When is the R = 1 epidemic stability threshold meaningful?

Kris V Parag[1,*], Mauricio Santillana[2,3,4], Anne Cori[1] and Uri Obolski[5]

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

[2]Machine Intelligence Group for the betterment of Health and the Environment, Network Science Institute, Northeastern University, Boston, MA, USA

[3]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Cambridge, MA, USA

[4]Network Science Institute, Northeastern University, Boston, MA, USA

[5]School of Public Health, Department of Epidemiology and Preventive Medicine, Faculty of Medical and Health Sciences, Tel Aviv University, Tel Aviv, Israel.

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

Abstract

The *effective reproduction number R* is a predominant statistic for tracking the transmissibility of infectious diseases and informing public health policies. An estimated $R=1$ is universally interpreted as indicating epidemic stability and is a critical threshold for deciding whether infections will grow ($R>1$) or fall ($R<1$). We demonstrate that this threshold, which is typically computed over coarse spatial scales, seldom signifies stability because those scales frequently average stochastic infections from groups with heterogeneous transmission characteristics. Groups with falling and rising infections counteract and early-warning signals from resurging groups are obscured by noisy fluctuations from stable groups with larger infections. We prove that an estimated $R=1$ is consistent with a vast space of epidemiologically diverse scenarios, often leading to false-positive stability signals that diminish its predictive and policymaking value. In contrast, we show that a popular, alternative definition of transmissibility, relating to the next-generation matrix of the groups, overcorrects for this issue and yields false-negative stability signals by maximising sensitivity to stochasticity. We find a recently developed statistic, E, derived from R using experimental design theory, rigorously constrains the space of scenarios corresponding to stability, while limiting noise sensitivity. We establish that $E=1$ is a more practical and meaningful stability threshold.

Keywords: infectious diseases; epidemic thresholds; reproduction numbers; public health indicators, next generation matrices; bias-variance trade-offs.

Introduction

Accurately tracking the transmissibility of infectious diseases is a longstanding and important problem. Timely indicators of the growth or decline of epidemics contribute valuable evidence for informing public health policy, assessing interventions, improving epidemic response and

supporting pandemic preparedness [1,2]. The *effective reproduction number*, *R*, is the most popular statistic for describing transmissibility. *R* estimates the ratio of expected new infections to actively circulating (past) infections [3,4]. Although there are many different approaches for computing *R* [1,5], its role as a key threshold statistic is a fixture across epidemiology. An estimated $R=1$ indicates that the incidence of new infections will remain roughly constant or stable, while values above or below this threshold foretell of rising or falling infection incidence.

This interpretation is ubiquitous and its simplicity underlies why *R* is prominent as a predictive statistic for guiding public health responses and communicating the state of an epidemic [6]. However, as *R* is typically estimated from incidence data at coarse spatial scales (frequently nationally or regionally) [1,7], it regularly averages over groups with heterogeneous dynamics. Consequently, finer scale variations that may contain critical signals for prediction and control are neglected. Group heterogeneities emerge from variations in behaviour, sociodemographic factors, immunity levels, location and other characteristic features [8,9]. Such heterogeneity-driven transmission patterns were frequently observed during the recent COVID-19 pandemic and across recurring Ebola virus, SARS, influenza and other outbreaks [9–14].

Many attempts have been made to refine effective reproduction numbers, both in formulation and estimation. These approaches generally incorporate known heterogeneities to limit biases in estimates or improve the construction of epidemic stability thresholds [8,15–18]. While these developments are valuable, they commonly require auxiliary data such as contact matrices or mobility indices or rely on complex models that reduce their immediate utility during real-time outbreak tracking. Effectively balancing practical surveillance and computational constraints against the need to capture population heterogeneity remains an open challenge to producing meaningful estimates of transmissibility in real time.

Here we provide new insights into this challenge by questioning the universal role of the $R=1$ threshold and exposing the limitations of this indicator of stability in heterogeneous settings. Using theory, simulations and empirical datasets, we highlight how averaging over groups, as employed by the most widely used formulations of *R,* mean that $R=1$ seldom signifies epidemic stability even under very mild assumptions. We further show that improving the formulation of the $R=1$ threshold to directly incorporate heterogeneity, in a manner analogous to popular next generation matrix methods [19], causes extreme sensitivity to stochasticity within the groups because it essentially replaces averaging with a maximising operation.

Consequently, *R* may have substantial disadvantages as both a prospective or retrospective statistic [20], especially at the large scales over which it is most often inferred and reported. We offer a simple solution to these issues by embedding both the averaging and maximising

operations inherent to different formulations of reproduction numbers in a common framework. We show that a recently introduced statistic, known as the risk averse reproduction number *E* [12], which was derived from the principles of statistical experimental design, sits within this framework and sensibly interpolates between these operations. As a result, we find that *E=1* is a more reliable and meaningful threshold for signifying epidemic stability in real time.

Results

Reproduction numbers may provide false positive indicators of stability

To interrogate the above claims, we consider $p>1$ heterogeneous groups with homogeneous mixing assumed within each group. We use R_j to denote the reproduction number of group *j*. This is the average number of secondary infections that a primary infection generates in that group. Using a renewal process framework (see Methods), we can define the infectious force Λ_j as the convolution of the past infections in group *j* with the generation time distribution of that group. The generation time distribution captures the random timings between primary and secondary infections [21]. We do not model interconnections among groups (e.g., travel or migrations) as we represent how infections are tracked during unfolding outbreaks across communities, where the auxiliary time-dependent data required to estimate these connections is likely unavailable, biased, or at best only snapshots in time are accessible [22].

The *R* that is frequently computed at regional or national scales, based on the aggregate infections occurring in all groups [3,23,24], is equivalent to the weighted mean in **Eq (1)** (see Methods for the derivation of this equivalence from conventional formulae). The weights represent the proportion to the relative infectious force of each location or group, which seems an intuitive way to compose the overall transmissibility.

$$R \stackrel{\text{def}}{=} \sum_{j=1}^{p} w_j R_j\,, \quad \text{with weights } w_j = \frac{\Lambda_j}{\sum_{i=1}^{p} \Lambda_i}. \quad (1)$$

However, by manipulating **Eq (1)**, we find that the *R=1* threshold condition is satisfied by any of the infinite number of solutions to $\sum_{j=1}^{p} \Lambda_j \left(R_j - 1\right) = 0$. Consequently, a vast and diverse space of epidemiologically important scenarios can all result in the same *R=1*.

This simple observation has important ramifications. Even if we additionally constrain, as in **Fig 1**, the mean of the group reproduction numbers to be 1 so that $p^{-1} \sum_{j=1}^{p} R_j = 1$, the *R=1* threshold still fails to distinguish among many divergent epidemic scenarios. At the simplest *p=2* setting, **Fig 1A** and **Fig 1B** demonstrate that these scenarios include numerous solutions with exponentially growing infection counts i.e., $R_1>1$. This discordance between the common

interpretation of *R=1* and the underlying group dynamics becomes harder to diagnose as the number of groups *p* becomes larger. **Fig 1C** confirms this discordance when we generalise **Eq (1)** (see Methods) to include the stochasticity of the infection counts observed in groups, in accordance with the most widely used estimators of *R* [3]. There we use X to generically refer to any estimator of transmissibility (some of which are derived in subsequent sections), with blue for $X=R$ and green for every $X=R_j$. We examine the expectation $E[R_j]$ across samples from the group reproduction numbers (histograms reflect uncertainty from the stochasticity) and observe numerous settings in which multiple $E[R_j]>1$ (and often no single $E[R_j]=1$). Yet *R* still confidently signals that the overall epidemic is stable ($E[R]=1$ and variance is small).

