Sub-spreading events limit the reliable elimination of heterogeneous epidemics

Kris V Parag[†, 1]

[1]MRC Centre for Global Infectious Disease Analysis, Imperial College London, London, W2 1PG, UK

[†]Email: k.parag@imperial.ac.uk

Abstract

We show that sub-spreading events i.e., transmission events in which an infection propagates to few or no individuals, can be surprisingly important for defining the lifetime of an infectious disease epidemic and hence its waiting time to elimination or fade-out, measured from the time-point of its last observed case. While limiting super-spreading promotes more effective control when cases are growing, we find that when incidence is waning, curbing sub-spreading is more important for achieving reliable elimination of the epidemic. Controlling super-spreading in this low-transmissibility phase offers diminishing returns over non-selective population-wide measures. By restricting sub-spreading we efficiently dampen remaining variations among event reproduction numbers, which minimises the risk of premature and late end-of-epidemic declarations. Because case-ascertainment rates can be modelled in exactly the same way as control policies, we concurrently show that the under-reporting of sub-spreading events during waning phases will engender overconfident assessments of epidemic elimination. While controlling sub-spreading may not be easily realised, the likely neglecting of these events by surveillance systems could result in unexpectedly risky end-of-epidemic declarations. Super-spreading controls the size of the epidemic peak but sub-spreading mediates the variability of its tail.

Key-words: infectious diseases, heterogeneity, overdispersion, super-spreading, elimination, reproduction numbers.

Background

Emerging infectious diseases are a major and recurring threat to both global health and economies. The ongoing COVID-19 pandemic has exemplified this, highlighting a need for improved understanding on how interventions might be applied and relaxed to jointly minimise public health risks and socio-economic costs. Key parameters, which characterise the impact of interventions on the transmission dynamics of an infectious disease are the time-varying reproduction number, denoted R_s at time s, with mean μ_s, [1] and the dispersion level, k, of the offspring

distribution of the epidemic [2]. The first describes transmission potential by measuring the number of secondary infections per primary case at s. The second defines transmission heterogeneity i.e., it captures the variation of possible R_s about mean μ_s. The values of these parameters can substantially influence intervention policy and much debate remains on their implications for both epidemic control and elimination [3], [4], [5].

A $\mu_s > 1$ forewarns of rising incidence (new cases), necessitating the application of controls, while $\mu_s < 1$ signifies that the epidemic is being controlled, potentially allowing the relaxation of some interventions [1]. However, the risk and effectiveness of policies are modulated by k. If $k \gg 1$ then transmission is homogeneous (μ_s is representative of the realised R_s) and we can apply simple, population-wide controls. Alternatively, if $k < 1$, then heterogeneity is large (the epidemic offspring distribution is overdispersed). This means that infrequent super-spreading events, in which each primary case causes $\gg \mu_s$ secondary ones, disproportionately drive overall transmission [2]. The majority of other transmission events, which we refer to as sub-spreading events, lead to fewer than μ_s or even 0 secondary cases. Heterogeneous epidemics may require more targeted, event-specific policy as population-wide measures can often be ineffectual [6], [2].

Many works have examined how k and μ_s interact to regulate the effectiveness of control measures applied to growing epidemics. When transmission is heterogeneous, the consensus is that interventions aimed at mitigating super-spreading events (e.g., by limiting large gatherings) can minimise both peak epidemic size and resource-usage [2], [7], [8]. While these insights inform on how to efficiently impose control measures, the converse problem of when to safely release interventions and de-escalate surveillance, for waning epidemics, has been understudied [9], [10]. The importance of this problem will only elevate as countries currently debate the value of elimination-based strategies for handling the ongoing COVID-19 pandemic [11], [3]. In this paper we investigate this relaxation problem in the context of achieving reliable and minimum-risk end-of-epidemic declarations.

The timing of an end-of-epidemic declaration is strategically crucial as it triggers the removal of restrictions or control measures, the re-allocation of resources and the resumption of trade, travel and other economic activities. However, getting this timing right is both non-trivial and consequential. Early declarations (and hence intervention relaxation) can elevate the risk of resurgence or second waves, whereas late ones can cause needless cost. Much is unknown about how the transmission dynamics of an infectious disease dictate the tail of an epidemic. Recent studies suggests that this knowledge-gap may have led to early declarations for Ebola virus disease (EVD) in West Africa [12] and late ones for MERS-CoV in South Korea [13].

Prime among these unknowns is the influence of heterogeneity and its associated targeted interventions. While some studies have started analysing the interacting roles of μ_s and k to decipher this influence [12], [14], a general framework for rigorously testing hypotheses about how heterogeneity and control choices impact end-of-epidemic declarations is lacking. We extend theory from [2], [15] to develop such a framework, which allows us to generate

novel and principled insight into how μ_s and k (and hence possible R_s values) mediate the tail of an epidemic. We take a bottom-up approach, tracing how heterogeneity in transmission, defined by variations in R_s due to k, causes fluctuations in incidence, which then manifest as noise in the probability of epidemic fade-out or elimination.

As this probability defines our confidence in any end-of-epidemic declaration [15], we obtain a measure of the risk of that declaration. This framework leads to several new results. First, we observe that while decreasing k (increasing heterogeneity) reduces the mean waiting time to elimination, measured from the last observed reported case (a standard reference point [16]), it also increases the maximum waiting time. This contextualises the common assertion that heterogeneity increases the likelihood of epidemic extinction [17], showing that while this may happen on average, the increased uncertainty it brings actually hampers safe end-of-epidemic declarations.

Second, we find that sub-spreading, surprisingly, plays a larger part than super-spreading in setting the uncertainty around the probability of elimination and hence end-of-epidemic declarations. We confirm this by testing three control strategies, which target (i) all transmission events, (ii) super-spreading and (iii) sub-spreading at a fixed level of control effort ρ (the mean reproduction number is then $\rho\mu_s$). Strikingly, we see that waning epidemics, which feature $\rho\mu_s < 1$, are most reliably eliminated by curbing sub-spreading events. Moreover, we find the benefits of super-spreading control relative to the simpler approach of (i) largely disappear in this epidemic phase.

Because ρ can equally model the mean case reporting or ascertainment rate [18], results from (i)-(iii) concurrently describe the impact of uniform, size-inverse (super-spreading events are under-ascertained) and size-biased (sub-spreading is under-ascertained) reporting, respectively. Consequently, we uncover that the under-reporting of sub-spreading events, which is likely, can engender overconfident and risky end-of-epidemic declarations. These issues are in addition to known biases that under-reporting cause for epidemic elimination [15]. Robustly assessing the endpoint of an outbreak may therefore require either targetting interventions at sub-spreading events (if possible) or increasing contact tracing surveillance to minimise the under-reporting of those events.

