

The basic reproduction number of SARS-CoV-2: a scoping review of available evidence

Ann Barber^{1,2}, John Griffin¹, Miriam Casey¹, Áine B. Collins^{1,7}, Elizabeth Ann Lane^{1,6},
Quirine Ten Bosch², Mart De Jong², David Mc Evoy³, Andrew W. Byrne⁴, Conor G.
McAloon⁵, Francis Butler⁸, Kevin Hunt⁸ and Simon J. More¹

¹ Centre for Veterinary Epidemiology and Risk Analysis, UCD School of Veterinary Medicine, University College Dublin, Belfield, Dublin D04 W6F6, Ireland.

² Department of Quantitative Veterinary Epidemiology, Wageningen University, 6700 AH, Wageningen, The Netherlands.

³ School of Public Health, Physiotherapy and Sports Science, University College Dublin, Belfield, Dublin 4, Ireland

⁴ One Health Scientific Support Unit, Department of Agriculture, Food and the Marine (DAFM), Kildare Street, Dublin 2, Ireland

⁵ Section of Herd Health and Animal Husbandry, UCD School of Veterinary Medicine, University College Dublin, Dublin D04 W6F6, Ireland.

⁶ Department of Agriculture, Food and the Marine, Backweston Campus, Co. Kildare, W23 X3PH.

⁷ Department of Agriculture, Food and the Marine, Kildare St., Dublin 2, Ireland.

⁸ School of Biosystems and Food Engineering, University College Dublin, Belfield, Dublin 4, Ireland

ABSTRACT

Background: The transmissibility of SARS-CoV-2 determines both the ability of the virus to invade a population and the strength of intervention that would be required to contain or eliminate the spread of infection. The basic reproduction number, R_0 , provides a quantitative measure of the transmission potential of a pathogen.

Objective: Conduct a scoping review of the available literature providing estimates of R_0 for SARS-CoV-2, provide an overview of the drivers of variation in R_0 estimates and the considerations taken in the calculation of the parameter.

Design: Scoping review of available literature between the 01 December 2019 and 07 May 2020.

Data sources: Both peer-reviewed and pre-print articles were searched for on PubMed, Google Scholar, MedRxiv and BioRxiv.

Selection criteria: Studies were selected for review if (i) the estimation of R_0 for SARS-CoV-2 represented either the initial stages of the outbreak or the initial stages of the outbreak prior to the onset of widespread population restriction (“lockdown”), (ii) the exact dates of the study period were provided and (iii) the study provided primary estimates of R_0 .

Results: A total of 20 R_0 for SARS-CoV-2 estimates were extracted from 15 studies. There was substantial variation in the estimates reported. Estimates derived from mathematical models fell within a wider range of 1.94-6.94 than statistical models which fell between the range of 2.2 to 4.4. Several studies made assumptions about the length of the infectious period which ranged from 5.8-20 days and the serial interval which ranged from 4.41-14 days. For a given set of parameters a longer duration of infectiousness or a longer serial interval equates to a higher R_0 . Several studies took measures to minimise bias in early case reporting, to account for the potential occurrence of super-spreading events, and to account for early sub-exponential epidemic growth.

Conclusions: The variation in reported estimates of R_0 reflects the complex nature of the parameter itself, including the context (i.e. social/spatial structure), the methodology used to estimate the parameter, and model assumptions. R_0 is a fundamental parameter in the study of infectious disease dynamics, however it provides limited practical applicability outside of the context in which it was estimated, and should be calculated and interpreted with this in mind.

STRENGTHS AND LIMITATIONS OF THE SCOPING REVIEW

- This study provides an overview of basic reproduction number estimates for SARS-CoV-2 across a range of settings, a fundamental parameter in gauging the transmissibility of an emerging infectious disease.
- The key drivers of variation in R_0 estimates and considerations in the calculation of the parameter highlighted across the reviewed studies are discussed.
- This evidence may be used to help inform modelling studies and intervention strategies.
- Given the need for rapid dissemination of information on a newly emerging infectious disease, several of the reviewed papers were in the pre-print phase yet to be peer-reviewed.

INTRODUCTION

On 31 December 2019 a series of cases of pneumonia of unknown cause were notified to the World Health Organization (WHO). The *Coronaviridae* Study Group (CSG) of the International Committee of Taxonomy of Viruses designated this novel virus as SARS-CoV-2, the etiologic agent of coronavirus disease 2019 (COVID-19)[1]. While the origin of COVID-19 in humans has been attributed to spillover from an unknown wildlife source, human-human transmission has been responsible for the rapid spread of SARS-CoV-2 across the globe. On 11 March 2020, the WHO characterised COVID-19 as a pandemic[2].

An understanding of the transmissibility of a newly emerging infectious disease is required to determine the ability of a pathogen to spread and establish within a population along with the strength of mitigation required to contain or eliminate infection[3]. The basic reproduction number, R_0 , is an indicator of the transmissibility of an infectious agent, defined as the expected number of new infections that are generated, on average, by a single infected individual, over the course of its infectious period, in an otherwise uninfected population[4]. It is a threshold parameter: with a value above one, an infection can spread and persist within a population, below one infection cannot be sustained. R_0 depends on the average contact rate between susceptible and infectious individuals (or infectious material of infectious individuals) per unit time, the (dimensionless) probability that the contact leads to infection and the average duration (time) of the infectious period[3]. The effective reproduction number, R , is a dynamic parameter that can chronicle the time-dependent (R_t) variation in transmission as a result of, for example; the intrinsic decline in the proportion of susceptibles (e.g. infection and natural immunity) and, extrinsic effects of intervention on effective contact rates (e.g. social distancing), the probability that upon contact infection occurs (e.g. face masks) or the effective duration of infectiousness (e.g. self-isolation)[5]. It is an important parameter for monitoring the

effectiveness of control measures and determining whether further measures are required. For the purpose of this review, we focus primarily on the basic reproduction number, R_0 .

The aim of this study was to conduct a scoping review of the available literature providing estimates on the basic reproduction number of SARS-CoV-2, to gain an understanding of the transmissibility of the virus and to aid the parameterisation of COVID-19 epidemiological models. R_0 is a complex parameter of which the calculation and interpretation requires significant consideration. As such, we discuss and aim to disentangle several relevant concepts highlighted across the reviewed papers including the context, methodological differences, the relationship between R_0 and the generation time/serial interval, the relationship between R_0 and the infectious period, managing surveillance bias, individual variation in transmission and exponential epidemic growth.