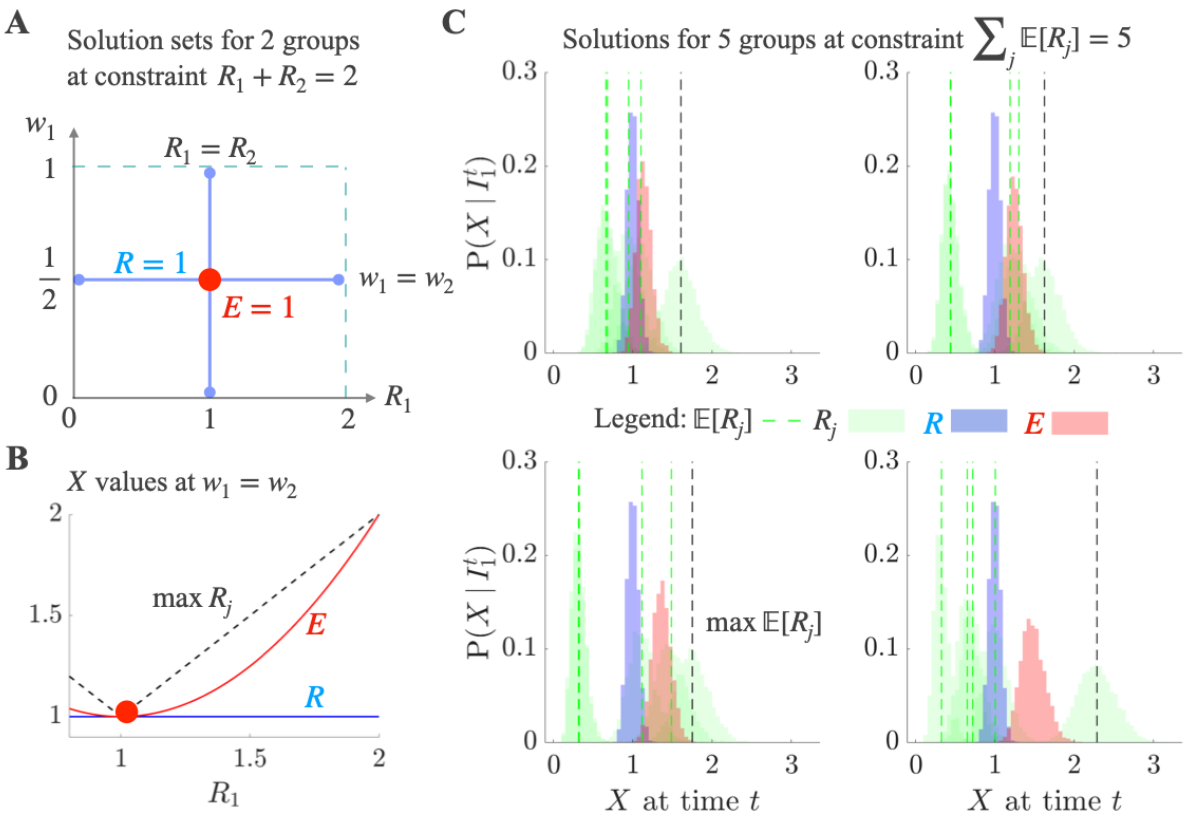

Fig 1: **Space of $R = 1$ and $E = 1$ solutions for varying group reproduction numbers R_j.** **A**: for $p = 2$ groups, we consider R_1 (x-axis) and its weight w_1 (y-axis). Weights always sum to 1 so $w_2 = 1 - w_1$. There are an infinite number of solutions yielding $R = 1$. We sketch the subset of those solutions (blue lines with dots at the end) satisfying the constraint $\sum_j R_j = p$, which means the R_j have an arithmetic mean of 1. One solution line sets $R_1 = R_2 = 1$ for all weights and the other sets $w_1 = w_2 = \frac{1}{2}$ for all values of $0 \leq R_1 \leq 2$. This line includes many scenarios in which $R = 1$ hides a resurging group ($R_j > 1$). In contrast, $E = 1$ has a unique solution (red dot) guaranteeing $R_1 = R_2 = 1$ (also $\max R_j = 1$). **B**: we plot global statistics (y-

axis) that we term X (see later **Eq (3)**) for the subsection of the $w_1 = w_2$ solution line from **A** over which R_1 is resurgent (x-axis). R (blue) is unresponsive to the growing R_1, E (red) indicates resurgence and $\max R_j$ (black dashed) is the most sensitive. **C**: for $p = 5$ groups we sample R_j values (green histograms) from gamma distributions (see Methods and [3,25] for simulation details). These use past incidence from times 1 to t, denoted as I_1^t. The mean of the samples is $\mathrm{E}[R_j]$ (green dashed with maximum $\mathrm{E}[R_j]$ in black) and we constrain the sum over groups to be 5 (the arithmetic mean of the $\mathrm{E}[R_j]$ is 1). The subplots, from top left to bottom right, show scenarios with increasing R_j heterogeneity among groups. We construct histograms for global statistics X from the R_j samples (reproduction numbers or X are on the x-axis). Across the subplots, we find that an overall $R = 1$ (blue) conceals several resurging groups. E (red) exposes these dynamics without being as sensitive as max $\mathrm{E}[R_j]$.

False positive signals of stability are misleading for two main reasons. First, they conceal the dynamics of groups experiencing resurging infections, which only become apparent when the infections in those groups grow substantially [25]. As the effectiveness of interventions is known to be tightly coupled to the earliness at which they are implemented [26], this delay can be detrimental for overall epidemic response. Second, these false indicators of stability, may inform premature release of existing controls or, when communicated to the public, engender risky behavioural responses. Both of these factors can accelerate the undetected growth and hence resulting burden of the disease. These detrimental effects only worsen under practical surveillance schemes, where incomplete reporting further obscures and delays the detection of resurgent dynamics [27].

Next generation matrices may provide false negative indicators of stability

A counterpoint to the standard *R* may be that, despite its frequent use in practice, it assumes homogeneity and is therefore not suited for assessing the stability of inherently heterogeneous systems. Next generation matrix and related methods [19,28] were proposed to rigorously compute epidemic thresholds for heterogeneous epidemic models. While these methods are mostly used to evaluate the basic reproduction number, R_0, of heterogeneous compartmental models, e.g., susceptible-infected-recovered (SIR) models with multiple infectious classes, we can still explore their relevance for our (renewal) modelling framework.

It is known that the eigenvalues of the next generation matrix (and associated approaches [8]) relate to the poles of the heterogeneous model [28]. Poles are central quantities for formally assessing the stability of linear dynamical systems [29] and are solutions to the denominator of what is termed the epidemic transfer function [30]. Transfer functions are standard tools for

describing linear systems that map how inputs (in our case imported infections) lead to outputs (the epidemic) and are computed in the frequency or Laplace domain as the ratio of outputs to inputs. We sketch the transfer function construction below (see Methods for a complete derivation, [29] for general transfer function theory and [30] for epidemic transfer functions).

The Laplace transform of the infections in group *j* is $I_j(s) = \left(1 - R_{0j}G_j(s)\right)^{-1} M_j(s)$ where s is a complex variable. $G_j(s)$ and $M_j(s)$ are the transforms for the generation time distribution and the imported infections for group *j,* respectively. The transfer function or ratio of output to input is $I_j(s)M_j(s)^{-1}$ with poles as the values of s solving $R_{0j}G_j(s) = 1$. If we assume a seed infection starts the epidemic in all groups, then $M_j(s) = 1$. We neglect susceptible depletion effects so that R_{0j} represents the basic reproduction number of group *j*. The stability of the epidemic is controlled by the dominant pole $s = r_j$, which is also the growth rate of group *j*. The other poles shape transient dynamics such as infection peaks and oscillations. At $r_j = 0$, critical stability occurs (infections plateau to a constant value) and $R_{0j} = 1$ as $G_j(0) = 1$ (this follows from the property of distributions summing to 1) [30].

