oversmoothing R_t estimates [26].

An important concept, basic to both methods, is the intrinsic generation time also referred to as the infectiousness profile. The intrinsic generation interval is a theoretical quantity derived from the renewal equation of Lotka and Euler [38, 60]. It describes the time distribution of potentially infectious contacts made by an index case, and is independent of population susceptibility [17]. In practice, the intrinsic generation interval is not observable, and it must be estimated carefully from observed serial intervals within contact tracing data [17, 46]. The serial interval is generally defined as the duration of time between onset of symptoms in an index case and in a secondary case [53]. In the early stages of an outbreak, accurate estimation should adjust for right truncation of observations, for changes over time in population susceptibility, and for interventions like case isolation, which may shorten the generation interval by limiting transmission events late in the course of infectiousness [2, 17, 46].

Both the Cori and Wallinga-Teunis methods are conceptually based on separating the infectiousness of an infective into two components, total amount and timing. The timing is expressed by the generation time distribution while the total amount is expressed by R_t. The variation of (average) infectivity over time is ascribed, at least in practical implementations of the methods, to changes in R_t, while the intrinsic generation time is assumed to remain fixed. This is a simplification that may lead to inaccurate estimation of R_t, since, in reality, the observed generation time distribution varies over time, both because of the epidemic dynamics [12, 53, 58], because of the epidemic affecting different subgroups of the population, with possibly different generation time distributions over time [36, 40], and, more importantly, because of interventions that affect the length or efficacy of the infectious period [2]. An additional complication is that the "intrinsic" generation interval of the Cori and Wallinga-Teunis estimators includes potentially infectious contacts with both susceptible and immune individuals, whereas only contacts with susceptible individuals cause new infections, and are observed in contact tracing [17, 46]. Even when using an accurately estimated, fixed generation time distribution, both R_t estimators are numerically sensitive to the specified mean and variance of the intrinsic generation

interval [16].

3.2 Data used to estimate R

Fundamentally, R_t is a measure of transmission. Ideally, it would be estimated from data on the total number of incident infections (i.e. transmission events) occurring each day. But in practice, only a small fraction of infections are observed, and notifications do not occur until days or weeks after the moment of infection. Temporally accurate R_t estimation requires adjusting for lags to observation, which can be estimated as the sum of the incubation period and delays from symptom onset to case observation [9, 16]. Delays not only shift observations into the future, they also blur infections incident on a particular date across many dates of observation. This blurring can be particularly problematic when working with long and variable delays (e.g. from infection to death), and when R_t is changing. Deconvolution [8, 24, 28, 42], or R_t estimation models that include forward delays [1] can be used to adjust lagged observations. Simpler approaches may be justifiable under some circumstances. If observation delays are relatively short and not highly variable, and if R_t is not rapidly changing, simply shifting unadjusted R_t estimates back in time by the mean delay can provide a reasonable approximation to the true value (see Challen *et al.*, in this volume, for an in-depth discussion [16]). The benefits and disadvantages of each approach are reviewed in [30]. Changes over time in case ascertainment can also bias R_t estimates, so ideally data should be drawn from structured surveillance (see, for example, the REACT study [34]) or adjusted for known changes in testing or reporting effort [34, 43].

In practice, R_t can be estimated from a time series of new symptom onset reports, cases, hospitalisations or deaths. Choosing an appropriate data stream involves weighing representativeness, timeliness of reporting, consistency of ascertainment, and length of lag. For example, reported deaths may be reasonably unaffected by changes over time in ascertainment, but adjusting for long lags to observation can be challenging, and deaths may not be representative of overall transmission (e.g. if the epidemic shifts toward younger age groups) [41, 51]. Extensions of existing statistical models for R_t estimation could potentially integrate multiple kinds of data, by assuming that (e.g.) cases, hospitalisations and deaths, arise from a shared, latent infection process, with different delays [30]. A mechanistic model can also pull multiple data streams together by modelling the different processes underlying each data stream. Problems can arise if different data streams disagree on the progress of the pandemic. However, if the disagreement is caused by a shift in delays from events to reporting in different data streams, a mechanistic model can highlight these changes. Sometimes different data streams can be used for model validation.

All methods used to estimate R_t must decide on the length of the time window over which it is to be estimated. All data used to estimate R_t are noisy.

The shorter the time window used for estimation, the higher will be the noise-to-signal ratio and, therefore, the uncertainty in the estimate of R_t. In contrast, longer time windows will produce estimates with lower uncertainty, but sudden changes in transmission may not be detected if the time window is too long.