MATERIALS AND METHODS

Search methodology, initial screening and categorisation:

A survey of the literature between 01 December 2019 and 07 May 2020 for all countries was implemented, as part of a larger research project characterising key parameters of COVID-19[6–10]. The following search strategy was used. Publications listed in the electronic databases PubMed, Google Scholar, MedRxiv and BioRxiv were searched with the following keywords: (“Novel coronavirus” OR “SARS-CoV-2” OR “2019-nCoV” OR “COVID-19”) AND “reproduction number”. The dynamic curated PubMed database “LitCovid” was also monitored, in addition to national and international government reports. No restrictions on language or publication status were imposed so long as an English abstract was available. On the 23 July 2020, pre-print studies were checked for changes in publication status and any subsequent changes in estimates (Supplementary Table 1.). Articles were evaluated for data relating to the aim of this review, and all relevant publications were considered for possible inclusion. Bibliographies within these publications were also searched for additional resources. This study is reported in compliance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses – Extension for Scoping Reviews (PRISMA-ScR) checklist[11].

Study appraisal and selection for review:

Studies were selected for review if they met the following criteria: (i) the estimation of R_0 represented either the initial stages of the outbreak or the initial stages of the outbreak prior to the onset of widespread population restriction (“lockdown”), (ii) the exact dates of the study period were provided, and (iii) the study provided primary estimates of R_0 . A period prior to lockdown was selected as a compulsory requirement to ensure estimates provided were calculated under the realisation of an entirely susceptible population and any imposed disturbance on effective contact rates or infectious

periods were minimised, as per the definition of R_0 . Study selection was undertaken by the first author. Parameter estimates for R_0 and confidence intervals (where provided) were recorded and assessed. The study period, location of the study, methodology and model assumptions were also extracted.

RESULTS

Study appraisal and selection for review:

There were 33 studies available for appraisal as of 07 May 2020 (Fig. 1).

- 14 papers were removed as the study period extended into a period of lockdown[12–25].
- 1 paper was excluded as the study period was not clearly defined[26].
- 1 paper estimated R_0 through phylogenetic analysis of 52 full genomes of viral strains sampled across different countries[27]. As this paper did not deal with a single population with defined study periods, it was not considered for review.
- 2 papers were excluded as they only considered the time-dependent reproduction number[28,29].

Following the removal of these papers, 15 papers were further evaluated, with estimates on the basic reproduction number provided in Table 1. Across these studies, estimates of R_0 are provided for populations within China ($n = 6$ studies), Italy ($n = 3$ studies), Iran ($n = 2$ studies), the Republic of Korea ($n = 1$ study), France ($n = 1$ study), the United Kingdom ($n = 1$ study), Spain ($n = 1$ study) and onboard the Diamond Princess cruise ship ($n = 1$ study).

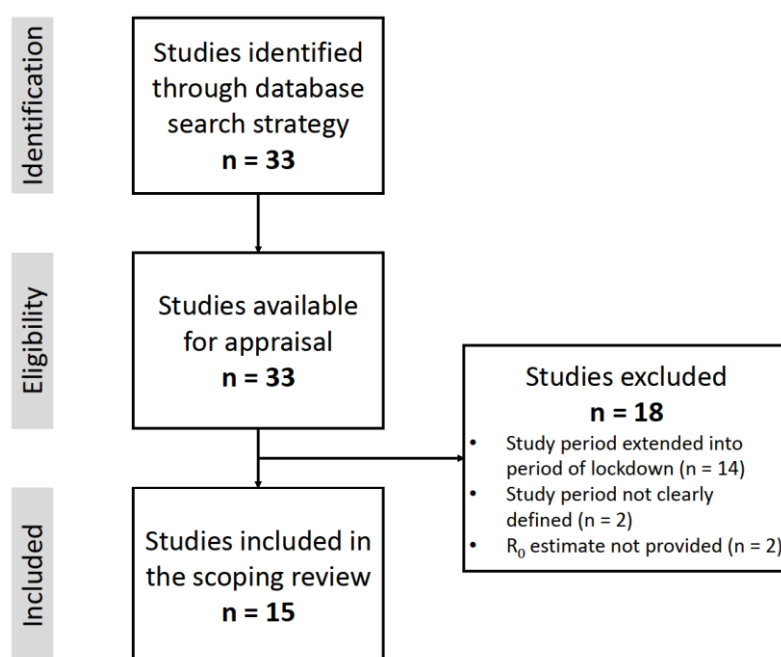


Figure 1. Flowchart for the studies included in the scoping review.

Overall findings:

From the 15 studies included, a total of 20 R_0 estimates were reported, ranging from 1.94 to 6.94 (Table 1). R_0 estimates varied both within countries, (for example, estimates in China, ranged between 1.94 and 4.71), and between countries (for example, an R_0 of 2.2 was reported for Spain, 3.2 for France and 6.94 for the UK) (Fig. 2).

Seven studies deployed statistical models and eight mathematical models (see Supplementary Table 2 for further methodological details). There were 2 types of statistical frameworks and 4 types of mathematical frameworks used across the studies. Overall, there were 10 types of inference methods used, 2 of which (Markov Chain Monte Carlo (MCMC) and Maximum Likelihood Estimation (MLE)) were used within both statistical and mathematical frameworks while 5 types of inference methods were used exclusively within statistical frameworks and three within mathematical frameworks.

Assumptions surrounding the duration of the infectious period and the serial interval were present across several of the studies. Assumptions about the length of the infectious period ranged between 5.8 to 20 days and assumptions on the serial interval ranged between 4.4 to 14 days. Three studies applied methods to control for surveillance bias in case reporting[30–32], the potential occurrence of individual variation in transmission was included in two studies[30,33], and one study accounted for sub-exponential growth in the early stages of the outbreak[34].

Table 1. Estimates of the basic reproduction number, R_0 , provided by the 15 reviewed studies.