One defining characteristic of heterogeneous transmission is an excess of sub-spreading events [2]. However, these events have received little attention in the literature. Only recently has the importance of sub-spreading been recognised, in related network epidemic models [19]. Our results draw attention to the unexpected influence of sub-spreading, clarifying its impact on epidemic elimination and adding context and detail to the subtleties of controlling and monitoring epidemics with overdispersed transmission. Hopefully, our framework will help inform guidelines for achieving safe and efficient end-of-epidemic declarations and intervention de-escalations in the face of heterogeneity and improve understanding of the complex dynamics driving real epidemics.

METHODS

Renewal models with heterogeneous transmission

We consider an outbreak observed daily over the time period (in days) $[1,\, t]$ with the number of infected cases on day $s \leq t$ as I_s. The incidence curve of this epidemic is denoted $I_1^t := \{I_1,\, I_2\, \ldots,\, I_t\}$. We model the time-varying transmission of a communicable disease within a population using a renewal process [20] [21]. This process describes how an infection spreads from a primary case to secondary ones at time s using two key variables: the effective reproduction number, R_s, and the generation time distribution, with probability w_u at time u. Here R_s is the number of secondary cases at time s every effective case at $s-1$ infects, while w_u gives the probability that the average time for a primary to secondary transmission is u days [20].

Under this renewal process framework, the observed incident cases at time s depend on R_s and past cases (over period $[1,\, s-1]$) according to the Poisson (Pois) relation on the left side of Eq. (1).

$$I_s \sim \mathrm{Pois}(R_s \Lambda_s), \qquad R_s \sim \mathrm{Gam}(k,\, \mu_s k^{-1}) \tag{1}$$

Here $\Lambda_s := \sum_{u=1}^{s-1} I_{s-u} w_u$, which depends on I_1^{s-1}, is known as the total infectiousness. It characterises how many past effective cases contribute to the next observed case-count at s. The generation time distribution (which we assume to be equal to the serial interval distribution) is central to defining the impact of each past case [20].

If we describe an epidemic as consisting of a sequence of spreading or transmission events, with the reproduction number of the spreading event at time s as R_s then standard renewal models assume a fixed R_s [21], [22], [23]. This formulation, while useful, does not account for possible heterogeneities in transmission, which are known features of many respiratory diseases such as the SARS and MERS coronaviruses [2]. If we define a distribution over R_s with mean $\mathbb{E}[R_s] = \mu_s$ then these models set $\mathbb{P}(R_s = \mu_s) = 1$. Heterogeneous transmission is a mean preserving spread of this condition i.e. events with fixed mean μ_s can have different R_s values.

We define super-spreading events as those driven by R_s significantly larger than μ_s. This characterisation differs slightly from the standard in [2], which directly uses numbers of secondary cases. Since the renewal model has long-term memory (i.e. factors in the age of infections via Λ_s) using I_s would not be as appropriate here. However, because I_s behaves like a noisy, scaled version of R_s (by the properties of Poisson mixtures [24]), these two definitions are largely consistent. In this work we consider the end of the epidemic, which follows the waning phase of the epidemic. This contrasts the development in [2], which focusses on the growth phase. We define sub-spreading events as having R_s notably smaller than μ_s. This type of event has received appreciably less attention (than super-spreading ones) and is our main topic of study.

If we make the usual assumption that R_s is gamma (Gam) distributed with shape k and scale $\mu_s k^{-1}$ i.e. the right side of Eq. (1), then we obtain the negative binomial (NB) relation $I_s \sim \mathrm{NB}\left(k,\, \frac{\mu_s \Lambda_s}{\mu_s \Lambda_s + k}\right)$. This is the most common

method for incorporating heterogeneity within renewal model frameworks [12], [25], [14]. As k gets smaller the likelihood of both super- and sub-spreading events increases [2], [24]. Special cases are at $k \to \infty$, $k = 1$ and $k \to 0$ for which I_s has a Poisson, geometric and logarithmic distribution respectively with mean $\mu_s \Lambda_s$ [18]. Many of the infectious diseases that feature significant heterogeneity have been found to exhibit $k < 1$ [2], [26]. Note that this NB model can also be used to describe reporting noise and other types of heterogeneity.

Variation in reproduction numbers and incidence

We explicitly characterise how heterogeneity in transmission can control the incidence of an epidemic and then assess the implications of this observation. Consider any arbitrary distribution over the effective reproduction number at time s, R_s, with mean μ_s. Using the law of total expectation we can write $\mathbb{E}_I[I_s] = \mathbb{E}_R \mathbb{E}_I[I_s \,|\, R_s]$ with the subscripts clarifying (when needed) about which variable we are taking expectations. Using the renewal model (left expression in Eq. (1)), we obtain the straightforward but general Eq. (2) below.

$$\mathbb{E}[I_s] = \Lambda_s \mathbb{E}[R_s] = \Lambda_s \mu_s \tag{2}$$

Eq. (2) shows that for any renewal model there is a direct relationship between mean incidence and μ_s.

Similarly, we apply the law of total variance to get: $\mathbb{V}[I_s] = \mathbb{E}_R[\mathbb{V}_I[I_s \,|\, R_s]] + \mathbb{V}_R[\mathbb{E}_I[I_s \,|\, R_s]]$ [24] (with $\mathbb{V}[.]$ indicating variance). Expanding this for renewal models we derive Eq. (3), which holds for any R_s distribution.

$$\mathbb{V}[I_s] = \mathbb{E}[I_s] + \Lambda_s^2 \, \mathbb{V}[R_s] \tag{3}$$

If we substitute the statistics from the standard gamma R_s distribution (right expression in Eq. (1)) we recover the usual variance of the NB incidence distribution i.e. $\Lambda_s \mu_s + \Lambda_s^2 \mu_s^2 k^{-1}$ [12]. The key takeaway from Eq. (3) is that, given the past (summarised by Λ_s), the variance of the reproduction numbers linearly controls that of the incidence at any time. Eq. (2) and Eq. (3) explain why we can generally map dynamics of R_s onto I_s.

As a result, we can decipher the influence of super- and sub-spreading events (which are respectively linked to the tail and head i.e the right and left portions of the R_s distribution) by investigating the level of heterogeneity or variation among reproduction numbers. This is useful because it is hard to reliably estimate secondary cases caused by past primary cases during the waning stages of the epidemic [13].