Applying this construction, we obtain the Laplace transform of the total infections across the *p* heterogeneous groups as the sum of $I_j(s)$ i.e., $I(s) = \sum_{j=1}^{p} I_j(s)$. This gives **Eq (2)**, which sums the transfer functions from all groups and reveals that the overall system poles are the union of all the poles from every group. We let r be the overall dominant pole of **Eq (2)**.

$$I(s) = \sum_{j=1}^{p} \frac{1}{1 - R_{0j}G_j(s)} \Rightarrow r = 0 \Leftrightarrow \max_j R_j = 1. \quad (2)$$

If $r = 0$, the overall system is critically stable and the epidemic threshold is achieved. This is only possible when every $r_j \leq 0$ with at least one of the $r_j = 0$. If any $r_j > 0$ the overall system inherits this unstable group pole. We can guarantee that **Eq (2)** does not become unstable by requiring $\max r_j = 0$, meaning that no $R_{0j} > 1$. As these reproduction numbers can be treated as piecewise-constant approximations of R_j [8], we get the condition in **Eq (2)** that max $R_j = 1$ is the correct threshold indicator for our deterministic, heterogeneous system. Note that next generation matrix and related approaches also result in maximum eigenvalue conditions and those eigenvalues precisely map to the poles of the epidemic.

Returning to the simplest *p=2* setting of **Fig 1A**, this maximum statistic shrinks the space of threshold solutions to one symmetrical point. This guarantees that there are no false positive

signals of stability as max $R_j = 1$ requires every $R_j = 1$. However, this statistic displays maximally sensitivity to the group dynamics, as we see in **Fig 1B**. Since real epidemics feature stochastic trajectories, we find in **Fig 1C** that using a max E[R_j] epidemic threshold leads to frequent false negative indications of stability (because this measure is too sensitive to noise or fluctuations). These overly conservative assessments can inform or support unnecessary deployment or maintenance of interventions and restrictions, leading to elevated economic and societal burden without any genuine counterbalancing benefit.

Balancing sensitivity and averaging in reproduction number statistics

Although R may average over group fluctuations to the degree that informative signals are lost, the max R_j threshold (from next generation matrix approaches) may instead amplify group fluctuations such that noise becomes misconstrued as signal. How can we moderate this issue? We want to sensibly leverage the information from the R_j to derive a global statistic X such that $X=1$ balances between averaging and upweighting local or group-level dynamics. Here we propose a solution by recognising that both R and max R_j are extremes within a continuum of weighted means. We construct these means as in **Eq (3)** with $\gamma \geq 0$ as a parameter that traverses this continuum.

$$X(\gamma) \stackrel{\text{def}}{=} \sum_{j=1}^{p} \omega_j(\gamma) R_j\,, \quad \text{with weights } \omega_j(\gamma) = \frac{R_j^{\gamma}}{\sum_{i=1}^{p} R_i^{\gamma}}. \quad (3)$$

Eq (3) defines a sequence of power means that provide baseline detection at small powers i.e., recover the smoother (baseline) trend from among the R_j . At $\gamma = 0$, we see that $X(0) = p^{-1} \sum_{j=1}^{p} R_j$ is the arithmetic mean, which matches R in **Eq (1)** when the groups have similar infectious forces. As γ rises, **Eq (3)** instead offers peak detection i.e., emphasises fluctuations. We observe this other extreme as $X(\infty) = \max R_j$. If we can optimise γ, we should be able to balance between the false positive stability signals that stem from over-averaging and the false negative ones from amplifying noise, to obtain a more reliable epidemic threshold. Note that if every group has identical transmissibility $R_j = \bar{R}$, then all the $X(\gamma) = \bar{R} = R = \max R_j$.

In the previous work from some of the authors, experimental design theory was used [12] to minimise the maximum uncertainty across all group level reproduction number estimates. This yielded the *risk-averse reproduction number E,* which solved what is known as an E-optimal design to improve resurgence detection [31]. Interestingly, this statistic also sits within the framework developed here, as we find that $E = X(1)$ as shown in **Eq (4)**.

$$E \stackrel{\text{def}}{=} \sum_{j=1}^{p} \omega_j(1) R_j\,, \quad \text{with weights } \omega_j(1) = \frac{R_j}{\sum_{i=1}^{p} R_i}. \quad (4)$$

Eq (4) weights how groups contribute to the overall measure of heterogeneous transmission using their relative transmissibility. This has similarities to *R*, replacing weighting by relative infectious potential in **Eq (1)** with relative transmissibility. **Eq (4)** is also a softer approximation to max R_j and has the ingredients to achieve the balance we seek.

We therefore test if the new stability threshold signified by *E=1* is more meaningful than *R=1* and max R_j=1. First, we observe that *E=1* requires $\sum_{j=1}^{p} R_j\left(R_j - 1\right) = 0$, which collapses the space of solutions in **Fig 1A** to the single point R_1=R_2=*1*. This matches the max R_j threshold but is less sensitive to changes in group reproduction numbers, as we see in **Fig 1B**. This is in contrast to the *R* threshold, which remains at 1 throughout a wide range of scenarios with R_1>*1*. In contrast to the max R_j threshold, *E=1* has important advantages when we consider group-level stochasticity. In **Fig 1C** we validate these points at larger *p*, demonstrating that unlike $X(0)$ (constrained to 1) and *R* (approximately 1), *E* does not falsely signal stability. It also does not amplify noise like $X(\infty) = \max R_j$, which is always significantly above 1.

Fig 2 verifies that *E* sensibly moderates how group variations inform overall stability. We test a range of weight vectors (runs) that result in an *R* of approximately 1. These runs determine the fraction of total infections contributed by each group and set overall heterogeneity among the R_j. We examine how distant the group R_j are from 1 using their combined mean squared error (MSE). This MSE is evaluated across random R_j samples at every run. MSEs are ordered by how different the R_j are from *R*. We use those same runs and samples to construct *E* and get its MSE about 1. We observe that *E* converges to 1 i.e., its MSE goes to 0, only when the overall MSE of the R_j tends to 0 and hence when all R_j converge towards 1. Consequently, *E* provides a global measure of stability that reflects the patterns of stability among the local groups while considering their heterogeneity. We also compute (but do not show) the MSE of max R_j, which is much larger and noisier than that of *E* but does reflect the R_j MSE trend.

