# **1** Angular reproduction numbers improve estimates of transmissibility

# 2 when disease generation times are misspecified or time-varying

- 3 Kris V Parag<sup>1,\*</sup>, Benjamin J Cowling<sup>2</sup> and Ben C Lambert<sup>3</sup>
- <sup>4</sup> <sup>1</sup>MRC Centre for Global Infectious Disease Analysis, Imperial College London, London, UK.
- 5 <sup>2</sup>WHO Collaborating Centre for Infectious Disease Epidemiology and Control, School of Public Health,
- 6 The University of Hong Kong, Hong Kong
- <sup>7</sup> <sup>3</sup>Department of Mathematics, College of Engineering, Mathematics and Physical Sciences, University
- 8 of Exeter, Exeter, UK.
- 9 \*For correspondence: <u>k.parag@imperial.ac.uk</u>.

# 10 Abstract

11 We introduce the angular reproduction number  $\Omega$ , which measures time-varying changes in 12 epidemic transmissibility resulting from variations in both the effective reproduction number R, 13 and generation time distribution w. Predominant approaches for tracking pathogen spread 14 either infer R or the epidemic growth rate r. However, R is biased by mismatches between the 15 assumed and true w, while r is difficult to interpret in terms of the individual-level branching 16 process underpinning transmission. *R* and *r* may also disagree on the relative transmissibility 17 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 18  $\Omega$  responds meaningfully to mismatches and time-variations in w while maintaining much of 19 the interpretability of R. We prove that  $\Omega > 1$  implies R > 1 and that  $\Omega$  agrees with r on the relative 20 transmissibility of pathogens. Estimating  $\Omega$  is no more difficult than inferring R, uses existing 21 software, and requires no generation time measurements. These advantages come at the 22 expense of selecting one free parameter. We propose  $\Omega$  as complementary statistic to R and 23 r that improves transmissibility estimates when w is misspecified or time-varying and better 24 reflects the impact of interventions, when those interventions concurrently change R and w or 25 alter the relative risk of co-circulating pathogens.

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

# 28 Introduction

29 Estimating the rate of spread or transmissibility of an infectious disease is a fundamental and

- 30 ongoing challenge in epidemiology [1]. Identifying salient changes in pathogen transmissibility
- 31 can contribute important information to policymaking, providing useful warnings of resurgent
- 32 epidemics, assessments of the efficacy of interventions and signals about the emergence of
- 33 new variants of concern [1–3]. The effective or instantaneous reproduction number. R and NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

34 time-varying growth rate, *r*, are commonly used to characterise pathogen transmissibility. The 35 former statistic is an estimate of the average number of new infections per active (circulating) 36 past infection, while the latter describes the exponential rate of new infection accumulation [4].

37 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 38 39 meaningfulness of R depends on our ability to measure the generation time distribution of the 40 infection under study, w. This distribution captures the inter-event times among primary and 41 secondary infections [5] and is convolved with the past infections to define  $\Lambda$ , the time-varying 42 total infectiousness of the disease. The total infectiousness serves as the denominator when 43 inferring R, which is the ratio of new infections to  $\Lambda$ . We illustrate all key notation in **Figure 1**. 44 However, infection times and hence w are difficult to measure, requiring detailed transmission 45 chain data from contact tracing or transmission studies [6]. Even if these data are available, 46 the estimated w (and hence  $\Lambda$ ) depends on how inter-event times are sampled or interpreted 47 (e.g., there are forward, backward, intrinsic and realised generation intervals) [7.8].



49 Figure 1: Definitions of transmissibility metrics. Panel A plots generation time distributions 50 that define how past infections cause later ones via the probabilities or weights w with support m. This involves convolving these weights with past infection incidence I. We show in panel B 51 52 that if we represent w and I as vectors then the convolution is equivalent to a projection of w 53 onto the vector of *I*. Panels C-D illustrate that standard reproduction numbers *R* implicitly apply 54 this projection to compute the denominator  $\Lambda$ . This projection and hence  $\Lambda$  is sensitive to the 55 w, meaning that if the distribution switches between the two from panel A, our estimates of Rbecome biased (often changes in generation time are difficult to measure). Our new metric  $\Omega$ 56 57 maximises the projection from panel B to reduce sensitivity (practically this involves a window 58 based on *m*) leading to a new denominator *M* in panel C. This maintains the branching process 59 interpretation of the epidemic in panel D, while improving transmissibility estimate robustness.

60 Workarounds, such as approximating w by the serial interval distribution [9], which describes 61 inter-event times between the onset of symptoms, or inferring w from this distribution [10], do exist but suffer from related problems [6]. Consequently, w and  $\Lambda$  can often be misspecified, 62 63 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 64 65 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 66 67 be suppressed (roughly  $R^{-1}$ ), herd immunity thresholds (related to  $1-R^{-1}$ ) or the probability of epidemic elimination and establishment (both linked to  $R^{-N}$  for N infections) [11–13]. The only 68 69 known means of attaining such information converts r into R using estimates of w [14].

70 Difficulties in accurately inferring generation times therefore cause practical bottlenecks that 71 constrain our ability to measure pathogen transmissibility. These problems are worsened as 72 recent studies have empirically found that generation times also vary substantially with time 73 (i.e., w is non-stationary) [15]. These variations may correspond to different epidemic phases 74 [16], emerging variants of concern [17] and coincide with the implementation of interventions 75 [18]. These are precisely the situations in which we also want to infer R. However, concurrent 76 changes in R and w are rarely identifiable, and r inextricably groups the effects of w and R on 77 transmissibility. While high quality, longitudinal contact tracing data [19] can potentially resolve 78 these identifiability issues, this is an expensive and logistically hard solution. Here we propose 79 another means of alleviating the above problems and complementing the insights provided by 80 R and r – the angular reproduction number,  $\Omega$ .

81 The angular reproduction number defines transmissibility as a ratio of new infections to M, the 82 root mean square number of past infections over a user-defined window  $\delta$ . Because it replaces 83  $\Lambda$  with *M*, a quantity that does not require knowledge of generation times,  $\Omega$  is more robust to 84 the problems of inferring w. We demonstrate that  $\Omega$  is able to measure the overall changes in 85 transmissibility caused by fluctuations in both R and w. Moreover, we prove that  $\Omega$  has similar threshold properties to R, maintains much of its individual-level interpretation and is a useful 86 87 metric for communicating transmissibility. This last point follows as we only need to quote  $\Omega$ 88 and the known window  $\delta$  to generalise our estimates of transmissibility to different settings. In 89 contrast, the meaningfulness of R is contingent on the unknown or uncertain w. Downstream 90 studies sometimes use R outside of its generation time context [20], while dashboards aiming 91 at situational awareness commonly quote R without w, introducing biases and interpretability 92 problems into how disease spread is communicated [21].

