Encoding generation time changes within reproduction numbers

Kris V Parag[1,*], Benjamin J Cowling[2] and Ben C Lambert[3]

[1]MRC Centre for Global Infectious Disease Analysis, Imperial College London, London, UK.

[2]WHO Collaborating Centre for Infectious Disease Epidemiology and Control, School of Public Health, The University of Hong Kong, Hong Kong

[3]Department of Mathematics, College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, UK.

*For correspondence: k.parag@imperial.ac.uk.

Abstract

We introduce the *angular reproduction number* Ω, which measures time-varying changes in epidemic transmissibility resulting from variations in both the effective reproduction number R, and generation time distribution w. Predominant approaches for tracking pathogen spread either infer R or the epidemic growth rate r. However, R is biased by mismatches between the assumed and true w, while r is difficult to interpret in terms of the individual-level branching process underpinning transmission. R and r may also disagree on the relative transmissibility of epidemics or variants (i.e., $r_A>r_B$ does not imply $R_A>R_B$ for variants A and B). We find that Ω responds meaningfully to mismatches in w while maintaining most of the interpretability of R. Additionally, we prove that $\Omega>1$ implies $R>1$ and that Ω agrees with r on the relative transmissibility of pathogens. Estimating Ω is no harder than inferring R, uses existing software, and requires no generation time measurement. These advantages come at the expense of selecting one free parameter. We propose Ω as a useful statistic for tracking and comparing infectious disease dynamics that may better reflect the impact of interventions when those interventions concurrently change both R and w or alter the relative risk of co-circulating pathogens.

Keywords: infectious diseases; epidemic models; reproduction numbers; generation times; growth rates; transmission dynamics.

Introduction

Estimating the rate of spread or transmissibility of an infectious disease is a fundamental and ongoing challenge in epidemiology [1]. Identifying salient changes in pathogen transmissibility can contribute important information to policymaking, providing useful warnings of resurgent epidemics, assessments of the efficacy of interventions and signals about the emergence of new variants of concern [1–3]. The effective or instantaneous reproduction number, R, and time-varying growth rate, r, are commonly used to characterise pathogen transmissibility. The

former statistic is an estimate of the average number of new infections per active (circulating) past infection, while the latter describes the exponential rate of new infection accumulation [4].

Although *R* and *r* are important and popular means of tracking the dynamics of epidemics, they suffer from key limitations that diminish their fidelity and interpretability. Specifically, the meaningfulness of *R* depends on our ability to measure the generation time distribution of the infection under study, *w*. This distribution captures the inter-event times among primary and secondary infections [5] and together with the history of infection times, defines Λ, the time-varying total infectiousness of the disease. The total infectiousness serves as the denominator when inferring *R*, which is effectively a ratio of new infections to Λ. However, infection times and hence *w* are difficult to measure and require the availability of detailed transmission chain data from contact tracing or transmission studies [6]. Even when these data are available, the estimated *w* (and hence Λ) depends on how inter-event times are sampled or interpreted (e.g., there are forward, backward, intrinsic and realised generation intervals) [7,8].

Workarounds, such as approximating *w* by the serial interval distribution [9], which describes inter-event times between the onset of symptoms, or inferring *w* from this distribution [10], do exist but also suffer from related problems [6]. Consequently, *w* and Λ are often misspecified, biasing *R* and likely misrepresenting the true branching process dynamics of epidemics. While *r* is more robust to *w* misspecification (it only depends on the log gradient of the smoothed infection time series) [4], it lacks the individual-level informativeness and interpretation of *R*. Given estimates of *r*, it is unclear how to derive the proportion of new infections that need to be suppressed (roughly R^{-1}), herd immunity thresholds (related to $1\text{-}R^{-1}$) or the probability of epidemic elimination and establishment (both linked to R^{-N} for *N* infections) [11–13]. The only known means of attaining such information converts *r* into *R* using estimates of *w* [14].

Difficulties in accurately inferring generation times therefore cause practical bottlenecks that constrain our ability to measure pathogen transmissibility. These problems are worsened as recent studies have empirically found that generation times also vary substantially with time (i.e., *w* is non-stationary) [15]. These variations may correspond to different epidemic phases [16], emerging variants of concern [17] and coincide with the implementation of interventions [18]. These are precisely the situations in which we also want to infer *R*. However, concurrent changes in *R* and *w* are rarely identifiable, and *r* inextricably groups the effects of *w* and *R* on transmissibility. While high quality, longitudinal contact tracing data [19] can potentially resolve these identifiability issues, this is an expensive and logistically hard solution. Here we propose another means of alleviating the above problems – the *angular reproduction number*, Ω.

The angular reproduction number defines transmissibility as a ratio of new infections to *M*, the root mean square number of past infections over a user-defined window δ. Because it replaces Λ with *M*, a quantity that does not require knowledge of generation times, Ω is not subject to the problems of inferring *w*. We demonstrate that Ω is able to measure the overall changes in transmissibility caused by fluctuations in both *R* and *w*. Moreover, we prove that Ω has similar threshold properties to *R*, maintains much of its individual-level interpretation and is potentially a better metric for communicating transmissibility. This last point follows as we only need to quote Ω and the known window δ to generalise our estimates of transmissibility to different settings. In contrast, the meaningfulness of *R* is contingent on the unknown or uncertain *w*. Downstream studies sometimes use *R* outside of its generation time context [20], while dashboards aiming at situational awareness often quote *R* without *w*, devaluing this statistic as a robust means of communicating disease spread [21].

Additionally, we demonstrate how *r* and *R* can easily disagree on relative transmissibility, both across time and for co-circulating variants. Unmeasured changes in *w* over time can cause *R* and *r* to vary in opposite directions (one signals an increase in transmissibility and the other a decrease). Similarly, co-circulating pathogens with different but stationary and known *w*, may possess contradictory *R* and *r* value rankings i.e., for variants A and B, $r_A > r_B$ does not imply $R_A > R_B$. These issues are further complicated when interventions (which can change *w*, *R* or both [18]) occur, obscuring notions of the relative risk of spread. However, we find that $r_A > r_B$ guarantees $\Omega_A > \Omega_B$ and that Ω is consistent with *r* across time even if *w* changes. Last, while we may also convert *r* into threshold statistics about 1 by using a free parameter together with a transformation from [14], we show that Ω is more robust than those statistics, which implicitly make stronger assumptions, to choices of its free parameter. The robustness and consistency properties of Ω reinforce its usefulness for tracking and comparing outbreak spread.

Results

Angular reproduction numbers

The epidemic *renewal model* [22] provides a general and flexible representation of disease transmission. It defines how the incidence of new infections at time *t*, denoted I_t, depends on the effective or *instantaneous reproduction number*, R_t, and the past incident time series of infections, $I_1^{t-1} \stackrel{\text{def}}{=} \{I_1, I_2, \ldots I_{t-1}\}$. This results in the conditional moment relationship in **Eq. (1)** [9]. Generally, we use X_a^b to denote the time series $\{X_a, X_{a+1}, \ldots, X_{b-1}, X_b\}$ and $\mathbf{E}[X|Y]$ for the expectation of X over possible epidemic trajectories given known variables Y. Where obvious, and for convenience, we sometimes drop Y in $\mathbf{E}[X|Y]$, writing $\mathbf{E}[X]$.

$$\mathbf{E}[I_t \mid I_1^{t-1}, w_1^m] = R_t \Lambda_t, \qquad \Lambda_t = \sum_{u=1}^{m} w_u I_{t-u}. \quad (1)$$

In this model Λ_t is known as the *total infectiousness* and summarises the weighted influence of past infections. The set of weights w_u for all u defines the *generation time distribution* of the infectious disease with $\sum_{u=1}^{m} w_u = 1$, and m as the support of this distribution, which we assume to be practically finite [14]. When the time series is shorter than m we truncate and renormalise the w_u. Commonly, the stochasticity around the expectation $R_t \Lambda_t$ is modelled using either Poisson or negative binomial count distributions [1,12].

Although **Eq. (1)** has successfully been applied to model many diseases including COVID-19, Ebola virus disease, pandemic influenza and measles, among others, it has one major flaw – it assumes that the generation time distribution is fixed or stationary and known [9]. If this assumption holds (we ignore surveillance biases [9,23] until the Discussion), **Eq. (1)** allows epidemic transmissibility to be summarised in fluctuations of the time-varying R_t parameters. This follows because the sign of $R_t - 1$ determines if I_t will increase or decline relative to the total infectiousness Λ_t. This reproduction number can be linked to the *instantaneous epidemic growth rate*, r_t, using the moment generating function of the generation time distribution [14].

Consequently, from R_t, we obtain temporal information about the rate of pathogen spread and its mechanism i.e., we learn how many new infections we can expect per circulating infection because $R_t = \mathbf{E}[I_t]\Lambda_t^{-1}$. Since R_t is a threshold parameter, we know that at a fraction of at least R_t^{-1} of new infections must be blocked to suppress epidemic growth ($R_t = 1$ signifies that $r_t = 0$). The time scale over which this suppression is achievable [14] and our ability to detect changes in R_t [24] in the first place, however, are determined by the generation times.

