Risk averse reproduction numbers improve resurgence detection

Kris V Parag[1,*] and Uri Obolski[2,3]

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

[2]School of Public Health, Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel.

[3]Porter School of the Environment and Earth Sciences, Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel.

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

Abstract

The *effective reproduction number R* is a prominent statistic for inferring the transmissibility of infectious diseases and the effectiveness of interventions. *R* purportedly provides an easy-to-interpret threshold for deducing if an epidemic will grow ($R > 1$) or decline ($R < 1$). We posit that this interpretation can be misleading and overconfident when applied to infections aggregated from groups with heterogeneous dynamics. In these settings, *R* implicitly weights the dynamics of groups by their number of circulating infections. This induces losses in sensitivity to the dynamics of high-risk groups that may spur resurgences, promotes premature indications of epidemic control, and renders the $R = 1$ threshold uninformative. Applying *E-optimal* experimental design theory, we derive the *risk averse reproduction number E* to ameliorate these issues. Using analytic approaches, simulations, and a real-world case study, we find that *E* achieves more meaningful consensus of the dynamics across groups without being overconfident in its estimates. An $E > 1$ generates timely resurgence signals (upweighting the impact of high-risk groups), while $E < 1$ ensures constituent sub-outbreaks are under control. We propose *E* as an alternative to *R* for informing policy and assessing transmissibility over large scales (e.g., state or nationwide), where well-mixed assumptions break down.

Keywords: infectious diseases; epidemic models; reproduction numbers; experimental design; multiscale dynamics.

Introduction

The *effective reproduction number*, R, summarises the time-varying transmissibility or spread of an infectious disease by the average number of secondary infections that it generates per effective primary infection [1]. A value of R above or below 1 is interpreted as a threshold signifying that the epidemic is, respectively, growing (supercritical) or under control (subcritical) [2]. This interpretation is widely used, and estimates of R computed at

various scales, ranging from the district to country level, have led to valuable insights into the transmission dynamics of numerous pathogens including pandemic influenza, Ebola virus, malaria and SARS-CoV-2 [3–5]. During epidemics, R is commonly monitored and reported in real time to assess the effectiveness of interventions [2,6], signal the emergence of pathogenic variants [7], estimate the probability of sustained outbreaks [8], increase public awareness [9] and inform public health policymaking [10].

The benefits of R mainly stem from two properties: it is easily interpretable, as a threshold parameter, and it is easily computable in real time, requiring only routine surveillance data such as epidemic curves of cases [11,12]. However, these must be balanced against its core limitation – R is commonly derived under a well-mixed assumption in which individuals are homogeneous and have equal probabilities of encountering one another. Generalising this assumption to account for the fact that realistic contact rates are heterogenous, leading to assortative and preferential mixing [13], necessitates some loss in either interpretability or computability [14]. As an example, models that explicitly include these complexities can become computationally intractable (e.g., agent or individual-based models [15]) or require extensive contact-network data [16].

An alternative that can minimise the severity of this complexity-interpretability trade-off is to model the epidemic as a metapopulation or multitype process, in which local scales are well-mixed but diverse. Global scales then enforce structural heterogeneity [14] and compose an overall R as a weighted function of the local reproduction numbers of each group, denoted R_j for group *j* [17]. Here we consider spatial scales, though analyses often equally apply to sociodemographic or other types of heterogeneity. In our context, local scales may describe regions and the global scale a country composed of those regions. While metapopulation approaches have allowed more informative generalisations of R (e.g., using next generation matrices) [18,19], two key questions remain understudied and form the focus of this paper.

First, can we derive an alternative global statistic that better captures the salient dynamics of local scales than R? Standard formulations of R lose sensitivity to key events such as local resurgences [17], because these often initially occur in groups with low incidence, which are down-weighted by R. Second, can we optimally trade-off the uncertainty among estimates at local and global scales? In estimating the dynamics of local groups, we reduce the data informing each estimate and are hence subject to a bias-variance trade-off. For example, if we directly use estimates of R_j to assess local resurgence, false positives due to increased

uncertainty may be more likely. However, if we ignore heterogeneity at local scales, then R can be overconfident in its assessment of transmissibility.

Using experimental design theory [20], which provides a framework for optimising and fusing estimates of R_j according to cost functions of interest, we derive the *risk averse reproduction number*, E. We prove that E upweights groups with likely resurging dynamics because it is the optimal solution of a minimax cost function over the uncertainties of R_j estimates (hence it is risk averse). We demonstrate via analytic arguments, simulations and analysis of empirical COVID-19 data from Israel that E attains more meaningful consensus over local-scale dynamics than R, with better uncertainty quantification and timelier resurgence detection. Given its transparent and optimal properties, we believe E can help inform policy at large scales where R may be spurious.

Results

Pitfalls of standard reproduction numbers

We consider the transmissibility of a pathogen at two scales: a local scale, in which we may model a well-mixed population (e.g., within specific geographic or sociodemographic groups) and a global scale, which integrates the dynamics of $p \geq 1$ local groups. Our local scale, for example, may refer to spread at a district level, whereas the corresponding global scale is countrywide and covers all districts that compose that country. We assume that subdivision into local groups is based on prior knowledge and logistical constraints. We denote the time-varying reproduction number within local group j as R_j and model the dynamics in this group with a Poisson (Pois) renewal model (see Supplement) [11] as on the left side of **Eq. (1)**.

$$I_j \sim \mathbf{Pois}\left(R_j \Lambda_j\right) \text{ for } 1 \leq j \leq p, \qquad \sum_{j=1}^{p} I_j \sim \mathbf{Pois}\left(\sum_{j=1}^{p} R_j \Lambda_j\right). \quad (1)$$

This is a widely used framework for modelling time-varying transmissibility [10], with I_j as the new infections and Λ_j as the total infectiousness within group j. Λ_j measures the circulating (active) infections as a weighted sum of past infections in group j, with the weights set by the generation time distribution of that group. We allow this distribution, which captures variations in the times between primary and secondary infections, to vary across groups [1]. All variables are functions of time – e.g., I_j is explicitly $I_j(t)$ – but we often disregard time indices to simplify notation. We also neglect correlations among the different R_j. An outbreak in group j as growing or controlled if the sign of $R_j - 1$ is, respectively, positive or negative.

This renewal model approach is also commonly applied over global scales (e.g., to compute national reproduction numbers during the COVID-19 pandemic [21]) by summing infections from every constituent group. This amounts to a well-mixed assumption at this global scale. If we define $I \stackrel{\text{def}}{=} \sum_{j=1}^{p} I_j$ and $\Lambda \stackrel{\text{def}}{=} \sum_{j=1}^{p} \Lambda_j$ as the new infections and total infectiousness on this global scale, then the transmission model used is $I \sim \mathbf{Pois}(R\Lambda)$, with R as the effective reproduction number on that scale. However, if we instead develop a global model from our local models, we obtain the right side of **Eq. (1)**. This simple observation has an important ramification – by assuming that this single R summarises global scale dynamics, we make an implicit judgment about the relative importance of the dynamics in different local groups. We can expose this judgment by simply equating both global models to get **Eq. (2)**.

$$R = \sum_{j=1}^{p} w_j R_j \ , \qquad w_j = \Lambda_j \left(\sum_{i=1}^{p} \Lambda_i \right)^{-1} . \quad (2)$$

We see that group j is assigned a weight w_j, which is the ratio of active infections in group j to the total active infections at the global scale. Note that $0 \leq w_j \leq 1$ and $\sum_{j=1}^{p} w_j = 1$. This weighting has two key consequences. First, groups with outsized infection loads dominate R. This means a group with a large Λ_j and small $R_j < 1$, can mask potentially important resurgent groups, which likely possess small Λ_j and $R_j > 1$, until those groups generate an appreciable number of infections [17]. Consequently, using R may lead to lagging indicators of resurgence i.e., late warnings. The alternative – to scrutinise each local region for signs of concentrated upticks in R_j – may also be suboptimal. Higher stochasticity is expected from data at smaller scales, potentially causing false positive resurgence alarms. This is similar to the classic bias-variance trade-off commonly encountered in statistical modelling [22].