The impact of control measures [2] or under-reporting [18] (which are often mathematical analogues) can be measured by their signature on the variance-to-mean ratio (VM) of the incidence: $\mathrm{VM}[I_s] := \mathbb{V}[I_s]/\mathbb{E}[I_s]$. Using Eq. (2) and Eq. (3) we obtain Eq. (4) connecting the VM ratios of R_s to I_s.

$$\mathrm{VM}[I_s] = 1 + \Lambda_s \mathrm{VM}[R_s] \tag{4}$$

Thus, once again the statistics of R_s strongly modulate those of I_s at any time, for a given Λ_s. While this result is simple, its ramifications, which are meaningful, have not been explored.

Understanding how properties of the R_s distribution map onto the related I_s one, which is a mixed Poisson distribution, provides insights into other epidemic properties. From mixed Poisson theory [24] we deduce that (a) if R_s is unimodal and continuous then I_s is also unimodal, (b) the shape of I_s will be similar to that of the mixing distribution describing R_s (e.g. when R_s is exponential, I_s becomes geometrically distributed) and (c) every mixed Poisson distribution corresponds to a unique mixing distribution i.e. multiple R_s distributions cannot map to the same I_s distribution [24]. Properties (a)–(c) establish that much about super- and sub-spreading events can be learned from R_s. We will exploit these relationships to better understand epidemic elimination and heterogeneity.

Elimination probabilities and epidemic lifetimes

We define an epidemic to be eliminated (or extinct) [5] at some time s if no future infected (local) cases are observed i.e. $I_{s+1} = I_{s+2} = \cdots = I_\infty = 0$ [15], [5]. If we initially assume that all future mean reproduction numbers, μ_{s+1}^∞, are known, then we can construct the probability of elimination given some sample R_{s+1}^∞ from μ_{s+1}^∞ (see Eq. (1)) as $z_s = \mathbb{P}(\sum_{j=s+1}^\infty I_j = 0 \,|\, R_{s+1}^\infty, I_1^s)$. As we condition on the sample R_{s+1}^∞ and because incidence is non-negative and I_j does not depend on R_{j+1}^∞, we can decompose z_s to get Eq. (5).

$$z_s := \prod_{j=s+1}^\infty \mathbb{P}\left(I_j = 0 \,|\, R_j,\, I_1^s\right) \tag{5}$$

From Eq. (1), $\mathbb{P}\left(I_j = 0 \,|\, R_j,\, I_1^s\right) = e^{-\Lambda_j R_j}$ (also see [3]) and we obtain Eq. (6), with future Λ_j computed by incorporating the $I_{s+1}^j = 0$ terms from above as pseudo-data (see [15], [27]).

$$\log z_s = -\sum_{j=s+1}^\infty \Lambda_j R_j \tag{6}$$

Hence, the elimination probability has an uncomplicated log-linear dependence on the sample sequence of reproduction numbers and accordingly depends on all future spreading events. This type of relationship is maintained even if we generalise Eq. (1) to include case observation noise (e.g due to importation or delays) [22], [28]. Methods for inferring z_s given these noise sources are currently being developed in [15], [10].

While Eq. (5) can be computed at any time, commonly elimination is only considered when zero-case days are observed in sequence [13], [16]. Keeping to convention, we will often present results in terms of Δs, which is the time relative to that at which incidence was last non-zero, t_0. We compute the relative time at which the epidemic is eliminated with $\alpha\%$ confidence, given the transmission event sample R_{s+1}^∞ , as t_α in Eq. (7).

$$t_\alpha = \arg\min_s z_s - t_0 \geq \tfrac{\alpha}{100} \tag{7}$$

Eq. (7) also gives the time that an epidemic, composed of spreading events R_{s+1}^{∞}, can be declared over with at least $\alpha\%$ confidence [29], [15]. This confidence reflects the remaining variability among possible epidemic trajectories despite conditioning on the fixed future R_{s+1}^{∞} sample and past observed incidence I_1^s.

If we think of the epidemic as a process that generates infections then its survival function, which captures the probability of the epidemic propagating at least 1 future case after s, is precisely $1 - z_s$. The mean lifetime of the epidemic then follows from survival theory (where it is called the mean time to failure) as $\sum_{s=1}^{\infty}(1 - z_s)$ [30]. Given R_{s+1}^{∞}, we can compute this lifetime exactly, as Λ_{s+1}^{∞} only depends on future incidence values, which are all 0 for this condition [15]. Hence z_s has an influential role in determining the risk of epidemic survival. As end-of-epidemic declarations times depend on z_s, this risk has important public health and economic implications.

However, we cannot know what sample from μ_{s+1}^{∞} will be realised, and must account for this uncertainty, which derives from the heterogeneity in transmission and hence depends on the levels of super- and sub-spreading. Since, practically, an authority would want a single time for declaring an epidemic to be eliminated or over [16], we need some way of summarising the range of z_s curves and hence t_α values that could result. We propose two such statistics: the mean $\bar{z}_s$ and the minimum $z_{s,\min}$. The resulting declaration times are then $\bar{t}_\alpha$ and $t_{\alpha,\max}$. The mean elimination probability was used in [15], while the worst case (minimum) was proposed in [29].

We derive $\bar{z}_s$ exactly from Eq. (6) as $\mathbb{E}[e^{-\sum_{j=s+1}^{\infty} R_j \Lambda_j}]$. This yields Eq. (8) with $\mathbb{M}_{R_j}$ as the moment generating function according to the distribution of R_j, which we evaluate at Λ_j [31].

$$\bar{z}_s = \mathbb{E}[z_s] = \prod_{j=s+1}^{\infty} \mathbb{M}_{R_j}(-\Lambda_j) \tag{8}$$

This formulation has immediate consequences [20]. First, $\bar{z}_s \geq \prod_{j=s+1}^{\infty} e^{-\Lambda_j \mu_j}$. Ignoring heterogeneity will therefore on average result in our underestimating the elimination probability and so overestimating when the outbreak is over with $\alpha\%$ confidence. This extends earlier works, which used simpler branching process epidemic models to link heterogeneity and extinction [2], to more realistic renewal epidemic descriptions [7].

Second, because $\mathbb{M}_{R_j}$ has a unique correspondence to the distribution over R_j [31], sub- and super-spreading events have direct roles in shaping the mean elimination probability. For example, given some threshold c, Chernoff's bound stipulates that $\mathbb{M}_{R_j}(-\Lambda_j) \geq e^{-c\Lambda_j}\mathbb{P}(R_j \geq c)$, with $\mathbb{P}(R_j \geq c)$ specifying our spreading event likelihoods. Last, under the gamma distribution in Eq. (1) we can explicitly compute Eq. (8) as Eq. (9) [31].

$$\bar{z}_s = \prod_{j=s+1}^{\infty} \left(1 + \Lambda_j \mu_j k^{-1}\right)^{-k} \tag{9}$$

Eq. (9) provides intuition into how the offspring dispersion parameter k controls the mean elimination probability and hence the epidemic lifetime. Applying the product rule, we can show that $\frac{\mathrm{d}\bar{z}_s}{\mathrm{d}k} < 0$, proving that as heterogeneity rises i.e. k gets smaller, $\bar{z}_s$ (and hence $\bar{t}_\alpha$) always increases (respectively decreases).