Location	Study Period	R_0 Estimate	95% CI	Framework	Inference method	Assumptions	Study
Rep. of Korea	31/01/20-05/03/20	2.6	2.3-2.9	Stochastic simulations of early outbreak trajectories	Maximum Likelihood Estimation	Assume exponential growth started on 31/02/20. Assume a mean serial interval of 4.5 days (SD: 3.1 days)	Zhuang <i>et al.</i> [33]
Rep. of Korea	05/02/20-05/03/20	3.2	2.9-3.5	Stochastic simulations of early outbreak trajectories	Maximum Likelihood Estimation	Assume exponential growth started on 05/02/20. Assume a mean serial interval of 4.5 days (SD: 3.1 days)	Zhuang <i>et al.</i> [33]
China	Up to 18/01/20	2.2	1.4-3.8	Stochastic simulations of early outbreak trajectories	Approximate Bayesian Computation	Assume a mean generation time between 7-14 days	Riou and Althaus.[30]
Wuhan, China	07/12/19-01/01/20	3.58	-	Network transmission model	Next Generation Matrix	Assume an infectious period of 5.8 days	Chen <i>et al.</i> [35]
Wuhan, China	10/12/19-03/01/20	1.94	1.83-2.06	SIR transmission model	Next Generation Matrix	-	Wu <i>et al.</i> [36]
Wuhan, China	10/12/19-04/01/20	2.2	1.4-3.9	Exponential Growth Model	R_0 derived from the growth rate of the	Assume a mean serial interval of 8.4 days (SD: 3.8 days)	Li <i>et al.</i> [37]

medRxiv preprint doi: <https://doi.org/10.1101/2020.07.28.20163535>; this version posted July 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

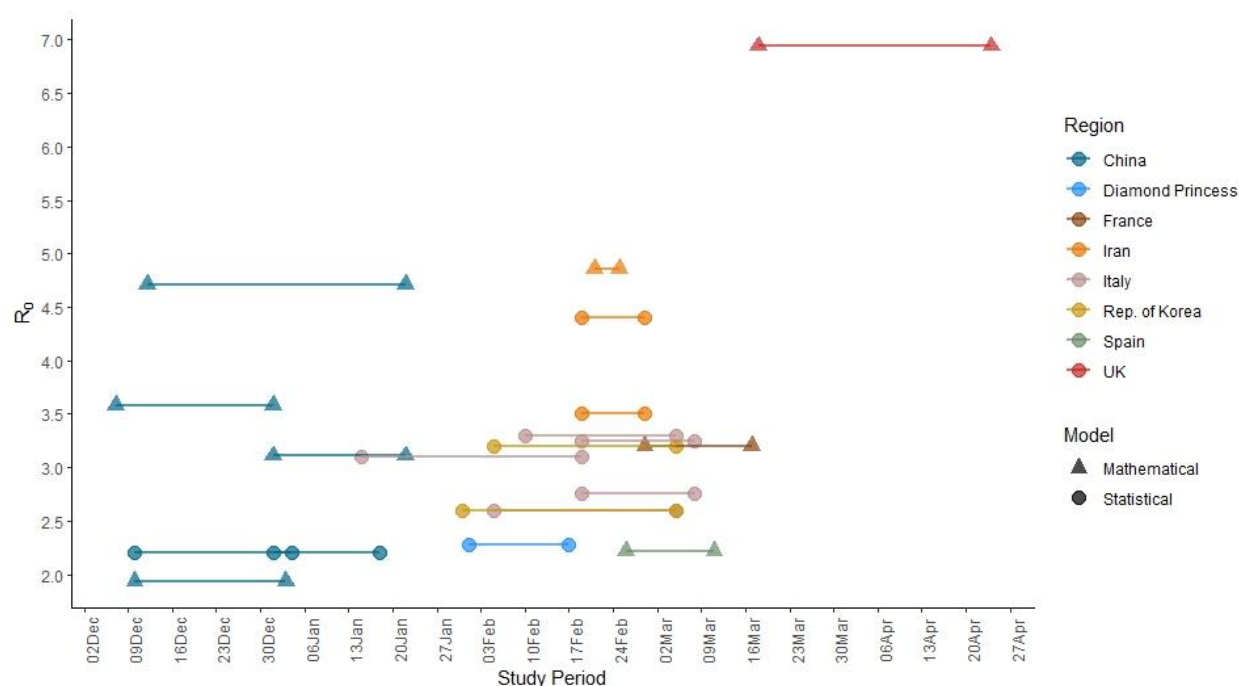


Figure 2. Relationship between start and end dates of the study period and estimated R_0 across region and model.

DISCUSSION

In this review we found a total of 20 reported R_0 estimates from 15 studies which ranged from 1.94 to 6.94. There was substantial variation in estimates both within and between countries. We found 12 values of R_0 were estimated within statistical frameworks and 8 values were estimated within mathematical frameworks. The duration of the infectious period and the serial interval were common assumptions ranging in value across several studies. We identified three key potential drivers of variation in reported estimates of R_0 including: i) the complex nature of R_0 itself, in that it is context specific, ii) methodological differences and iii) the model assumptions (including the duration of the infectious period and serial interval). Further we discuss how studies sought to minimise surveillance bias, the importance of individual variation in transmission, and the occurrence of exponential growth at the early stages of an outbreak.

Drivers of variation:

i) Context

A key component of R_0 is the effective contact rate (a function of the average contact rate and probability that contact leads to infection), which can vary over space or time. Traditional epidemiological models assume homogeneous mixing of individuals, where equal contact rates are assumed between individuals within a population. Applying this simplifying assumption allows for more tractable modelling and analysis but in reality heterogeneities in contact patterns are introduced by, for example; population density and social organisation[45]. In density dependent systems, transmission increases linearly with population density, and a threshold density is required for infection to persist[46]. For example; density dependent infections are more likely to propagate and persist in urban settings compared to rural. Social structure also plays an important role in infection transmission, where infectious individuals are more likely to infect members of the same social group[47]. R_0 is therefore considered a function of the factors capable of influencing contact rates along with the biological components of the pathogen (e.g. the probability that contact leads to infection and the length of the infectious period)[48]. R_0 provides limited practical applicability outside of the context in which it was estimated; therefore, R_0 for one geographical/social setting may not be applicable to that of another[49]. Even within the same region variation in estimates occur. To illustrate, 4 estimates were reported for Wuhan, which ranged from 1.94-4.71.

ii) Methodological differences

Statistical and mathematical models in epidemiology play a fundamental role in providing an understanding of transmission dynamics and evaluating the effectiveness of control strategies. Statistical methods to estimate R_0 which rely on incidence data and are generally in the form of either descriptive non-dynamical models or stochastic models. At the early phase of an epidemic, the number of new infections per unit time increases exponentially and R_0 can be estimated in descriptive non-dynamical models, for example exponential growth models, from the empirically observed growth rate p of the epidemic curve[50,51]. In contrast, stochastic models, for example branching process models, simulate early outbreak trajectories based on the generation concept whereby each infected individual i is associated with a generation time T_G . At the end of T_G , a random number N_i of new infections ('offspring') have been generated, where R_0 is the offspring mean[52–54]. Furthermore, where data on chains of transmission (i.e. who infected whom) are available, the number of secondary infections generated by each infected individual can be simply counted. In the current study values of R_0 estimated by means of statistical methods fell within the range of 2.2-4.4.