Second, R is only truly a representative summary of the dynamics at the local scales at two boundary conditions – when all the R_j are highly similar and when there is only one active group (i.e., effectively $p = 1$). The latter case is trivial, while the former is unlikely because epidemics usually traverse connected regions in waves [23] and different groups tend to possess heterogeneous contact patterns and risks of infection. These all result in diverse R_j time series and desynchronised epidemic curves [24]. Hence, we argue that this commonly estimated R [9,10] presents neither a sufficient nor representative summary statistic for communicating transmission risk or for informing policymaking. We bolster our argument by demonstrating that not only is R a misleading summary statistic across scales but it is also overconfident in the summary it provides (i.e., its estimates have underestimated variance).

We analyse the properties of R by computing maximum likelihood estimates (MLEs) and Fisher information (FI) values. The left side of **Eq. (3)** defines $\hat{R}$, the MLE of R, in terms of the MLEs of the R_j of every group. These are $\hat{R}_j = I_j \Lambda_j^{-1}$ [1] under **Eq. (1)** (see Supplement for derivations). The smallest asymptotic uncertainty around these MLEs (or any consistent estimator) is delineated by the inverse of the FI i.e., larger FI values imply smaller estimate uncertainties [25]. For the renewal models studied here, we know that $\mathbf{FI}[R_j] = \Lambda_j R_j^{-1}$ [26].

When comparing across scales, it is easier to work under the robust or variance stabilising transform $2\sqrt{R_j}$, as it yields $\mathbf{FI}[2\sqrt{R_j}] = \Lambda_j$ (see [26] and Supplement for details). We will often switch between the FI of R_j and $2\sqrt{R_j}$ as needed to clarify comparisons, but ultimately will provide key results in the standard R_j formulation. Substituting our FI expressions into **Eq. (2)**, we find the global FI sums local FI values as on the right side of **Eq. (3)**:

$$\hat{R} = \sum_{j=1}^{p} w_j \hat{R}_j = I\Lambda^{-1}, \qquad \mathbf{FI}[2\sqrt{R}] = \sum_{j=1}^{p} \mathbf{FI}[2\sqrt{R_j}] = \Lambda. \quad (3)$$

The uncertainty around $\hat{R}$ is therefore likely to be substantially smaller than that around any $\hat{R}_j$. This formulation underestimates overall uncertainty because the FI acts as a weight that is inversely proportional to the variance of the R_j estimates. Estimates of R are therefore not only insufficient but also overconfident as measures of global epidemic transmissibility.

The goal of our study is to design alternatives to R that attain better consensus across heterogeneous dynamics, with defined properties over diverse locales and without inflated estimate confidence. To achieve this, we must make a principled bias-variance trade-off among signals from R and every R_j, deciding how to best emphasise actionable dynamics from local groups without overamplifying noise. Next, we apply optimal design theory to develop new consensus reproduction numbers with these properties.

D and E optimal reproduction numbers and their properties

The consensus problem of deriving a statistic that is representative of local dynamics can be reframed as an optimal design on the weights mapping the R_j to that statistic, according to some cost function of interest. The uncertainty around estimates of R_j, encoded (inversely) by $\mathbf{FI}[R_j] = \Lambda_j R_j^{-1}$, fundamentally relates to the important dynamics of the epidemic (e.g., resurgence events likely occur at small Λ_j and large R_j, minimising $\mathbf{FI}[R_j]$ [17]). Hence, we focus our designs on the Fisher information matrix $\mathbf{FI}_{\boldsymbol{R}}$ of **Eq. (4)**, which summarises the

uncertainty from all local reproduction number estimates. There we replace Λ_j with a factor $\alpha_j > 0$ that redistributes the information across the p groups, subject to a constraint on the total information, which is equal to Λ (see **Eq. (3)**).

$$\mathbf{FI}_{\boldsymbol{R}} = \begin{bmatrix} \alpha_1 R_1^{-1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \alpha_p R_p^{-1} \end{bmatrix}, \text{ such that } \sum_{j=1}^{p} \alpha_j = \Lambda. \quad (4)$$

This particular formulation facilitates the description of several important scenarios. When $\alpha_j = \Lambda_j$, we recover the standard formulation of R (**Eq. (2)**). If we model migrations among groups (resulting in probabilities of transporting active infections, as in [27]), then α_j measures the active infections that are informative about R_j – infections that are generated in group j, including those exported to other groups. If we assume that infections observed in group j are actually a random sample from multiple groups according to a multinomial distribution, then α_j corresponds to the fraction of Λ assigned by that distribution to group j. There are also analogues among α_j and models of epidemic surveillance biases [28].

Since the α_j are design variables subject to the conservation constraint in **Eq. (4)**, we can leverage experimental design theory [20,29] to derive novel consensus statistics to replace the default formulation from **Eq. (2)**. We examine *A*, *D* and *E-optimal* designs, which have standard definitions of how the total uncertainty on our p parameters is optimised. If $p = 2$, this uncertainty can be circumscribed by an ellipse in the space spanned by R_1 and R_2, and designs have a geometric interpretation as in **Figure 1**. *A*-optimal designs minimise the bounding box of the ellipse, while *D* and *E*-optimal designs minimise its area (or volume, when extending to higher dimensions) and largest chord respectively [29,30]. These designs yield optimal versions of Λ_j, $\Lambda_j^* = \alpha_j$, computed as shown in **Eq. (5)**, where $\mathrm{tr}[.]$, $\det[.]$ and $\mathrm{eig}[.]$ indicate the trace, determinant and eigenvalue of their input matrix.

$$\Lambda_j^* | A = \max_{\{\alpha_j\}} \mathrm{tr}[\mathbf{FI}_{\boldsymbol{R}}], \qquad \Lambda_j^* | D = \max_{\{\alpha_j\}} \det[\mathbf{FI}_{\boldsymbol{R}}], \qquad \Lambda_j^* | E = \max_{\{\alpha_j\}} \min_j \mathrm{eig}[\mathbf{FI}_{\boldsymbol{R}}]. \quad (5)$$