Recent works emphasise that the assumption of a known or fixed generation time distribution is often untenable, with appreciable fluctuations caused by interventions [15,18] and emerging pathogenic variants [17] or occurring as the epidemic progresses through various stages of its lifetime [5]. Substantial biases in R_t can result (because its denominator Λ_t is incorrectly specified [4]), which even impede optimal Bayesian inference algorithms [25]. As R_t is a predominant metric of transmissibility, contributing key evidence towards infectious disease policymaking [1], this may potentially obscure situational awareness or misinform intervention planning. While improved and intensive contact tracing can provide updated generation time information, this is usually difficult and expensive. We propose a robust alternative.

We redefine the total infectiousness by recognising that it is a dot product between the vector of generation time probabilities $\vec{w} \stackrel{\text{def}}{=} w_1^m$ and the past incidence $\vec{I} \stackrel{\text{def}}{=} I_{t-m}^{t-1}$ over the support of the generation time distribution, m. This gives the left equality of **Eq. (2)** with the Euclidian norm of $\vec{X}$ as $\left\|\vec{X}\right\| \stackrel{\text{def}}{=} (\sum_{u=1}^{m} X_u^2)^{\frac{1}{2}}$ and θ_t as the time-varying angle between $\vec{w}$ and $\vec{I}$. This equality holds for non-stationary generation times i.e., both $\vec{w}$ and $\vec{I}$ can have elements that change over time). **Eq. (2)** implies that the count of new infections (for any given R_t) becomes maximised when the angle between $\vec{w}$ and $\vec{I}$ is minimised i.e., when the temporal profile of past infections matches the shape of the generation time distribution.

$$\Lambda_t = \|\vec{w}\|\left\|\vec{I}\right\| \cos\theta_t\,, \qquad \mathbf{E}[I_t] = \left(\frac{\|\vec{w}\|}{\|\vec{w}_{\max}\|} R_t \cos\theta_t\right) M_t. \quad (2)$$

We can compute the root mean square incidence across the support of the generation time distribution as $M_t \stackrel{\text{def}}{=} \frac{1}{\sqrt{m}}\left\|\vec{I}\right\|$. Under the constraint that $\sum_{u=1}^{m} w_u = 1$ (if $t-1 < m$ we truncate this distribution to sum to 1 – this is an edge effect of the epidemic) then the maximum possible value of the generation time norm is $\|\vec{w}_{\max}\| = \frac{1}{\sqrt{m}}$. This is achieved by the maximum entropy generation time distribution of $\vec{w}$, which is uniform (has m entries of $\frac{1}{m}$).

Combining these definitions with **Eq. (1)**, we derive the second expression in **Eq. (2)** for the expected number of new infections at time *t*. This may seem an unnecessarily complicated manipulation of the standard renewal model, but it admits a novel and important insight – we can separate the influences of the reproduction numbers and the generation time distribution (together with its changes) on epidemic transmissibility. These multiply M_t, which defines a new denominator – the root mean square number of past infections (this is also the average signal power of the past infection time series) – that replaces the total infectiousness Λ_t.

Consequently, we define a new metric in **Eq. (3)**, the *angular reproduction number* Ω_t, which multiples R_t by the scaled projection of the generation time distribution, $\frac{\|\vec{w}\|}{\|\vec{w}_{\max}\|}\cos\theta_t$, onto $\vec{I}$, the past incidence vector. This means that Ω_t is a time-varying reproduction number between the expected infection incidence and the past root mean square incidence M_t.

$$\Omega_t \stackrel{\text{def}}{=} \frac{\|\vec{w}\|}{\|\vec{w}_{\max}\|} R_t \cos\theta_t \quad \Longrightarrow \quad \Omega_t = \mathbf{E}[I_t] M_t^{-1}. \quad (3)$$

This metric captures all possible variations that impact the ability of the epidemic to transmit. It responds to both changes in R_t and the generation time distribution. The latter would scale $\|\vec{w}\|$ and rotate cos θ_t, which is why we term this angular. The benefit of compactly describing both types of transmissibility changes does come with a trade-off in interpretability as it may be harder to intuit the meaning behind $\mathbf{E}[I_t] = \Omega_t M_t$ than the more usual $\mathbf{E}[I_t] = R_t \Lambda_t$.

We argue that this is not the case practically because Λ_t is frequently and easily misspecified [15,26], obscuring the meaning of R_t. In contrast, M_t does not depend on generation time assumptions (beyond characterising its support *m*), and Ω_t is always defined as a reproduction number relative to root mean square incidence. We remove structural uncertainty induced by the often unknown w_u since M_t is a maximum entropy version of Λ_t i.e., $M_t = \max_{\|\vec{w}\| \cos\theta_t} \Lambda_t = \|\vec{w}_{\max}\| \left\|\vec{I}\right\|$ subject to $\sum_{u=1}^{m} w_u = 1$. We find that $M_t = \Lambda_t$ and hence $\Omega_t = R_t$, when the past incidence is flat (as then $\Lambda_t = M_t$ and w_u has no effect). This defines the important and universal equilibrium condition $\Omega_t = R_t = 1$. There is also convergence for branching process models [27] with timesteps at its fixed generation time, as then trivially $w_1 = 1$.

Relationship to popular transmissibility metrics

Having defined the angular reproduction number above, we explore its properties and show why it is an interesting and viable measure of transmissibility. We examine an exponentially growing epidemic with incidence $I_t = I_0 e^{rt}$ and constant growth rate r. This model matches the dynamics of fundamental *compartmental models* such as the SIR and SEIR (in the limit of an excess of susceptible individuals) and admits the equation $gr = (R - 1)$ [28], with g as the mean generation time. We assume growth occurs over a period of δ and compute Ω_t as in **Eq. (3)** by noting that $\mathbf{E}[I_t] = I_t$ and $M_t = \left(\delta^{-1} \int_{t-\delta}^{t} I_s^2 \, ds\right)^{\frac{1}{2}}$ ($\delta = m$ and for continuous time models $\left\|\vec{I}\right\|$ involves an integral). After some algebra we get the left relation in **Eq. (4)**.

$$\Omega_t^2 = \frac{2\delta r}{1 - e^{-2\delta r}} \geq 1, \qquad \Omega_t^2 = \frac{2\delta g^{-1}(R-1)}{1 - e^{-2\delta g^{-1}(R-1)}}. \quad (4)$$

Several important points follow. First, as $x \geq 1 - e^{-x}$ for every $x \geq 0$, then $\Omega_t - 1$ and r are positive too (an analogous argument proves the negative case). Second, we substitute the compartmental *R-r* relationship $gr = (R - 1)$ to get the right-side relation of **Eq. (4)**. Applying L' Hopital's rule we find $\lim_{R \to 1} \Omega_t = 1$. We hence confirm the threshold behaviour of Ω_t i.e., the sign of $\Omega_t - 1$ and $R_t - 1$ are always consistent (for all values of $\delta > 0$).

Third, we see that constant growth rates imply constant angular reproduction numbers. The converse is also true, and we may input time-varying growth rates, r_t, into **Eq. (4)** to estimate Ω_t. These properties hold for any δ, which is now a piecewise-constant window. Later sections show that the relationship between r_t and Ω_t has important consequences when comparing outbreaks. We plot key *R*-*r*-Ω relationships in **Figure 1**. We may also invert this relatioship to estimate r_t from Ω_t (see Methods) .This involves solving **Eq. (5)**, where $W_k(x)$ is the Lambert W function with index $k \in [0, -1]$ (this range results from the indicator $\mathbf{1}(y)$) [29].

$$\frac{d \log I_t}{dt} = r_t = 2\delta^{-1}\left(\Omega_t^2 + W_{-\mathbf{1}(\Omega<1)}\left(-\Omega_t^2 e^{-\Omega_t^2}\right)\right). \quad (5)$$

A central implication of **Eq. (4)** and **Eq. (5)** is that we can infer angular reproduction numbers directly from growth rates or vice versa, without requiring knowledge of the generation times.

We also comment on the relationship of angular and effective reproduction numbers using a deterministic *branching process* model, which is also foundational in epidemiology. We again focus on growth, which is geometric since this is a discrete time process with time steps scaled in multiples of the mean generation time g. We can write incidence as $I_t = R^t$ leading to $\Omega_t = R^t\left(\delta^{-1}\sum_{s=t-\delta}^{t-1} R^{2s}\right)^{-\frac{1}{2}}$, where window δ is in units of g. If $\delta = 1$ we recover $\Omega_t = R$. Further, if $R = 1$, then $\Omega_t = R$ for all δ. For growing epidemics, as δ increases, $\Omega_t > R$ because we reference present incidence to smaller past infections (or denominators). The opposite occurs if the epidemic declines. This may seem undesirable, but we argue that Ω_t improves overall practical transmissibility measurement because g will likely be misspecified or vary with time.

Any g mismatches will bias *R*, limiting its interpretation and meaningfulness as well as making comparisons among outbreaks or pathogenic variants difficult, since we cannot be certain that our denominators correspond. This can be particularly problematic if estimates of *R* obtained from a modelling study are incorporated as parameters into downstream studies without accounting for the generation time context on which those estimates depend. However, by communicating Ω and δ, we are sure that denominators match and, further, that we properly include the influences of any g mismatches. Choosing δ is also no worse (and more explicit) than equivalent window assumptions made when inferring *R* and *r* [4].