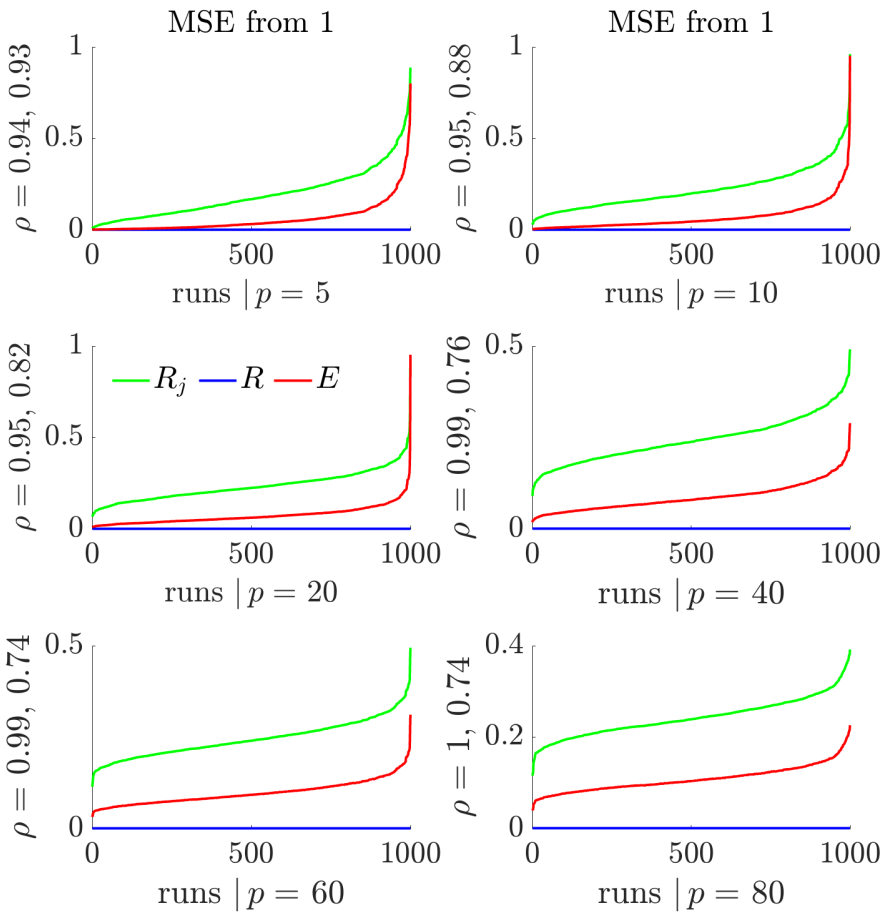

Fig 2: Mean squared error (MSE) curves about 1 for R, E and the groups R_j. In each panel, we simulate R_j values that are consistent with $R \approx 1$ for differing numbers of groups p. We sample 1000 weightings, termed runs, from Dirichlet distributions. The runs set various levels of group heterogeneity and for each run we generate 5000 samples of every R_j using the framework described in the Methods and applied in **Fig 1C**. The combined MSE of the R_j across every run relative to 1 i.e. the mean of $\sum_{j=1}^{p}\left(R_j - 1\right)^2$ over the R_j samples (green), is computed and ordered by how different the combined R_j are from R. The MSE tends to 0 as all the R_j get closer to R. We also calculate the means of $(R-1)^2$ (blue) and $(E-1)^2$ (red) under the above ordering to obtain global statistic MSEs about 1. We find that E reflects the overall behaviour of the R_j well, following the pattern of its MSE variation about 1, whereas R ignores these local variations. We do not show $\left(\max R_j - 1\right)^2$ but note that while its curves do follow the pattern of the local combined R_j, they are substantially above those for E (larger

MSE) and are very noisy. We also present correlation coefficients ρ between the combined MSE of R_j and the MSEs of E (first value) and max R_j (second value).

Important signals that are occluded by common stability thresholds

Having proposed *E=1* as a more representative and balanced stability threshold, we simulate and enumerate important instances where these distinctions matter and can be occluded by routine use of available data. We simulate epidemics in two regions or groups (*p=2*) with reported Ebola virus disease generation time distributions and time-varying transmissibility in **Fig 3**. We construct these simulations to highlight some important and different scenarios that lead to an inferred *R=1* in real time. In scenarios A-C both the estimated one-step-ahead incidence, which is the best estimate of infections at the next time step using all information before that time and accounting for stochasticity [32], as well as the time-varying reproduction number *R* inferred from total incidence, suggest overall stability. Moreover, *R* is strongly confident in its assessment of stability (narrow credible intervals).

However, the individual group one-step-ahead estimates present diverse trends that are also revealed by the time-varying R_j estimates. The risk averse reproduction number *E* converges to the standard *R* only when both groups are consistent and stable themselves. This happens in scenario A. In scenarios B-C there is discordance, indicating that *R* presents false indicators of stability. Note that in all the scenarios *R* features small credible intervals, whereas those of *E* widen for B and C, where there is disagreement on whether the epidemic is stable. In **Fig 4** we investigate these scenarios and their real-time ramifications in detail.

We investigate real-time epidemic stability, as it would be practically calculated as new data emerge, using the completely informed posterior estimate at our scenario timepoints τ and the expected probability of growth. The posterior distribution given all data up to the time τ is $\mathrm{P}(X|I_1^\tau)$, where I_1^τ is the vector of all incidence data from time 1 to τ. All estimators process this data differently. The posterior estimate of R_j uses the portion of $I(t) = I_1(t) + I_2(t)$ that informs on group *j*, *R* uses $I(t)$ and *E* uses both R_j estimates. We plot data and posteriors respectively in the left and central panels of **Fig 4**. We use the $\mathrm{P}(X|I_1^\tau)$ to infer the probability that the total incidence $I(t)$ is growing $\mathrm{P}(I(\tau + h) > I(\tau))$ for a horizon h. These are the right panels of **Fig 4** and are computed for both *E* and *R* from the total incidence and for the groups by summing their local projections (the ground truth).

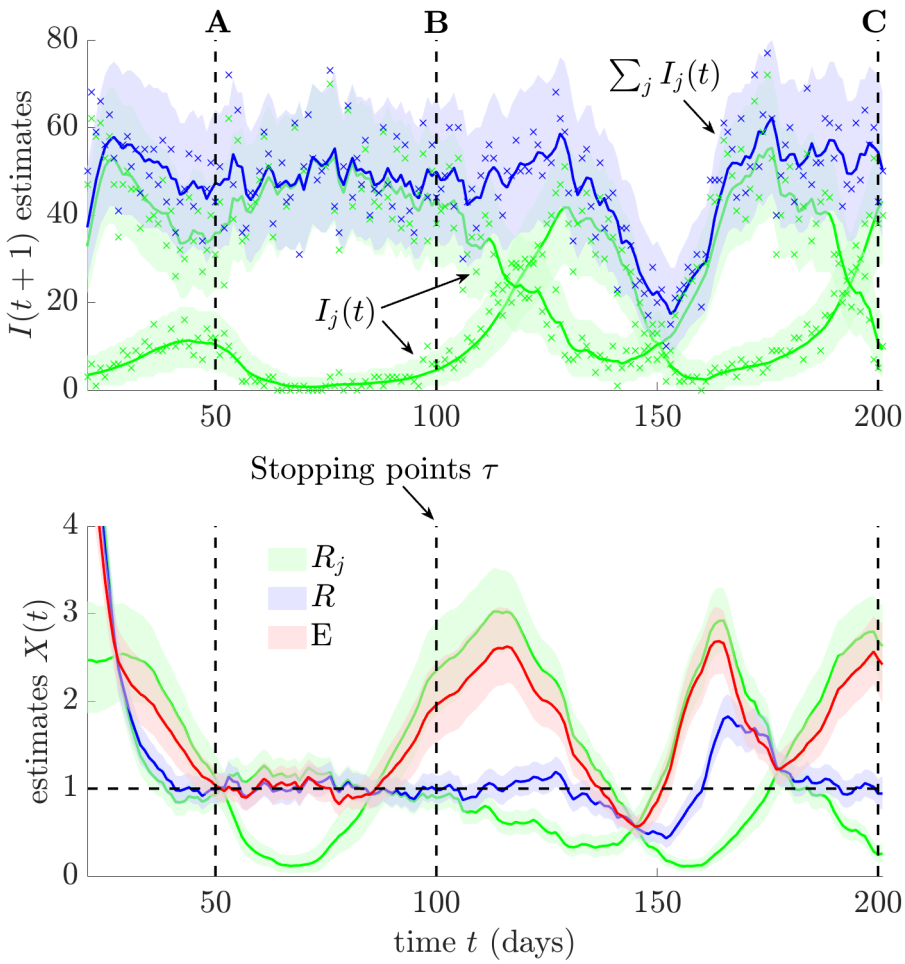

Fig 3: Estimates of time-varying reproduction numbers with diverse $R = 1$ scenarios. We use renewal models (see Methods) to simulate epidemics in two regions ($p = 2$) with Ebola virus dynamics under generation time distributions from [33]. We plot (top panel) the time-series of the inferred one-step-ahead incidence for each region (green) and in total (blue), with actual incidence data shown as crosses. A flat one-step-ahead estimate means that the epidemic appears (critically) stable at that time. We identify three regions of apparent overall stability (black, dashed). We present time-varying reproduction number estimates (bottom panel) for the groups R_j (green), the overall epidemic (based on the total incidence, blue) and estimates of the risk averse E (red). In the flat total incidence regions of A-C, we have $R \approx 1$. We investigate each of the three regions of stability as the scenarios in **Fig 4**. We derive all estimates from the EpiFilter package [24] with ribbons representing 95% credible intervals.