These designs can also be performed using robust transforms by replacing $\mathbf{FI}_{\boldsymbol{R}}$ with $\mathbf{FI}_{\mathbf{2}\sqrt{\boldsymbol{R}}}$, yielding the diagonal FI matrix $\mathbf{FI}_{\mathbf{2}\sqrt{\boldsymbol{R}}} = \mathrm{diag}[\alpha_1, \dots, \alpha_p]$. Under this transform, we observe that $\mathrm{tr}\left[\mathbf{FI}_{\mathbf{2}\sqrt{\boldsymbol{R}}}\right] = \sum_{j=1}^{p} \alpha_j = \Lambda$. The default allocation of $\Lambda_j^* = \Lambda_j$ is therefore, trivially, an *A*-optimal design under this transform. We compute *D* and *E*-optimal designs without robust transforms as we want to work with R_j directly. As $\det[\mathbf{FI}_{\boldsymbol{R}}] = \left(\prod_{j=1}^{p} R_j^{-1}\right) \prod_{j=1}^{p} \alpha_j$, with the

bracketed term as a constant, we find that, subject to our constraint, $\Lambda_j^*|D = p^{-1}\Lambda$. This follows from majorization theory and solutions are adapted from [30,31] (see Supplement). Since $\mathbf{FI}_{\boldsymbol{R}}$ is diagonal, its eigenvalues are $\alpha_j R_j^{-1}$ and deriving $\Lambda_j^*|E$ equates to solving the condition $R_1^{-1}\Lambda_1^* = R_2^{-1}\Lambda_2^* = \cdots R_p^{-1}\Lambda_p^*$ (see [30,31] and Supplement for derivations). We hence obtain the *E* optimal design as $\Lambda_j^*|E = \Lambda R_j\left(\sum_{j=1}^{p} R_j\right)^{-1}$.

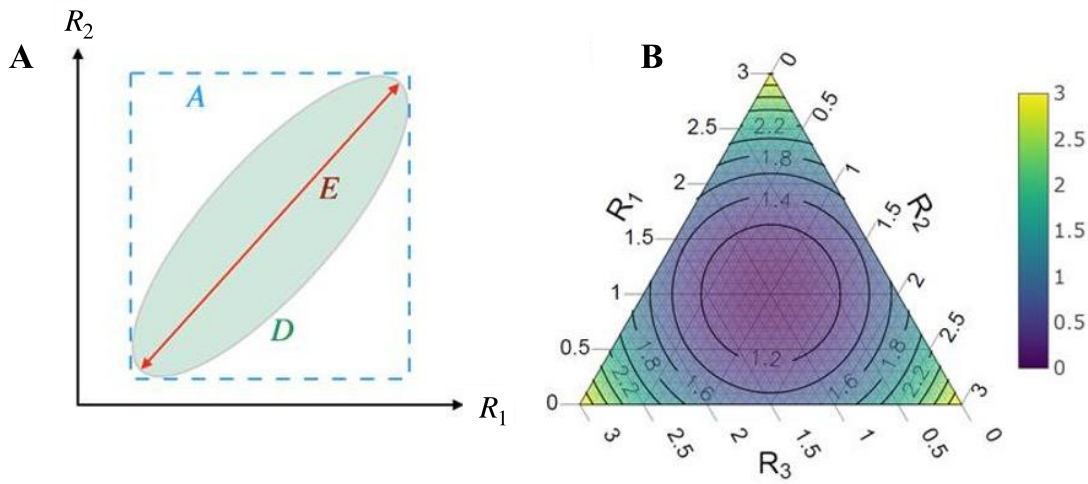

Figure 1: Illustrations of optimal experimental designs and local reproduction number combinations. (A) The geometric interpretation of *A*, *D* and *E*-optimal designs for the $p = 2$ parameter scenario. The overall uncertainty of the parameters is defined by an uncertainty ellipse in the space spanned by possible values of the local reproduction numbers. The ellipse is centred on the MLEs of the parameters and its shape is determined by the inverse of the FI around those estimates. Each design minimises a different characteristic of the ellipse. *A* minimises the bounding box, *D* minimises the ellipse area, and *E* minimises the largest chord (coloured respectively). (B) A ternary plot demonstrating the trajectories of the consensus statistics D and E as a function of different group reproduction numbers R_j. These are for $p = 3$ and constrained so $R_1 + R_2 + R_3 = 3$. The colour and contour lines represent E at each combination of R_j. We see that E is maximised at the edges, when only one R_j is non-zero. D is at the centre of the triangle, as it is the arithmetic mean of the R_j.

We formulate the new consensus reproduction numbers, D and E, by substituting the above optimised Λ_j^* into **Eq. (2)** to derive **Eq. (6)**, which forms our main result. This corresponds to using optimised weights $w_j^*|D = p^{-1}$ and $w_j^*|E = R_j\left(\sum_{j=1}^{p} R_j\right)^{-1}$ in **Eq. (2)**. In **Eq. (6)** we compute these statistics as convex sums $D = \sum_{j=1}^{p}(w_j^*|D)R_j$ and $E = \sum_{j=1}^{p}(w_j^*|E)R_j$.

$$D = \frac{1}{p}\sum_{j=1}^{p} R_j\,, \qquad E = \left(\sum_{j=1}^{p} R_j^2\right)\left(\sum_{j=1}^{p} R_j\right)^{-1}. \quad (6)$$

We refer to D as the *mean reproduction number* because it is the first moment or arithmetic mean of the effective reproduction numbers of each group i.e., it weights the dynamics of each group equally. This construction ensures that the overall uncertainty volume over the estimates of every R_j is minimised. We define E as a *risk averse reproduction number*, and derive it as the ratio of the second to first raw moment of the group reproduction numbers. This is also known as the contraharmonic mean. E weights each group reproduction number by the fraction of the total reproduction number sum attributable to that group. This weighting emphasises groups with large R_j, which are considered to be high risk. Both D and E are also informed by the active infections Λ_j of each group. These contribute to the uncertainty of their estimates, which feeds into $\widehat{D}$ and $\widehat{E}$ (see Supplement).

All three statistics possess important similarities that define them as proxies for reproduction numbers. Because they are all convex sums of the local R_j, the value of each statistic lies inside a simplex with vertices at the R_j. At the boundary conditions of one dominant group (i.e., essentially $p = 1$) or of highly similar group dynamics (i.e., the R_j are roughly the same over time) this simplex collapses and we find $R = D = E$. Moreover, if all the $R_j = 1$, then $R = D = E = 1$, signifying convergence to the reproduction number threshold. Thus, D and E are alternative strategies to R for combining local reproduction numbers with different properties that may offer benefits when making decisions at large scales. We visualise how these statistics can determine our global view of transmissibility via the simplex in **Figure 1**.

We refer to E as risk averse because it ensures that the most uncertain R_j is upweighted as compared to the standard formulation of R in **Eq. (2)**. This protects against known losses of sensitivity to resurgent dynamics [17] that occur due to averaging across groups, because the FI is expected to be smallest for resurgent groups. As opposed to simply interrogating the individual R_j estimates to identify resurgent groups, *E* weights those groups, while also accounting for the uncertainty in their estimates, to obtain a consensus on overall epidemic transmissibility risk. Consequently, E reduces false positives that may occur due to the noise in individual resurgent groups and provides a framework for interpreting situations where some groups may be concurrently resurgent while others are under control. We illustrate these points in **Figure 2** below and explore their ramifications in the next section.