Last, we illustrate how Ω_t relates to other key indicators of epidemic dynamics such as herd immunity and elimination probabilities. As our derivation replaces **Eq. (1)** with $\mathbf{E}[I_t \mid I_1^{t-1}, \delta] = \Omega_t M_t$ for the same observed incidence, these indicators are also readily obtained. Assuming

Poisson noise, the elimination probability $\prod_{s=t}^{\infty}\mathbf{P}[I_t = 0|\, I_1^{t-1}, R_1^{t-1}] = e^{-\sum_{s=t}^{\infty}\Lambda_t R_t}$ is replaced by $e^{-\sum_{s=t}^{\infty}M_t\Omega_t}$, and has analogous properties [30]. Herd immunity, which traditionally occurs when a fraction $1-R^{-1}$ of the population is immune is approximated by $1-\Omega^{-1}$ (since both metrics possess the same threshold behaviour) [11]. In a subsequent section we demonstrate that one-step-ahead incidence predictions from both approaches are also comparable.

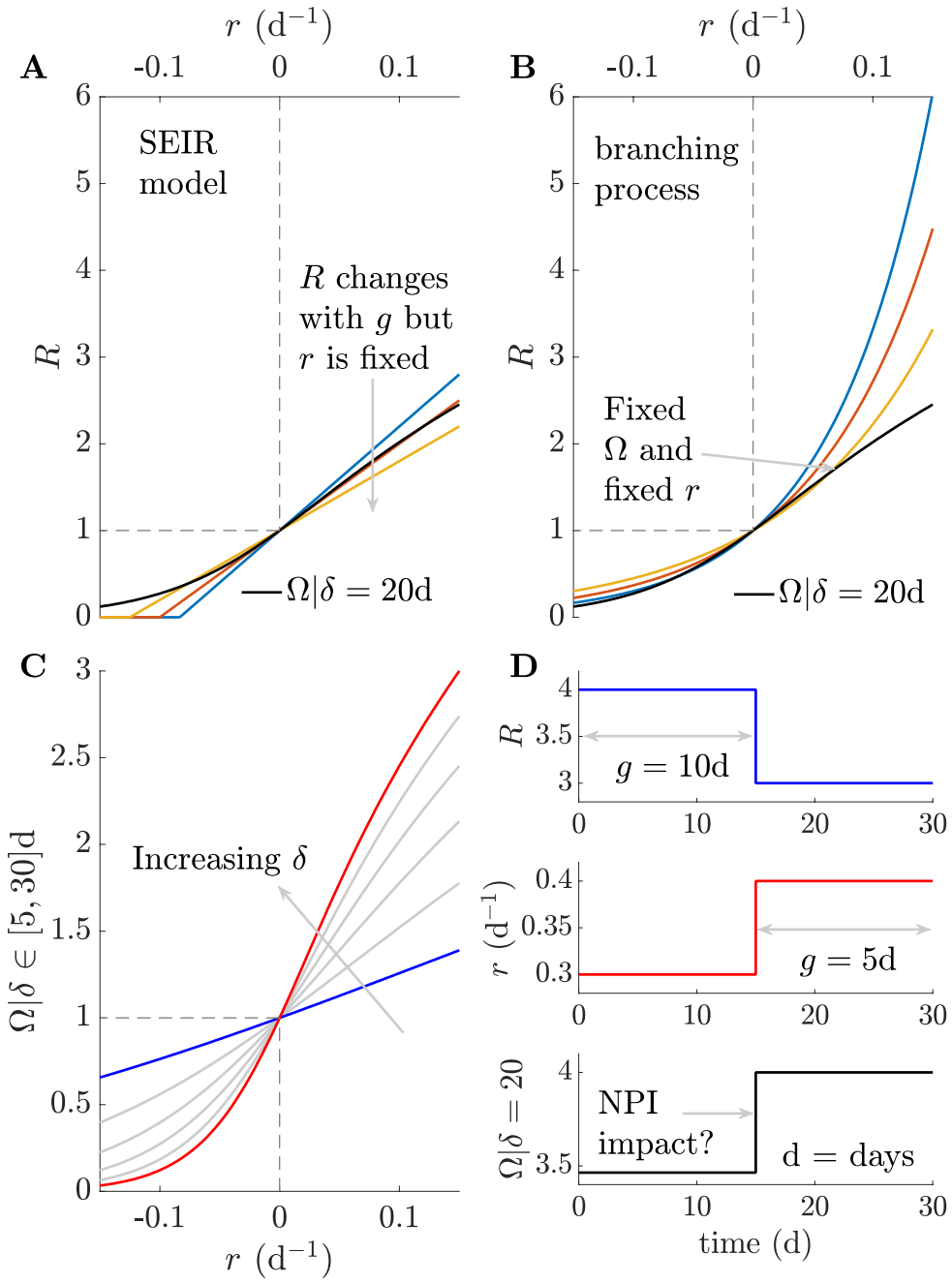

Figure 1: Relationships among transmissibility metrics. Panel A and B show how growth rates (*r*) and reproduction numbers (*R*) have diverse functional relationships (see [14]) for

SEIR models with an excess of susceptible individuals (A) and branching processes (B). Coloured lines indicate *R* at different mean generation times (*g*). Black lines highlight a single functional relationship between angular reproduction numbers Ω and *r* at all *g*, using a window δ of 20d. Panel C shows that while Ω varies with choice of δ (increasing from blue to red), we have a bijective relationship with *r*. Panel D demonstrates that *R* and *r* do not correspond e.g., if an NPI reduces *R* and *g* (see [15,18]), but Ω properly converts *r* into a transmissibility metric.

Responding to variations in generation time distributions

We demonstrate the practical benefits of Ω_t using simulated epidemics with non-stationary or time-varying generation time distributions. Such changes lead to misspecification of Λ_t in **Eq. (1)**, making estimates of the effective reproduction number R_t, denoted $\hat{R}_t$, a poor reflection of the true underlying R_t. In contrast variations in the estimated $\widehat{\Omega}_t$ are a feature (see **Eq. (3)**) and not a bug (for some chosen δ we control M_t, which is not misspecified). We simulate epidemics with Ebola virus or COVID-19 generation times from [31,32] using renewal models with Poisson noise [9]. We estimate both the time-varying R_t and Ω_t using *EpiFilter* [25], which applies Bayesian algorithms that minimise mean square estimation error.

Inferring Ω_t from incident infections, I_1^t, requires only that we replace the input Λ_t with M_t in the estimation function and that we choose a window δ for computing M_t. We provide software for general estimation of Ω_t at https://github.com/kpzoo/EpiFilter. Code for reproducing this and all other analyses in this paper is also freely available at https://github.com/kpzoo/Omega. We heuristically set $\delta \approx 2g_0$ as our window with g_0 as the original mean generation time of each disease from [31,32]. We find (numerically) that this δ ensures $\sum_{u=0}^{\delta} w_u \geq 0.86$ over many possible gamma distributed generation times i.e., we cover most of the probability mass of likely but unknown changes to the generation time distributions, which cause time-varying means g_t, without expanding much beyond their supports or incurring large edge effects.

Our results are plotted in **Figure 2**. We show that $\widehat{\Omega}_t$ responds as expected to both changes in the true R_t and w_1^m, subject to the limits on what can be inferred [24]. In **Figure 2** we achieve changes in w_1^m by altering the mean generation time g_t by ratios that are similar in size to those reported from empirical data [15]. In contrast, we observe that $\hat{R}_t$ provides incorrect and overconfident transmissibility estimates, which emerge because its temporal fluctuations also have to encode structural differences due to misspecification of w_1^m. These can strongly mislead our interpretation and understanding of the risk posed by a pathogen.