4 Summary: Cautions and Recommendations

During the early phase of the epidemic:

- R_0 estimates in the early phase may not be representative for the population as a whole if the group of initial transmitters is atypical.
- R_0 may be incorrectly estimated in the early phase if infected but asymptomatic individuals are not counted or recognised, and their epidemiologically relevant behaviour differs from that of symptomatic individuals.

When the epidemic is established in the population:

- R_t can differ for different population groups, and the value of R_t is dominated by the group in which most transmission occurs. To improve targeted containment measures, where possible additional information should be reported alongside case data, such as demographic, socio-economic and occupational information.
- The estimated value of R_t and its associated uncertainty depend on the data stream(s) used and the time window over which R_t was estimated, and these should be reported alongside the estimates. This will make it possible to draw more robust conclusions when considering results from different models.
- Model components that are likely to change over the time course of the epidemic (e.g. the generation time distribution) should be updated regularly, and sensitivity to changing assumptions should be kept under consideration.

When the ongoing epidemic is fragmented:

- R_t estimates from local outbreaks, if they can be contained, cannot inform on the progress of the epidemic and efficacy of interventions at the national level. They may inform local interventions. Other descriptors should be considered to assess the progress of the epidemic, such as the number of new cases per capita per day in a defined area, the number of hospitalisations and the spare hospital and intensive care capacity.

- Imported cases that are effectively quarantined should not be counted towards R_t estimates as they do not contribute to the local transmission potential in the community.

Vaccination and herd immunity:

- If the available vaccine supply is limited, optimal vaccination strategies should be designed that take into account population structure and the transmission potential within different groups and other public health priorities, e.g. protection of the vulnerable groups.

In conclusion, estimated R values do not exactly correspond to the theoretically defined quantities. In statistical terms, model uncertainty, sampling variability, and data accuracy affect the estimates. Nevertheless, R_0 and R_t are useful quantities to assess the potential and progress of an epidemic. Their usefulness for decision making varies depending on the phase of the epidemic (early, established, fragmented). Clearly defining the context, the data streams and the statistical methods used to estimate R can improve its value for the management of an epidemic.

The authors would like to thank the Isaac Newton Institute for Mathematical Sciences, Cambridge, for support and hospitality during the programme Infectious Dynamics of Pandemics where work on this paper was undertaken. This work was supported by EPSRC grant no EP/R014604/1. EBP acknowledges funding from the Medical Research Council (MRC) (MC/PC/19067) and the NIHR Health Protection Research Unit in Behavioural Science and Evaluation at the University of Bristol. Support for RC's research is provided by the EPSRC via grant EP/N014391/1, RC is also funded by the NHS Global Digital Exemplar programme (GDE). JMH acknowledges funding from the Natural Science and Engineering Research Council of Canada (NSERC), Canadian Institutes for Health Research (CIHR). MGR is supported by the Marsden Fund under contract MAU1718. GPST acknowledges the MIUR Excellence Department Project awarded to the Department of Mathematics, University of Rome Tor Vergata, CUP E83C18000100006. LP acknowledges the Wellcome Trust and the Royal Society (grant 202562/Z/16/Z) for funding. EK was supported by National Institute of Allergy and Infectious Diseases (NIAID) grant R01 AI116770. The contents are solely the responsibility of the author and do not necessarily represent the official views of NIAID or the US National Institutes of Health.

References

[1] S Abbott, J Hellewell, RN Thompson, K Sherratt, HP Gibbs, NI Bosse, JD Munday, S Meakin, EL Doughty, JY Chun, YWD Chan, F Finger, P Campbell,

A Endo, CAB Pearson, A Gimma, T Russell, null null, S Flasche, AJ Kucharski, RM Eggo, and S Funk. Estimating the time-varying reproduction number of SARS-CoV-2 using national and subnational case counts [version 1; peer review: awaiting peer review]. *Wellcome Open Research*, 5(112), 2020.

[2] Sheikh Taslim Ali, Lin Wang, Eric H. Y. Lau, Xiao-Ke Xu, Zhanwei Du, Ye Wu, Gabriel M. Leung, and Benjamin J. Cowling. Serial interval of SARS-CoV-2 was shortened over time by non-pharmaceutical interventions. *Science*, 369(6507):1106–1109, 2020.

[3] Frank Ball and Owen Lyne. Optimal vaccination schemes for epidemics among a population of households, with application to variola minor in Brazil. *Statistical Methods in Medical Research*, 15(5):481–497, 2006.

[4] Frank Ball, Denis Mollison, and Gianpaolo Scalia Tomba. Epidemics with two levels of mixing. *Annals of Probability*, 7(1):46–89, 1997.