93 Additionally, we demonstrate how r and R can easily disagree on relative transmissibility, both 94 across time and for co-circulating variants. Unmeasured changes in w over time can cause R 95 and r to vary in opposite directions (one signals an increase in transmissibility and the other a 96 decrease). Similarly, co-circulating pathogens with different but stationary and known w, may 97 possess contradictory R and r value rankings i.e., for variants A and B,  $r_A > r_B$  does not imply 98  $R_A > R_B$ . These issues are amplified when interventions (which can change w, R or both [18]) 99 occur, obscuring notions of the relative risk of spread. However, we find  $r_A > r_B$  guarantees  $\Omega_A$ 100 >  $\Omega_B$  and that  $\Omega$  is consistent with *r* across time even if *w* changes.

101 Last, while we may also convert r into threshold statistics about 1 by using a free parameter 102 together with a transformation from [14], we show that  $\Omega$  is more robust to choices of its free 103 parameter than those statistics, which implicitly make stronger assumptions (Supplementary 104 **Information**). These robustness and consistency properties of  $\Omega$  reinforce its usefulness for 105 tracking and comparing outbreak spread and emerge from its maximum entropy approach to 106 managing uncertain generation time distributions. We propose  $\Omega$  as a complementary statistic 107 that can be integrated with R and r to present a more comprehensive perspective on epidemic 108 transmissibility, especially when w is poorly specified or varying with time.

# 109 Results

# 110 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,  $I_1^{t-1} \stackrel{\text{def}}{=} \{I_1, I_2, \dots I_{t-1}\}$ . This results in the conditional moment relationship in **Eq. (1)** [9]. Generally, we use  $X_a^b$  to denote the time series  $\{X_a, X_{a+1}, \dots, X_{b-1}, X_b\}$  and **E**[X|Y] for the

116 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  $\mathbf{E}[X]$ .

118 
$$\mathbf{E}[I_t | I_1^{t-1}, w_1^m] = R_t \Lambda_t, \qquad \Lambda_t = \sum_{u=1}^m w_u I_{t-u}.$$
 (1)

In this model  $\Lambda_t$  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].

125 Although Eq. (1) has successfully been applied to model many diseases including COVID-19. 126 Ebola virus disease, pandemic influenza and measles, among others, it has one major flaw -127 it assumes that the generation time distribution is fixed or stationary and known [9]. If this 128 assumption holds (we ignore surveillance biases [9,23] until the Discussion), Eq. (1) allows 129 epidemic transmissibility to be summarised by fluctuations of the time-varying  $R_t$  parameters. 130 This follows because the sign of  $R_t - 1$  determines if  $I_t$  will increase or decline relative to the 131 total infectiousness  $\Lambda_t$ . 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]. 132

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  $R_t = \mathbf{E}[I_t]\Lambda_t^{-1}$ . As  $R_t$  is a threshold parameter, we know that we must block at least a fraction  $1 - R_t^{-1}$  of new infections 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 these 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  $\Lambda_t$  is incorrectly specified [4]), which even impede optimal Bayesian inference algorithms [25]. As  $R_t$  is the 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 timeinformation, this is usually difficult and expensive. We propose a robust alternative.

148 We redefine  $\Lambda_t$  by recognising it as a dot product between the vectors 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 149 distribution, *m*. This gives the left equality of **Eq. (2)** with the Euclidian norm of  $\vec{X}$  as  $\|\vec{X}\| \leq \|\vec{X}\|$ 150  $(\sum_{u=1}^{m} X_{u}^{2})^{\frac{1}{2}}$  and  $\theta_{t}$  as the time-varying angle between  $\vec{w}$  and  $\vec{l}$ . This equality holds for non-151 stationary generation times i.e., both  $\vec{w}$  and  $\vec{l}$  can have elements that change over time. We 152 illustrate this notation and elements of the subsequent derivation in Figure 1. Eq. (2) implies 153 that the count of new infections (for any  $R_t$ ) is maximised when  $\theta_t$  is minimised i.e., when the 154 155 temporal profile of past infections matches the shape of the generation time distribution.

156 
$$\Lambda_t = \|\vec{w}\| \|\vec{l}\| \cos \theta_t, \qquad \mathbf{E}[I_t] = \left(\frac{\|\vec{w}\|}{\|\vec{w}_{\max}\|} R_t \cos \theta_t\right) M_t.$$
(2)

We can compute the root mean square of the incidence across the support of the generation time distribution as  $M_t \stackrel{\text{def}}{=} \frac{1}{\sqrt{m}} \|\vec{I}\|$ . 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  $\frac{1}{m}$ ).

162 Combining these definitions with **Eq. (1)**, we derive the second expression in **Eq. (2)** for the 163 expected number of new infections at time *t*. This may seem an unnecessarily complicated 164 manipulation of the standard renewal model, but it admits a novel and important insight – we 165 can separate the influences of the reproduction numbers and the generation time distribution 166 (together with its changes) on epidemic transmissibility. These multiply  $M_t$ , which defines a 167 new denominator – the root mean square number of past infections (this is also the average 168 signal power of the past infection time series) – that replaces the total infectiousness  $\Lambda_t$ .

169 Consequently, we define a new metric in **Eq. (3)**, the *angular reproduction number*  $\Omega_t$ , which 170 multiplies  $R_t$  by the scaled projection of the generation time distribution,  $\frac{\|\vec{w}\|}{\|\vec{w}_{\max}\|} \cos \theta_t$ , onto  $\vec{I}$ , 171 the past incidence vector (see **Figure 1**). This means that  $\Omega_t$  is a time-varying ratio between 172 the expected infection incidence and the past root mean square incidence  $M_t$ . We use the

term reproduction number for  $\Omega_t$  due to its relation to  $R_t$ , the similarity of **Eq. (2)** and **Eq. (3)** and because of its threshold properties, which we explore in the next section.

175 
$$\Omega_t \stackrel{\text{\tiny def}}{=} \frac{\|\vec{w}\|}{\|\vec{w}_{\max}\|} R_t \cos \theta_t \implies \Omega_t = \mathbf{E}[I_t] M_t^{-1}. \tag{3}$$

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  $\theta_t$ , 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  $\mathbf{E}[I_t] = R_t \Lambda_t$ .

181 We argue that this is not the case practically because  $\Lambda_t$  is frequently misspecified [15,26], 182 obscuring the meaning of  $R_t$ . In contrast,  $M_t$  does not depend on generation time assumptions (beyond characterising its support m). We remove structural uncertainty induced by the often 183 unknown  $w_u$  because  $M_t$  is a maximum entropy version of  $\Lambda_t$  i.e.,  $M_t = \max_{\|\overline{w}\| \cos \theta_t} \Lambda_t =$ 184  $\|\vec{w}_{\max}\|\|\vec{I}\|$  subject to  $\sum_{u=1}^{m} w_u = 1$ . We also find that  $M_t = \Lambda_t$  and hence  $\Omega_t = R_t$ , when 185 the past incidence is flat (as then  $\Lambda_t = M_t$  and  $w_u$  has no effect). This defines the important 186 and universal equilibrium condition  $\Omega_t = R_t = 1$ . There is further convergence for branching 187 188 process models [27] with timesteps at its fixed generation time, as then trivially  $w_1 = 1$ .