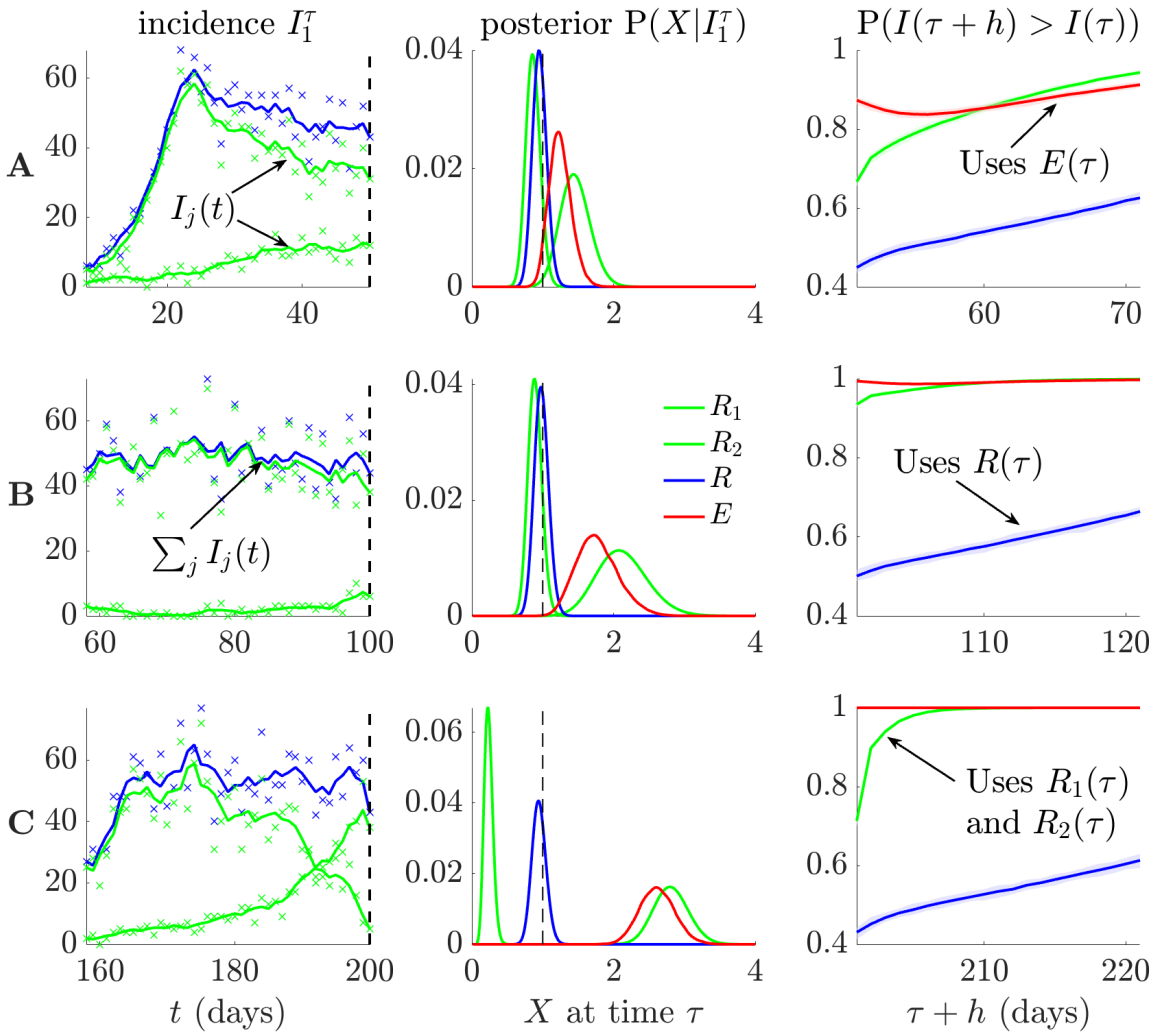

Fig 4: **Growth projections from diverse epidemic scenarios with estimated $R \approx 1$.** We examine the scenarios **A-C** from **Fig 3**. For each scenario, we plot incidence values on the left panels (crosses represent data points and solid lines plot smoothed means) for each group (green) and their sum (blue). Total incidence is approximately stable for all scenarios, but the groups showcase diverse infection patterns. We compute posterior estimates given all past incidence data up to the endpoints τ, denoted I_1^τ (central panels) for the group reproduction numbers R_j (green) and resulting global statistics X of the overall (blue) R and risk-averse (red) E reproduction numbers. In all scenarios the posterior of R is close to 1 (dashed black). This occludes important resurgences for groups in **B** and **C**. In the right panels, we explore ramifications of these posteriors by computing probabilities that the total incidence projected over a horizon h, $I(\tau + h)$ is larger than past incidence $I(\tau)$ (averaged over the last half-week from τ to reduce noise). Group-based projections (green), obtained by summing the $I_j(\tau + h)$ projected for every group from its R_j posterior, define the expected ground truth dynamics.

These dynamics are more reliably signalled at coarse scales by projections from E (red) than from R (blue) (these both use total incidence (blue) from the left panels). Estimates are from the EpiFilter package [24] and ribbons show 95% credible intervals.

We find that scenario A is (roughly) genuinely stable. Both groups have their posterior mass near 1 and projections grow relatively slowly (we set a growth probability of 0.5 as a baseline for stability). In contrast, B-C present two key settings in which the *R* threshold is misleading. B and C have *R* around 1 but also one group with large $R_j > 1$ and a growth probability that is sizably higher than 0.5 and substantially above what *R* indicates. The mechanisms underlying B-C are different. In B, the group with larger incidence is stable and this, together with infection stochasticity, masks the growth from the other group. Scenario C instead has groups with contrasting dynamics. The larger incidence group is declining and the smaller one is resurging such that their sum is stable. Notably and unlike *R*, $E=1$ accurately reflects stability amid the group uncertainties and heterogeneities in B-C but agrees with *R* when all $R_j=1$ as in A.

Empirical data illustrate false stability thresholds

We have provided mathematical and simulation-based evidence that the popular thresholds of R=1 and max R_j =1 may often provide misleading assessments of epidemic stability. Here we demonstrate that these problems are apparent in real-world, empirical datasets and that $E=1$ practically serves as an improved threshold for signifying overall or global stability. We consider a case study of the COVID-19 pandemic across 7 provinces of the Veneto region of Italy (data taken from [17]) and present our findings in **Fig 5**. There we identify and analyse three key periods in which the estimated R=1, around Aug 30th, Sep 15th and Nov 30th, 2020. We compute global statistics X and the probability of X>1 in alignment with previous analyses in this study and assess the resulting real-time indicators of stability.

Interestingly, we find that these three periods represent real-world examples of the scenarios discussed in **Fig 4**. During Aug 30th we infer approximate $R=1$ but a clear $E>1$. From the group level incidence, we see that the province with the largest infections is declining but the majority of the provinces with small incidence are growing. This generalises scenario C of **Fig 4**, where the dominant group obscures resurgent signals. *R* and *E* similarly disagree around Sep 15th, but now the provinces have comparable incidence levels and declining provinces counteract growing ones, mirroring scenario B from **Fig 4**. In both periods max $R_j > 1$, overly sensitive to noise and unsuitable for guiding decisions. The Nov 30th period features genuine stability with $E=R=1$ and even the max R_j statistic overlaps 1. This reflects scenario A from of **Fig 4**. **Fig 5** confirms that $E=1$ can be a viable real-time epidemic stability threshold.