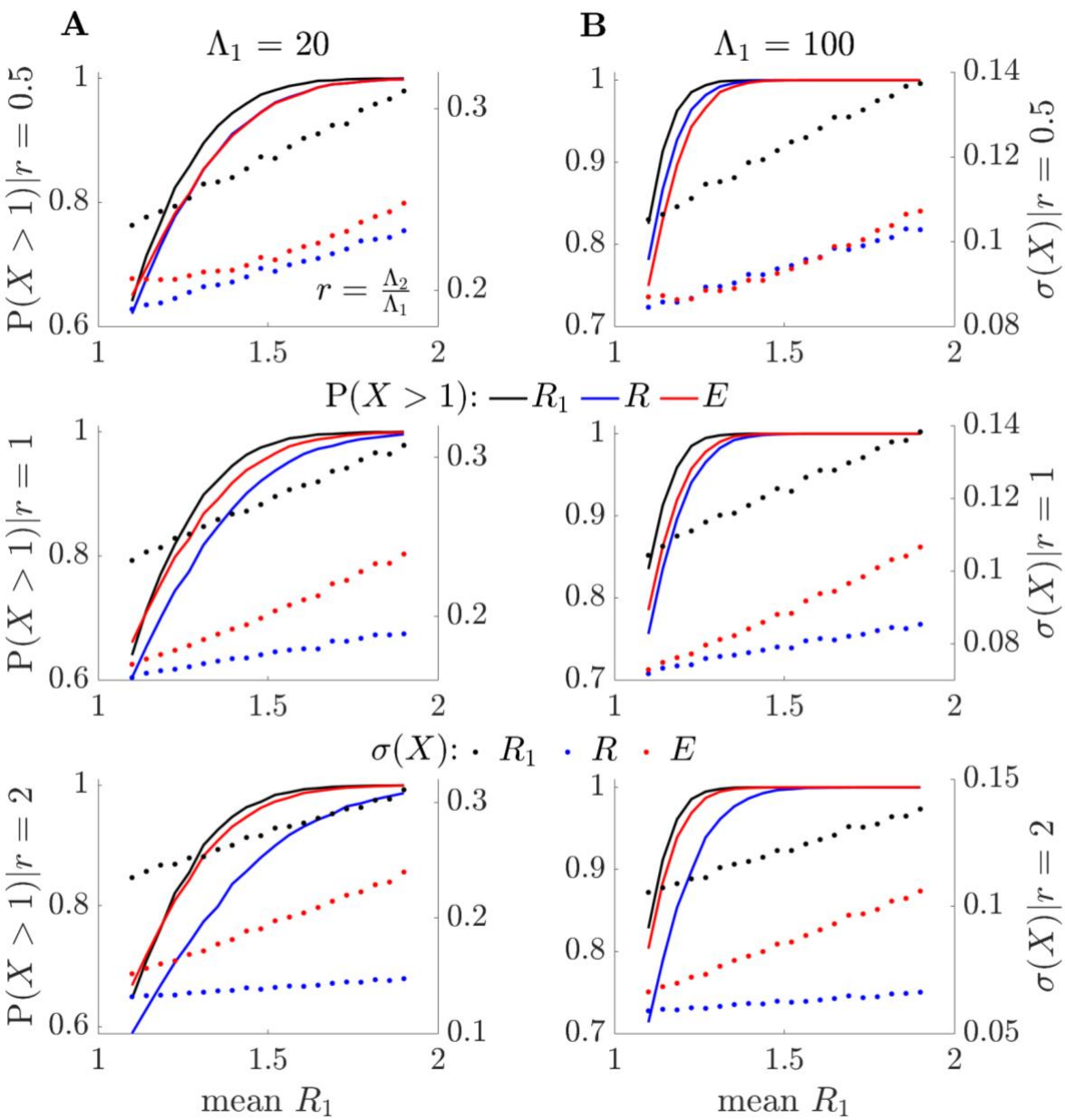

Figure 2: Relative sensitivity of *E* and *R* to resurgence dynamics. By sampling from the posterior gamma distributions of [11], we simulate $p = 2$ local groups, varying the values of R_1, while keeping R_2 at mean value of 1. We plot sensitivities to resurgence of the effective, R, and risk averse, E, reproduction numbers relative to that of the local reproduction number of group 1, R_1. These are indicated by the values of $\boldsymbol{P}(X > 1)$ for $X = R, E$ and R_1 (solid blue, red, and black, respectively). (A) and (B) show these resurgence probabilities on left y-axes over a range of mean R_1 values for scenarios with small (A) and large (B) numbers of active group 1 infections, Λ_1. We can assess sensitivity by how quickly $\boldsymbol{P}(X > 1)$ rises and describe the impact of active infections in group 2 using $r = \frac{\Lambda_2}{\Lambda_1}$. We find that E balances the sensitivity between R_1 and R. The latter loses sensitivity as the active infections in group 2

become larger relative to that of group 1 (i.e., as r increases). This occurs despite group 2 having stable infection counts. We plot the standard deviation of the reproduction number estimates on right-y axes as $\boldsymbol{\sigma}(X)$ for $X = R, E$ and R_1 (dotted blue, red, and black, respectively). We see that the local R_1 is noisiest (largest uncertainty), while R has the smallest uncertainty (overconfidence). E again, provides a useful balance.

In **Figure 2** we apply the renewal model framework from [11,17], which is based on **Eq. (1)** but models estimates of the local reproduction numbers according to the posterior gamma (Gam) distribution $R_j \sim \mathbf{Gam}(I_j, \Lambda_j^{-1})$. This yields a mean estimate of R_j that is equal to the MLE $\hat{R}_j = I_j \Lambda_j^{-1}$ with standard deviation $\boldsymbol{\sigma}(R_j) = I_j^{0.5} \Lambda_j^{-1}$. We consider two local groups $(p = 2)$ and compute the probabilities of resurgence $\mathbf{P}(X > 1)$ for $X = R_1$, R and E for scenarios that likely represent a resurgence (i.e., small Λ_1 and increasing I_1 with the second group stable at $\Lambda_2 = I_2$). In these cases, E is able to signal resurgence substantially earlier than R but is distinct from simply observing dynamics in the resurging group 1. We confirm this by noting that $\boldsymbol{\sigma}(R_1)$ is appreciably larger than $\boldsymbol{\sigma}(E)$, the standard deviation from E. In contrast R is overconfident, with usually a much smaller $\boldsymbol{\sigma}(R)$.

Risk averse reproduction numbers are more representative of key dynamics

We test our D and E reproduction numbers against R on epidemics that are simulated from renewal models with Ebola virus generation times from [32] in **Figure 3** and **Figure 4**. We consider $p = 3$ groups with various true R_j dynamics (black, dashed) that fluctuate across controlled and resurgent stages. We use the EpiFilter package [33], which applies Bayesian smoothing algorithms, to obtain local estimates (blue) from the incidence curves I_j. Similarly, we estimate the overall R (blue) from the total incidence $\sum_{j=1}^{3} I_j$, which is how this statistic is obtained in practice. We infer D (green) and E (red) by sampling from posterior distributions of all R_j estimates and combining them according to **Eq. (6)**. All our estimates include 95% equal tailed Bayesian credible intervals, and we use default EpiFilter settings.