While the worst case elimination probability, $z_{s,\min}$, does not admit such analytic development and so is computed via simulation, we gain some insight about its behaviour from the variance among the z_s curves: $\mathbb{V}[z_s] = \mathbb{V}\left[\prod_{j=s+1}^{\infty} e^{-\Lambda_j R_j}\right]$. Each term in this product is independent so we decompose this to $\prod_{j=s+1}^{\infty} \mathbb{E}[e^{-2\Lambda_j R_j}] - \prod_{j=s+1}^{\infty} \mathbb{E}[e^{-\Lambda_j R_j}]^2$ to get Eq. (10) with $\bar{z}_s$ taken from Eq. (8).

$$\mathbb{V}[z_s] = \prod_{j=s+1}^{\infty} \mathbb{M}_{R_j}(-2\Lambda_j) - \bar{z}_s^2 \tag{10}$$

We obtain an explicit form for Eq. (10) by substituting the moment generating functions for gamma distributions as in Eq. (9). It follows that $\lim_{k\to\infty} \mathbb{V}[z_s] = 0$. We generally find that as k falls, variation among z_s curves and hence possible t_α increases (see Results). The gap between $t_{\alpha,\max}$ and $\bar{t}_\alpha$ widens with heterogeneity.

Understanding how variation in R_s maps to uncertainty in elimination times is the main focus of this work. Specifically, we examine how changes to this variation, due to different control or case ascertainment strategies, express themselves in our ability to reliably adjudge the endpoint of the epidemic. In the Results we exploit the mathematical framework developed here to investigate how $\bar{z}_s$ and $z_{s,\min}$, which are two measures of the aggregate risk of an epidemic, depend on super- and sub-spreading events. Based on these, we attempt to derive policies for reliably eliminating an epidemic i.e. reducing its lifetime with minimum risk. This moment generating function approach to elimination or fade-out is novel.

RESULTS

Reducing variation among reproduction numbers

In Eq. (4) we showed how VM ratios of the event reproduction numbers, R_s, directly control those of the incidence values, I_s. We first verify this relationship on numerous epidemics simulated according to the heterogeneous renewal model of Eq. (1). We consider epidemics characterised by an initial exponential growth followed by drastic control (e.g. a lockdown measure) and compute the VM ratios of both I_s and R_s for all times s of the epidemics. Fig. 1 shows that as heterogeneity increases (i.e. k falls) the VM of I_s and R_s both rise, despite different mean values across the possible epidemics as well as varying total infectiousness Λ_s.

This solidifies the idea of modulating $\mathrm{VM}[I_s]$ via the control of $\mathrm{VM}[R_s]$. We consider three main strategies for implementing control in a heterogeneous epidemic. To make notation simpler we drop the subscript s and constrain all control protocols to reduce the mean reproduction number from μ to $\rho\mu$ with $\rho \leq 1$. Consequently, after control the new distribution describing R satisfies $\int_0^\infty \mathbb{P}(R)\,\mathrm{d}R = 1$ and $\int_0^\infty \mathbb{P}(R)R\,\mathrm{d}R = \rho\mu$. These descriptions also model case reporting or ascertainment schemes, where ρ is the sample or reporting fraction instead [18].

(i) Uniform control (constant reporting). Also called population-wide control and introduced in [18], this protocol assumes that all event reproduction numbers are reduced by the constant fraction ρ. This is achieved by rescaling

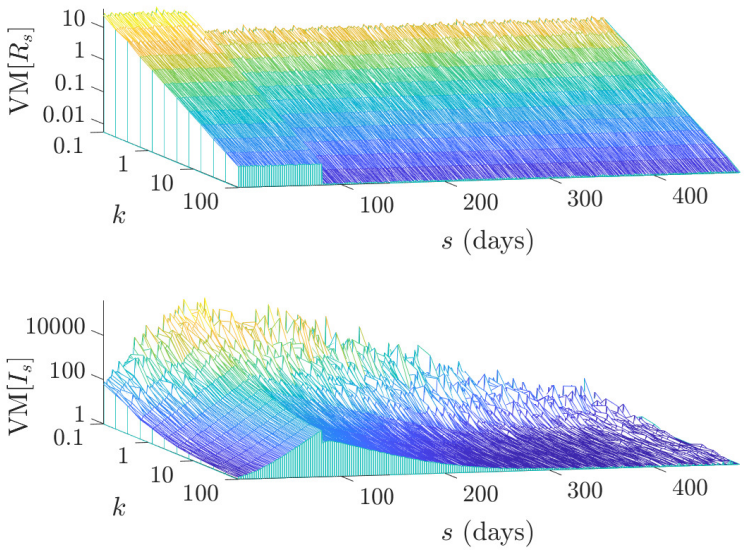

Fig. 1: **VM ratios of incidence and reproduction numbers.** We simulate 2000 epidemics under a reproduction number profile that has fixed mean of 2.5 up to $s = 80$ and then a step change fall to 0.5 until elimination, using the generation time distribution from [14]. We vary the offspring dispersion k of the renewal model used to generate the epidemics and compute the VM ratio of R_s and I_s at every time s. As expected from Eq. (4), there is a strong correspondence between $\mathrm{VM}[R_s]$ and $\mathrm{VM}[I_s]$ (mediated by Λ_s) and smaller k inflates both values.

Eq. (1) so that $R \sim \mathrm{Gam}(k,\, \rho\mu k^{-1})$ and is the simplest type of control that can be applied. Uniform control admits the analytic $\mathrm{VM}[R] = \rho\mu k^{-1}$, which falls linearly as $\rho \to 0$. It is equivalent to constant (or Bernoulli) reporting with the probability of observation or sampling any type of spreading event set at ρ.

(ii) Super-spreading control (size-inverse reporting). Also called targeted control, and formulated in [7], this strategy disproportionately reduces super-spreading events and is within the class of individual-specific control measures detailed in [2]. It is implemented by upper-truncating the distribution of R at some maximum value b and rescaling so that $\int_0^b \mathbb{P}(R)\,\mathrm{d}R = 1$ and $\int_0^b \mathbb{P}(R)R\,\mathrm{d}R = \rho\mu$. Here b defines the event R we consider associated with super-spreading. This is equivalent to size-inverse case ascertainment where super-spreading events are under-sampled. In this analogy b is the right under-reporting point i.e. we never observe events with $R > b$.