#### 189 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 known relation gr = (R - 1) [28], with g as the mean generation time. We assume growth occurs over some period of  $\delta$  and compute  $\Omega_t$  as the ratio  $\mathbf{E}[I_t]M_t^{-1}$  from **Eq. (3)**. Since this model is deterministic  $\mathbf{E}[I_t] = I_t = I_0e^{rt}$ .

197 We evaluate  $M_t$  from its definition above as  $\frac{1}{\sqrt{m}} \|\vec{I}\|$  with  $\delta = m$  and using the continuous-time 198 expression for  $\|\vec{I}\| = \left(\int_{t-\delta}^t I_s^2 ds\right)^{\frac{1}{2}}$ . This yields  $M_t = \left(\delta^{-1} \int_{t-\delta}^t I_s^2 ds\right)^{\frac{1}{2}}$  with  $I_s^2 = I_0^2 e^{2rs}$  from 199 the exponential incidence equation and evaluates to  $\sqrt{2\delta r I_0 (e^{2rt} - e^{2r(t-\delta)})}$ . Substituting 200 this into  $\Omega_t = (I_0 e^{rt}) M_t^{-1}$  results in the left relation in **Eq. (4)**.

201 
$$\Omega_t^2 = \frac{2\delta r}{1 - e^{-2\delta r}} \ge 1, \qquad \Omega_t^2 = \frac{2\delta g^{-1}(R-1)}{1 - e^{-2\delta g^{-1}(R-1)}}.$$
 (4)

Several important points follow. First, as  $x \ge 1 - e^{-x}$  for every  $x \ge 0$ , then  $\Omega_t - 1$  and r are positive too (an analogous argument proves the negative case). Second, we substitute for rusing 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  $\Omega_t$ i.e., the sign of  $\Omega_t - 1$  and  $R_t - 1$  are always consistent (for all values of  $\delta > 0$ ).

207 Third, we see that constant growth rates imply constant angular reproduction numbers. The 208 converse is also true, and we may input time-varying growth rates,  $r_t$ , into Eq. (4) to estimate 209  $\Omega_t$ . These properties hold for any  $\delta$ , which is now a piecewise-constant time window. We plot 210 the ramifications of Eq. (4) in Figure 2. Further, in Table 1 we summarise how  $\Omega_t$  relates to 211 predominant  $R_t$  and  $r_t$  metrics. We explore some properties in this table in later sections (in 212 addition to reinforcing our analyses with stochastic models) and demonstrate that relationships 213 among  $r_t$ ,  $R_t$  and  $\Omega_t$  have important consequences when comparing outbreaks subject to 214 interventions and variations in generation times.

Metric property	Growth <i>r</i>	Effective R	Angular $\Omega$
Definition of transmissibility	$r_t \stackrel{\text{\tiny def}}{=} \frac{d\log \mathbf{E}[I_t]}{dt}$	$R_t \stackrel{\text{\tiny def}}{=} rac{\mathbf{E}[I_t]}{\Lambda_t}$	$\Omega_t \stackrel{\text{\tiny def}}{=} rac{\mathbf{E}[I_t]}{M_t}$
Pathogen spread threshold	$r_t > 0$	$R_t > 1$	$\Omega_t > 1$
Biased by generation time $\vec{w}$ assumed, given curve $I_1^t$	reration time $\vec{w}$ Insensitive to the Biased when $\vec{w}$ ven curve $I_1^t$ assumed $\vec{w}$ 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	$r_A > r_B  eq R_A >$ $R_B$ (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 $\delta$

**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 ( $\Omega$ ) and assess their value as measures of transmissibility.

Note that we may also invert the relationship in **Eq. (4)** to estimate  $r_t$  from  $\Omega_t$  (see Methods for details). 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].

221 
$$\frac{d\log I_t}{dt} = r_t = 2\delta^{-1} \left(\Omega_t^2 + W_{-1(\Omega<1)} \left(-\Omega_t^2 e^{-\Omega_t^2}\right)\right).$$
(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.

224 We further comment on connections between angular and effective reproduction numbers 225 using a deterministic branching process model, which is also foundational in epidemiology. 226 We again focus on growth, which is geometric as this is a discrete-time process with time steps scaled in multiples of the mean generation time g. Here incidence is  $I_t = R^t$  and  $\Omega_t =$ 227  $\mathbf{E}[I_t]M_t^{-1} = R^t \left(\delta^{-1} \sum_{s=t-\delta}^{t-1} R^{2s}\right)^{-\frac{1}{2}}, \text{ with window } \delta \text{ in units of } g. \text{ If } \delta = 1 \text{ we recover } \Omega_t = R.$ 228 If R = 1, then  $\Omega_t = R$  for all  $\delta$ . For growing epidemics, as  $\delta$  increases,  $\Omega_t > R$  because we 229 230 reference present incidence to smaller past infections (or denominators). The opposite occurs 231 if the epidemic declines. This may seem undesirable, but we argue that  $\Omega_t$  improves overall

232 practical transmissibility measurement because g will likely be misspecified or vary with time.

233 Any g mismatches bias R, limiting its interpretation, meaningfulness and making comparisons 234 among outbreaks or pathogenic variants difficult, because we cannot be certain that our 235 denominators correspond. This is particularly problematic when estimates of *R* obtained from 236 a modelling study are incorporated as parameters into downstream studies without accounting 237 for the generation time context on which those estimates depend. However, by additionally 238 communicating  $\Omega$  and  $\delta$ , we are sure that denominators match and that we properly include 239 the influences of any g mismatches. Choosing  $\delta$  is also no worse (and more explicit) than 240 equivalent window assumptions made when inferring R and r [4,30] In the **Supplementary** 

241 Information we perform analyses of window choices for  $\Omega$  and other threshold metrics



242

Figure 2: Relationships among transmissibility metrics. Panel A and B show how growth rates (r) and reproduction numbers (R) have diverse functional relationships (see [14,30]) for SEIR models with an excess of susceptible individuals and branching processes. Coloured lines indicate R at different mean generation times (g). Black lines highlight a single functional relationship between angular reproduction numbers  $\Omega$  and r at all g, using a window  $\delta$  of 20d. Panel C shows that while  $\Omega$  varies with choice of  $\delta$  (increasing from blue to red and computed

- from **Eq. (4)**), we have a bijective relationship with *r*. Panel D indicates that *R* and *r* can signify
- inverted changes e.g., an NPI reducing *R* and *g* may increase *r*, raising questions about impact
- 251 (see [15,18]). Here  $\Omega$  converts *r* into a consistent transmissibility metric (also from Eq. (4)).