[5] Frank Ball and Peter Neal. Network epidemic models with two levels of mixing. *Mathematical Biosciences*, 212(1):69–87, 2008.

[6] Frank Ball, Lorenzo Pellis, and Pieter Trapman. Reproduction numbers for epidemic models with households and other social structures. II. Comparisons and implications for vaccination. *Mathematical Biosciences*, 274:108–139, 04 2016.

[7] NG Becker and K Dietz. The effect of household distribution on transmission and control of highly infectious diseases. *Mathematical Biosciences*, 127(2):207–219, 1995.

[8] Niels G Becker, Lyndsey F Watson, and John B Carlin. A method of non-parametric back-projection and its application to AIDS data. *Statistics in Medicine*, 10(10):1527–1542, 1991.

[9] Q Bi, Y Wu, S Mei, C Ye, X Zou, Z Zhang, X Liu, L Wei, SA Truelove, T Zhang, W Gao, C Cheng, X Tang, X Wu, Y Wu, B Sun, S Huang, Y Sun, J Zhang, T Ma, J Lessler, and T Feng. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *Lancet Infectious Diseases*, 20:911–919, 2020.

[10] Paul J Birrell, Joshua Blake, Edwin van Leeuwen, , Nick Gent, and Daniela De Angelis. Real-time nowcasting and forecasting of COVID-19 dynamics in England: the first wave? *medRxiv*, 2020.

[11] Tom Britton, Frank Ball, and Pieter Trapman. A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2. *Science*, 369(6505):846–849, 2020.

[12] Tom Britton and Gianpaolo Scalia Tomba. Estimation in emerging epidemics: biases and remedies. *Journal of the Royal Society Interface*, 16(150):20180670, 2018.

[13] Rose-Marie Carlsson, Lauren M Childs, Zhilan Feng, John W Glasser, Jane M Heffernan, Jing Li, and Gergely Röst. Modeling the waning and boosting of immunity from infection or vaccination. *Journal of Theoretical Biology*, 497:110265, 2020.

[14] Simon Cauchemez, Pierre-Yves Boëlle, Christl A Donnelly, Neil M Ferguson, Guy Thomas, Gabriel M Leung, Anthony J Hedley, Roy M Anderson, and Alain-Jacques Valleron. Real-time estimates in early detection of SARS. *Emerging Infectious Diseases*, 12(1):110, 2006.

[15] Simon Cauchemez, Pierre-Yves Boëlle, Guy Thomas, and Alain-Jacques Valleron. Estimating in real time the efficacy of measures to control emerging communicable diseases. *American Journal of Epidemiology*, 164(6):591–597, 2006.

[16] Robert Challen, Ellen Brooks-Pollock, Leon Danon, and Krasimira Tsaneva-Atanasova. Impact of uncertainty in serial interval, generation interval, incubation period and delayed observations in estimating the reproduction number for COVID-19. *Statistical Methods in Medical Research*, this volume, 2020.

[17] David Champredon and Jonathan Dushoff. Intrinsic and realized generation intervals in infectious-disease transmission. *Proceedings of the Royal Society B*, 282(1821):20152026, 2015.

[18] Anne Cori, Neil M Ferguson, Christophe Fraser, and Simon Cauchemez. A new framework and software to estimate time-varying reproduction numbers during epidemics. *American Journal of Epidemiology*, 178(9):1505–1512, 09 2013.

[19] S Davis, P Trapman, Herwig Leirs, Michael Begon, and JAP Heesterbeek. The abundance threshold for plague as a critical percolation phenomenon. *Nature*, 454(7204):634–637, 2008.

[20] Kendall DG. Deterministic and stochastic epidemics in closed populations. *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability*, 4:149–165, 1956.

[21] O Diekmann, JAP Heesterbeek, and MG Roberts. The construction of next-generation matrices for compartmental epidemic models. *Journal of the Royal Society Interface*, 7:873–885, 2010.

[22] Odo Diekmann, JAP Heesterbeek, and Tom Britton. *Mathematical Tools for Understanding Infectious Disease Dynamics*. Princeton University Press, 2013.

[23] Luca Ferretti, Chris Wymant, Michelle Kendall, Lele Zhao, Anel Nurtay, Lucie Abeler-Dörner, Michael Parker, David Bonsall, and Christophe Fraser. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science*, 368(6491):eabb6936, 2020.

[24] Swiss National COVID-19 Science Task Force. Effective reproductive number. https://ibz-shiny.ethz.ch/covid-19-re/, August 2020. accessed 20 August 2020.

[25] C Fraser. Estimating individual and household reproduction numbers in an emerging epidemic. *PLOS ONE*, 2(8):e758, 2007.