(iii) Sub-spreading control (size-biased reporting). This a novel intervention that we introduce here. It focuses on removing the sub-spreading events and is the converse of (ii). The gamma distribution of R is lower-truncated at some minimum value a and rescaled so that $\int_a^\infty \mathbb{P}(R)\,\mathrm{d}R = 1$ and $\int_a^\infty \mathbb{P}(R)R\,\mathrm{d}R = \rho\mu$. We use a to define sub-spreading events. This scheme is analogous to a size-biased case reporting strategy in which events producing few cases are under-sampled, with a as the left under-reporting point of the R distribution i.e. we never sample $R < a$. As (ii) and (iii) do not admit simple VM expressions and we investigate them through simulation.

As $a \to 0$ and $b \to \infty$ all three control (or reporting) measures (and the scale parameter of their respective gamma R distributions) converge. They also converge as $k \to \infty$ since the R distribution becomes degenerate at

$\rho\mu$ under these conditions and there is no heterogeneity. For simplicity, from this point we will usually refer to schemes (i)-(iii) via their control classification, switching to their reporting analogue only later when discussing results. We treat uniform control as a baseline since it ignores the specific form of the R distribution. While preferentially limiting super-spreading is sensible, and has been shown to have superior performance relative to uniform control [2], sub-spreading control has, to our knowledge, not been investigated. This is likely because it seems counter-intuitive to focus on events with low transmission potential.

However, we propose sub-spreading control based on the observation that a key source of the extra variability that over-dispersed distributions possess results from the increased probability of observing a zero sample i.e. zero inflation [32]. In an epidemic with heterogeneous transmission an excess of zero secondary cases would likely result from the sub-spreading event reproduction numbers. Here we investigate, for a fixed mean control effort ρ, whether the excess zeros or the extreme-tail events (i.e. the super-spreading ones) are more critical for shaping $\mathrm{VM}[I]$ (later we map $\mathrm{VM}[I]$ onto our elimination probabilities). We examine various k and compute the VM ratios for control strategies (i)–(iii) in Fig. 2 for increasing control effort i.e. decreasing ρ.

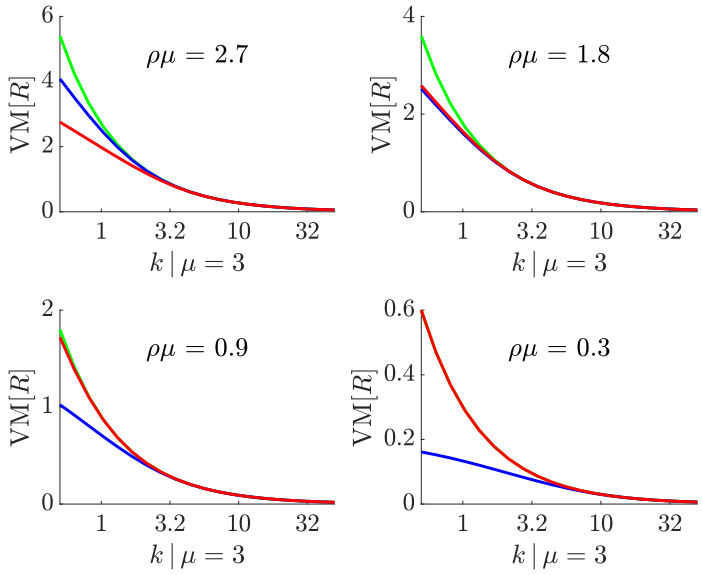

Fig. 2: **Control strategies and their VM ratios.** We compute the VM ratios of the R distributions resulting from uniform (green), super-spreading (red) and sub-spreading (blue) control measures for various k with $a = 1/10$ and $b = 10$. Before control $\mathbb{E}[R] = \mu = 3$ and after control $\mathbb{E}[R] = \rho\mu$ with smaller ρ indicating increased average control effort. We find that when $\rho\mu$ is large limiting super-spreading is most effective in reducing $\mathrm{VM}[R]$. However, as $\rho\mu$ decreases interventions targeting sub-spreading becomes the most effective.

Intriguingly, we find that when the controlled mean R i.e. $\rho\mu$ is large, super-spreading control is most effective in reducing the VM ratio of the reproduction numbers. However, as the epidemic is better controlled and so $\rho\mu$ falls, sub-spreading controls becomes the best intervention with respect to the VM ratio. This would make sense if as the mean controlled reproduction number falls the tail (super-spreading) events become increasingly improbable,

meaning that most of the variability derives from the sub-spreading events. We confirm this notion in Fig. 3, which examines the cumulative distribution function $F(R)$ at the extreme values of Fig. 2.

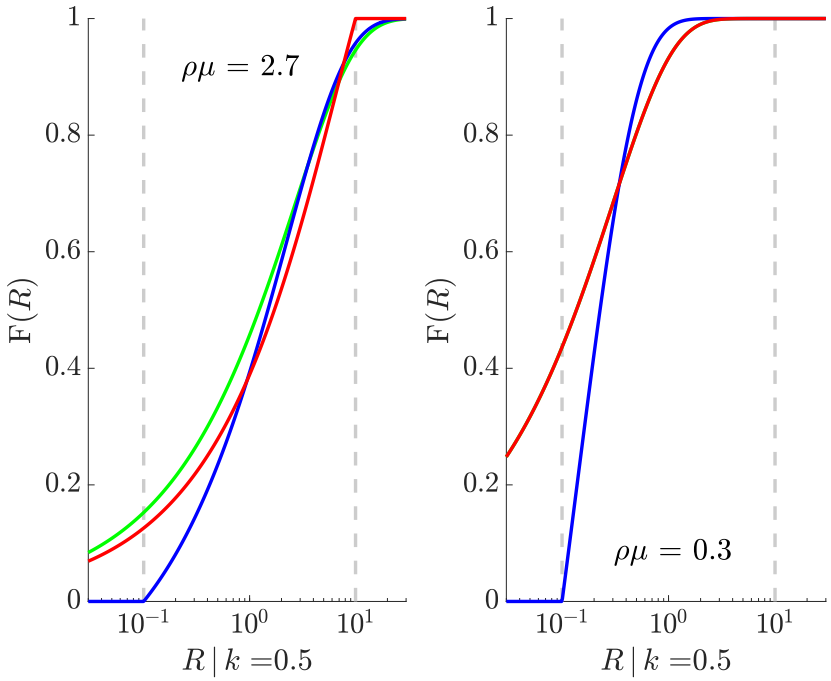

Fig. 3: **Control strategies shape R distributions.** We plot the cumulative distribution function $F(R)$ for uniform (green), super-spreading (red) and sub-spreading (blue) control measures for a renewal model with $k = 0.5$ at large and small $\rho\mu$ from Fig. 2. We find that super-spreading control is best (in terms of VM ratios) at large $\rho\mu$ because there is a notable probability of super-spreading events which becomes truncated under this measure i.e. it rises to 1 first. However, when $\rho\mu$ is small this becomes vastly less important (the super-spreading and uniform controls converge) and the sub-spreading control rises more quickly.