The simulations in **Figure 3** examine abrupt changes in transmissibility. Every group is first controlled before resurging ($j = 1,3$) or being driven towards elimination ($j = 2$). Because of its weighting by active infections, R proposes a false period of control at $t \approx 70$, even though the majority of groups feature an $R_j > 1$. This also causes R to be slow to indicate resurgence both at $t \approx 100$ and $t \approx 220$. In contrast, E is quick to signal resurgence at $t \approx 220$, without losing the ability to indicate control at $t \approx 140$. D largely interpolates between

R and E, serving as a mean of all the R_j. Moreover, as expected from the theory, R is overconfident about its transmissibility estimates (seen from its narrow credible intervals).

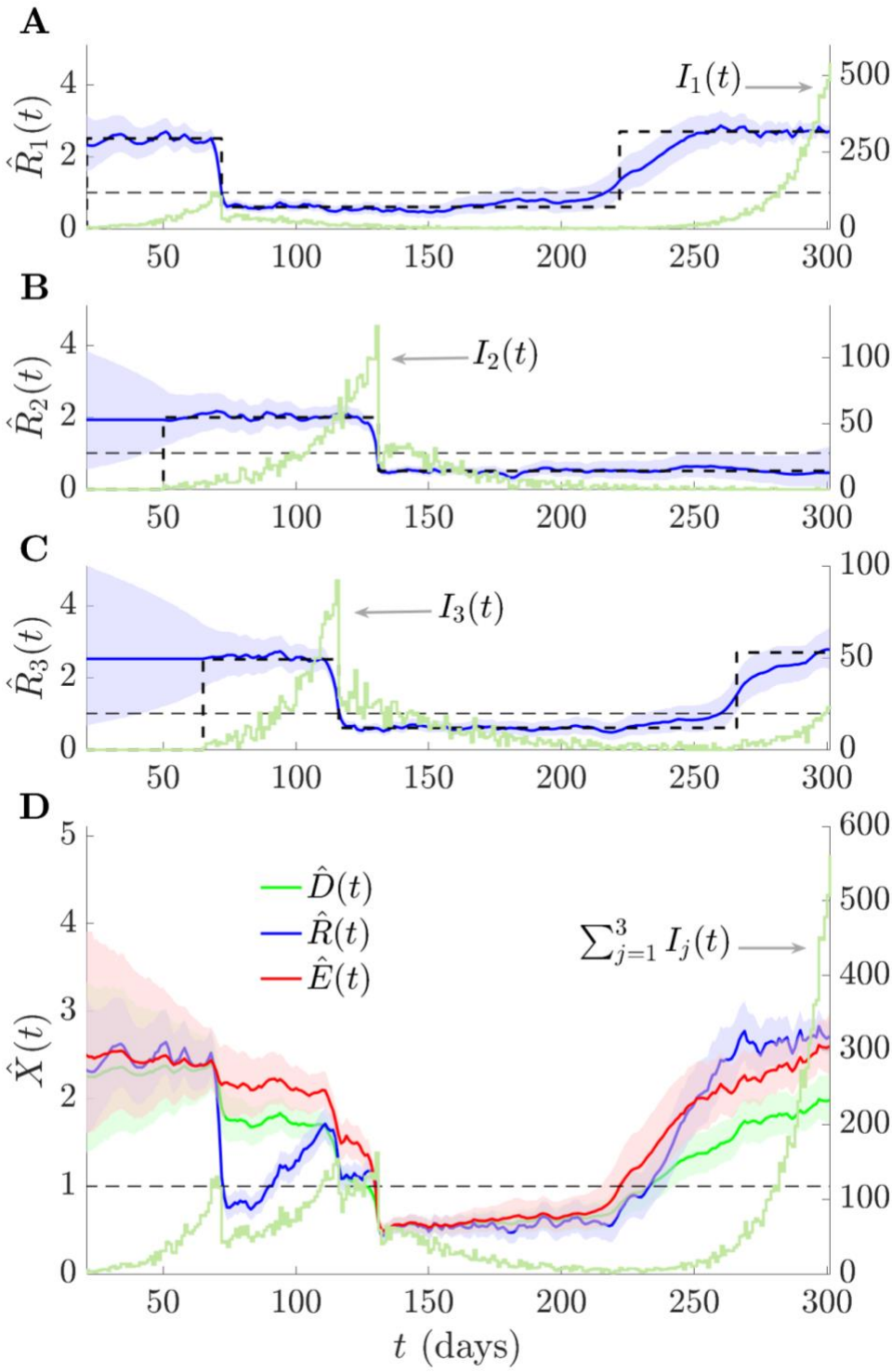

Figure 3: Consensus statistics for resurging and controlled epidemics. We simulate local epidemics $I_j(t)$ (dark green) across time t using renewal models with Ebola virus generation times from [32] and true local reproduction numbers with step-changing profiles (dashed black). Estimates of these are in (A) – (C) as $\hat{R}_j(t)$ together with 95% credible

intervals (blue curves with shaded regions). (D) provides consensus and summary statistic estimates (also with 95% credible intervals), which we calculate by combining the $\hat{R}_j(t)$. Variations in the standard reproduction number $\hat{R}(t)$ are also reflected in the total incidence $\sum_{j=1}^{3} I_j(t)$. Risk averse $\hat{E}(t)$ and mean $\hat{D}(t)$ reproduction numbers do not signal control around $t \approx 70$ and $\hat{E}(t)$ is most sensitive to resurgence signals. We use EpiFilter [33] to estimate all local and consensus reproduction numbers.

We further investigate fluctuating but anti-synchronised epidemics ($j = 1,2$) against the backdrop of a much larger monotonically increasing and then decreasing outbreak ($j = 3$) in **Figure 4**. The two out-of-phase groups roughly average to a constant value in both their infections and R_j. Consequently, we infer a monotonic R and D, with R being overconfident in its assessment of transmissibility. In contrast, estimates of E highlight the transmission potential from the fluctuating but smaller epidemics in other groups, while incorporating their uncertainties. It recognises the overall risk across $170 \leq t \leq 270$ posed by the fluctuating groups. Additionally, *E* rapidly signals the transmissibility risk that dominates from $t > 270$, which is only indicated by R after a substantial delay.