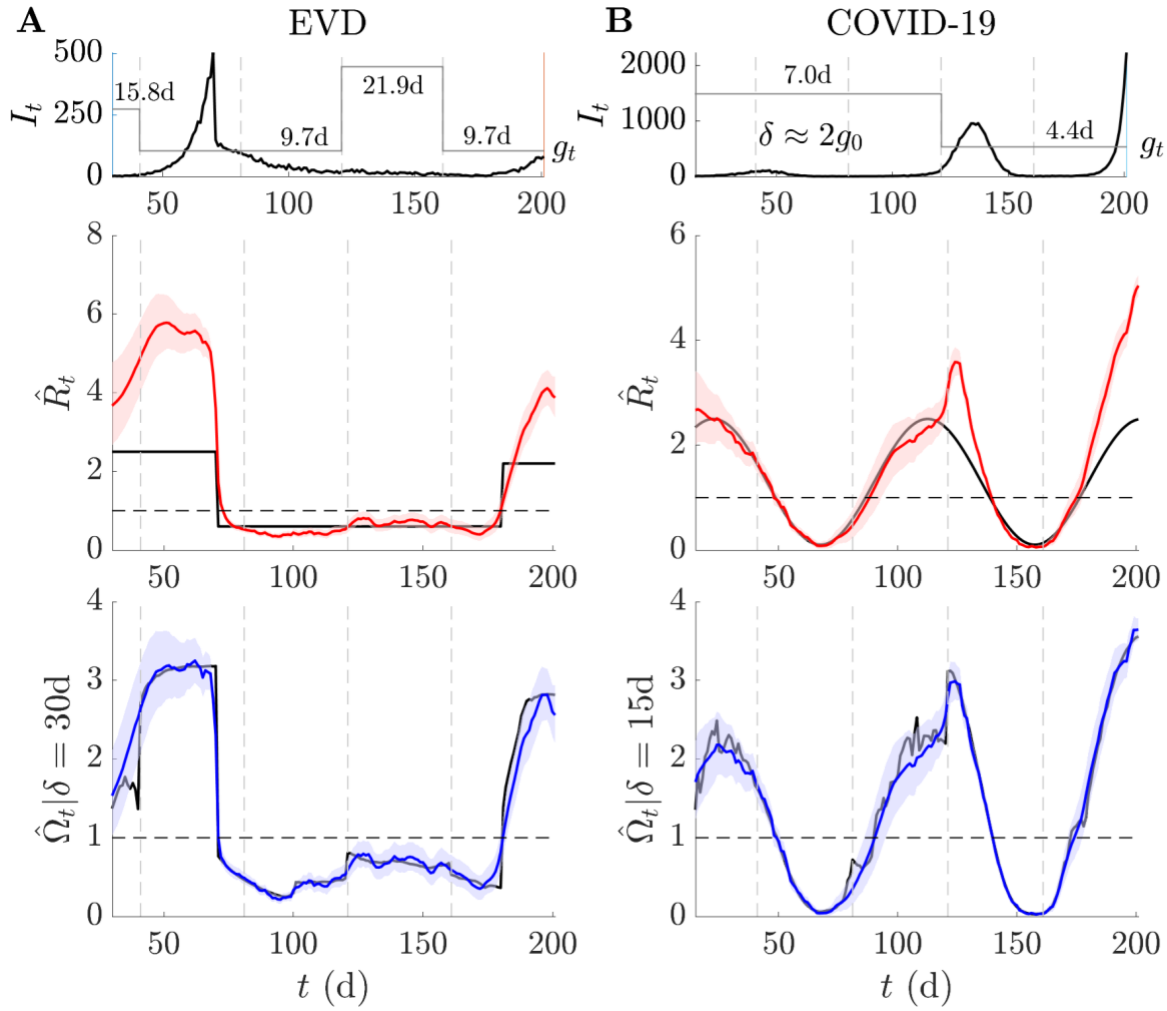

Figure 2: Estimating transmissibility under temporal variations in generation times. We simulate epidemics using generation time distributions of Ebola virus disease (EVD) [31] and COVID-19 [32] in panels A and B. The means of these distributions (*g*) vary over time (grey), but we fix their variance at their original values. We find substantial bias in estimates of *R* (red with 95% credible intervals, true value in black), with estimates attempting to compensate for generation time mismatches in an uncontrolled manner that obscures interpretation. However, Ω responds as we expect (blue with 95% credible intervals, window δ, true value in black) and we infer change-points due to both *R* and *g* fluctuations (subject to bounds induced by noise i.e., at low incidence inference is more difficult [24]). Our estimates derive from EpiFilter [25] with default settings and we truncate the time series to start from δ to remove any edge effects.

We can derive alternative threshold statistics that relate to r_t and do not explicitly depend on the generation time by applying monotonic transformations from [14]. In theory these should have comparable behaviour around the critical point of 1 as both R_t and Ω_t. We investigate

these statistics in the **Supplementary Information**, computing them across the simulations of **Figure 2**. We find that they require stronger assumptions than Ω_t (i.e., they fix distributional formulae for generation times), possess at least as many free parameters as Ω_t and are less robust to changes in those parameters (often strongly over-estimating transmissibility), than Ω_t is to fluctuations in δ. This confirms the practical utility of angular reproduction numbers for tracking disease spread, especially when generation times are changing or unknown.

Ranking epidemics or variants by transmissibility

Misspecification of generation time distributions, and corresponding misestimation of *R* as in **Figure 2**, also plays a crucial role when assessing the relative transmissibility of pathogens, variants of concern or even outbreaks (where we may want to contrast the spread of contagion among key demographic or spatial groups). As shown in **Figure 1**, these variations can mean that increases in the growth rate r_t actually signify decreases in the effective reproduction number R_t or that a pathogen with a larger r_t can have a smaller R_t. Here we illustrate that these issues can persist even if the generation time distributions of pathogens are correctly specified and remain static, obscuring our understanding of relative transmission risk.

In **Figure 3** we simulate epidemics under two hypothetical variants of two pathogens. We use EVD and COVID-19 generation time distributions from [31,32] to define our respective base variants. For both pathogens we specify the other variant by reducing the mean generation of each base but hold the generation time variance fixed. Reductions such as these are plausible and have been measured for COVID-19 variants [17]. All w_1^m distributions are stationary and known in this analysis. We discover that changes in R_t alone can initiate inversions in the relative growth rate of different variants or epidemics. As far as we can tell, this phenomenon has not been explicitly investigated. Given that interventions can change R_t in isolation or in combination with w_1^m [15,18], this effect has the potential to be widespread.

Interestingly, the angular reproduction numbers of **Figure 3** do preserve an ordering that is consistent with the relative growth rates, while maintaining the interpretability (e.g., threshold properties) of a reproduction number. Hence, we argue that Ω_t blends advantages from both R_t and r_t [4] and serves as a useful outbreak analytic for understanding and conveying the relative risk of spread of differing pathogens or pathogen strains, or of spread among different spatial and demographic groups. As recent research has only started to disentangle the component drivers of transmission, including the differing influences that interventions may introduce (e.g., by defining notions of the strength and speed of control measures [33]) and

the diverse properties of antigenic variants [17], we believe that Ω can play a distinctive and key role in accelerating these investigations.