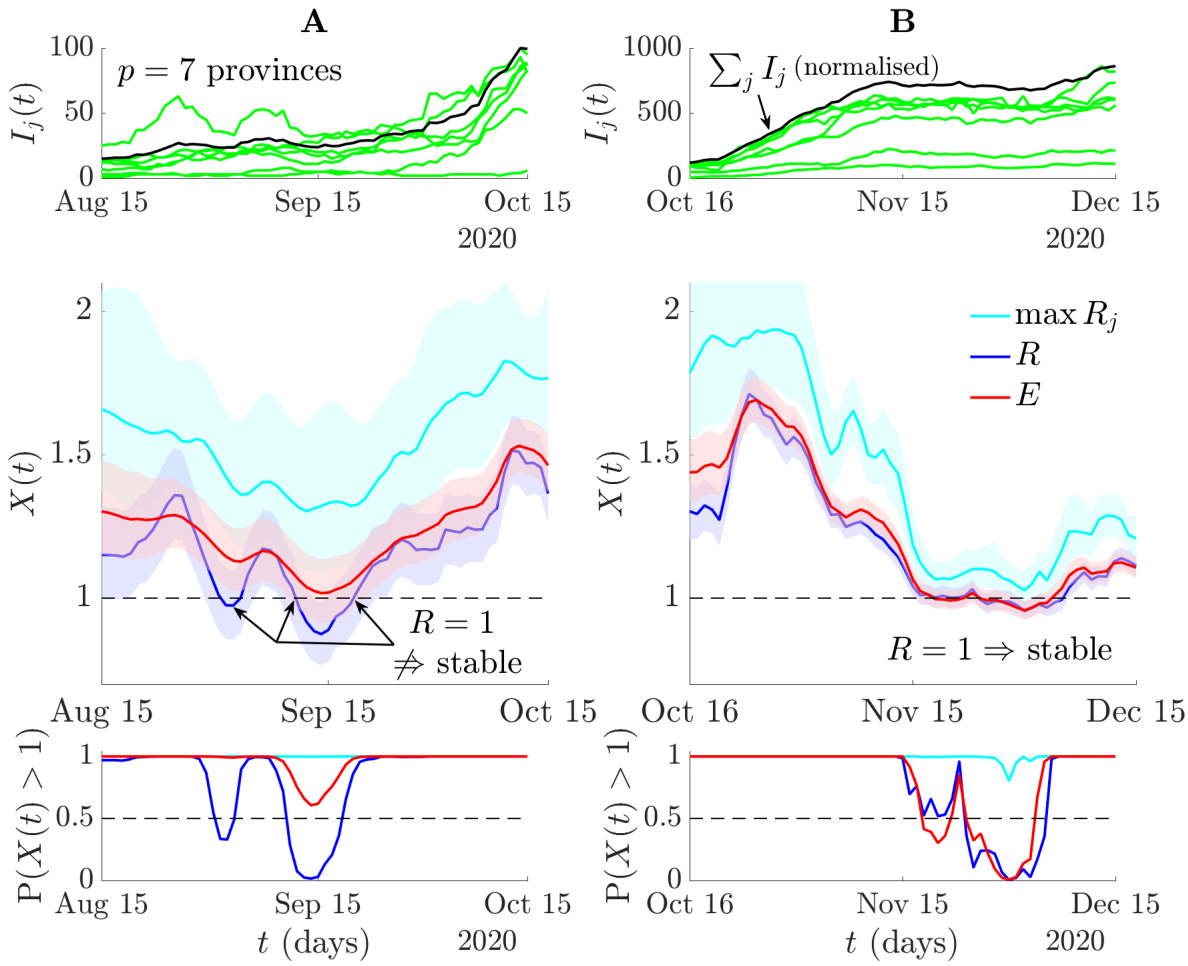

Fig 5: **False stability thresholds for COVID-19 in Italy.** We examine COVID-19 dynamics across $p = 7$ groups, which are the provinces of Veneto, Italy. Data are the incidence of new positive tests from [17], which we split into two time-regions of interest (A-B) and use renewal models with generation times from [34]. Top panels plot incidence $I_j(t)$ across the provinces (green) and the total incidence (black) normalised to the maximum $I_j(t)$. We estimate the next generation type max $R_j(t)$ (cyan), standard $R(t)$ (blue) and the risk averse $E(t)$ reproduction number (red) using EpiFilter and compute the probability that these statistics (broadly denoted $X(t)$) are above 1. A: we observe two periods with $R \approx 1$ and $E > 1$. In the first (about Aug 30th), most of the provinces show growing infections, but the one with the largest incidence declines. This mirrors scenario B from **Fig 4**. The second period (about Sep 15th) is analogous to C from **Fig 4**, with both declining and growing infections in provinces summing to a near constant incidence. The max R_j statistic is too noise sensitive and always signals epidemic growth. Across both periods, E correctly infers the stability trends, balancing the misleading estimates from R and $\max R_j$. B: a period of genuine stability. The provinces all have relatively

stable incidence (around Nov 30th) and all global statistics largely agree (though max R_j is still too sensitive). This occurs because epidemics are synchronised and is akin to A from **Fig 4**.

In the supplement, we provide an additional empirical study of COVID-19 cases in the USA state of Texas. We consider incidence curves of the counties that compose that state (we take the top 24 counties by total cases) and estimate both *R* and *E*. We again find clear periods in which the estimated $R=1$ but several counties have growing incidence (and $E>1$). We also identify periods where the estimated $E=1$ but $R<1$. Here the risk-averse reproduction number offers a better stability indicator as most counties are stable (with a few weakly growing). There is one county with a large, declining incidence, which is overly shaping the estimated *R,* forcing it confidently below 1. We do not show max R_j as it does not help distinguish stability. Our supplemental results confirm that the phenomena we identify are not anomalies but rather a general risk when aggregating cases across heterogeneous regions.

Discussion

Questions have been raised before about the accuracy and meaningfulness of reproduction numbers for tracking disease transmissibility [20,35–37]. Various works have appraised the formulation, estimation and adequacy of these statistics, especially under heterogeneous group dynamics and contact patterns [18,36,38] or when using data of varying availability and quality [27,35,39]. Despite these critiques, the interpretation that a reproduction number of 1 is the threshold of epidemic stability, whether derived from a ratio over infections (*R* or similar) [40,41] or the maximum eigenvalue of a next generation matrix approach (related to max R_j) [8,15], has remained a fixture across epidemiology. Even studies exposing how stated values of reproduction numbers can be deceptive, strongly model-dependent and even meaningless in some instances [37,40], often agree on this threshold.

Here we have revealed that common formulations of reproduction numbers, particularly in the context of real-time outbreak response, yield misleading stability thresholds. This extends the above critiques and explicitly considers when longstanding measures of epidemic stability can be untrue. We showed that even if total infections are roughly constant, *R* in heterogeneous settings may cause biased and false negative projections of growth, misrepresenting overall epidemic risk. We exposed two consequential scenarios where $R=1$ can conceal growth: if the resurgent groups (i) are masked by infections from stable groups or (ii) counteracted by groups with declining infections. We found that scenarios (i)-(ii) can easily occur in practice (appearing in empirical data) and make *R* a lagging indicator of exponential growth [9,10,17].

In contrast, we demonstrated that next generation matrix type statistics like max R_j, which consider group heterogeneities, can be highly sensitive to group stochasticity, leading to false positive stability thresholds. Although an inferred max $R_j = 1$ guarantees overall stability (and is theoretically correct), achieving this is often impractical given the varying noise in infections, group estimates and uncertainty from data availability. Improved methods that combine next generation matrix theory with standard R type averaging (as in [8]) do improve inference, but in those cases the maximum eigenvalue type approach tends to somewhat agree with R at the threshold of 1. Consequently, public health policy decisions informed by $R=1$, max $R_j = 1$ and related variants may be at risk of being mistimed and inefficient [26].

By reframing both R and max R_j as extremes within a continuum of possible epidemic stability statistics, we uncovered a suitable compromise – the risk averse reproduction number E. This was introduced in [12] to minimise the worst case uncertainty among group R_j via experimental design but emerged here as an intermediate member of this continuum. Crucially, we found E (and likely other members of the continuum) can better distinguish between genuinely stable scenarios (in which we recover $E=R=1$) and scenarios where group or location heterogeneity only appears to lead to stability. The latter is obvious when $R=1$ but $E>1$ and max $R_j > 1$.

Consequently, experimentally designed statistics that leverage both the local (group level) and global (aggregate level) signals can generate more reliable thresholds of stability for informing outbreak response, exposing patterns obscured by R but moderating dynamics amplified by max R_j. However, these benefits require identifying the heterogeneous groups (and stratifying the infections at the group level) in real time. While grouping by geographical or demographic features is possible, provided data are collected at those scales, it is difficult to disaggregate groupings by immunity levels, behaviour or other less conspicuous heterogeneities. In limited data settings or when grouping is unknown, both E and next generation matrix type methods cannot be computed. Accordingly, R then remains the best means of tracking transmissibility.

As we need group-level information to infer E, one may question the added value of global statistics for decision-making relative to the already computed R_j. If groups are homogeneous, then $R_j = 1$ correctly indicates stability. However, issues arise when assessing wider patterns that effect interventions. For example, if $p=5$, how many groups should have $R_j > 1$ before we can reliably justify a widescale action (e.g., lockdowns)? This information is also valuable for already stable groups as failing to act globally given the likely connectivity and mobility among individuals can be problematic for curbing spread. Global knowledge also helps shape local strategies such as when the stable groups or locations need to enforce travel restrictions [42].