Fitting mechanistic models to epidemic data can provide useful insights about transmission dynamics. These models are capable of describing the unobservable mechanism of transmission between

individuals whereby dynamics at the population level are composed of the sum of individual-level processes of which the observed data are generated. Furthermore, they provide a suitable framework to study the effects of control measures on the spread of infection[52,55,56]. R_0 can be calculated directly from model parameters or coupled with statistical techniques[57]. In the current study, R_0 values estimated by mathematical models fell within the range of 1.94 to 6.94, a substantially wider range compared to that of estimates generated within statistical frameworks. Transmission models range in complexity depending on the types of questions addressed, subject to trade-off between structuring the model to conform with biological reality and the ability to sensibly parameterise and generalise the system[58]. As such the wider variation across these models may be attributed to model structure or assumptions, along with the context in which R_0 was estimated.

Several studies deployed a set of ordinary differential equations to model disease spread[31,32,35,36,38–40,43]. Parsimonious models may be robust to difficulties in model parameterisation associated with limited data availability[32,40]. However, as a consequence, they lack the ability to capture more realistic dynamics such as pre- and asymptomatic transmission, susceptibility to different levels of clinical severity in infection, or age-related infection dynamics. In contrast, a reservoir-people transmission network model simulating the spread of infection captured spill-over from a wildlife reservoir and human-human transmission as well as incorporating variation in transmission between symptomatic and asymptomatic infected individuals[35]. Although this model describes the detailed process of infection spread associated with SARS-COV-2, model parameterisation is heavily dependent on limited data sourced from the literature, unreliable mobility data, and assumptions surrounding the proportion and relative transmissibility of asymptomatic infection. Complex models may provide a good fit to the data, however, any relationship with the underlying process may not necessarily be interpretable. Further, complex models may overfit the data and thus; generalise poorly. In general, parsimonious models are preferred to complex ones and models should not contain more than the minimum necessary assumptions[59]. While model structure may, in part, account for variation in R_0 estimates, even within models with similar structure variation in estimates occur, reflecting differences in the context or model assumptions.

iii) Model assumptions

On the relationship between R_0 and the duration of the infectious period: The infectious period, the time between the end of the pre-infectious period to when an individual can no longer pass on infection, is a key component determining R_0 . It is a clinical characteristic which does not necessarily differ across settings. With a newly emerging infectious disease it is often difficult to quantify the duration of the infectious period where relatively little is known in relation to the occurrence or relative occurrence of events such as asymptomatic, pre-symptomatic and symptomatic infectiousness. As such, several studies made assumptions about the duration of the infectious period which ranged

from 5.8 to 20 days. One study derived R_0 from the growth rate of the epidemic curve and the duration of the infectious period which was assumed to be either 15 or 20 days[42]. For a given growth rate, an infectious period of 20 days resulted in a higher R_0 ($R_0 = 3.25$) than a shorter infectious period of 15 days ($R_0 = 2.76$). A longer infectious period relates to a longer time for potential transmission opportunities to occur. The inverse of the average duration of infectiousness is the recovery rate (that is, how quickly an individual exits a state of infectiousness) and so a shorter infectious period will result in individuals exiting the infectious state at a faster rate than that of a longer infectious period. Therefore, for a given set of parameters, a longer duration of infectiousness equates to a higher expected R_0 than that associated with a shorter infectious period.

On the relationship between R_0 and the generation time/serial interval: Several studies rely on the distribution of the generation time or its proxy the serial interval (time between the infection events or symptom onset of primary and secondary cases, respectively) in the estimation of R_0 [30,34,37,41,44]. The serial interval is a context-specific parameter that is sensitive to factors such as contact patterns. For example, if transmission following symptom onset is reduced due to isolation of symptomatic individuals, there may be relatively more pre-symptomatic transmission events and shorter serial intervals[9]. Ideally, the serial interval should be estimated from the corresponding population of interest. However, in reality this may not always be feasible, particularly at the early stages of an outbreak where data may be limited in availability. When there is a scarcity of data, studies often make assumptions about the distribution of the serial interval. This included assuming a serial interval distribution based on previous reports[34], and exploring a range of parameter combinations that include the distribution of the serial interval[30]. Two studies directly estimated the distribution of the serial interval by fitting a gamma distribution to data from the population under study[37,41]. One study investigated the sensitivity of the time-dependent reproduction number, R_t , to the serial interval and showing that although R_t followed the same trend across values of the serial interval, a longer serial interval resulted in a higher R_t estimate[41]. For a given rate of epidemic growth, a higher R will require fewer (so longer) generations of transmission chains to realise the same population of infection spread that would occur with a lower R (i.e. a shorter generation interval equates to faster transmission at the individual level)[60].

Minimising surveillance bias:

There are several challenges associated with incidence data at the early epidemic stages of a newly emerging infectious disease. Case reporting is often unstable during this period due to lack of knowledge or awareness of, or confusion with, case definitions, which can limit the ability of health officials to identify infected individuals. Further, during the early stages of an epidemic, diagnostic facilities are often lacking in quality and/or quantity. These factors may contribute to unreliable or

incomplete data limiting the ability to accurately characterise important parameters, such as R_0 , with high levels of precision early in an outbreak. If the rate of reporting remains stable over time, estimates of the reproduction number would be unaffected by underreporting[61]. However, as awareness increases, case definitions change and diagnostic facilities improve, variations in reporting over time may lead to erroneous interpretations of the reproduction number. Several of the reviewed studies took measures to minimise surveillance bias, including smoothing case and death data with a moving average over 5 days[32], gaining an indirect estimate of epidemic size through cases identified outside of the study population[30], and relying on hospital death data[31], which may be more stable than case reporting. These methods provide alternatives to relying on unreliable or unstable surveillance data at the early stages of an outbreak.

Individual variation in transmission:

Superspreading events (SSEs) describe situations where a small proportion of infectious individuals account for more than the expected number of transmission events. As R_0 describes a population-average of the number of secondary cases generated by a primary case, individual-level variation in transmission is not expressed. Greater levels of individual heterogeneity in transmission lead to a less frequent occurrence of outbreaks, however outbreaks that do occur, have a tendency to occur explosively[62,63]. Several cases of COVID-19 SSEs have been identified[64–66]. While SSEs often remain rare events, they can have important implications for the epidemic trajectory and control efforts[30]. Following the approach described by Lloyd-Smith *et al.*, (2005)[62], in the estimation of R_0 , two studies incorporated a dispersion parameter, k , that measures the likelihood of occurrence of SSEs[30,33]. The lower the value of k , the greater the level of individual heterogeneity in transmission. Both studies indicated that homogenous patterns of transmission events were more likely, although suggested that the potential for SSEs to occur should not be discounted[30]. These results are inconsistent with a study modelling overdispersion in transmission of COVID-19, which suggested that approximately 10% of infectious individuals were responsible for 80% of secondary transmissions where R_0 values ranged between 2 and 3[67]. Overdispersed transmission (high individual-level variation) has important implications for control measures, for example if transmission is highly overdispersed control efforts would be most effective by targeting situations where SSEs are likely to occur.