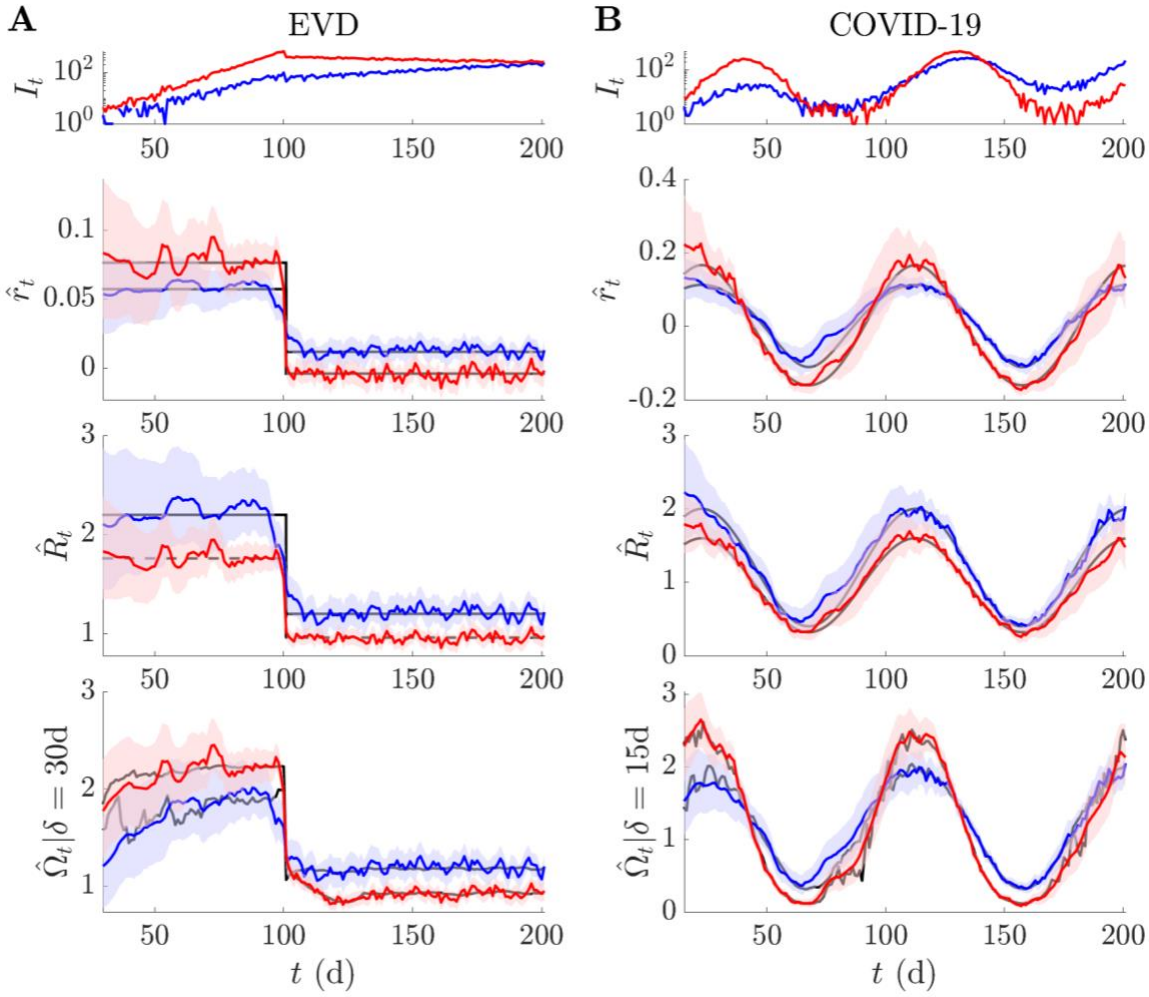

Figure 3: Comparing transmissibility across outbreaks, variants or even diseases. We simulate epidemics in blue (with estimates of metrics also in blue) under standard generation time distributions of Ebola virus disease (EVD) [31] and COVID-19 [32] in panels A and B. In red (with estimates also in red) we overlay simulations in which the generation time of these diseases is 40% and 50% shorter, which may indicate a new co-circulating variant or another epidemic with different properties (e.g., due to being in a higher risk group). We show (for the first time to our knowledge) that changes in *R* due to an intervention (or release of one) may alter the relative growth rates (*r*) of the epidemics. The mismatches in the *R*-*r* rankings alter perceptions of relative risk, making comparisons of transmissibility difficult. However, Ω is able to classify the risk of these epidemics in line with their realised growth rates, while still offering the individual-level interpretability of a reproduction number. True values are in black and all

estimates (with 95% credible intervals) are outputs from EpiFilter [25] with default settings. We truncate the time series to start from δ to remove any edge effects.

Reproduction numbers for explanation or prediction?

We highlight an important but underappreciated subtlety when inferring the transmissibility of epidemics – that the value of accurately estimating *R*, *r* and Ω largely depends on if our aim is to explain or predict [34] the dynamics of epidemics. The above analyses have focussed on characterising transmissibility to explain mechanisms of spread and design interventions. For these problems, misestimation of parameters, such as *R*, can affect our understanding of outbreak risk and consequently might misinform the implementation of control measures. An important concurrent problem aims to predict the likely incidence of new infections from these estimates. This involves projecting the epidemic dynamics forward in time to infer upcoming patterns in incidence of infections.

Here we present evidence that the solution of this problem, at least over short projection time horizons, is robust to misspecification of generation times provided both the incorrect estimate and the misspecified denominator are used in conjunction. We repeat the analyses of **Figure *2*** for 200 replicate epidemics and apply EpiFilter [25] to obtain the one-step-ahead predictive distributions $\mathbf{P}(I_t|\, I_1^{t-1})$ for every t. We compute the predicted mean square error (PMSE) and the accumulated predictive error (APE). These scores, which we denote as $D(I_t|\, I_1^{t-1})$, average square errors between mean predictions and true incidence and sum log probabilities of observing the true incidence from the predicted distribution respectively [35,36]. We plot the distributions of scores over replicates and illustrate individual predictions in **Figure *4***.

We find only negligible differences among the one-step-ahead predictive accuracies of the *R* estimated given knowledge of the changing generation times (*R*|*w*), the *R* estimated assuming an unchanged (and hence wrongly specified) *w* and our inferred Ω. As APE and PMSE also measure model suitability, their similarity across the three estimates demonstrate that, if the problem of prediction is of interest, then incorrect generation time choices are not important as long as the erroneous denominator (Λ_t) and estimate (R_t) are used together. If this estimate is however used outside of the context of its denominator (e.g., if it is simply input into other studies), then inaccurate projections will occur (in addition to poor estimates). As multi-step-ahead predictions can be composed from iterated one-step-ahead ones [37], we conjecture that subtleties between prediction and explanation are likely to also apply on longer horizons.

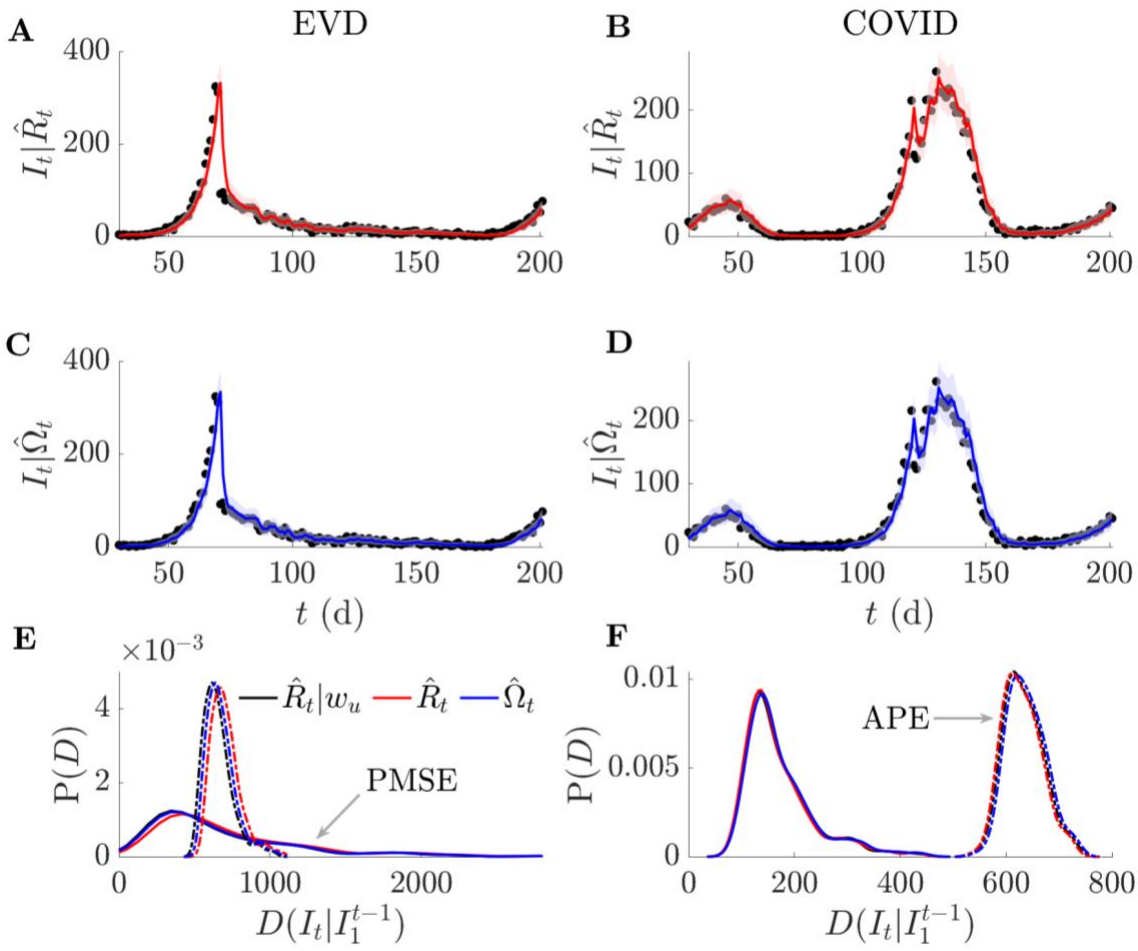

Figure 4: One-step-ahead prediction accuracy and model mismatch. We simulate 200 replicates of the epidemics from **Figure 2**, which involve non-stationary changes to EVD and COVID-19 generation times. We use estimates of effective, R, and angular, Ω, reproduction numbers to produce successive one-step-ahead predictions and assess their accuracy to the simulated (true) incidence. Panels A-D provide a representative example of a single simulated epidemic (true incidence shown as black dots) and the R and Ω one-step ahead predictions (red and blue respectively with 95% credible intervals). In panels E-F we formally compute accuracy using distance metrics, D, based on accumulated prediction errors (APE, dashed) and prediction mean square errors (PMSE, solid) for all 200 replicates from R, Ω and R given knowledge of the generation time changes i.e., $R|w$. We obtain distributions of D by applying kernel smoothing. We find negligible differences in predictive power from all approaches.

Empirical example: COVID-19 in mainland China

We complete our analysis by illustrating the practical usability of Ω on an empirical case study where generation time changes are known to have occurred. In [15], the dynamics of COVID-19 in mainland China are tracked across January and February 2020. Transmission pair data indicated that the serial interval of COVID-19 shortened across this period leading to biases

in the inferred *R* if updated serial intervals are not used. Here serial intervals, which measure the lag between the symptom onset times of an infector and infectee are used as a proxy for the generation time. **Figure *5*** presents our main results. We find Ω (blue), which requires no serial interval information, behaves similarly to the *R* (red) inferred from the time-changing *w*. Both metrics appear less biased than estimates of *R* (green) that assume a fixed serial interval. This is largely consistent with the original investigation in [15].