252 Last, we illustrate how  $\Omega_t$  relates to other key indicators of epidemic dynamics such as herd immunity and elimination probabilities. As our derivation replaces Eq. (1) with  $\mathbf{E}[I_t | I_1^{t-1}, \delta] =$ 253 254  $\Omega_t M_t$  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 255 by  $e^{-\sum_{s=t}^{\infty} M_t \Omega_t}$ , and has analogous properties [31]. Herd immunity, which traditionally occurs 256 when a fraction  $1 - R^{-1}$  of the population is immune is approximated by  $1 - \Omega^{-1}$  (since both 257 metrics possess the same threshold behaviour) [11]. In a subsequent section we demonstrate 258 259 that one-step-ahead incidence predictions from both approaches are also comparable.

#### 260 **Responding to variations in generation time distributions**

- 261 We demonstrate the practical benefits of  $\Omega_t$  using simulated epidemics with non-stationary or 262 time-varying generation time distributions. Such changes lead to misspecification of  $\Lambda_t$  in **Eq.** (1), making estimates of the effective reproduction number  $R_t$ , denoted  $\hat{R}_t$ , a poor reflection 263 of the true underlying  $R_t$ . In contrast variations in the estimated  $\hat{\Omega}_t$  are a feature (see Eq. (3)) 264 and not a bug (for some chosen  $\delta$  we control  $M_t$ , which is not misspecified). We simulate 265 epidemics with Ebola virus or COVID-19 generation times from [32,33] using renewal models 266 with Poisson noise [9]. We estimate both the time-varying  $R_t$  and  $\Omega_t$  using EpiFilter [25], which 267 268 applies Bayesian algorithms that minimise mean square estimation error.
- Inferring  $\Omega_t$  from incident infections,  $I_1^t$ , requires only that we replace the input  $\Lambda_t$  with  $M_t$  in 269 270 the estimation function and that we choose a window  $\delta$  for computing  $M_t$ . We provide software 271 for general estimation of  $\Omega_t$  and code for reproducing this and all other analyses in this paper 272 at <u>https://github.com/kpzoo/Omega</u>. We heuristically set  $\delta \approx 2g_0$  as our window with  $g_0$  as 273 the original mean generation time of each disease from [32,33]. We find (numerically) that this  $\delta$  ensures  $\sum_{\mu=0}^{\delta} w_{\mu} \ge 0.86$  over many possible gamma distributed generation times i.e., it is 274 275 long enough to cover most of the likely probability mass of unknown changes to the generation 276 time distributions, which cause time-varying means  $g_t$ . In general, we find that an overly small 277  $\delta$  tends to neglect important dynamics, while too large a  $\delta$  induces edge effects. The Methods 278 and **Supplementary Information** provide for more information on choosing  $\delta$ .

Our results are plotted in **Figure 3**. We show that  $\hat{\Omega}_t$  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 3** 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 the misspecification of  $w_1^m$ . These can strongly mislead our interpretation and understanding of the risk posed by a pathogen.





Figure 3: Estimating transmissibility under temporal variations in generation times. We simulate epidemic incidence curves (black) using generation time distributions of Ebola virus disease (EVD) [32] and COVID-19 [33] in panels A and B. The means of these distributions (g) vary over time (grey piecewise, starting from original mean g<sub>0</sub>), but we fix their variance at their original values. We find substantial bias in *R* estimated from the initial EVD and COVID-19 generation times (red with 95% credible intervals, true value in black). These estimates try to compensate for generation time mismatches and changes in an uncontrolled manner that obscures interpretation. However,  $\Omega$  responds as we expect (blue with 95% credible intervals, window  $\delta$ , 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 time series to start from  $\delta$  to remove any edge effects. Vertical dashed lines highlight times at which we change *g* or keep it fixed. When it is fixed,  $\Omega$  infers no spurious changes.

300 We can derive alternative threshold statistics that relate to  $r_t$  and do not explicitly depend on 301 the generation time by applying monotonic transformations from [14]. In theory these should 302 have comparable behaviour around the critical point of 1 to both  $R_t$  and  $\Omega_t$ . We investigate 303 these statistics in the **Supplementary Information**, computing them across the simulations 304 of **Figure 3**. We find that they require stronger assumptions than  $\Omega_t$  (i.e., they fix distributional 305 formulae for generation times), possess at least as many free parameters as  $\Omega_t$  and are less 306 robust to changes in those parameters (often strongly over-estimating transmissibility), than 307  $\Omega_t$  is to fluctuations in  $\delta$ . This confirms that angular reproduction numbers can complement 308 standard metrics, improving transmissibility estimates when generation times are changing, 309 or unknown and forming part of a more comprehensive suite of outbreak diagnostics.

### 310 Ranking epidemics or variants by transmissibility

311 Misspecification of generation time distributions, and corresponding misestimation of R as in 312 **Figure 3**, also plays a crucial role when assessing the relative transmissibility of pathogens, 313 variants of concern or even outbreaks (where we may want to contrast the spread of contagion 314 among key demographic or spatial groups). As shown in **Figure 2**, these variations can mean 315 that increases in the growth rate  $r_t$  actually signify decreases in the effective reproduction 316 number  $R_t$  or that a pathogen with a larger  $r_t$  can have a smaller  $R_t$ . Here we illustrate that 317 these issues can persist even if the generation time distributions of pathogens are correctly 318 specified and remain static, obscuring our understanding of relative transmission risk.

319

320 In Figure 4 we simulate epidemics under two hypothetical variants of two pathogens. We use 321 EVD and COVID-19 generation time distributions from [32,33] to define our respective base 322 variants. For both pathogens we specify the other variant by reducing the mean generation of 323 each base but fixing the variance of the generation times. Reductions of this type are plausible and have been measured for COVID-19 variants [17]. All  $w_1^m$  distributions are stationary and 324 325 known in this analysis. We discover that changes in  $R_t$  alone can initiate inversions in the 326 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 327

328 combination with  $w_1^m$  [15,18], this effect has the potential to be widespread. We determine the 329 mathematical conditions for this inversion in the **Supplementary Information**.

330 Interestingly, the angular reproduction numbers of Figure 4 do preserve an ordering that is 331 consistent with growth rates, while maintaining the interpretability (e.g., threshold properties) 332 of reproduction numbers. Hence, we argue that  $\Omega_t$  blends advantages from both  $R_t$  and  $r_t$  [4] 333 and serves as a useful outbreak analytic for understanding and conveying the relative risk of 334 spread of differing pathogens or pathogen strains, or of spread among different spatial and 335 demographic groups. Recent studies have only begun to disentangle component drivers of 336 transmission, including the differing effects that interventions can introduce (e.g., by defining the strength and speed of control measures [34]) and the diverse properties of antigenic 337 338 variants [17]. We believe that  $\Omega$  can play a distinctive role in accelerating these investigations. 339





Figure 4: Comparing transmissibility across outbreaks, variants or even diseases. We
 simulate epidemics of variant 1 in blue (with estimates of metrics also in blue) under standard
 generation time distributions of Ebola virus disease (EVD) [32] and COVID-19 [33] in panels

344 A and B. In red (with estimates also in red) we overlay simulations in which the generation 345 time of these diseases is 40% and 50% shorter (than the blue epidemics), which may indicate 346 a new co-circulating variant 2 or another epidemic with different properties (e.g., in a higher 347 risk group). We demonstrate (for the first time to our knowledge) that changes in R due to an 348 intervention (or release of one) may invert the relative growth rates (r) of the epidemics (see 349 **Supplementary Information** for mathematical intuition for this inversion). The mismatches in 350 the *R*-*r* rankings alter perceptions of relative risk, making transmissibility comparisons difficult. 351 However,  $\Omega$  classifies the risk of these epidemics in line with their realised growth rates, while 352 still offering the individual-level interpretability of a reproduction number. True values are in 353 black and all estimates (with 95% credible intervals) are outputs from EpiFilter [25] with default 354 settings. We truncate the time series to start from  $\delta$  to remove any edge effects.