Exponential growth:

At the early stages of an epidemic, while the depletion of susceptibles is negligible the number of new cases per unit time grows exponentially. The growth rate of the exponential curve measures how quickly infection spreads and can be used to estimate R_0 . However, if the initial growth phase is slower than exponential (i.e. sub-exponential growth), R_0 estimates from exponential growth models may be inflated. The generalised growth model is capable of explicitly accounting for a sub-exponential growth phase[68,69]. One study applied this framework incorporating a scaling factor (the deceleration of growth parameter[68]) to account for the sub-exponential growth phase[34]. Although the effect of different scaling factor values on R_0 was not explored, it is expected that the further that epidemic growth is from exponential, the lower R_0 would be[68]. Consideration should therefore be given as to whether the early stage of the epidemic is truly growing exponentially when applying exponential growth models to estimate R_0 .

CONCLUSION

We found 20 reported R_0 estimates which ranged from 1.94 to 6.95. We identified several key drivers of variation in estimates including the context (i.e. social/geographic structure), methodology (i.e. mathematical vs statistical frameworks and model structure) and model assumptions (i.e. duration of the infectious period and serial interval). Notably, there was a wider variation in estimates generated within a mathematical framework (1.94-6.94) compared to a statistical framework, which fell within a narrower range of 2.2 to 4.4. However, variation across estimates is unlikely to be a result of a single driver in variation rather estimates differ across drivers or combinations of drivers. Here we have highlighted key potential sources of variation in R_0 and attempted to disentangle the effects these drivers have on the estimates generated. Further, we identified several key considerations accounted for, including: approaches to minimise bias introduced by unstable case reporting, the importance of individual variation in transmission and the early epidemic characterisation of exponential or sub-exponential growth. R_0 is a fundamental parameter in the study of infectious disease dynamics however the calculation and interpretation is not a straightforward exercise and careful consideration is required for both. Although this review details the some of these considerations, we have by no means exhausted all potential drivers of variation and we have not discussed all of the complexities associated with R_0 . However, we have discussed the key considerations that occurred within the reviewed studies in the context of COVID-19 and the theory is widely applicable across the board of infectious diseases which may help inform future studies attempting to calculate R_0 .

FOOTNOTES

Funding: All investigators are full-time employees (or retired former employees) of Wageningen University, University College Dublin or the Irish Department of Food and the Marine (DAFM). No additional funding was obtained for this research.

Author contributions: AB conducted the eligibility screening of shortlisted studies, extracted the data and reviewed the parameter estimates with input from all authors; ÁC, KH and FB conducted the initial literature searches; AB completed the initial drafts of the manuscript; All authors read earlier manuscript versions and approved the final manuscript.

Competing interests: None to declare.

Patient and public involvement statement: It was not appropriate or possible to involve patients or the public in the design, or conduct, or reporting, or dissemination plans of our research.

REFERENCES

- 1 Gorbalenya AE, Baker SC, Baric RS, *et al.* The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.* 2020. doi:10.1038/s41564-020-0695-z
- 2 WHO. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. 2020. <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>
- 3 Dietz K. The estimation of the basic reproduction number for infectious diseases. *Stat Methods Med Res* 1993;2:23–41. doi:10.1177/096228029300200103
- 4 Diekmann O, Heesterbeek JAPAP, Metz JAJAJ. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *J Math Biol* 1990;28:365–82. doi:10.1007/BF00178324
- 5 Nishiura H, Chowell G. The effective reproduction number as a prelude to statistical estimation of time-dependent epidemic trends. *Math Stat Estim Approaches Epidemiol* 2009;:103–21. doi:10.1007/978-90-481-2313-1_5
- 6 McAloon CG, Collins A, Hunt K, *et al.* The incubation period of COVID-19: A rapid systematic review and meta-analysis of observational research. 2020. doi:10.1101/2020.04.24.20073957
- 7 Byrne AW, Mcevoy D, Collins ÁB, *et al.* Inferred duration of infectious period of SARS-CoV-2: rapid scoping review 1 and analysis of available evidence for asymptomatic and symptomatic. *medRxiv* Published Online First: 2020. doi:10.1101/2020.04.25.20079889
- 8 Casey M, Griffin J, McAloon CG, *et al.* Estimating pre-symptomatic transmission of COVID-19: a secondary analysis using published data. 2020. doi:10.1101/2020.05.08.20094870
- 9 Griffin JM, Collins AB, Hunt K, *et al.* A rapid review of available evidence on the serial interval and generation time of COVID-19. 2020. doi:10.1101/2020.05.08.20095075

- 10 Lane EA, Barrett DJ, Casey M, *et al.* Country differences in hospitalisation, length of stay, admission to Intensive Care Units, and mortality due to SARS-CoV-2 infection at the end of the first wave in Europe: a rapid review of available literature. *medRxiv* 2020;:2020.05.12.20099473. doi:10.1101/2020.05.12.20099473
- 11 Tricco AC, Lillie E, Zarin W, *et al.* PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann Intern Med* 2018;**169**:467. doi:10.7326/M18-0850
- 12 Rovetta A, Bhagavathula AS. Modelling the epidemiological trend and behavior of COVID-19 in Italy. *medRxiv* Published Online First: 2020. doi:<https://doi.org/10.1101/2020.03.19.20038968>
- 13 Tindale L, Coombe M, Stockdale JE, *et al.* Transmission interval estimates suggest pre-symptomatic spread of COVID-19. *medRxiv* 2020;:2020.03.03.20029983. doi:10.1101/2020.03.03.20029983
- 14 Du Z, Xu X, Wu Y, *et al.* Serial Interval of COVID-19 among Publicly Reported Confirmed Cases. *Emerg Infect Dis* 2020;**26**:1–7. doi:10.3201/eid2606.200357
- 15 Liu T, Hu J, Kang M, *et al.* Time-varying transmission dynamics of Novel Coronavirus Pneumonia in China. *bioRxiv* 2020;:2020.01.25.919787. doi:10.1101/2020.01.25.919787
- 16 Wen Y, Wei L, Tang X, *et al.* Epidemiological and Clinical Characteristics of Coronavirus Disease 2019 in Shenzhen, the Largest Migrant City of China Authors: *medRxiv Prepr* 2020;**21**:1–9. doi:<https://doi.org/10.1101/2020.03.22.20035246>
- 17 Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet* 2020;**395**:689–97. doi:10.1016/S0140-6736(20)30260-9
- 18 Sanche S, Lin YT, Xu C, *et al.* The Novel Coronavirus, 2019-nCoV, is Highly Contagious and More Infectious Than Initially Estimated. *medRxiv* Published Online First: 2020. doi:<https://doi.org/10.1101/2020.02.07.20021154>
- 19 Khan MA, Atangana A. Modeling the dynamics of novel coronavirus (2019-nCoV) with fractional derivative. *Alexandria Eng J* Published Online First: 2020. doi:10.1016/j.aej.2020.02.033
- 20 Majumder M, Mandl KD. Early Transmissibility Assessment of a Novel Coronavirus in Wuhan, China. *SSRN Electron J* Published Online First: 2020. doi:10.2139/ssrn.3524675
- 21 Meng W, Jingtiao QI. A deterministic epidemic model for the emergence of COVID-19 in China. *medRxiv Prepr* 2020;:1–11. doi:<https://doi.org/10.1101/2020.03.08.20032854>
- 22 Qianqian S, Han Z, Liqun F, *et al.* Estimation of Epidemiological Parameters of Early Infectious Diseases of New Coronavirus Pneumonia. *Chinese J Epidemiol* 2020;**41**:461–5. doi:10.1088/1674-1056/23/3/035203
- 23 Wangping J, Ke H, Yang S, *et al.* Extended SIR prediction of the epidemics trend of COVID-19 in Italy and compared with Hunan , China. *medRxiv* 2020;**10**. doi:<https://doi.org/10.1101/2020.03.18.20038570>
- 24 Zhao S, Lin Q, Ran J, *et al.* Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. *Int J Infect Dis* 2020;**92**:214–7. doi:10.1016/j.ijid.2020.01.050