Here we see that, at large $\rho\mu$, limiting super-spreading forces $F(R)$ towards 1 at the fastest rate i.e. the clipping of super-spreading events closes the effective support of the R distribution earlier than the other measures. In this regime the tail events of the distribution are important. Elimination is not practically possible at such a large reproduction number, but this provides an interesting perspective for comparing to the low $\rho\mu$ regime. There, because the tail events are strongly improbable, super-spreading and uniform control converge in behaviour. Interestingly, sub-spreading control forces $F(R)$ more quickly to 1 than other measures, by truncating the small reproduction numbers. This may explain its performance in Fig. 2.

This assessment of the VM ratios and the effective reproduction number support (i.e. how quickly $F(R)$ converges to 1) does not invalidate the results of [2], where a variant of super-spreading control increased VM ratios. We will resolve these apparent contradictions in a later section, highlighting the difference between renewal models and the branching processes used in [2]. We next investigate whether these reductions in $\mathrm{VM}[R]$ and $\mathrm{VM}[I]$ actually improve our ability to constrain z_s and the epidemic lifetime and hence to achieve reliable end-of-epidemic declaration times.

Reducing variation among elimination times

We examine two data-justified measures of the end of an epidemic: the mean elimination time, $\bar{t}_\alpha$ and the maximum elimination time $t_{\alpha,\max}$. Both have previously been used for assessing end-of-epidemic declarations [13], [15]. As is convention, we focus on 95% confidence and set $\alpha = 95$. Times $\bar{t}_\alpha$ and $t_{\alpha,\max}$ are obtained by finding when the mean and minimum of the z_s curves i.e. $\bar{z}_s$ and $z_{s,\min}$, generated from possible future epidemic trajectories, first cross 0.95, respectively. We generate possible z_s by drawing samples from the distributions of R^{∞}_{s+1} and then computing Eq. (6). All z_s curves are conditioned on some fixed past incidence I^s_1.

For our development here we set I^s_1 to the daily new infections observed over the MERS-CoV epidemic in South Korea in 2015, which was investigated in [29]. We then project zero case-days forward in time as in [15] but for various hypothetical k values and assess the elimination statistics for each of these scenarios and under the control measures (i)–(iii) from above. In the appendix we perform complementary analyses using the incidence of the SARS 2003 epidemic in Hong Kong from [22] and simulated EVD data from [33], obtaining consistent results.

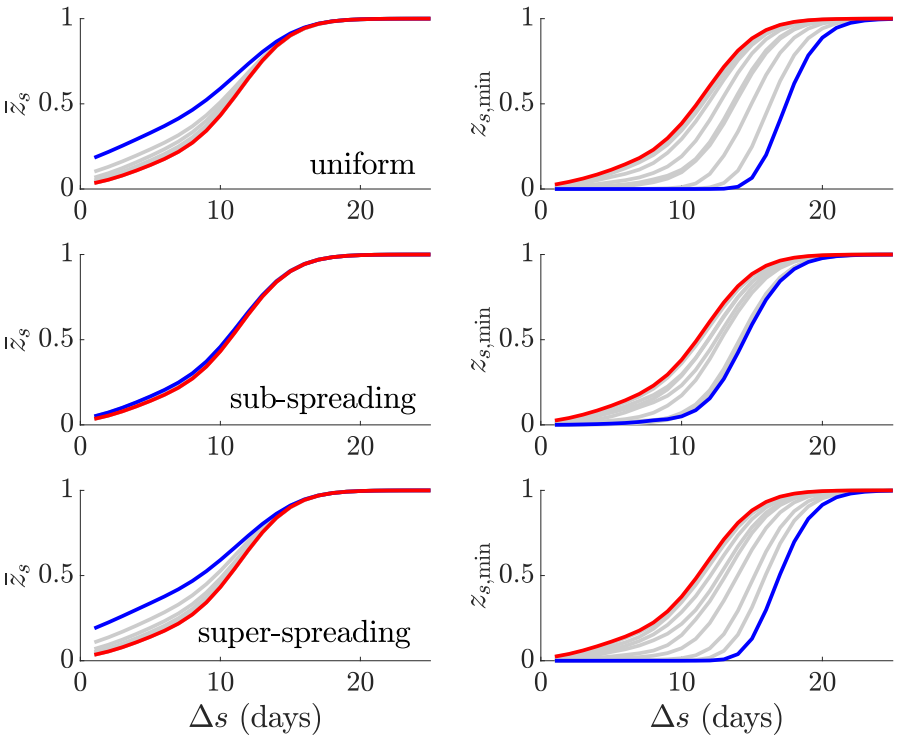

Fig. 4: **Mean and worst case elimination probabilities.** We compute the mean ($\bar{z}_s$) and worst case ($z_{s,\min}$) elimination probability curves conditioning on incidence data from the MERS-CoV epidemic in South Korea in 2015 as in [29]. These curves are for various k rising from 0.1 (blue) to 2 (red) under a mean controlled reproduction number of $\rho\mu = 0.5$. We find that increasing heterogeneity (smaller k) leads to the earliest mean elimination time (the $\bar{z}_s$ is largest) and the latest maximum elimination time (the $z_{s,\min}$ is smallest). These statistics are obtained over 2000 simulated future epidemic trajectories for each control scenario listed.

In Fig. 4 we present a range of mean and worst case z_s curves for various k with blue indicating the smallest k and red the largest. All times are given relative to the last observed non-zero case day. For all control measures we

notice an interesting trend. First, and as is expected from Eq. (8), we see that k ranks the mean z_s curves, with more heterogeneity (smaller k) leading to a larger $\bar{z}_s$. This behaviour is also consistent with [30], [2]. Second, we find that this ranking is inverted when we consider the worst case $z_{s,\min}$ instead i.e. epidemics with more heterogeneous transmission have larger worst case extinction or fade-out times.

This result is striking. It means that control actions which increase heterogeneity, and have been proposed as the most effective type of interventions [2], may lead to a false sense of confidence in the end of the epidemic. This follows from their larger worst case elimination times. Heterogeneity has the contrasting effect of contracting the mean lifetime of the epidemic but prolonging its maximum lifetime (counted in zero-case days). Since authorities would want a single decision time for safely declaring an epidemic to be over [16] and relaxing interventions (e.g. to re-open trade or travel), this creates a practical, potentially costly and unexpected complication.