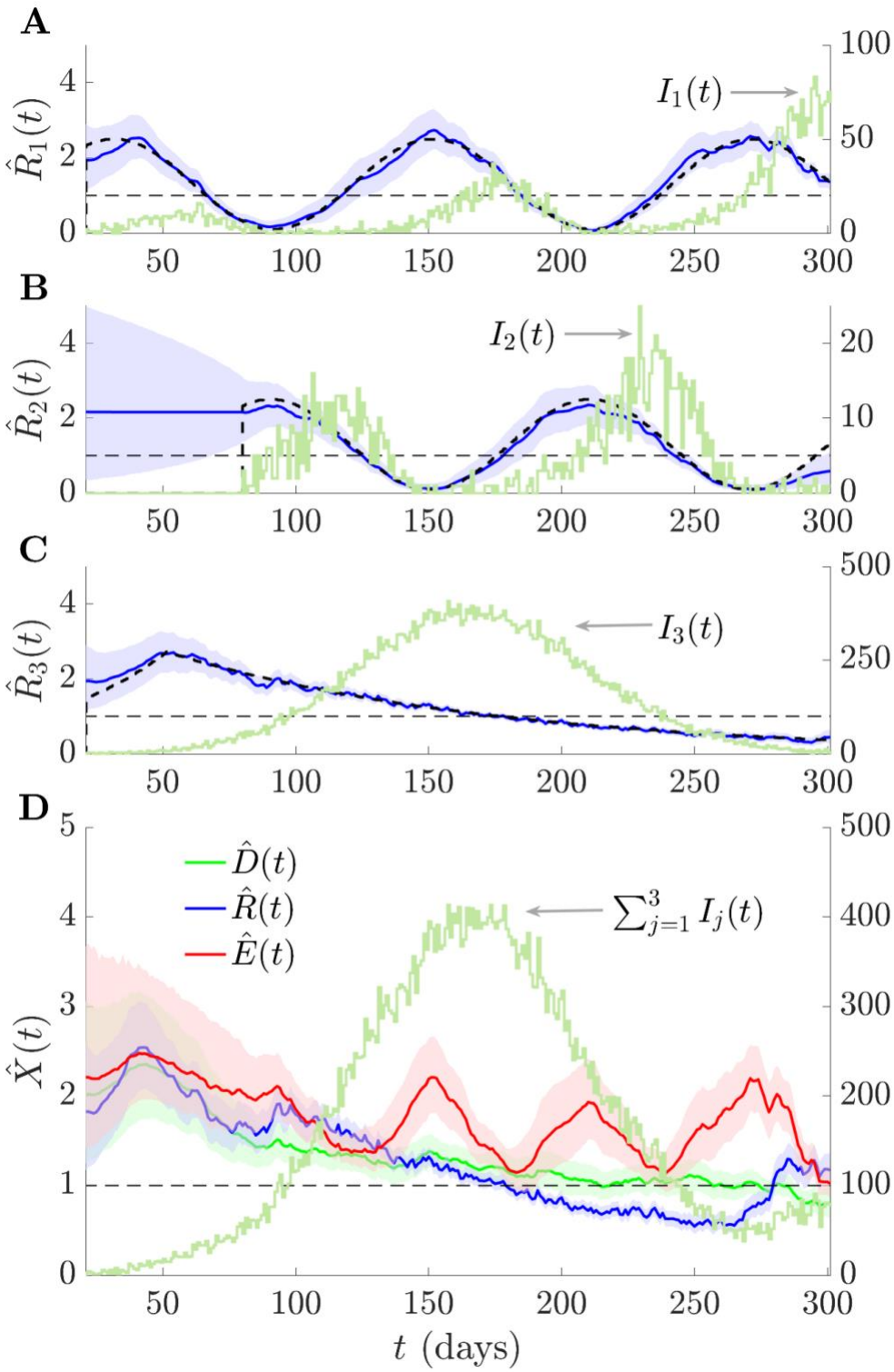

Figure 4: Consensus statistics for fluctuating and monotonic epidemic dynamics. We simulate local epidemics $I_j(t)$ over time t from renewal models with Ebola virus generation times as in [32] and true local reproduction numbers with either sinusoidal or monotonically increasing and then decreasing profiles (dashed black). Estimates of these are in (A) – (C) as $\hat{R}_j(t)$ together with 95% credible intervals. (D) plots consensus and summary statistics

(also with 95% credible intervals), which we compute by combining those $\hat{R}_j(t)$. Variations in the standard reproduction number $\hat{R}(t)$ are also reflected in the total incidence $\sum_{j=1}^{3} I_j(t)$. Both $\hat{R}(t)$ and the mean $\hat{D}(t)$ reproduction number do not average over the fluctuating transmissibility of resurging groups but the risk averse $\hat{E}(t)$ is sensitive to these potentially important signals. Only $\hat{R}(t)$ deems the epidemic to be controlled around $t \approx 200$. We use EpiFilter [33] to estimate all local and consensus reproduction numbers.

Empirical application to COVID-19 across 20 cities in Israel

We compare our consensus statistics with the standard reproduction number on empirical data from the Delta strain outbreak of COVID-19 in Israel across May–December 2021. This dataset provides a convenient case study because daily positive tests results are available from different cities in Israel and both non-pharmaceutical interventions and restrictions were mild during this period. The main intervention deployed, which was highly successful at reducing cases, was the booster vaccine campaign [34,35]. This campaign started July 30 and gradually extended to all ages across August. We examine COVID-19 incidence curves from [36] by date of test for the $p = 20$ cities with the most cases of this wave. These cities account for 49% of the entire caseload in Israel and are plotted in log scale in **Figure *5***.

We estimate the standard, mean and risk averse reproduction numbers (generically labelled as $\hat{X}(t)$) and the probability of resurgence at any time ($\mathbf{P}(\hat{X}(t) > 1)$) as in above sections but with the serial interval distribution in [37], which is consistent with the Israel-specific Delta wave parameters estimated in [35]. We assume case reporting is perfect and that this serial interval is a good approximation for the generation times. These are reasonable given high fidelity surveillance in Israel during this wave and are consistent with the analyses of [34,35]. As we only focus on relative trends, we make no further corrections to the dataset but note that accounting for issues such as testing delays generally cause incidence curves to be back-shifted and increase uncertainty, but necessitate auxiliary data [28,38]. **Figure *5*** demonstrates that all reproduction numbers agree that the wave was curbed across the booster period and that the epidemic was eventually controlled.