- 25 You C, Deng Y, Hu W, *et al.* Estimation of the Time-Varying Reproduction Number of COVID-19 Outbreak in China. *Lancet Public Heal* Published Online First: 2020. doi:<https://doi.org/10.1101/2020.02.08.20021253>
- 26 Li M, Chen P, Yuan Q, *et al.* Transmission characteristics of the COVID-19 outbreak in China: a study driven by data. *medRxiv* 2020;:2020.02.26.20028431. doi:10.1101/2020.02.26.20028431
- 27 Lai A, Bergna A, Acciarri C, *et al.* Early phylogenetic estimate of the effective reproduction number of SARS-CoV-2. *J Med Virol* 2020;:14–8. doi:10.1002/jmv.25723
- 28 Tang, Wang, Li, *et al.* Estimation of the Transmission Risk of the 2019-nCoV and Its Implication for Public Health Interventions. *J Clin Med* 2020;9:462. doi:10.3390/jcm9020462
- 29 Kucharski AJ, Russell TW, Diamond C, *et al.* Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect Dis* Published Online First: 2020. doi:10.1016/s1473-3099(20)30144-4
- 30 Riou J, Althaus CL. Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. *Euro Surveill* 2020;25:1–5. doi:10.2807/1560-7917.ES.2020.25.4.2000058
- 31 Dropkin G. COVID-19 UK Lockdown Forecasts and R0. *Front Public Heal* Published Online First: 2020. doi:10.3389/fpubh.2020.00256
- 32 Roques L, Klein EK, Papaix J, *et al.* Using early data to estimate the actual infection fatality ratio from covid-19 in france. *Biology (Basel)* Published Online First: 2020. doi:10.3390/biology9050097
- 33 Zhuang Z, Zhao S, Lin Q, *et al.* Preliminary estimates of the reproduction number of the coronavirus disease (COVID-19) outbreak in Republic of Korea and Italy by 5 March 2020. *Int J Infect Dis* Published Online First: 2020. doi:10.1016/j.ijid.2020.04.044
- 34 Muniz-Rodriguez K, Fung ICH, Ferdosi SR, *et al.* Severe Acute Respiratory Syndrome Coronavirus 2 Transmission Potential, Iran, 2020. *Emerg Infect Dis* Published Online First: 2020. doi:10.3201/eid2608.200536
- 35 Chen TM, Rui J, Wang QP, *et al.* A mathematical model for simulating the phase-based transmissibility of a novel coronavirus. *Infect Dis Poverty* 2020;9:1–8. doi:10.1186/s40249-020-00640-3
- 36 Wu JT, Leung K, Bushman M, *et al.* Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nat Med* 2020;:1–5. doi:10.1038/s41591-020-0822-7
- 37 Li Q, Guan X, Wu P, *et al.* Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *N Engl J Med* 2020;:1–9. doi:10.1056/nejmoa2001316
- 38 Shen M, Peng Z, Xiao Y, *et al.* Modelling the epidemic trend of the 2019-nCoV outbreak in Hubei Province, China. *medRxiv* 2020;:2020.01.30.20019828. doi:10.1101/2020.01.30.20019828
- 39 Read JM, Bridgen JRRE, Cummings DATA, *et al.* Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. *medRxiv* 2020;53:11. doi:10.1017/CBO9781107415324.004
- 40 Sahafizadeh E, Sartoli S. Estimating the reproduction number of COVID-19 in Iran using