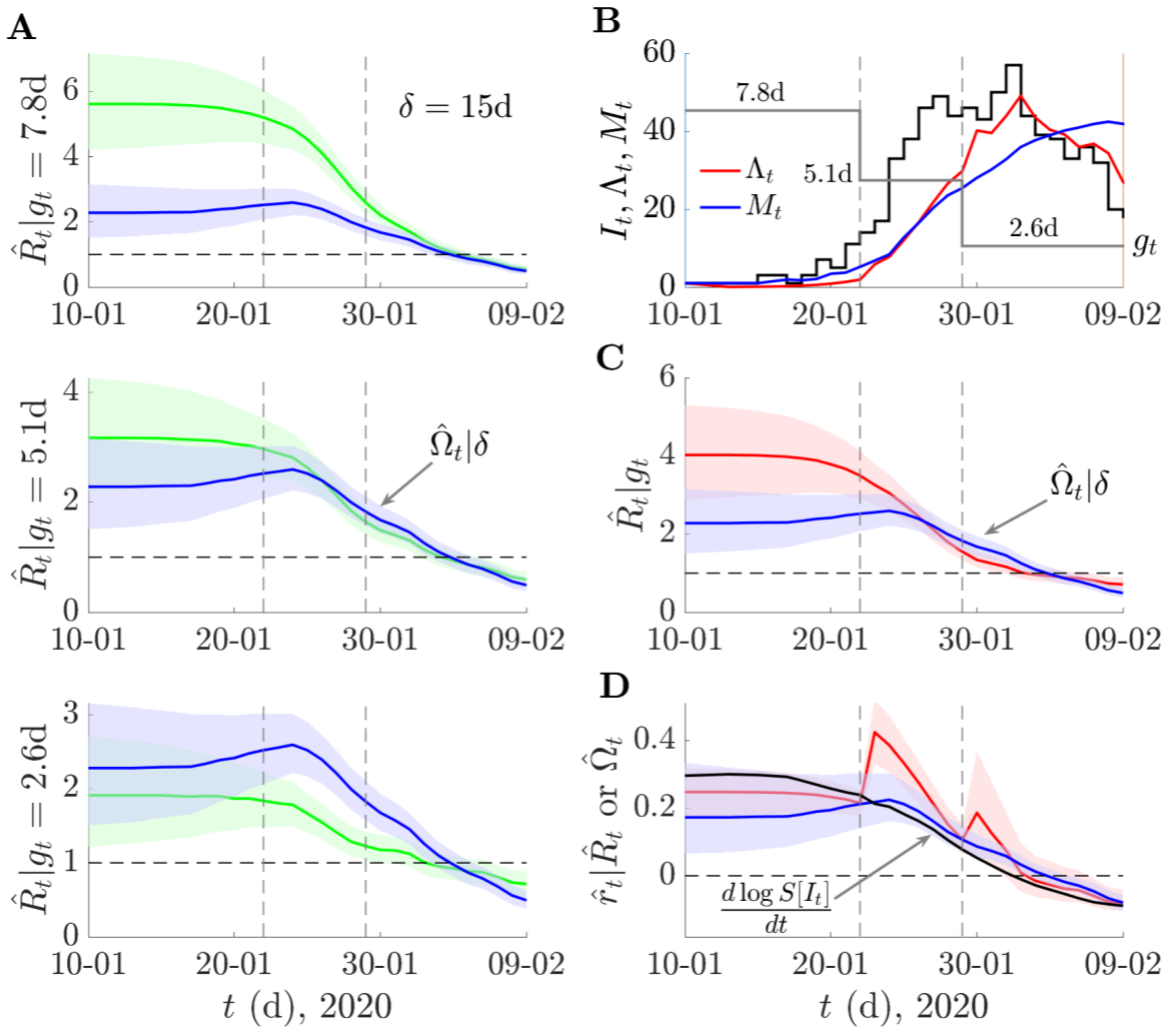

Figure 5: COVID-19 transmissibility in China under non-stationary generation times. We analyse COVID-19 data from [15], which spans 9th January 2020 to 13 February 2020 and is known to feature a serial interval distribution that shortened in mean substantially from 7.8d to 2.6d (change times are shown as grey vertical lines). We assume that the serial interval approximates the generation time well and replicate the analysis from Figure 2 of [15]. In panel A, we compare estimates (green) of effective reproduction numbers, *R*, using fixed generation time distributions inferred in [15] (specified by their means *g*) against those of our angular reproduction number Ω (blue). We use EpiFilter [25] to obtain all estimates (means shown with 95% credible intervals) and find relative trends similar to those in Figure 2 of [15]. In panel

B we plot the incidence (black) and the denominators we use to compute an *R* that does account for the generation time changes (Λ, red) and for Ω (*M*, blue). This *R* uses the different distributions inferred at the grey vertical change times (their means are in panel B and are also the fixed distributions of panel A in sequence). We plot these *R* and Ω estimates in panel C. In panel D we show the growth rates that are inferred from the *R* and Ω estimates of C (red and blue respectively) against that obtained from taking the smoothed log derivative (black).

We see that Ω provides a lower assessment of the initial transmissibility as compared to the *R* that is best informed by the changing *w* but that both agree in general and in particular at the important threshold between super- and subcritical spread. Interestingly, Ω indicates no sharp changes at the *w* change-times. This follows because the incidence is too small for those changes to influence overall transmissibility and matches the gradual *w* changes originally inferred in [15]. The distributions used in **Figure *5*** provide a piecewise approximation to these variations. We also compare *r* estimates derived from *R* (red, from [14]), Ω (blue, from **Eq. (5)**) and the empirical log gradient of smoothed incidence (black, $\frac{d \log S[I_t]}{dt}$ [4]). We find that the *r* from Ω agrees more closely with the empirical growth rate than the *r* from *R*, which somewhat by design shows jumps at the *w* change-points. While this analysis is not meant as a detailed study of COVID-19 in China, it does demonstrate the practical usefulness of Ω.

Discussion

Quantifying the time-varying transmissibility of a pathogen remains an enduring challenge in infectious disease epidemiology. Changes in transmissibility may signify shifts in the dynamics of an epidemic of relevance to both preparedness and policymaking. While this challenge has been longstanding, the statistics that we use to summarise transmissibility have evolved from dispersibility [38] and incidence to prevalence ratios [39] to cohort [40] and instantaneous [22] reproduction numbers. While the last, which we have denoted *R*, has become the predominant metric of transmissibility, all of these proposed statistics ultimately involve a ratio between new infections and a measure of active infections (i.e., the denominator). Deciding on appropriate denominators necessitates some notion (implicit or explicit) of a generation time [41].

Difficulties in characterising these generation times and their changes substantially bias [6] estimates of transmissibility and have motivated recent works to propose the instantaneous growth rate, *r*, as a more reliable approach for inferring pathogen spread [20]. However, on its own, *r* is insufficient to resolve many of the transmission questions that *R* can answer and its computation may employ smoothing assumptions that are in some instances equivalent to the generation time ones behind *R* [4]. We formulated a novel statistic, the angular reproduction

number Ω, to merge advantages from both R and r and to provide a more comprehensive view of transmissibility. By applying basic vector algebra (**Eqs. 1-3**), we encoded both changes to R and the generation time distribution, w, into a single time-varying metric, deriving Ω.

We found that Ω maintains the threshold properties and individual-level interpretability of R but responds to variations in w, in a manner consistent with r (**Figure *1***). Moreover, Ω indicates variations in transmissibility caused by R and w without requiring measurement of generation times (**Figure *2***). This is a consequence of its denominator, which is the root mean square of infections over a user-specified window δ that is relatively simple to tune (see Methods). We can interpret $\Omega = a > 1$ as indicating that infections across δ need to be reduced by a^{-1}. This reduces mean and root mean square infections by a^{-1} and causes Ω to equal 1. Further, Ω circumvents identifiability issues surrounding the joint inference or R and w [42] by refocussing on estimating the net changes produced by both. This improves our ability to *explain* the shifts in transmissibility underpinning observed epidemic dynamics and means Ω is a reproduction number that provides an individual-level interpretation of growth rates (**Eqs. 4-5**).

The benefits of this r-Ω correspondence are twofold. First, as interventions may alter R, w or R and w concurrently [15,18] situations can arise where r and R disagree across time on both the drivers and magnitude of transmissibility. While it may seem possible to minimise this issue by constructing alternative threshold statistics by directly combining r with assumed generation time structures, we find these statistics exhibit worse performance than Ω (**Supplementary Information**). Second, this disagreement can also occur when comparing pathogenic variants or epidemics (e.g., from diverse spatial or sociodemographic groups) with different but known and unchanging w. As far as we can tell, this study is among the earliest to highlight these discrepancies. Realistic transmission landscapes possess all of these complexities, meaning that more conventional measures of relative transmissibility can be fraught with contradictions.

In contrast to R, we found that Ω consistently orders epidemics by growth rate while capturing notions of the average new infections per past infection (**Figure *3***). This suggests Ω blends the advantages of R and r, with clearer assumptions (choice of window δ). However, Ω offers no advantage if our goal is to *predict* epidemic dynamics (see [34] for more on the prediction-explanation distinction). For this problem even an R inferred with a misspecified denominator performs equally well (**Figure *4***). This follows because only the product of any reproduction number and its denominator matter when determining the next incidence value, and iterations of this underpin multi-step ahead predictions [37]. This may be the reason why autoregressive models, which ignore some characteristics of w, can serve as useful predictive models [43]. Other instances where Ω will not improve analysis are at times earlier than δ (due to edge

effects [9]) and in periods of near zero incidence (there is no information to infer *R* either [24]). We summarise and compare key properties of *R*, *r* and Ω in **Table *1*** below.