Stability thresholds are fundamental benchmarks for implementing or releasing interventions and pinpointing epidemiologically important shifts. As these shifts frequently emerge at group levels, we recommend that future response efforts should prioritise (a) robustly detecting the heterogeneous subgroups within a population (b) estimating transmissibility within those well-mixed groups (to obtain R_j) and (c) merging those estimates into balanced global statistics like *E*. An *E=1* more reliably indicates overall epidemic stability and practically achieves what *R=1,* commonly and overconfidently, only estimates theoretically.

Methods

Renewal models and averaging inherent in R numbers

The effective reproduction number *R* of an infectious disease is commonly inferred over large scales, for example across a country or nation, by modelling the epidemic at this scale as a renewal process that ignores all heterogeneity [1]. This model assumes that the incidence *I* is the noisy product of *R* and the total infectious force of the disease *Λ*. For Poisson (Pois) noise this model is $I \sim \text{Pois}(R\Lambda)$. Here, *Λ* is the convolution of past incidence with the generation time distribution i.e., $\Lambda = \sum_{x=1}^{t-1} G(x)I(t-x)$, with $G(x)$ as the probability that an infection is transmitted in *x* time units, and defines how many infectious individuals are able to generate secondary infections in the population. The predominant method for inferring *R* [3] results in a gamma (Gam) posterior estimate distribution for *R* (ignoring prior distribution terms) of $\text{Gam}(I, \Lambda)$. The posterior mean estimate of *R* is then $I\Lambda^{-1}$ and its variance is $I\Lambda^{-2}$ [3].

A simple extension of this renewal model has also been applied to describe the epidemic at smaller scales, such as when we consider local epidemics in regions that compose our study country [1]. Assuming *p* groups or regions, this approach describes the incidence of infections independently in each group *j* as $I_j \sim \text{Pois}(R_j\Lambda_j)$. The effective reproduction number of group *j* has posterior estimate distribution $\text{Gam}(I_j, \Lambda_j)$. The mean estimate of R_j is $I_j\Lambda_j^{-1}$ [3]. This extension is commonly used to provide more localised reproduction number estimates [23,25]. However, it is uncommon to consider what these two descriptions of the epidemic imply.

Aggregating the incidence from all *p* groups yields $I = \sum_{j=1}^{p} I_j \sim \text{Pois}\left(\sum_{j=1}^{p} R_j\Lambda_j\right)$. Because this method often assumes a fixed generation time distribution among groups [23,25], it also follows that $\Lambda = \sum_{j=1}^{p} \Lambda_j$. Equating this with the earlier expression for I we see that R is by definition a weighted mean of the R_j as in **Eq. (1)**, with weights $\Lambda_j\left(\sum_{i=1}^{p} \Lambda_i\right)^{-1}$. Consequently, irrespective of the inference method applied, standard *R* estimates will possess this weighted-average relationship. The popular approach of [3] mentioned above also reveals the inherent

averaging within *R* because if $R_j \sim \text{Gam}(I_j, \Lambda_j)$ then $R = \sum_{j=1}^{p} \Lambda_j \left(\sum_{i=1}^{p} \Lambda_i\right)^{-1} R_j \sim \text{Gam}(I, \Lambda)$. This representation of *R* underlies **Eq (1)** and subsequent simulations in **Figs 1-4**.

Epidemic transfer functions and the max R_j statistic

Control theory is a field concerned with regulating the stability and performance of dynamical systems. The renewal processes defined above are examples of linear systems and their local and global stability properties are completely measured by the locations of their poles in the complex *s* plane. Poles are solutions to the denominator of the transfer function of the system [29]. We construct this following [30]. If a system has at least one pole with a positive real part then it is unstable (i.e., infections grow unboundedly). Critical stability i.e. the situation that is conventionally described by *R=1* and *r=0*, occurs only when the real parts of all poles are non-positive with at least one pole having zero real part.

Consider a group with an initial imported infection represented by the delta function $\delta(t)$ and basic reproduction number R_{0j} (or equivalently some constant approximation to the effective reproduction number). The deterministic, continuous time version of the renewal process that describes the infections in that group is $I_j(t) = \delta(t) + R_{0j} \int_0^t G_j(t-x) I_j(x)\, \mathrm{d}x$. Here $G(x)$ over all *x* composes the disease generation time distribution of the group. We take the Laplace transform of this equation. For a signal $v(t)$ in the time domain, this is $\int_0^t v(x) e^{-sx}\, \mathrm{d}x$ in the *s* domain. We rearrange to get $I_j(s) M_j(s)^{-1} = \left(1 - R_{0j} G_j(s)\right)^{-1}$ [30]. We use argument *s* to denote the signals now in the complex frequency domain and use $M_j(s)$ to show that any import time series is usable. This ratio of output infection to input is the transfer function.

As we only apply a delta function $M_j(s) = 1$ and this also defines the infections in the group directly. Poles (which are the same regardless of the chosen $M_j(s)$) solve $R_{0j} G_j(s) = 1$. The dominant pole of this system is always real, has the largest real part and equals the epidemic growth rate [28,30]. Hence if this pole or the growth rate is 0, we have critical stability. This can be shown to require that $R_{0j} = 1$ [21]. The other poles shape other characteristics of the infection trajectory (e.g., peaks and oscillations in response to interventions or other imports) and this framework completely described the renewal epidemic dynamics, assuming that the depletion of susceptibles is sufficiently slow that its effect is within R_{0j}.

The benefit of this framework is for pinpointing the conditions for critical stability when we want to consider the global dynamics of all *p* groups. We assume, as in many real-world and real-time situations, that we do not know the connectivity among groups and so can easily define

$I(s) = \sum_{j=1}^{p} I_j(s)$ as the transfer function of the entire heterogeneous system. If we did know the connectivity, this transfer function would be modified to include group interaction terms [43]. However, even under this generalisation the poles still completely describe the system dynamics and our approach is broadly applicable. This framework leads to **Eq (2)** and because we need all our poles to have non-positive real parts for critical stability, we find the maximum group growth rate must be 0 and hence $\max R_j = 1$ for critical stability.

This approach while grounded in control systems theory overlaps the next generation matrix and related approaches such as [8,19], which focus on eigenvalues instead of poles. In many instances these are exactly the same but the transfer function also pays attention to the other poles and the Laplace transforms are valuable for describing all the system dynamics (e.g., not only do we examine the other poles but can assess how they influence e.g., the size of the endemic infection load when the system is critically stable). Another benefit is that since we clarify how local poles contribute to global ones, we know conditions for stability for various architectures of groups (e.g., groups connected in series or parallel) immediately [43].

General global statistics X and risk averse E

There are a number of ways of examining the stability properties of epidemics that comprise heterogeneous groups. The above subsections considered two of the most popular: R (or E[R] when infection stochasticity is included with E denoting expectation) and max R_j (or max E[R_j]). The first is used directly for real-time analyses at broad spatial and socio-demographic scales and is the predominant time-varying measure of stability in epidemiology. The latter, often in a related next generation matrix form, is more frequently used for basic reproduction numbers and when connectivity is known but is also an important measure of heterogeneous stability. As we applied both under the same assumptions, we aimed to extract the key operation that each measure might be applying in relation to the group level R_j.

Our suite of global statistics, X, are obtained by recognising that most statistics are essentially a type of weighted average over the R_j. The central distinction is how weights are chosen, with R performing averaging based on the relative total infectiousness of groups and max R_j using a limiting weighting that is representable as a power mean with a power that tends to infinity. This insight led to the generic formula of X in **Eq (3)**. The risk averse reproduction number E was originally obtained in [12] by solving the E-optimal experimental design of minimising the maximum uncertainty among R_j estimates. This involves constructing the Fisher information matrix for the overall (global) renewal model, which has p eigenvalues of form $\alpha_j R_j^{-1}$. E results by maximising the smallest $\alpha_j R_j^{-1}$ subject to a constraint of $C = \sum_{j=1}^{p} \alpha_j$. This yields a statistic

of $\sum_{j=1}^{p} \frac{\alpha_j}{C} R_j$ with the optimised $\alpha_j = \Lambda R_j \left(\sum_{i=1}^{p} R_i\right)^{-1} = \Lambda \omega_j(1)$. Combining these gives **Eq (4)**, which emphasises resurgent group dynamics but acknowledges that resurgence at small infection counts (where less data are available) leads to larger estimate uncertainty [12].