- epidemic modeling. 2020;:0–4.
- 41 Cereda D, Tirani M, Rovida F, *et al.* The early phase of the COVID-19 outbreak in Lombardy, Italy. Published Online First: 2020.<http://arxiv.org/abs/2003.09320>
- 42 Remuzzi A, Remuzzi G. COVID-19 and Italy: what next? *Lancet* 2020;**2**:10–3. doi:10.1016/s0140-6736(20)30627-9
- 43 Garcia-Iglesias D, Juez FJ de C. Early behavior of Madrid Covid-19 disease outbreak: A mathematical model. *medRxiv* 2020;**093**:2020.03.30.20047019. doi:10.1101/2020.03.30.20047019
- 44 Zhang S, Diao MY, Yu W, *et al.* Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A data-driven analysis. *Int J Infect Dis* 2020;**93**:201–4. doi:10.1016/j.ijid.2020.02.033
- 45 Anderson R, May R. *Infectious Diseases of Humans: Dynamics and Control*. Oxford: : Oxford University Press 1992.
- 46 Ryder JJ, Miller MR, White A, *et al.* Host-parasite population dynamics under combined frequency- and density-dependent transmission. *Oikos* Published Online First: 2007. doi:10.1111/j.2007.0030-1299.15863.x
- 47 Britton T, Deijfen M, Lagerås AN, *et al.* Epidemics on random graphs with tunable clustering. *J Appl Probab* Published Online First: 2008. doi:10.1239/jap/1222441827
- 48 Delamater PL, Street EJ, Leslie TF, *et al.* Complexity of the basic reproduction number (R0). *Emerg Infect Dis* Published Online First: 2019. doi:10.3201/eid2501.171901
- 49 Ridenhour B, Kowalik JM, Shay DK. Unraveling R0: Considerations for public health applications. *Am J Public Health* Published Online First: 2018. doi:10.2105/AJPH.2013.301704
- 50 Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc R Soc B Biol Sci* Published Online First: 2007. doi:10.1098/rspb.2006.3754
- 51 Vynnycky E, White R. *An Introduction to Infectious Disease Modelling*. Oxford University Press 2010.
- 52 Jacob C. Branching processes: Their role in epidemiology. *Int J Environ Res Public Health* Published Online First: 2010. doi:10.3390/ijerph7031204
- 53 Yan P. Distribution Theory, Stochastic Processes and Infectious Disease Modelling. In: Brauer F, van den Driessche P, Wu J, eds. *Mathematical Epidemiology*. Berlin, Heidelberg: : Springer Berlin Heidelberg 2008. 229–93. doi:10.1007/978-3-540-78911-6_10
- 54 Becker N. On parametric estimation for mortal branching processes. *Biometrika* Published Online First: 1974. doi:10.1093/biomet/61.2.393
- 55 Kretzschmar M. Measurement and Modeling: Infectious Disease Modeling. *Int Encycl Public Heal* 2016;:579–85. doi:10.1016/B978-0-12-803678-5.00229-0
- 56 Roberts M, Heesterbeek H. Bluff your way in epidemic models. *Trends Microbiol* 1993;**1**:343–8. doi:10.1016/0966-842X(93)90075-3

- 57 Cauchemez S, Ferguson NM. Methods to infer transmission risk factors in complex outbreak data. *J R Soc Interface* Published Online First: 2012. doi:10.1098/rsif.2011.0379
- 58 O'Neill PD. A tutorial introduction to Bayesian inference for stochastic epidemic models using Markov chain Monte Carlo methods. *Math Biosci* Published Online First: 2002. doi:10.1016/S0025-5564(02)00109-8
- 59 May RM. Uses and Abuses of Mathematics in Biology. *Science* (80-.). 2004. doi:10.1126/science.1094442
- 60 Park SW, Champredon D, Weitz JS, *et al.* A practical generation-interval-based approach to inferring the strength of epidemics from their speed. *Epidemics* 2019;**27**:12–8. doi:10.1016/j.epidem.2018.12.002
- 61 Thompson RN, Stockwin JE, van Gaalen RD, *et al.* Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics* Published Online First: 2019. doi:10.1016/j.epidem.2019.100356
- 62 Lloyd-Smith JO, Schreiber SJ, Kopp PE, *et al.* Superspreading and the effect of individual variation on disease emergence. *Nature* 2005;**438**:355–9. doi:10.1038/nature04153
- 63 Lipsitch M, Cohen T, Cooper B, *et al.* Transmission Dynamics and Control of Severe Acute Respiratory Syndrome. *Science* (80-) 2003;**300**:1966–70. doi:10.1126/science.1086616
- 64 Hodcroft EB. Preliminary case report on the SARS-CoV-2 cluster in the UK, France, and Spain. *Swiss Med Wkly* Published Online First: 2020. doi:10.4414/smw.2020.20212
- 65 Shim E, Tariq A, Choi W, *et al.* Transmission potential of COVID-19 in South Korea Authors: *medRxiv* Published Online First: 2020. doi:https://doi.org/10.1101/2020.02.27.20028829
- 66 Hu K, Zhao Y, Wang M, *et al.* Identification of a super-spreading chain of transmission associated with COVID-19. *medRxiv* Published Online First: 2020. doi:https://doi.org/10.1101/2020.03.19.20026245
- 67 Endo A, Abbott S, Kucharski AJ, *et al.* Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. *Wellcome Open Res* Published Online First: 2020. doi:10.12688/wellcomeopenres.15842.1
- 68 Chowell G, Viboud C, Simonsen L, *et al.* Characterizing the reproduction number of epidemics with early subexponential growth dynamics. *J R Soc Interface* Published Online First: 2016. doi:10.1098/rsif.2016.0659
- 69 Viboud C, Simonsen L, Chowell G. A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks. *Epidemics* Published Online First: 2016. doi:10.1016/j.epidem.2016.01.002

SUPPLEMENTARY MATERIAL

Supplementary Table 1: All studies initially reviewed (extracted up to 07 May 2020) in the pre-print phase were revisited on 23 July 2020 to check for updates in publication status and/or changes in reported estimates. 5 pre-print studies remain in the pre-print phase without changes to estimates of R_0 .(1–5) 1 pre-print study was subsequently peer-reviewed with no changes to the initial estimate.(6) 2 pre-print studies were subsequently peer-reviewed with changes to the initial reported estimates – in both cases there were changes to the study period and thus, data informing the analyses.(7,8) One study initially provided estimates for both the Republic of Korea and France, the peer-reviewed version did not contain the Rep. of Korea analysis and there was a change in the estimate reported for France (initially assumed an infectious period of 20 days, subsequently reduced to 10 days).(9)

Supplementary Table 2: Studies were categorised into 2 frameworks based on the modelling approach: 1) mathematical models and 2) statistical models. Mathematical models were either in the form of compartmental epidemic models (e.g. Susceptible-Infectious-Recovered models) or network models (e.g. Reservoir-Human and Human-Human models). Statistical frameworks included those of epidemic growth models (e.g. exponential/generalised growth models) and stochastic simulations of early outbreak trajectories (e.g. branching process models). The inference method refers to the method of calculation of R_0 . Several approaches were used to calculate R_0 , for example; directly obtaining the parameter from models (e.g. Maximum Likelihood Estimation (MLE) of the mean (R_0) of the offspring distribution) or deriving R_0 from model parameters (e.g. the transmission rate parameter divided by the recovery rate parameter or $1 + \text{growth rate} \times \text{serial interval}$).