Metric property	**Growth *r***	**Effective *R***	**Angular Ω**
Definition of transmissibility	$r_t \stackrel{\text{def}}{=} \frac{d \log \mathbf{E}[I_t]}{dt}$	$R_t \stackrel{\text{def}}{=} \frac{\mathbf{E}[I_t]}{\Lambda_t}$	$\Omega_t \stackrel{\text{def}}{=} \frac{\mathbf{E}[I_t]}{M_t}$
Pathogen spread threshold	$r_t > 0$	$R_t > 1$	$\Omega_t > 1$
Biased by generation time $\vec{w}$ assumed, given curve I_1^t	Insensitive to the assumed $\vec{w}$	Biased when $\vec{w}$ is misspecified	Signals changes in $\vec{w}$ and R_t
Ranking risk of outbreaks or variants by spreading rate	$r_A > r_B \Rightarrow$ variant *A* spreads faster	$r_A > r_B \not\Rightarrow R_A > R_B$ (inconsistent)	$r_A > r_B \Rightarrow \Omega_A > \Omega_B$ (consistent)
Short-term predictive power	Negligible differences among metrics in prediction quality		
Non-dimensional metric	No, inverse time	Yes, both have no units, scalable	
Individual-level interpretability	Not obvious	New infections per circulating ones	
Computability if $\vec{w}$ unknown	Yes (smooth I_1^t)	Not possible	Yes, for any δ

Table 1: Summary of transmissibility metrics. We list important relationships among the instantaneous growth rate (*r*), the instantaneous or effective reproduction number (*R*) and the angular reproduction number (Ω) and assess their value as measures of transmissibility.

There are several limitations to our study. First, we only examined biases inherent to *R* due to the difficulty of measuring the generation time accurately and across time. While this is a major limitation of existing transmissibility metrics [15], practical surveillance data are also subject to under-reporting and delays, which can severely diminish the quality of any transmissibility estimates [23,42,44]. While Ω ameliorates issues due to generation time mismatch, it is as susceptible as *R* and *r* to surveillance biases and corrective algorithms (e.g., deconvolution methods [45]) should be applied before inferring Ω. Second, our analysis depends on renewal and compartmental epidemic models [22]. These assume random mixing and cannot account for realistic contact patterns. Despite this key structural uncertainty, there is evidence that well-mixed and network models are comparable when estimating transmissibility [46].

Although the above limitations can, in some instances, reduce the added value of improving the statistics summarising transmissibility, we believe that Ω will be of practical and theoretical benefit. Its similarity in formulation to *R* means it is as easy to compute using existing software and therefore can be deployed on dashboards and updated in real time to improve situational awareness. Further, Ω makes comparison and communication of the relative risk of circulating variants or epidemics among diverse groups more reliable because it avoids *R*-*r* contradictions provided a known parameter, δ, is fixed. This is in contrast to *R*, which is hard to contextualise [20], since its value depends on an often-unknown *w*, and hence compare across groups, as each group may have distinct and correspondingly poorly specified denominators. Last, Ω can help probe analytical questions about how changes in *R* and *w* interact because it presents a common framework for testing how variations in either influence overall transmissibility.

Methods

Inferring angular reproduction numbers across time

We outline how to estimate Ω_t given a time series of incident infections I_1^T, with T defining the present or last available data timepoint i.e., $1 \leq t \leq T$. Because Ω_t simply replaces the total infectiousness Λ_t, used for computing R_t, with the root mean square of the new infection time series, M_t, we can obtain Ω_t from standard R_t estimation packages with minor changes. This requires evaluating M_t over some user-defined backward sliding window of size δ. Under a Poisson (Pois) renewal model this follows as in **Eq. (6)** for timepoint t.

$$\mathbf{P}(I_t \mid I_1^{t-1}, \delta) \equiv \mathrm{Pois}(\Omega_t M_t), \qquad M_t = \left(\frac{1}{\delta}\sum_{u=t-\delta}^{t-1} I_u^2\right)^{\frac{1}{2}}. \quad (6)$$

The choice of δ is mostly arbitrary but should be sufficiently long to capture most of the likely probability mass of the unknown generation time but not overly long since it induces an edge effect (similar to the windows in [9,36]). We found a suitable heuristic to be twice or thrice the initial expected mean generation time (g_0). We can then input M_t and I_t into packages such as EpiEstim [9] or EpiFilter [25] to estimate Ω_t with 95% credible intervals.

Due to the similarity between computing R_t and Ω_t we only specify the latter but highlight that replacing M_t with Λ_t yields the expressions for evaluating any equivalent quantities from R_t. The only difference relates to how the growth rates r_t are computed. We estimate r_t from R_t by applying the generation time, $\vec{w}$, based transformation from [14]. For a correctly specified $\vec{w}$ this gives the same result as the smoothed derivative of the incidence curve [4]. We derive

r_t from Ω_t using **Eq. (5)**, which follows by rearranging **Eq. (4)** into $\left(2\delta r_t - \Omega_t^2\right) e^{2\delta r_t - \Omega_t^2} = -\Omega_t^2 e^{-\Omega_t^2}$. This expression then admits the Lambert W function solutions. In all estimates of r_t we propagate uncertainty from the posterior distributions (see below) over R_t or Ω_t.

We applied EpiFilter in this study due to its improved extraction of information from I_1^T. This method assumes a random walk state model for our transmissibility metric as in **Eq. (7)** with ϵ_{t-1} as a normally distributed (Norm) noise term and η as a free parameter (default 0.1).

$$\Omega_t = \Omega_{t-1} + \left(\eta\sqrt{\Omega_{t-1}}\right)\epsilon_{t-1}, \qquad \mathbf{P}(\epsilon_{t-1}) \equiv \mathrm{Norm}(0,1). \quad (7)$$

The EpiFilter approach utilises Bayesian smoothing algorithms incorporating the models of **Eq. (6)-(7)** and outputs the complete posterior distribution $\mathbf{P}(\Omega_t | I_1^T, \delta)$ with T as the complete length of all available data (i.e., $1 \leq t \leq T$). We compute our mean estimates $\widehat{\Omega}_t$ and 95% credible intervals from this posterior distribution and these underlie our plots in **Figures 2-3**.

EpiFilter also outputs the one-step-ahead predictive distributions $\mathbf{P}(I_t | I_1^{t-1}, \delta)$, which we use in **Figure *4***. There we also quantify predictive accuracy using the predicted mean square error PMSE and the accumulated prediction error APE, defined as in **Eq. (8)** [35,36] with $\hat{I}_t$ as the posterior mean estimate from $\mathbf{P}(I_t | I_1^{t-1}, \delta)$ and I_t^* as the true simulated incidence. These are computed with $\mathbf{P}(\Omega_{t-1} | I_1^{t-1}, \delta)$ and not $\mathbf{P}(\Omega_t | I_1^T, \delta)$, ensuring no future information is used.

$$\mathrm{PMSE} = \frac{1}{T-\delta}\sum_{t=\delta+1}^{T}\left(I_t^* - \hat{I}_t\right)^2, \quad \mathrm{APE} = \sum_{t=\delta+1}^{T} -\log \mathbf{P}(I_t = I_t^* | I_1^{t-1}, \delta). \quad (8)$$

We collectively refer to these as distance metrics $D(I_t | I_1^{t-1})$ and construct their distributions, $\mathbf{P}(D)$, over many replicates of simulated epidemics. Last, we use $\mathbf{P}(\Omega_t | I_1^T, \delta)$ to compute the posterior distribution of the growth rate $\mathbf{P}(r_t | I_1^T, \delta)$ and hence its estimates as in **Eq. (5)**. More details on the EpiFilter algorithms are available at [25,30,47]. We supply open source code to reproduce all analyses at https://github.com/kpzoo/Omega and a function in MATLAB and R to allow users to estimate Ω_t from their own data at https://github.com/kpzoo/EpiFilter.