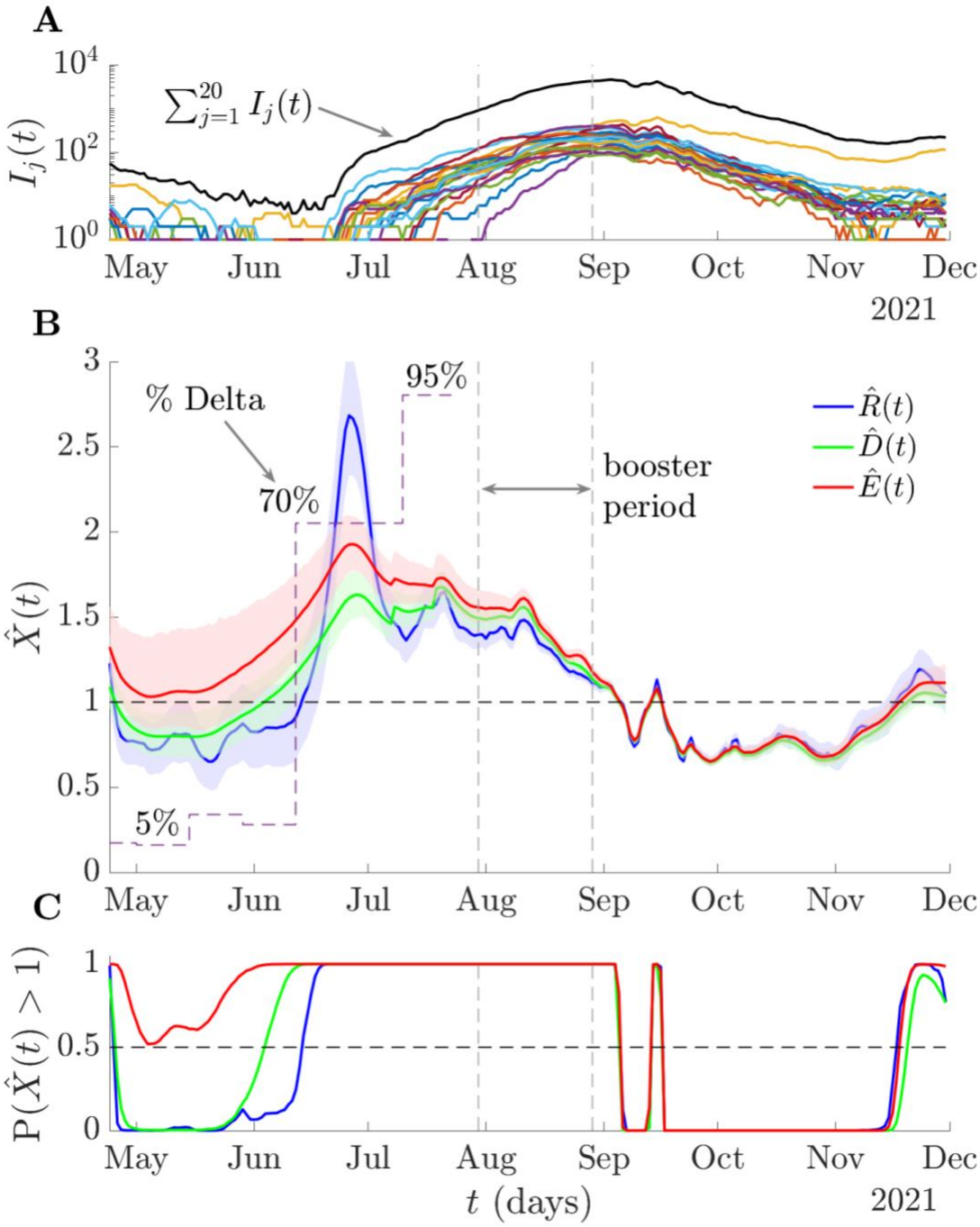

Figure 5: Risk averse reproduction numbers for COVID-19 in Israel. We plot the cases by date of positive test (and in log scale) in (A) for $p = 20$ cities in Israel during the Delta wave of COVID-19 from [36]. These constitute 49% of all cases in Israel (summed incidence in black) and have been smoothed with a weekly moving average. We infer the standard, $\hat{R}(t)$, mean, $\hat{D}(t)$, and risk averse, $\hat{E}(t)$, reproduction numbers (with their 95% credible intervals) using EpiFilter [33] in (B) under the serial interval distribution estimated in [37]. We also plot the proportion of cases attributable to the Delta strain from [35] (dark blue, dashed). We assume perfect reporting and that generation times are well approximated by the serial

intervals. (C) converts the estimates from (B) into resurgence probabilities i.e., $\boldsymbol{P}(\hat{X}(t) > 1)$. While all reproduction numbers indicate effectiveness of the vaccination campaign in curbing spread, $\hat{R}(t)$ is the slowest to signal resurgence across June, at which point the Delta strain has a 70% share in all cases. $\hat{E}(t)$ is more aligned with signalling Delta emergence.

There is substantial disagreement among the three statistics prior to the booster campaign. The standard $\hat{R}$ suggests that the wave is under control in May due to the decrease in total COVID-19 incidence. However, the risk averse $\hat{E}$ (and to some extent also $\hat{D}$) indicates resurgence and may have contributed evidence to prompt an earlier start to the booster campaign. $\hat{E}$ also better aligns with the emergence of the Delta strain or variant, which is signalled by $\hat{R}$ with a substantial delay. Given that counterfactual analyses from [34] showed that the success of the campaign was highly sensitive to the timing of its implementation, this earlier signalling of resurgence could have important ramifications as part of the policy response. The convergence of statistics across September follows as the epidemic curves of many cities became synchronised.

Discussion

The value of reproduction numbers or similar measures of transmissibility (e.g., growth rates [12]) as statistics for providing actionable information about the state of an epidemic, lies in their ability to accurately identify changepoints between subcritical and supercritical spread [39]. However, the meaning of a changepoint across scales is ambiguous and understudied. For example, if we have p local groups, how many need to resurge before we decide the epidemic has become supercritical? Is it a changepoint if these groups resurge at different times? A related question is, if those local groups are heterogeneous, is there any meaning in an overall average [40] such as the standard effective reproduction number R? Here we have explored such question and their implications for describing epidemics at large scales.

We modelled epidemics at two scales: a local scale, over which the well-mixed assumption likely holds, and a global scale, where this assumption almost surely breaks down. Reproduction numbers are commonly computed and reported at global scales. Using this framework, we analysed how changepoints in local reproduction numbers, R_j, influence the properties of global statistics. We showed that, due to its weighting of each R_j by the infections circulating in that group, $\Lambda_j\left(\sum_{j=1}^{p} \Lambda_j\right)^{-1}$, R is generally controlled by the dynamics of the groups with the most extant infections. This causes loss of sensitivity to resurgence changepoints (which may often occur at small Λ_j) and means that estimates of R are usually

overconfident or oversmoothed (**Eq. (3)**). We attempted to counter these undesirable properties by formally optimising the weights of the R_j using experimental design theory.

We derived a novel reproduction number, E, which is the optimal solution to a minimax cost function over the uncertainties in our R_j estimates. E has the property of upweighting more uncertain estimates (often associated with resurgent groups [17]) and incorporates local circulating infections according to their impact on overall estimate uncertainty. This prevents estimate overconfidence and presents a principled approach for combining the changepoints of the R_j. An $E > 1$ ensures resurging groups are emphasised without being too sensitive to individual group noise (**Figure *2***), while $E < 1$ indicates that groups are under control with a high likelihood. Interestingly, E weights each R_j by its transmissibility ratio, $R_j\left(\sum_{j=1}^{p} R_j\right)^{-1}$, which results in a formulation (**Eq. (6)**) that seems consistent with that derived from network epidemic models when individuals have heterogeneous contact rates, as discussed in [40].

We illustrated and validated these properties of E on simulated and empirical data (**Figure *3*–5**), demonstrating how E presents a better consensus than R of the local group dynamics. The earlier resurgence signalling of E may be important given the sensitivity of the cost and effectiveness of interventions on their implementation times [34,41]. Note that if this earlier resurgence is ignored then eventually $\Lambda_j\left(\sum_{j=1}^{p} \Lambda_j\right)^{-1}$ will be similar to $R_j\left(\sum_{j=1}^{p} R_j\right)^{-1}$ and R will also indicate that a change to supercritical spread has occurred. As a consequence of its risk averse property, E will signal subcritical spread at global scales when there is a higher likelihood of groups being under control. This avoids premature relaxations of interventions, which may be supported by R, and may be useful for informing exit strategies [42].

While E is a promising addition to the suite of infectious disease outbreak statistics, it is not perfect. First, its formulation depends on Poisson noise models (**Eq. (1)**). While such models are commonly applied [10], in some cases they may only offer simplistic representations of the stochasticity of epidemics. Second, we defined indicators of resurgence and control based on the reproduction number threshold value of 1. Although this is an almost universal definition, other measures (e.g., the early warning signals of [43]) may circumvent some of the problems of R as a metric for informing decision-making. Third, E requires infection time and incidence data (or proxies [28]) at the resolution of the local scale. While this is becoming the norm with improved surveillance [44], it is not guaranteed, and may be scarce for emerging infectious diseases. In this scenario R is still directly computable from global scale data. Last, the added value of E in informing real time decision-making is controlled by

quality of data. Practical biases can preclude timely responses to transmission changepoints [38], though E may still be of use retrospectively.