### 355 **Reproduction numbers for explanation or prediction?**

356 We highlight an important but underappreciated subtlety when inferring the transmissibility of 357 epidemics – that the value of accurately estimating R, r and  $\Omega$  largely depends on if our aim 358 is to explain or predict [35] the dynamics of epidemics. The above analyses have focussed on 359 characterising transmissibility to explain mechanisms of spread and design interventions. For 360 these problems, misestimation of parameters, such as R, can bias our assessment of outbreak 361 risk and hence misinform the implementation of control measures. An important concurrent 362 problem aims to predict the likely incidence of new infections from these estimates. This 363 involves projecting the epidemic dynamics forward in time to infer upcoming infection patterns.

364 Here we present evidence that the solution of this problem, at least over short projection time 365 horizons, is robust to misspecification of generation times provided both the incorrect estimate 366 and the misspecified denominator are used in conjunction. We repeat the analyses of Figure 367 **3** for 200 replicate epidemics and apply EpiFilter [25] to obtain the one-step-ahead predictive distributions  $P(I_t | I_1^{t-1})$  for every t. We compute the predicted mean square error (PMSE) 368 and the accumulated predictive error (APE). These scores, which we denote as  $D(I_t | I_1^{t-1})$ , 369 370 average square errors between mean predictions and true incidence and sum log probabilities 371 of observing the true incidence from the predicted distribution respectively [36,37]. We plot the 372 distributions of scores over replicates and illustrate individual predictions in Figure 5.

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  $\Omega$ . 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 ( $\Lambda_t$ ) 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-stepahead predictions can be composed from iterated one-step-ahead ones [38], we conjecture that subtleties between prediction and explanation are likely to also apply on longer horizons.



383

384 Figure 5: One-step-ahead prediction accuracy and model mismatch. We simulate 200 385 replicates of the epidemics from Figure 3, which involve non-stationary changes to EVD and 386 COVID-19 generation times. We use estimates of effective, R, and angular,  $\Omega$ , reproduction 387 numbers to produce successive one-step-ahead predictions and assess their accuracy to the 388 simulated (true) incidence. Panels A-D provide a representative example of a single simulated 389 epidemic (true incidence shown as black dots) and the R and  $\Omega$  one-step ahead predictions 390 (red and blue respectively with 95% credible intervals). In panels E-F we formally compute 391 accuracy using distance metrics, *D*, based on accumulated prediction errors (APE, dashed) 392 and prediction mean square errors (PMSE, solid) for all 200 replicates from R,  $\Omega$  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.

### 395 Empirical example: COVID-19 in mainland China

396 We complete our analysis by illustrating the practical usability of  $\Omega$  on an empirical case study 397 where generation time changes are known to have occurred. In [15], the dynamics of COVID-398 19 in mainland China are tracked across January and February 2020. Transmission pair data 399 indicated that the serial interval of COVID-19 shortened across this period leading to biases 400 in the inferred R if updated serial intervals are not used. Here serial intervals, which measure 401 the lag between the symptom onset times of an infector and infectee are used as a proxy for 402 the generation time. Figure 6 presents our main results. We find  $\Omega$  (blue), which requires no 403 serial interval information, behaves similarly to the R (red) inferred from the time-changing w. 404 Both metrics appear less biased than estimates of R (green) that assume a fixed serial interval. 405 This is largely consistent with the original investigation in [15].



406

407 Figure 6: COVID-19 transmissibility in China under non-stationary generation times. We analyse COVID-19 data from [15], which spans 9<sup>th</sup> January 2020 to 13 February 2020 and is 408 409 known to feature a serial interval distribution that shortened in mean substantially from 7.8d 410 to 2.6d (change times are shown as grey vertical lines). We assume that the serial interval 411 approximates the generation time well and replicate the analysis from Figure 2 of [15]. In panel 412 A, we compare estimates (green) of effective reproduction numbers, R, using fixed generation 413 time distributions inferred in [15] (specified by their means q) against those of our angular 414 reproduction number  $\Omega$  (blue). We use EpiFilter [25] to obtain all estimates (means shown 415 with 95% credible intervals) and find relative trends similar to those in Figure 2 of [15]. In panel 416 B we plot the incidence (black) and the denominators we use to compute an R that does 417 account for the generation time changes ( $\Lambda$ , red) and for  $\Omega$  (*M*, blue). This *R* uses the different 418 distributions inferred at the grey vertical change times (their means are in panel B and are 419 also the fixed distributions of panel A in sequence). We plot these R and  $\Omega$  estimates in panel 420 C. In panel D we show the growth rates that are inferred from the R and  $\Omega$  estimates of C (red 421 and blue respectively) against that obtained from taking the smoothed log derivative (black).

422 We see that  $\Omega$  provides a lower assessment of the initial transmissibility as compared to the 423 R that is best informed by the changing w but that both agree in general and in particular at 424 the important threshold between super- and subcritical spread. Interestingly,  $\Omega$  indicates no 425 sharp changes at the w change-times. This follows because the incidence is too small for 426 those changes to shape overall transmissibility and matches the gradual w changes originally 427 inferred in [15]. The distributions used in **Figure 6** provide a piecewise approximation to these 428 variations. We also compare *r* estimates derived from *R* (red, from [14]),  $\Omega$  (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 429 430 from  $\Omega$  agrees more closely with the empirical growth rate than the r from R, which somewhat 431 by design shows jumps at the w change-points. While this analysis is not meant as a detailed 432 study of COVID-19 in China, it does demonstrate the practical usefulness of  $\Omega$ .

### 433 **Discussion**

434 Quantifying the time-varying transmissibility of a pathogen remains an enduring challenge in 435 infectious disease epidemiology. Changes in transmissibility may signify shifts in the dynamics 436 of an epidemic of relevance to both preparedness and policymaking. While this challenge has 437 been longstanding, the statistics that we use to summarise transmissibility have evolved from 438 dispersibility [39] and incidence to prevalence ratios [40] to cohort [41] and instantaneous [22] 439 reproduction numbers. While the last, which we have denoted *R*, has become the predominant 440 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 appropriatedenominators necessitates some notion (implicit or explicit) of a generation time [42].