Bibliography

1. Anderson R, Donnelly C, Hollingsworth D, Keeling M, Vegvari C, Baggaley R. Reproduction number (R) and growth rate (r) of the COVID-19 epidemic in the UK: methods of. The Royal Society. 2020;
2. Li Y, Campbell H, Kulkarni D, Harpur A, Nundy M, Wang X, et al. The temporal association of introducing and lifting non-pharmaceutical interventions with the time-varying reproduction number (R) of SARS-CoV-2: a modelling study across 131

countries. Lancet Infect Dis. 2021;21: 193–202. doi:10.1016/S1473-3099(20)30785-4
3. Volz E, Mishra S, Chand M, Barrett JC, The COVID-19 Genomics UK (COG-UK) consortium, Johnson R, et al. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. Nature. 2021; doi:10.1038/s41586-021-03470-x
4. Parag KV, Thompson RN, Donnelly CA. Are epidemic growth rates more informative than reproduction numbers? J Royal Statistical Soc A. 2022; doi:10.1111/rssa.12867
5. Svensson A. A note on generation times in epidemic models. Math Biosci. 2007;208: 300–311. doi:10.1016/j.mbs.2006.10.010
6. Britton T, Scalia Tomba G. Estimation in emerging epidemics: biases and remedies. J R Soc Interface. 2019;16: 20180670. doi:10.1098/rsif.2018.0670
7. Champredon D, Dushoff J. Intrinsic and realized generation intervals in infectious-disease transmission. Proc Biol Sci. 2015;282: 20152026. doi:10.1098/rspb.2015.2026
8. Nishiura H. Time variations in the generation time of an infectious disease: implications for sampling to appropriately quantify transmission potential. Math Biosci Eng. 2010;7: 851–869. doi:10.3934/mbe.2010.7.851
9. Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. Am J Epidemiol. 2013;178: 1505–1512. doi:10.1093/aje/kwt133
10. Ganyani T, Kremer C, Chen D, Torneri A, Faes C, Wallinga J, et al. Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020. Euro Surveill. 2020;25. doi:10.2807/1560-7917.ES.2020.25.17.2000257
11. Hethcote HW. The Mathematics of Infectious Diseases. SIAM Rev. 2000;42: 599–653. doi:10.1137/S0036144500371907
12. Parag KV. Sub-spreading events limit the reliable elimination of heterogeneous epidemics. J R Soc Interface. 2021;18: 20210444. doi:10.1098/rsif.2021.0444
13. Anderson R, May R. Infectious diseases of humans: dynamics and control. Oxford University Press; 1991.
14. Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. Proc R Soc B. 2007;274: 599–604.
15. Ali ST, Wang L, Lau EHY, Xu X-K, Du Z, Wu Y, et al. Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions. Science. 2020;369: 1106–1109. doi:10.1126/science.abc9004
16. Kenah E, Lipsitch M, Robins JM. Generation interval contraction and epidemic data analysis. Math Biosci. 2008;213: 71–79. doi:10.1016/j.mbs.2008.02.007
17. Hart WS, Miller E, Andrews NJ, Waight P, Maini PK, Funk S, et al. Generation time of the alpha and delta SARS-CoV-2 variants: an epidemiological analysis. Lancet Infect Dis. 2022;22: 603–610. doi:10.1016/S1473-3099(22)00001-9
18. Favero M, Scalia Tomba G, Britton T. Modelling preventive measures and their effect on generation times in emerging epidemics. J R Soc Interface. 2022;19: 20220128. doi:10.1098/rsif.2022.0128
19. Kraemer MUG, Pybus OG, Fraser C, Cauchemez S, Rambaut A, Cowling BJ. Monitoring key epidemiological parameters of SARS-CoV-2 transmission. Nat Med. 2021;27: 1854–1855. doi:10.1038/s41591-021-01545-w
20. Pellis L, Scarabel F, Stage HB, Overton CE, Chappell LHK, Fearon E, et al. Challenges in control of COVID-19: short doubling time and long delay to effect of interventions. Philos Trans R Soc Lond B, Biol Sci. 2021;376: 20200264. doi:10.1098/rstb.2020.0264
21. The R value and growth rate - GOV.UK [Internet]. [cited 1 Jul 2021]. Available: https://www.gov.uk/guidance/the-r-value-and-growth-rate
22. Fraser C. Estimating individual and household reproduction numbers in an emerging epidemic. PLoS One. 2007;2: e758. doi:10.1371/journal.pone.0000758
23. Parag KV, Donnelly CA, Zarebski AE. Quantifying the information in noisy epidemic curves. Nat. Comput. Sci. 2022; 2: 584-594. doi: 10.1038/s43588-022-00313-1
24. Parag KV, Donnelly CA. Fundamental limits on inferring epidemic resurgence in real time using effective reproduction numbers. PLoS Comput Biol. 2022;18: e1010004.

doi:10.1371/journal.pcbi.1010004
25. Parag KV. Improved estimation of time-varying reproduction numbers at low case incidence and between epidemic waves. PLoS Comput Biol. 2021;17: e1009347. doi:10.1371/journal.pcbi.1009347
26. Torneri A, Libin P, Scalia Tomba G, Faes C, Wood JG, Hens N. On realized serial and generation intervals given control measures: The COVID-19 pandemic case. PLoS Comput Biol. 2021;17: e1008892. doi:10.1371/journal.pcbi.1008892
27. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. Nature. 2005;438: 355–359. doi:10.1038/nature04153
28. Bettencourt LMA, Ribeiro RM. Real time bayesian estimation of the epidemic potential of emerging infectious diseases. PLoS One. 2008;3: e2185. doi:10.1371/journal.pone.0002185
29. Lehtonen J. The Lambert W function in ecological and evolutionary models. Methods Ecol Evol. 2016;7: 1110–1118. doi:10.1111/2041-210X.12568
30. Parag KV, Cowling BJ, Donnelly CA. Deciphering early-warning signals of SARS-CoV-2 elimination and resurgence from limited data at multiple scales. J R Soc Interface. 2021;18: 20210569. doi:10.1098/rsif.2021.0569
31. Van Kerkhove MD, Bento AI, Mills HL, Ferguson NM, Donnelly CA. A review of epidemiological parameters from Ebola outbreaks to inform early public health decision-making. Sci Data. 2015;2: 150019. doi:10.1038/sdata.2015.19
32. Ferguson N, Laydon D, Nedjati-Gilani G, Others. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID- 19 mortality and healthcare demand. Imperial College London; 2020.
33. Dushoff J, Park SW. Speed and strength of an epidemic intervention. Proc Biol Sci. 2021;288: 20201556. doi:10.1098/rspb.2020.1556
34. Shmueli G. To Explain or to Predict? Stat Sci. 2010;25: 289–310. doi:10.1214/10-STS330
35. Wagenmakers E-J, Grünwald P, Steyvers M. Accumulative prediction error and the selection of time series models. J Math Psychol. 2006;50: 149–166. doi:10.1016/j.jmp.2006.01.004
36. Parag KV, Donnelly CA. Using information theory to optimise epidemic models for real-time prediction and estimation. PLoS Comput Biol. 2020;16: e1007990. doi:10.1371/journal.pcbi.1007990
37. Marcellino M, Stock JH, Watson MW. A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. J Econom. 2006;135: 499–526. doi:10.1016/j.jeconom.2005.07.020
38. Nishiura H, Chowell G, Heesterbeek H, Wallinga J. The ideal reporting interval for an epidemic to objectively interpret the epidemiological time course. J R Soc Interface. 2010;7: 297–307. doi:10.1098/rsif.2009.0153
39. White PJ, Ward H, Garnett GP. Is HIV out of control in the UK? An example of analysing patterns of HIV spreading using incidence-to-prevalence ratios. AIDS. 2006;20: 1898–1901. doi:10.1097/01.aids.0000244213.23574.fa
40. Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. Am J Epidemiol. 2004;160: 509–516. doi:10.1093/aje/kwh255
41. Yan P. Separate roles of the latent and infectious periods in shaping the relation between the basic reproduction number and the intrinsic growth rate of infectious disease outbreaks. J Theor Biol. 2008;251: 238–252. doi:10.1016/j.jtbi.2007.11.027
42. Azmon A, Faes C, Hens N. On the estimation of the reproduction number based on misreported epidemic data. Stat Med. 2014;33: 1176–1192. doi:10.1002/sim.6015
43. Bracher J, Held L. Endemic-epidemic models with discrete-time serial interval distributions for infectious disease prediction. Int J Forecast. 2020; doi:10.1016/j.ijforecast.2020.07.002
44. Dalziel BD, Lau MSY, Tiffany A, McClelland A, Zelner J, Bliss JR, et al. Unreported

cases in the 2014-2016 Ebola epidemic: Spatiotemporal variation, and implications for estimating transmission. PLoS Negl Trop Dis. 2018;12: e0006161. doi:10.1371/journal.pntd.0006161
45. Goldstein E, Dushoff J, Ma J, Plotkin JB, Earn DJD, Lipsitch M. Reconstructing influenza incidence by deconvolution of daily mortality time series. Proc Natl Acad Sci USA. 2009;106: 21825–21829. doi:10.1073/pnas.0902958106
46. Liu Q-H, Ajelli M, Aleta A, Merler S, Moreno Y, Vespignani A. Measurability of the epidemic reproduction number in data-driven contact networks. Proc Natl Acad Sci USA. 2018;115: 12680–12685. doi:10.1073/pnas.1811115115
47. Sarrka S. Bayesian Filtering and Smoothing. Cambridge, UK: Cambridge University Press; 2013.

Funding

KVP acknowledges funding from the MRC Centre for Global Infectious Disease Analysis (reference MR/R015600/1), jointly funded by the UK Medical Research Council (MRC) and the UK Foreign, Commonwealth & Development Office (FCDO), under the MRC/FCDO Concordat agreement and is also part of the EDCTP2 programme supported by the European Union. The funders had no role in study design, data collection and analysis, decision to publish, or manuscript preparation.

Data availability statement

All data and code underlying the analyses and figures of this work are freely available (in R and MATLAB) at: https://github.com/kpzoo/Omega