[26] T Gasser and M Rosenblatt. *Smoothing Techniques for Curve Estimation*. Springer, 1979.

[27] E Goldstein, K Paur, C Fraser, E Kenah, J Wallinga, and M Lipsitch. Reproductive numbers, epidemic spread and control in a community of households. *Mathematical Biosciences*, 221(1):11–25, 2009.

[28] Edward Goldstein, Jonathan Dushoff, Junling Ma, Joshua B Plotkin, David JD Earn, and Marc Lipsitch. Reconstructing influenza incidence by deconvolution of daily mortality time series. *Proceedings of the National Academy of Sciences of the United States of America*, 106(51):21825–9, 2009.

[29] M Gabriela M Gomes, Rodrigo M Corder, Jessica G King, Kate E Langwig, Caetano Souto-Maior, Jorge Carneiro, Guilherme Goncalves, Carlos Penha-Goncalves, Marcelo U Ferreira, and Ricardo Aguas. Individual variation in susceptibility or exposure to SARS-CoV-2 lowers the herd immunity threshold. *medRxiv*, 2020.

[30] Katelyn M Gostic, Lauren McGough, Edward Baskerville, Sam Abbott, Keya Joshi, Christine Tedijanto, Rebecca Kahn, Rene Niehus, James A Hay, Pablo M. De Salazar, Joel Hellewell, Sophie Meakin, James Munday, Nikos Bosse, Katharine Sherratt, Robin N Thompson, Laura F White, Jana Huisman, Jérémie Scire, Sebastian Bonhoeffer, Tanja Stadler, Jacco Wallinga, Sebastian Funk, Marc Lipsitch, and Sarah Cobey. Practical considerations for measuring the effective reproductive number, R_t. *medRxiv*, 2020.

[31] JAP Heesterbeek. R_0. PhD thesis, University of Leiden, 1992.

[32] JAP Heesterbeek. A brief history of R_0 and a recipe for its calculation. *Acta Biotheoretica*, 50:189–204, 2002.

[33] JM Heffernan and MJ Keeling. Implications of vaccination and waning immunity. *Proceedings of the Royal Society B*, 276(1664):2071–2080, 2009.

[34] REACT Study Investigators. Real-time Assessment of Community Transmission (REACT) Study. https://www.imperial.ac.uk/medicine/research-and-impact/groups/react-study/, 2020. accessed 28 September 2020.

[35] Mark Jit, Thibaut Jombart, Emily S Nightingale, Akira Endo, Sam Abbott, LSHTM Centre for Mathematical Modelling of Infectious Diseases COVID-19 Working Group, and W John Edmunds. Estimating number of cases and spread of coronavirus disease (COVID-19) using critical care admissions, United Kingdom, February to March 2020. *Eurosurveillance*, 25(18):2000632, 2020.

[36] Eben Kenah, Marc Lipsitch, and James M Robins. Generation interval contraction and epidemic data analysis. *Mathematical Biosciences*, 213(1):71–79, 2008.

[37] IZ Kiss, J Miller, and PL Simon. *Mathematics of Epidemics on Networks*. Springer, 2017.

[38] M Kot. The Lotka integral equation. In *Elements of Mathematical Ecology*, pages 353–364. Cambridge University Press, 2001.

[39] Alessia Lai, Annalisa Bergna, Carla Acciarri, Massimo Galli, and Gianguglielmo Zehender. Early phylogenetic estimate of the effective reproduction number of SARS-CoV-2. *Journal of Medical Virology*, 92:675–679, 2020.

[40] Quan-Hui Liu, Marco Ajelli, Alberto Aleta, Stefano Merler, Yamir Moreno, and Alessandro Vespignani. Measurability of the epidemic reproduction number in data-driven contact networks. *Proceedings of the National Academy of Sciences of the United States of America*, 115(50):12680–12685, 2018.

[41] Judith Malmgren, Boya Guo, and Henry G Kaplan. COVID-19 confirmed case incidence age shift to young persons age 0-19 and 20-39 years over time: Washington State March - April 2020. *medRxiv*, 2020.

[42] IC Marschner. Back-projection of COVID-19 diagnosis counts to assess infection incidence and control measures: analysis of Australian data. *Epidemiology and Infection*, 148:e97, 2020.

[43] R Omori, K Mizumoto, and G Chowell. Changes in testing rates could mask the novel coronavirus disease (COVID-19) growth rate. *International Journal of Infectious Diseases*, 94:116–118, 2020.