443 Difficulties in characterising these generation times and their changes substantially bias [6] 444 estimates of transmissibility and have motivated recent works to propose the instantaneous 445 growth rate, r, as a more reliable approach for inferring pathogen spread [20]. However, on its 446 own, r is insufficient to resolve many of the transmission questions that R can answer and its 447 computation may employ smoothing assumptions that are in some instances equivalent to the 448 generation time ones behind R [4]. We formulated the novel angular reproduction number  $\Omega$ , 449 to merge some advantages from both R and r and to contribute to a more comprehensive view 450 of transmissibility. By applying basic vector algebra (Eqs. 1-3), we encoded both changes to 451 *R* and the generation time distribution, *w*, into a single time-varying metric, deriving  $\Omega$ .

452 We found that  $\Omega$  maintains the threshold properties and individual-level interpretability of R 453 but responds to variations in w, in a manner consistent with r (Figure 2). Moreover,  $\Omega$  indicates 454 variations in transmissibility caused by R and w without requiring measurement of generation 455 times (Figure 3). This is a consequence of its denominator, which is the root mean square of 456 infections over a user-specified window  $\delta$  that is relatively simple to tune (see Methods). We 457 can interpret  $\Omega = a > 1$  as indicating that infections across  $\delta$  need to be reduced by  $a^{-1}$ . This 458 reduces mean and root mean square infections by  $a^{-1}$  and causes  $\Omega$  to equal 1. Further,  $\Omega$ 459 circumvents identifiability issues surrounding the joint inference or R and w[43] by refocussing 460 on estimating the net changes produced by both. This improves our ability to explain the shifts 461 in transmissibility underpinning observed epidemic dynamics and means  $\Omega$  is essentially a 462 reproduction number that provides individual-level interpretation of growth rates (Egs. 4-5).

463 The benefits of this *r*- $\Omega$  correspondence are twofold. First, as interventions may alter *R*, *w* or 464 R and w concurrently [15,18] situations can arise where r and R disagree across time on both 465 the drivers and magnitude of transmissibility. While it may seem possible to minimise this issue 466 by constructing alternative threshold statistics by directly combining r with assumed generation 467 time structures, we find these statistics often exhibit worse performance and larger bias than 468  $\Omega$  (**Supplementary Information**). Second, this disagreement can also occur when comparing pathogenic variants or epidemics (e.g., from diverse spatial or sociodemographic groups) with 469 470 different but known and unchanging w. This study appears to be among the earliest to highlight 471 these discrepancies, which can occur in many settings (see **Supplementary Information**). 472 Realistic transmission landscapes possess all of the above complexities, meaning that relying 473 solely on conventional measures of relative transmissibility can lead to contradictions.

474 We found that  $\Omega$  consistently orders epidemics by growth rate while capturing notions of the 475 average new infections per past infection (**Figure 4**). This suggests  $\Omega$  blends advantages from 476 R and r, with clearer assumptions (choice of window  $\delta$ ). However,  $\Omega$  offers no advantage if we want to predict epidemic dynamics (see [35] for more on prediction-explanation distinctions). 477 478 For this problem even an *R* inferred using a misspecified denominator performs equally well 479 (Figure 5). This follows as only the product of any reproduction number and its denominator 480 matter when determining the next incidence value. Iterations of this product underpin multi-481 step ahead predictions [38]. This may explain why autoregressive models, which ignore some 482 characteristics of w, can serve as useful predictive models [44]. Other instances where  $\Omega$  will 483 not improve analysis are at times earlier than  $\delta$  (due to edge effects [9]) and in periods of near 484 zero incidence (there is no information to infer R either [24]). We summarised and compared 485 key properties of *R*, *r* and  $\Omega$  in **Table 1**.

486 There are several limitations to our study. First, we only examined biases inherent to R due to 487 the difficulty of measuring the generation time accurately and across time. While this is a major 488 limitation of existing transmissibility metrics [15], practical surveillance data are also subject 489 to under-reporting and delays, which can severely diminish the quality of any transmissibility 490 estimates [23,43,45]. While  $\Omega$  ameliorates issues due to generation time mismatch, it is as 491 susceptible as R and r to surveillance biases and corrective algorithms (e.g., deconvolution 492 methods [46]) should be applied before inferring  $\Omega$ . Second, our analysis depends on renewal 493 and compartmental epidemic models [22]. These assume random mixing and cannot account 494 for realistic contact patterns. Despite this key structural uncertainty, there is evidence that well-495 mixed and network models are comparable when estimating transmissibility [47].

496 Although the above limitations can, in some instances, reduce the added value of improving 497 the statistics summarising transmissibility, we believe that  $\Omega$  will be of practical and theoretical 498 benefit, offering complementary insights to R and r and forming part of a more comprehensive 499 epidemic analytic toolkit. Its similarity in formulation to R means it is as easy to compute using 500 existing software and therefore can be deployed on dashboards and updated in real time to 501 improve situational awareness. Further,  $\Omega$  improves comparison and communication of the 502 relative risks of circulating variants or epidemics among diverse groups, avoiding R-r 503 contradictions provided the known parameter,  $\delta$ , is fixed. This supplements *R*, which is hard 504 to contextualise [20] when w is misspecified or varying and hence compare across groups, as 505 each group may have distinct and correspondingly poorly specified denominators. Last,  $\Omega$  can 506 help probe analytical questions about how changes in R and w interact because it presents a 507 common framework for testing how variations in either influence overall transmissibility.

#### 508 Methods

#### 509 Inferring angular reproduction numbers across time

510 We outline how to estimate  $\Omega_t$  given a time series of incident infections  $I_1^T$ , with T defining the 511 present or last available data timepoint i.e.,  $1 \le t \le T$ . Because  $\Omega_t$  simply replaces the total 512 infectiousness  $\Lambda_t$ , used for computing  $R_t$ , with the root mean square of the new infection time 513 series (see **Figure 1**),  $M_t$ , we can obtain  $\Omega_t$  from standard  $R_t$  estimation packages with minor 514 changes. This requires evaluating  $M_t$  over some user-defined, backward sliding window of 515 size  $\delta$ . Under a Poisson (Pois) renewal model this follows as in **Eq. (6)** for timepoint t.

516 
$$\mathbf{P}(I_t | I_1^{t-1}, \delta) \equiv \text{Pois}(\Omega_t M_t), \qquad M_t = \left(\frac{1}{\delta} \sum_{u=t-\delta}^{t-1} I_u^2\right)^{\frac{1}{2}}.$$
 (6)

517 The choice of  $\delta$  is mostly arbitrary but should be sufficiently long to capture most of the likely 518 probability mass of the unknown generation time but not overly long since it induces an edge 519 effect (similar to the windows in [9,37]). We found a suitable heuristic to be twice or thrice the 520 initial expected mean generation time ( $g_0$ ). We can then input  $M_t$  and  $I_t$  into packages such 521 as EpiEstim [9] or EpiFilter [25] to estimate  $\Omega_t$  with 95% credible intervals.