However, we observe that non-uniform control measures can help alleviate this problem. The sub-spreading control scheme considerably shrinks the variation among z_s curves. This proceeds from the results of the previous section, where we found that, at small mean reproduction numbers (which is realistic when we are near the tail or end of an epidemic), limiting sub-spreading significantly reduces $\mathrm{VM}[R_s]$ and hence $\mathrm{VM}[I_s]$ (see Eq. (4)). We assess this in more detail by examining all the z_s curves, which led to the mean and minima in Fig. 4, and the distributions of the 95% declaration times, t_{95}, that result. We provide these in Fig. 5.

As k becomes larger the difference among all control measures expectedly shrinks. For epidemics with significant heterogeneity ($k < 1$) we observe that both uniform and super-spreading control result in large variations in z_s (top red and green curves in Fig. 5, which mostly overlay each other). This manifests in a notable spread of 95% declaration times (bottom red and green bars of Fig. 5, also overlaid). Sub-spreading control is, however, able to suppress much of this variation yielding more deterministic elimination probabilities and declaration times (blue curves and bars in Fig. 5) and thus minimising the possibility of early or late declarations.

We observe consistency in this trend for both EVD and SARS incidence curves in Fig. A.1 of the appendix. We further confirm the ability of sub-spreading control as a means of making heterogeneous incidence curves more deterministic in behaviour by simulating complete epidemic trajectories under each control measure in the bottom panel of Fig. 6. There we see that the VM ratios of the incidence are indeed minimised by sub-spreading control measures (the mean incidence from all control measures there is approximately the same).

At this point, we re-interpret these results from the perspective of case ascertainment. The similarity of super-spreading and uniform control mean that reporting strategies that sample possible spreading events with constant probability or inversely to their reproduction numbers have approximately the same effect on elimination waiting times. This likely follows from the vanishingly low probability of super-spreading. However, size-biased sampling, which is not only the analogue to sub-spreading control, but also likely to occur in practice [18], has an important

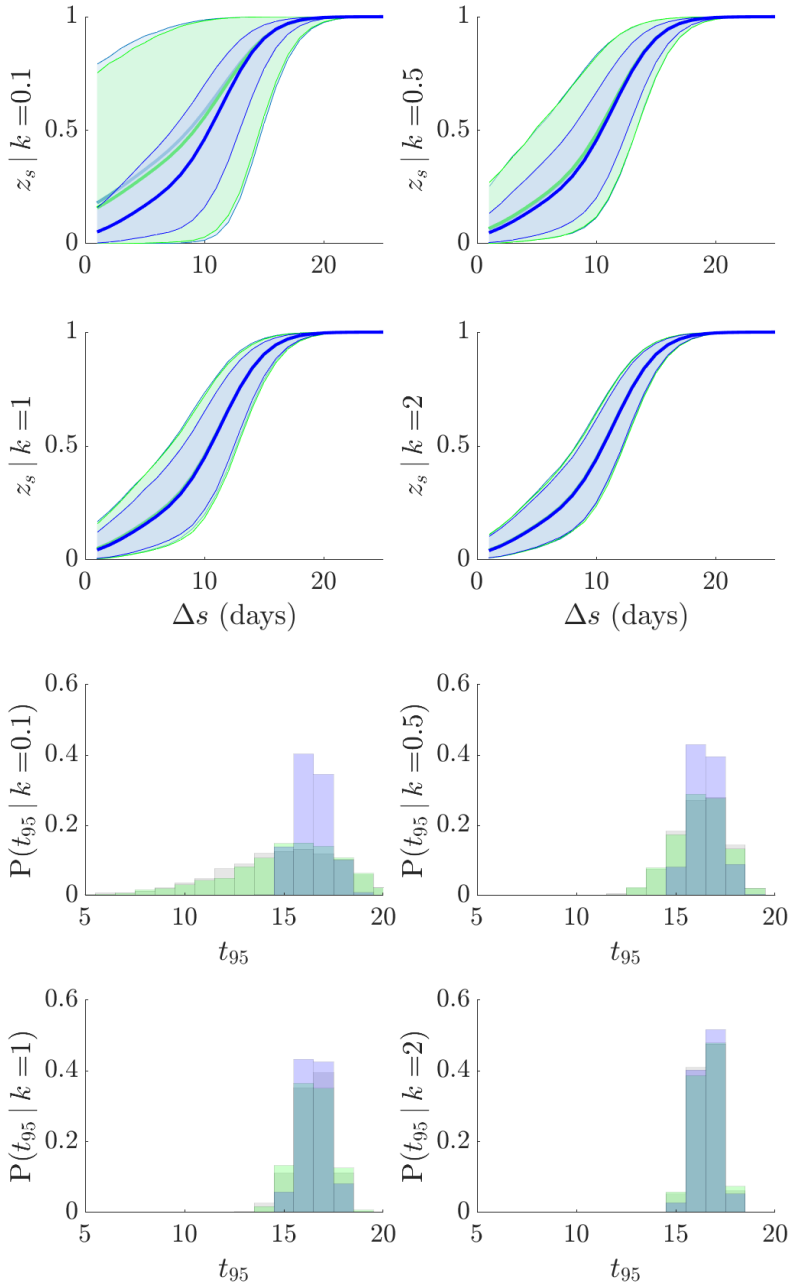

Fig. 5: **Elimination curves and declaration times for various control strategies.** We simulate 2000 future trajectories of zero case-days, conditioned on the incidence data of the MERS-CoV epidemic in South Korea in 2015 [29]. Each trajectory is formed by sampling from the R_{s+1}^{∞} distributions for uniform (green), super-spreading (red) and sub-spreading (blue) control measures. The top panel shows elimination probabilities (z_s) computed under these trajectories for various k and the bottom panel provides corresponding 95% declaration times (t_{95}). All times are relative to that of the last observed case and assume $\rho\mu = 0.5$ at every future time. We find that sub-spreading control is most effective at reducing the variability and hence in increasing the reliability of both z_s and t_{95}.

effect. Failing to observe sub-spreading events leads to a strongly overconfident view of elimination. The epidemic tail appears far less variable when those events are excluded, leading to risky end-of-epidemic declarations.