Supplementary Table 1. Initial and updated estimates of the basic reproduction number, R_0.							
Paper	Initial R_0 Estimate	Initial 95% CI	Initial Status	Updated R_0 Estimate	Updated 95% CI	Updated Status	Comments
Garcia-Iglesias and Juez.(1)	2.22	± 1.21 SD	Pre-print	2.22	± 1.21 SD	Pre-print	No changes to initial estimates
Read <i>et al.</i> (2)	3.11	2.39-4.13	Pre-print	3.11	2.39-4.13	Pre-print	No changes to initial estimates
Shen <i>et al.</i> (3)	4.71	4.50-4.92	Pre-print	4.71	4.50-4.92	Pre-print	No changes to initial estimates
Cereda <i>et al.</i> (4)	3.1	2.9-3.2	Pre-print	3.1	2.9-3.2	Pre-print	No changes to initial estimates
Sahafizadeh and Sartoli.(5)	4.86	-	Pre-print	4.86	-	Pre-print	No changes to initial estimates
Zhuang <i>et al.</i> (6)	2.6	2.3-2.9	Pre-print	2.6	2.3-2.9	Peer-reviewed	No changes to initial estimates
	3.2	2.9-3.5		3.2	2.9-3.5		
	2.6	2.3-2.9		2.6	2.3-2.9		
	3.3	3.0-3.6		3.3	3.0-3.6		
Muniz-Rodriguez <i>et al.</i> (8)	3.6	3.2-4.2	Pre-print	4.4	3.9-4.9	Peer-reviewed	Data updated from 19 Feb - 29 Feb to 19 Feb - 01 Mar
	3.58	1.29-8.46		3.5	1.28-8.14		
Dropkin.(7)	5.81	5.08-6.98	Pre-print	6.94	6.52-7.39	Peer-reviewed	Data updated from 28 Feb - 23 Mar to 18 Mar - 24 Apr
Roques <i>et al.</i> (9)	2.6	-	Pre-print	-	-	-	Korean analysis excluded.
	4.8	-	Pre-print	3.2	3.1-3.3	Peer-reviewed	Initial estimate based on infectious period of 20 days, current estimate based on infectious period of 10 days.

Supplementary Table 2. Summary of the framework and inference method adopted across the 15 reviewed studies.		
Study	Framework	Inference method
Zhuang <i>et al.</i> (6)	Statistical framework: Cases generated by a primary case followed a Negative-Binomial distribution with mean R_0 and dispersion parameter k , representing the likelihood of superspreading events occurring. Values reported for Republic of Korea and Italy across different starting points of exponential growth.	Mean R_0 and dispersion parameter k obtained by means of maximum likelihood estimation (MLE).
Riou and Althaus.(10)	Statistical framework: Number of secondary cases generated by a primary case followed a Negative-Binomial distribution with mean R_0 and dispersion parameter k , representing the likelihood of superspreading events occurring.	Approximate Bayesian computation (ABC) to estimate the posterior distributions of the parameters that led to the expected outbreak trajectories.
Chen <i>et al.</i> (11)	Mathematical framework: Reservoir-people transmission network model. Includes spill-over from a wildlife reservoir as well as human-human transmission.	Next Generation Matrix (NGM) describing transmission across symptomatic persons, asymptomatic persons and a reservoir host.
Wu <i>et al.</i> (12)	Mathematical framework: Susceptible-Infected-Recovered (SIR) transmission model.	NGM describing the transmission possibilities across 9 age cohorts.
Li <i>et al.</i> (13)	Statistical framework: Epidemic curve derived from transmission model fitted through renewal equations.	R_0 estimated using the epidemic curve and distribution of the serial interval.
Shen <i>et al.</i> (3)	Mathematical framework: SEIJR model where the J compartment represents individuals who have been isolated and received treatment.	MCMC Metropolis-Hastings sampling to obtain estimates of the model parameters. R_0 calculated as the transmission rate parameter divided by the rate of identification and isolation of infectious individuals.
Read <i>et al.</i> (2)	Mathematical framework: SEIR metapopulation model simulating transmission between and within major Chinese cities.	The transmission rate parameter was obtained by means of maximum likelihood estimation and R_0 was calculated as the transmission rate divided by the recovery rate.
Muniz-Rodriguez <i>et al.</i> (8)	Statistical framework: Method 1: Fit daily incidence to a generalised growth model (GGM). Method 2: Epidemic doubling time obtained from the cumulative curve of daily reported cases to estimate the epidemic growth rate ($\ln(2)/\text{doubling time}$).	R_0 obtained via a renewal equation informed by 1) the daily reported incidence characterised by a generation growth model and 2) the distribution of the serial interval. Given the growth rate and a gamma-distributed serial interval, R_0 was calculated as $1 + \text{growth rate} \times \text{serial interval}$.
Sahafizadeh and Sartoli.(5)	Mathematical framework: SIR transmission model.	Tuned the transmission rate and the recovery rate parameters to get the best fit plot on the reported data. R_0 was calculated as the transmission rate divided by the recovery rate.
Cereda <i>et al.</i> (4)	Statistical framework: Daily incidence approximated via a Poisson process following a renewal equation with parameters $R(t)$ and $\phi(s)$. Where $R(t)$ is the reproduction number at time t and $\phi(s)$ is the distribution of the generation time at time s .	Based on the observed timeseries, the posterior distribution of the $R(t)$ was obtained by MCMC Metropolis-Hastings sampling and R_0 was estimated as $R(t)$ during the early exponential growth phase of the epidemic.
Remuzzi and Remuzzi.(14)	Statistical framework: Fitted daily incidence to exponential growth model (EGM).	R_0 obtained from the growth rate of the exponential curve and the duration of the infectious period which was assumed to range from 15 to 20 days. Actual calculation of R_0 used is unclear.
Roques <i>et al.</i> (9)	Mathematical framework: Susceptible-Infectious-Recovered (SIR)- compartmental model.	MCMC to estimate the posterior distribution of the transmission rate parameter, a partial parameter of the maximum likelihood estimator applied to the SIR model. R_0 was calculated as the transmission rate divided by the recovery rate (inverse of the assumed infectious period).
Dropkin.(7)	Mathematical framework: Susceptible-Exposed-Infected-Removed (SEIR) compartmental model. With 3 infected compartments representing the progression of clinical severity (mild, severe, critical).	$R_0 = N \frac{1}{p_1 + \gamma_1} (\beta_1 \frac{p_1}{p_2 + \gamma_2} (\beta_2 + \beta_3 \frac{p_2}{\mu + \gamma_3}))$ <p>Where N is the total population size, p_i is the rate of progression from compartment I_i to compartment I_{i+1}, γ_i is recovery rate from compartment I_i, β_i is the transmission rate parameter for compartment I_i, μ is the death rate for individuals in the critical compartment. Parameters estimated via Metropolis-Hastings sampling.</p>
Garcia-Iglesias and Juez.(1)	Mathematical framework: SEIR compartmental model.	Sequential Monte Carlo simulations ran over different values of R_0 . Estimated outcomes compared to reported cases to calculate the most probable R_0 .
Zhang <i>et al.</i> (15)	Statistical framework: Daily incidence approximated via a branching process with a Poisson likelihood.	R_0 obtained via MLE based on an assumed distribution of the serial interval and incidence data.