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. Hethcote HW. The Mathematics of Infectious Diseases. SIAM Rev. 2000;42: 599–653. doi:10.1137/S0036144500371907
3. 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
4. Gostic KM, McGough L, Baskerville EB, Abbott S, Joshi K, Tedijanto C, et al. Practical considerations for measuring the effective reproductive number, Rt. PLoS Comput Biol. 2020;16: e1008409. doi:10.1371/journal.pcbi.1008409
5. Cori A, Kucharski A. Inference of epidemic dynamics in the COVID-19 era and beyond. Epidemics. 2024;48: 100784. doi:10.1016/j.epidem.2024.100784
6. 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
7. 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
8. Green W, Ferguson N, Cori A. Inferring the reproduction number using the renewal equation in heterogeneous epidemics. J R Soc Interface. 2022;19: 20210429. doi:10.1098/rsif.2021.0429
9. Donnat C, Holmes S. Modeling the heterogeneity in COVID-19's reproductive number and its impact on predictive scenarios. J Appl Stat. 2021; 1–29. doi:10.1080/02664763.2021.1941806
10. Backer JA, Wallinga J. Spatiotemporal analysis of the 2014 ebola epidemic in west africa. PLoS Comput Biol. 2016;12: e1005210. doi:10.1371/journal.pcbi.1005210
11. 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
12. Parag KV, Obolski U. Risk averse reproduction numbers improve resurgence detection. PLoS Comput Biol. 2023;19: e1011332. doi:10.1371/journal.pcbi.1011332
13. Meyers LA, Pourbohloul B, Newman MEJ, Skowronski DM, Brunham RC. Network theory and SARS: predicting outbreak diversity. J Theor Biol. 2005;232: 71–81. doi:10.1016/j.jtbi.2004.07.026
14. Cauchemez S, Bhattarai A, Marchbanks TL, Fagan RP, Ostroff S, Ferguson NM, et al. Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza. Proc Natl Acad Sci USA. 2011;108: 2825–2830. doi:10.1073/pnas.1008895108
15. van den Driessche P. Reproduction numbers of infectious disease models. Infect Dis Model.

2017;2: 288–303. doi:10.1016/j.idm.2017.06.002

16. Cushing JM, Diekmann O. The many guises of R0 (a didactic note). J Theor Biol. 2016;404: 295–302. doi:10.1016/j.jtbi.2016.06.017

17. Trevisin C, Bertuzzo E, Pasetto D, Mari L, Miccoli S, Casagrandi R, et al. Spatially explicit effective reproduction numbers from incidence and mobility data. Proc Natl Acad Sci USA. 2023;120: e2219816120. doi:10.1073/pnas.2219816120

18. Birello P, Re Fiorentin M, Wang B, Colizza V, Valdano E. Estimates of the reproduction ratio from epidemic surveillance may be biased in spatially structured populations. Nat Phys. 2024; doi:10.1038/s41567-024-02471-7

19. Diekmann O, Heesterbeek JAP, Roberts MG. The construction of next-generation matrices for compartmental epidemic models. J R Soc Interface. 2010;7: 873–885. doi:10.1098/rsif.2009.0386

20. Li J, Blakeley D, Smith RJ. The failure of R0. Comput Math Methods Med. 2011;2011: 527610. doi:10.1155/2011/527610

21. Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. Proc R Soc B. 2007;274: 599–604.

22. Swallow B, Birrell P, Blake J, Burgman M, Challenor P, Coffeng LE, et al. Challenges in estimation, uncertainty quantification and elicitation for pandemic modelling. Epidemics. 2022;38: 100547. doi:10.1016/j.epidem.2022.100547

23. Abbott S, Hellewell J, Thompson RN, Sherratt K, Gibbs HP, Bosse NI, et al. Estimating the time-varying reproduction number of SARS-CoV-2 using national and subnational case counts. Wellcome Open Res. 2020;5: 112. doi:10.12688/wellcomeopenres.16006.2

24. 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

25. 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

26. Morris DH, Rossine FW, Plotkin JB, Levin SA. Optimal, near-optimal, and robust epidemic control. Commun Phys. 2021;4: 78. doi:10.1038/s42005-021-00570-y

27. Parag KV, Lambert B, Donnelly CA, Beregi S. Asymmetric limits on timely interventions from noisy epidemic data. medRxiv. 2025; doi:10.1101/2025.01.28.25321278

28. Boldin B, Diekmann O, Metz JAJ. Population growth in discrete time: a renewal equation oriented survey. Journal of Difference Equations and Applications. 2024;30: 1062–1090. doi:10.1080/10236198.2023.2265499

29. Ogata K. Modern Control Engineering, Volume 1. 3, illustrated ed. Prentice Hall; 1997.

30. Parag KV. How to Measure the Controllability of an Infectious Disease? Physical Review X. 2024;

31. Atkinson A, Donev A. Optimal Experimental Designs. Oxford University Press; 1992.

32. 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

33. 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

34. Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. Int J Infect Dis. 2020;93: 284–286. doi:10.1016/j.ijid.2020.02.060

35. Vegvari C, Abbott S, Ball F, Brooks-Pollock E, Challen R, Collyer BS, et al. Commentary on the use of the reproduction number R during the COVID-19 pandemic. Stat Methods Med Res. 2021; 9622802211037080. doi:10.1177/09622802211037079

36. Mercer GN, Glass K, Becker NG. Effective reproduction numbers are commonly overestimated early in a disease outbreak. Stat Med. 2011;30: 984–994. doi:10.1002/sim.4174

37. De Leenheer P, Gregg Z, McCaslin J. Some limitations on the use of the basic reproduction number. Journal of Difference Equations and Applications. 2024;30: 1147–1171. doi:10.1080/10236198.2024.2308110

38. May RM. Network structure and the biology of populations. Trends Ecol Evol (Amst). 2006;21: 394–399. doi:10.1016/j.tree.2006.03.013

39. 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

40. Parag KV, Cowling BJ, Lambert BC. Angular reproduction numbers improve estimates of transmissibility when disease generation times are misspecified or time-varying. Proc Biol Sci. 2023;290: 20231664. doi:10.1098/rspb.2023.1664

41. Fraser C. Estimating individual and household reproduction numbers in an emerging epidemic. PLoS One. 2007;2: e758. doi:10.1371/journal.pone.0000758

42. Karatayev VA, Anand M, Bauch CT. Local lockdowns outperform global lockdown on the far side of the COVID-19 epidemic curve. Proc Natl Acad Sci USA. 2020;117: 24575–24580. doi:10.1073/pnas.2014385117

43. Parag KV, Hoscheit P. Why do border controls fail to stabilise epidemics? 64th IEEE CDC. 2025; In Press. Preprint: doi:10.1101/2025.05.12.25327440

Code and Data Availability

We provide MATLAB code at https://github.com/kpzoo/stabilityR to reconstruct the figures and reproduce all the analyses.

Funding

KVP and AC acknowledge funding from the MRC Centre for Global Infectious Disease Analysis (reference MR/X020258/1), funded by the UK Medical Research Council. This UK-funded grant is carried out in the frame of the Global Health EDCTP3 Joint Undertaking. UO was supported by a grant from Tel Aviv University Center for AI and Data Science 417 (TAD) in collaboration with Google, as part of the initiative of AI and DS for social good. The funders had no role in study design, data collection and analysis, decision to publish, or manuscript preparation. For the purpose of open access, the author has applied a 'Creative Commons Attribution' (CC BY) licence to any Author Accepted Manuscript version arising from this submission.