522 Due to the similarity between computing  $R_t$  and  $\Omega_t$  we only specify the latter but highlight that 523 replacing  $M_t$  with  $\Lambda_t$  yields the expressions for evaluating any equivalent quantities from  $R_t$ . 524 The only difference relates to how the growth rates  $r_t$  are computed. We estimate  $r_t$  from  $R_t$ 525 by applying the generation time,  $\vec{w}$ , based transformation from [14]. For a correctly specified 526  $\vec{w}$  this gives the same result as the smoothed derivative of the incidence curve [4]. We derive  $r_t$  from  $\Omega_t$  using Eq. (5), which follows from rearranging Eq. (4) into  $(2\delta r_t - \Omega_t^2)e^{2\delta r_t - \Omega_t^2} =$ 527  $-\Omega_t^2 e^{-\Omega_t^2}$ . This expression then admits Lambert W function solutions. In all estimates of  $r_t$  we 528 propagate uncertainty from the posterior distributions (see below) over  $R_t$  or  $\Omega_t$ . 529

530 We applied EpiFilter in this study due to its improved extraction of information from  $I_1^T$ . This 531 method assumes a random walk state model for our transmissibility metric as in **Eq. (7)** with 532  $\epsilon_{t-1}$  as a normally distributed (Norm) noise term and  $\eta$  as a free parameter (default 0.1).

533 
$$\Omega_t = \Omega_{t-1} + \left(\eta \sqrt{\Omega_{t-1}}\right) \epsilon_{t-1}, \quad \mathbf{P}(\epsilon_{t-1}) \equiv \operatorname{Norm}(0, 1). \quad (7)$$

534 The EpiFilter approach utilises Bayesian smoothing algorithms incorporating the models of 535 **Eq. (6)-(7)** and outputs the complete posterior distribution  $\mathbf{P}(\Omega_t | I_1^T, \delta)$  with *T* as the complete

- 536 length of all available data (i.e.,  $1 \le t \le T$ ). We compute our mean estimates  $\hat{\Omega}_t$  and 95% 537 credible intervals from this posterior distribution and these underlie our plots in **Figures 3-4**.
- 538 EpiFilter also outputs the one-step-ahead predictive distributions  $P(I_t | I_1^{t-1}, \delta)$ , which we use 539 in **Figure 5**. There we quantify predictive accuracy using the predicted mean square error
- 540 PMSE and the accumulated prediction error APE, defined as in Eq. (8) [36,37] with  $\hat{I}_t$  as the
- 541 posterior mean estimate from  $\mathbf{P}(I_t | I_1^{t-1}, \delta)$  and  $I_t^*$  as the true simulated incidence. These are
- 542 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.

543 
$$PMSE = \frac{1}{T - \delta} \sum_{t=\delta+1}^{T} (I_t^* - \hat{I}_t)^2, \quad APE = \sum_{t=\delta+1}^{T} -\log \mathbf{P}(I_t = I_t^* | I_1^{t-1}, \delta). \quad (8)$$

We collectively refer to these as distance metrics  $D(I_t | I_1^{t-1})$  and construct their distributions, **P**(*D*), over many replicates of simulated epidemics. Last, we use **P**( $\Omega_t | I_1^T, \delta$ ) to compute the posterior distribution of the growth rate **P**( $r_t | I_1^T, \delta$ ) and hence its estimates as in **Eq. (5)**. More details on the EpiFilter algorithms are available at [25,31,48]. We supply open source code to reproduce all analyses at <u>https://github.com/kpzoo/Omega</u> as well as functions in MATLAB and R to allow users to estimate  $\Omega_t$  from their own data.

# 550 **Bibliography**

- 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;
- 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 timevarying 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
- 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
- 561 4. Parag KV, Thompson RN, Donnelly CA. Are epidemic growth rates more informative 562 than reproduction numbers? J Royal Statistical Soc A. 2022; doi:10.1111/rssa.12867
- 563 5. Svensson A. A note on generation times in epidemic models. Math Biosci. 2007;208: 564 300–311. doi:10.1016/j.mbs.2006.10.010
- 5656.Britton T, Scalia Tomba G. Estimation in emerging epidemics: biases and remedies. J566R Soc Interface. 2019;16: 20180670. doi:10.1098/rsif.2018.0670
- 567 7. Champredon D, Dushoff J. Intrinsic and realized generation intervals in infectious568 disease transmission. Proc Biol Sci. 2015;282: 20152026. doi:10.1098/rspb.2015.2026
- 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
- 572 9. Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to

573		estimate time-varying reproduction numbers during epidemics. Am J Epidemiol.
574 575	10	2013,176, 1505–1512, 001,10,1093/aje/kwi153
576	10.	convertion interval for coronavirus disease (COVID 10) based on symptom opset data
570		March 2020, Euro Surveill, 2020:25, doi:10.2807/1560.7017 ES 2020.25.17.2000257
578	11	Hotheste HW/ The Mathematics of Infectious Diseases, SIAM Pay, 2000;42: 500, 652
570		doi:10 1137/\$0036144500371007
580	12	Parag KV/ Sub-spreading events limit the reliable elimination of beterogeneous
581	12.	enidemics I.R. Soc Interface 2021:18: 20210444 doi:10.1008/rsif 2021.0444
582	13	Anderson R May R Infectious diseases of humans: dynamics and control. Oxford
583	10.	University Press: 1991
584	14	Wallinga J. Lipsitch M. How generation intervals shape the relationship between
585		growth rates and reproductive numbers. Proc R Soc B 2007.274. 599–604
586	15.	Ali ST, Wang L, Lau EHY, Xu X-K, Du Z, Wu Y, et al. Serial interval of SARS-CoV-2
587		was shortened over time by nonpharmaceutical interventions. Science, 2020:369:
588		1106–1109. doi:10.1126/science.abc9004
589	16.	Kenah E, Lipsitch M, Robins JM. Generation interval contraction and epidemic data
590		analysis. Math Biosci. 2008;213: 71–79. doi:10.1016/j.mbs.2008.02.007
591	17.	Hart WS, Miller E, Andrews NJ, Waight P, Maini PK, Funk S, et al. Generation time of
592		the alpha and delta SARS-CoV-2 variants: an epidemiological analysis. Lancet Infect
593		Dis. 2022;22: 603-610. doi:10.1016/S1473-3099(22)00001-9
594	18.	Favero M, Scalia Tomba G, Britton T. Modelling preventive measures and their effect
595		on generation times in emerging epidemics. J R Soc Interface. 2022;19: 20220128.
596		doi:10.1098/rsif.2022.0128
597	19.	Kraemer MUG, Pybus OG, Fraser C, Cauchemez S, Rambaut A, Cowling BJ.
598		Monitoring key epidemiological parameters of SARS-CoV-2 transmission. Nat Med.
599		2021;27: 1854–1855. doi:10.1038/s41591-021-01545-w
600	20.	Pellis L, Scarabel F, Stage HB, Overton CE, Chappell LHK, Fearon E, et al.
601		Challenges in control of COVID-19: short doubling time and long delay to effect of
602		interventions. Philos Trans R Soc Lond B, Biol Sci. 2021;376: 20200264.
603		doi:10.1098/rstb.2020.0264
604	21.	The R value and growth rate - GOV.UK [Internet]. [cited 1 Jul 2021]. Available:
605		https://www.gov.uk/guidance/the-r-value-and-growth-rate
606	22.	Fraser C. Estimating individual and household reproduction numbers in an emerging
607	~~	epidemic. PLoS One. 2007;2: e758. doi:10.1371/journal.pone.0000758
608	23.	Parag KV, Donnelly CA, Zarebski AE. Quantifying the information in hoisy epidemic
609	0.4	curves. medRxiv. 2022; doi:10.1101/2022.05.16.22275147
610 611	24.	Parag KV, Donnelly CA. Fundamental limits on inferring epidemic resurgence in real
011 612		dei:10.1271/jeurnel.pebi.1010004
612	25	Darag KV/ Improved estimation of time varying reproduction numbers at low case
61 <i>1</i>	25.	incidence and between enidemic wayes. PLoS Comput Riel, 2021:17: o1000347
615		doi:10.1371/journal.pcbi.1009347
616	26	Torneri A Libin P. Scalia Tomba G. Faes C. Wood IG. Hens N. On realized serial and
617	20.	deperation intervals given control measures: The COVID-19 pandemic case PLoS
618		Comput Biol 2021:17: e1008892 doi:10.1371/iournal.pcbi.1008892
619	27	Llovd-Smith JO. Schreiber SJ. Kopp PF. Getz WM. Superspreading and the effect of
620		individual variation on disease emergence. Nature, 2005:438: 355–359