[44] Christopher E Overton, Helena B Stage, Shazaad Ahmad, Jacob Curran-Sebastian, Paul Dark, Rajenki Das, Elizabeth Fearon, Timothy Felton, Martyn Fyles, Nick Gent, Ian Hall, Thomas House, Hugo Lewkowicz, Xiaoxi Pang, Lorenzo Pellis, Robert Sawko, Andrew Ustianowski, Bindu Vekaria, and Luke Webb. Using statistics and mathematical modelling to understand infectious disease outbreaks: COVID-19 as an example. *Infectious Disease Modelling*, 5:409–441, 2020.

[45] Jasmina Panovska-Griffiths, Cliff C Kerr, Robyn M Stuart, Dina Mistry, Daniel J Klein, Russell M Viner, and Chris Bonell. Determining the optimal strategy for reopening schools, the impact of test and trace interventions, and

the risk of occurrence of a second COVID-19 epidemic wave in the UK: a modelling study. *Lancet Child & Adolescent Health*, 0(0), 2020.

[46] Sang Woo Park, David Champredon, and Jonathan Dushoff. Inferring generation-interval distributions from contact-tracing data. *Journal of the Royal Society Interface*, 17(167):20190719, 2020.

[47] Lorenzo Pellis, Frank Ball, and Pieter Trapman. Reproduction numbers for epidemic models with households and other social structures. I. Definition and calculation of R_0. *Mathematical Biosciences*, 235(1):85–97, 2012.

[48] Steven Riley, Ken Eames, Valerie Isham, Denis Mollison, and Pieter Trapman. Five challenges for spatial epidemic models. *Epidemics*, 10:68–71, 2015.

[49] M Roberts and JAP Heesterbeek. Model-consistent estimation of the basic reproduction number from the incidence of an emerging infection. *Journal of Mathematical Biology*, 55(803):803–816, 2007.

[50] Alan Sariol and Stanley Perlman. Lessons for COVID-19 immunity from other coronavirus infections. *Immunity*, 53(2):248–263, 2020.

[51] S Seaman and D De Angelis. Update on estimates of numbers of COVID-19 deaths accounting for reporting delay. https://www.mrc-bsu.cam.ac.uk/wp-content/uploads/2020/06/Adjusting-COVID-19-deaths-to-account-for-reporting-delay-.pdf, 2020. accessed 28 September 2020.

[52] CE Smith. Factors in the transmission of virus infections from animal to man. *Scientific Basis of Medicine Annual Review*, pages 125–150, 1964.

[53] Åke Svensson. A note on generation times in epidemic models. *Mathematical Biosciences*, 208(1):300–311, 2007.

[54] Biao Tang, Xia Wang, Qian Li, Nicola Luigi Bragazzi, Sanyi Tang, Yanni Xiao, and Jianhong Wu. Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions. *Journal of Clinical Medicine*, 9(2):462, 2020.

[55] RN Thompson. Novel coronavirus outbreak in wuhan, china, 2020: Intense surveillance is vital for preventing sustained transmission in new locations. *Journal of Clinical Medicine*, 9:498, 2020.

[56] Robin N Thompson, T Deirdre Hollingsworth, Valerie Isham, Daniel Arribas-Bel, Ben Ashby, Tom Britton, Peter Challoner, Lauren H K Chappell, Hannah Clapham, N J Cunniffe, A Philip Dawid, Christl A Donnelly, Rosalind M Eggo, Seb Funk, N Gilbert, Julia R Gog, Paul Glendinning, William S Hart, JAP Heesterbeek, Thomas House, Matt Keeling, and et al. Key questions for modelling COVID-19 exit strategies. *Proceedings of the Royal Society B*, 287:20201405, 2020.

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

[58] Andrea Torneri, Amin Azmon, Christel Faes, Eben Kenah, Gianpaolo Scalia Tomba, Jacco Wallinga, and Niel Hens. Realized generation times: contraction and impact of infectious period, reproduction number and population size. *bioRxiv*, page 568485, 2019.

[59] Pieter Trapman, Frank Ball, Jean-Stéphane Dhersin, Viet Chi Tran, Jacco Wallinga, and Tom Britton. Inferring ¡i¿r¡/i¿¡sub¿0¡/sub¿ in emerging epidemics—the effect of common population structure is small. *Journal of The Royal Society Interface*, 13(121):20160288, 2016.

[60] Jacco Wallinga and Marc Lipsitch. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B*, 274(1609):599–604, 2007.

[61] Jacco Wallinga and Peter Teunis. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of Epidemiology*, 160(6):509–516, 09 2004.

[62] Ahmed Yaqinuddin. Cross-immunity between respiratory coronaviruses may limit COVID-19 fatalities. *Medical Hypotheses*, 144:110049, 2020.