Overall, we propose E as a consensus statistic that better encapsulates salient dynamics across groups without losing the interpretability or computability of R. With late responses to epidemic resurgence associated with larger epidemic burden [41], increasing interest in early warning signals [43] and reproduction numbers often being computed on vast scales in real time [9], the risk averse properties of E may be impactful. Public health policymaking is a complex process combining inputs from diverse data and models spanning epidemiology, economics and behavioural science. Given this complexity, we think that statistics designed with optimal and deliberate properties, such as E, can facilitate more transparent and robust data-driven decision-making.

Data and code availability

We provide open-source software to reproduce all analyses at https://github.com/kpzoo/risk-averse-R-numbers. While the main code for generating the figures in this text is written in MATLAB, we also include functions in R to compute *E* on user-defined datasets.

Bibliography

1. Fraser C. Estimating individual and household reproduction numbers in an emerging epidemic. PLoS One. 2007;2: e758. doi:10.1371/journal.pone.0000758
2. Cauchemez S, Boëlle P-Y, Thomas G, Valleron A-J. Estimating in real time the efficacy of measures to control emerging communicable diseases. Am J Epidemiol. 2006;164: 591–597. doi:10.1093/aje/kwj274
3. Team WER. Ebola Virus Disease in West Africa – The First 9 Months of the Epidemic and Forward Projections. N Engl J Med. 2014;371: 1481–95.
4. Churcher T, Cohen J, Ntshalintshali N, Others. Measuring the path toward malaria elimination. Science. 2014;344: 1230–32.
5. Fraser C, Donnelly CA, Cauchemez S, Hanage WP, Van Kerkhove MD, Hollingsworth TD, et al. Pandemic potential of a strain of influenza A (H1N1): early findings. Science. 2009;324: 1557–1561. doi:10.1126/science.1176062
6. 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
7. 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
8. Thompson RN, Jalava K, Obolski U. Sustained transmission of Ebola in new locations: more likely than previously thought. Lancet Infect Dis. 2019;19: 1058–1059. doi:10.1016/S1473-3099(19)30483-9
9. 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
10. 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;
11. 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
12. Parag KV, Thompson RN, Donnelly CA. Are epidemic growth rates more informative than reproduction numbers? J Royal Statistical Soc A. 2022; doi:10.1111/rssa.12867
13. Kiss IZ, Miller JC, Simon PL. Mathematics of epidemics on networks. Cham: Springer International Publishing; 2017. doi:10.1007/978-3-319-50806-1
14. Watts DJ, Muhamad R, Medina DC, Dodds PS. Multiscale, resurgent epidemics in a hierarchical metapopulation model. Proc Natl Acad Sci USA. 2005;102: 11157–11162. doi:10.1073/pnas.0501226102
15. Ferguson NM, Cummings DAT, Fraser C, Cajka JC, Cooley PC, Burke DS. Strategies for mitigating an influenza pandemic. Nature. 2006;442: 448–452. doi:10.1038/nature04795
16. Chang S, Pierson E, Koh PW, Gerardin J, Redbird B, Grusky D, et al. Mobility network models of COVID-19 explain inequities and inform reopening. Nature. 2021;589: 82–87. doi:10.1038/s41586-020-2923-3
17. 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
18. 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
19. 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
20. Atkinson A, Donev A. Optimal Experimental Designs. Oxford University Press; 1992.
21. 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
22. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning. 2nd ed. New York, NY: Springer New York; 2009. pp. 106–119. doi:10.1007/978-0-387-84858-7
23. Grenfell BT, Bjørnstad ON, Kappey J. Travelling waves and spatial hierarchies in measles epidemics. Nature. 2001;414: 716–723. doi:10.1038/414716a
24. Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. Am J Epidemiol. 2004;160: 509–516. doi:10.1093/aje/kwh255
25. Lehmann EL, Casella G. Theory of point estimation. New York: Springer-Verlag; 1998. doi:10.1007/b98854
26. Parag KV, Donnelly CA. Adaptive estimation for epidemic renewal and phylogenetic skyline models. Syst Biol. 2020;69: 1163–1179. doi:10.1093/sysbio/syaa035
27. Bhatia S, Lassmann B, Cohn E, Desai AN, Carrion M, Kraemer MUG, et al. Using digital surveillance tools for near real-time mapping of the risk of infectious disease spread. npj Digital Med. 2021;4: 73. doi:10.1038/s41746-021-00442-3
28. Parag KV, Donnelly CA, Zarebski AE. Quantifying the information in noisy epidemic curves. Nat. Comput. Sci. 2022; In Press
29. Banks H, Davidian M. Generalized Sensitivities and Optimal Experimental Design. North Carolina State University; 2009.

30. Parag KV, Pybus OG. Robust design for coalescent model inference. Syst Biol. 2019;68: 730–743. doi:10.1093/sysbio/syz008
31. Marshall AW, Olkin I, Arnold BC. Inequalities: theory of majorization and its applications. New York, NY: Springer New York; 2011. doi:10.1007/978-0-387-68276-1
32. 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
33. Parag KV. Improved estimation of time-varying reproduction numbers at low case incidence and between epidemic waves. PLoS Comput Biol. 2021;17: e1009347. doi:10.1371/journal.pcbi.1009347
34. Gavish N, Yaari R, Huppert A, Katriel G. Population-level implications of the Israeli booster campaign to curtail COVID-19 resurgence. Sci Transl Med. 2022;14: eabn9836. doi:10.1126/scitranslmed.abn9836
35. Feng A, Obolski U, Stone L, He D. Modelling COVID-19 Vaccine Breakthrough Infections in Highly Vaccinated Israel – the effects of waning immunity and third vaccination dose. medRxiv. 2022; doi:10.1101/2022.01.08.22268950
36. Israeli MoH corona dashboard [Internet]. [cited 23 Aug 2022]. Available: https://datadashboard.health.gov.il/COVID-19/general
37. 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
38. 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
39. Dehning J, Zierenberg J, Spitzner FP, Wibral M, Neto JP, Wilczek M, et al. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. Science. 2020;369. doi:10.1126/science.abb9789
40. 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
41. 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
42. Ruktanonchai NW, Floyd JR, Lai S, Ruktanonchai CW, Sadilek A, Rente-Lourenco P, et al. Assessing the impact of coordinated COVID-19 exit strategies across Europe. Science. 2020;369: 1465–1470. doi:10.1126/science.abc5096
43. Drake JM, Brett TS, Chen S, Epureanu BI, Ferrari MJ, Marty É, et al. The statistics of epidemic transitions. PLoS Comput Biol. 2019;15: e1006917. doi:10.1371/journal.pcbi.1006917
44. Buckee C. Improving epidemic surveillance and response: big data is dead, long live big data. Lancet Digit Health. 2020;2: e218–e220. doi:10.1016/S2589-7500(20)30059-5

Funding

KVP acknowledges funding from the MRC Centre for Global Infectious Disease Analysis (reference MR/R015600/1), jointly funded by the UK Medical Research Council (MRC) and the UK Foreign, Commonwealth & Development Office (FCDO), under the MRC/FCDO Concordat agreement and is also part of the EDCTP2 programme supported by the European Union. 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.