621		doi:10.1028/paturo0/1152
622	28	Bettencourt I MA Ribeiro RM Real time bayesian estimation of the enidemic notential
623	20.	of emerging infectious diseases. PLoS One 2008:3: e2185
624		doi:10.1371/journal.pone.0002185
625	20	Leptonen I. The Lambert W function in ecological and evolutionary models. Methods
626	23.	Ecol Evol 2016:7: 1110-1118 doi:10.1111/20/1-210X 12568
627	30	Costic KM McGough L Baskerville EB Abbott S Joshi K Tedijanto C et al Practical
628	50.	considerations for measuring the effective reproductive number. Rt. PL oS Comput
629		Biol 2020:16: e1008409 doi:10.1371/journal.pcbi.1008409
630	31	Parag KV/ Cowling B.I. Donnelly CA. Deciphering early-warning signals of SARS-CoV-
631	01.	2 elimination and resurgence from limited data at multiple scales . I.R. Soc Interface
632		2021.18. 2021.0569 doi:10.1098/rsif 2021.0569
633	32	Van Kerkhove MD, Bento AL Mills HL, Ferguson NM, Donnelly CA, A review of
634	02.	epidemiological parameters from Ebola outbreaks to inform early public health
635		decision-making Sci Data 2015;2: 150019. doi:10.1038/sdata 2015.19
636	33.	Ferguson N. Lavdon D. Nediati-Gilani G. Others, Impact of non-pharmaceutical
637		interventions (NPIs) to reduce COVID- 19 mortality and healthcare demand. Imperial
638		College London: 2020.
639	34.	Dushoff J. Park SW. Speed and strength of an epidemic intervention. Proc Biol Sci.
640	-	2021:288: 20201556. doi:10.1098/rspb.2020.1556
641	35.	Shmueli G. To Explain or to Predict? Stat Sci. 2010;25: 289–310. doi:10.1214/10-
642		STS330
643	36.	Wagenmakers E-J, Grünwald P, Steyvers M. Accumulative prediction error and the
644		selection of time series models. J Math Psychol. 2006;50: 149–166.
645		doi:10.1016/j.jmp.2006.01.004
646	37.	Parag KV, Donnelly CA. Using information theory to optimise epidemic models for real-
647		time prediction and estimation. PLoS Comput Biol. 2020;16: e1007990.
648		doi:10.1371/journal.pcbi.1007990
649	38.	Marcellino M, Stock JH, Watson MW. A comparison of direct and iterated multistep AR
650		methods for forecasting macroeconomic time series. J Econom. 2006;135: 499–526.
651		doi:10.1016/j.jeconom.2005.07.020
652	39.	Nishiura H, Chowell G, Heesterbeek H, Wallinga J. The ideal reporting interval for an
653		epidemic to objectively interpret the epidemiological time course. J R Soc Interface.
654		2010;7: 297–307. doi:10.1098/rsif.2009.0153
655	40.	White PJ, Ward H, Garnett GP. Is HIV out of control in the UK? An example of
656		analysing patterns of HIV spreading using incidence-to-prevalence ratios. AIDS.
657		2006;20: 1898–1901. doi:10.1097/01.aids.0000244213.23574.fa
658	41.	Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory syndrome
659		reveal similar impacts of control measures. Am J Epidemiol. 2004;160: 509–516.
660		doi:10.1093/aje/kwh255
661	42.	Yan P. Separate roles of the latent and infectious periods in shaping the relation
662		between the basic reproduction number and the intrinsic growth rate of infectious
663		disease outbreaks. J Theor Biol. 2008;251: 238–252. doi:10.1016/j.jtbi.2007.11.027
664	43.	Azmon A, Faes C, Hens N. On the estimation of the reproduction number based on
665		misreported epidemic data. Stat Med. 2014;33: 1176–1192. doi:10.1002/sim.6015
666	44.	Bracher J, Held L. Endemic-epidemic models with discrete-time serial interval
667		distributions for infectious disease prediction. Int J Forecast. 2020;
668		doi:10.1016/j.ijforecast.2020.07.002

45. 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

- 673 46. Goldstein E, Dushoff J, Ma J, Plotkin JB, Earn DJD, Lipsitch M. Reconstructing
  674 influenza incidence by deconvolution of daily mortality time series. Proc Natl Acad Sci
  675 USA. 2009;106: 21825–21829. doi:10.1073/pnas.0902958106
- 47. 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
- 679 48. Sarrka S. Bayesian Filtering and Smoothing. Cambridge, UK: Cambridge University
  680 Press; 2013.
- 681 682
- 683 Funding

684 KVP acknowledges funding from the MRC Centre for Global Infectious Disease Analysis 685 (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

687 Concordat agreement and is also part of the EDCTP2 programme supported by the European

- 688 Union. The funders had no role in study design, data collection and analysis, decision to
- 689 publish, or manuscript preparation.

# 690 Data availability statement

691 All data and code underlying the analyses and figures of this work are freely available (in R

692 and MATLAB) at: <u>https://github.com/kpzoo/Omega</u>

693

694