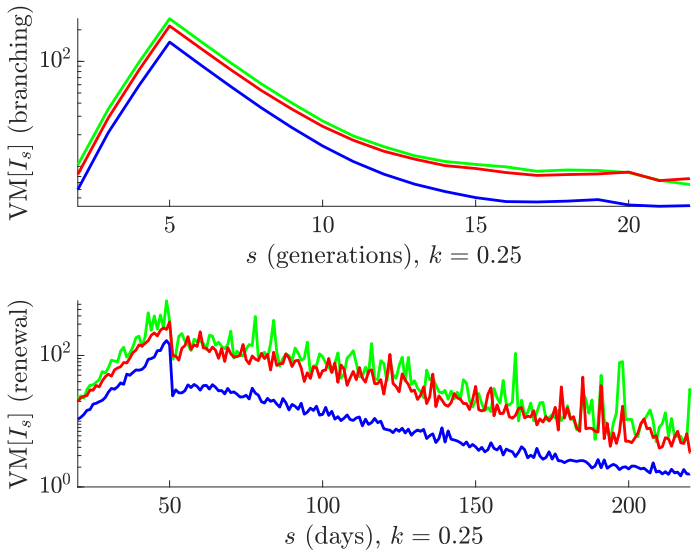

Fig. 6: **VM ratios for simulated epidemics.** Using either a branching process [2] (top) or renewal model [21] (bottom) we simulate 10^4 trajectories under a step-change in mean effective reproduction number, $\rho\mu$, from 2.5 to 0.6, with $k = 0.25$. At each time point we draw event reproduction numbers from the uniform (green), super-spreading (red) and sub-spreading (blue) distributions with these means. We consistently find, for waning epidemics, that sub-spreading minimises VM ratios. The mean incidence from all methods is approximately the same.

Last, we comment on differences between our renewal model approach and the Galton-Watson (GW) branching process used in [2]. Both models have significant dynamical disparities. Specifically, GW processes are truly only valid early in an epidemic, use a deterministic generation time and assume that the number of infected in the next generation only depend on those in the current generation. These characteristics make the GW process unsuitable for the analysis of elimination statistics [7], especially when the generation time distribution is known (a) to be non-degenerate, (b) to have long memory and (c) to determine the epidemic growth rate [20], [12].

Further, our problem of interest is ill-defined for GW processes since, under these models, we always obtain $t_{95} = 1$ (time is now in generations). This follows as $\mathbb{P}(\sum_{j=s+1}^{\infty} I_j = 0 \,|\, I_s = 0) = 1$. Our results therefore do not apply to GW processes and should not be expected to correspond with computations of the probability of extinction from [2]. Nevertheless, we make some key comparisons. In [2], increasing heterogeneity elevated the extinction probability of an epidemic by promoting early burn-out. This partially holds true for renewal models, but (a) mostly affects t_0 and (b) is not meaningful here as we condition on a significant epidemic having existed.

Since elimination measures only become an important consideration after t_0 and for epidemics of notable size [16], [29], burn-out due to overdispersion does not benefit our analysis. The extra dynamics beyond t_0, which do not exist for GW processes, form our problem of interest. However, the VM ratio behaviour of our control schemes (i)–(iii) is quite general and still works under GW models. In Fig. 6 we show that the GW process (top

panel) conserves the ordering of VM ratios from the renewal model (bottom panel), hence confirming the relative importance of sub-spreading control for achieving reliable elimination.

DISCUSSION

Understanding how transmission heterogeneity shapes the dynamics of diverse phases of an emerging epidemic is crucial if we are to safely and efficiently apply or relax interventions during those phases. While much research has focussed on super-spreading and the growing phases of epidemics, sub-spreading and the dynamics of waning infectious diseases have been understudied [9], [30]. In this paper we investigated both, by developing a general framework to measure how the variation or heterogeneity (embodied by k) around mean transmission at some time s (summarised by μ_s, the mean of R_s), engenders risk in the probability of epidemic elimination, z_s.

We extended the popular theory of transmission heterogeneity from [2] to precisely describe how variations (via VM ratios) in event reproduction numbers, R_s, manifest in fluctuations of the incidence, I_s (see Fig. 1). The resulting proportional relationship meant that the R_s distribution directly modulates the observed I_s curve [24]. We then generalised the formulae of [15], [3] to derive how these variations propagate uncertainty onto z_s. Since end-of-epidemic declarations depend on when z_s crosses a confidence threshold, this uncertainty makes adjudging when an epidemic is over complex, risky and potentially costly [13], [16].

Using our framework, we found that when epidemics are strongly heterogeneous, i.e. $k < 1$, the maximum and mean declaration times, relative to that of the last observed case, depend contrastingly on k (see Fig. 4). Although the average time to declaration decreases with k, supporting previous work linking heterogeneity to epidemic extinction [14], the concomitant increase in variability means that the safest declaration time actually increases and the risk of early or late declarations can be severely amplified. Variation originating from transmission heterogeneity is therefore not beneficial for achieving safe and reliable end-of-epidemic declarations.

Consequently, we investigated if targeted control can ameliorate this end-of-epidemic volatility. We considered three control schemes (for the same mean level of control ρ), which were non-selective or targeted either super- or sub-spreading (see Fig. 2 and Fig. 3). Intriguingly, we found that, because the controlled mean of the event reproduction numbers $\rho\mu_s$ is below 1 at pre-elimination settings, curbing super-spreading only marginally improved on non-selective approaches. However, sub-spreading appeared to be the main contributor to end-of-epidemic declaration risk, meaning limiting those events can significantly reduce that risk; forcing the epidemic tail to be more deterministic and increasing the reliability of resulting intervention relaxation policies.

This result, while new and surprising, is supported by current understanding. Overdispersed distributions have two defining characteristics: tall heads and fat tails [24], where head and tail refer to the left and right ends of our R_s distributions. Since $\rho\mu_s < 1$ near elimination and because we only start measuring time once a sequence of days with no cases appear (in accordance with official guidelines [16]) fat tail events are extremely unlikely to

affect epidemic dynamics. Consequently, it is the tall head i.e. the excess of events with $R_s \ll \rho\mu_s$, which are responsible for sustained overdispersion. Curbing sub-spreading suppresses this remaining source of risk.

We validated these theories on empirical MERS-CoV, SARS and simulated EVD epidemics (where we conditioned on past incidence curves and simulated forward) and observed significant reductions in end-of-epidemic declaration time spread (see Fig. 5 and Fig. A.1) when sub-spreading was controlled. The conditions under which these effects were prominent (when $k < 0.5$) are realistic for these and many other infectious diseases [2], [26]. An efficient strategy for controlling and then reliably eliminating an epidemic might therefore re-direct interventions from targeting super-spreading to sub-spreading transmission as infections stop growing and start waning.

Unfortunately, achieving the benefits of such strategies may be largely unrealistic. Predicting or simply identifying specific spreading event-types has proven difficult and hampered attempts at targeting super-spreading [8]. As sub-spreading frequently involves zero secondary cases, these events may be significantly harder to investigate. Intensive contact tracing and isolation schemes can help ameliorate some of these issues. Ongoing studies into associations among measurable traits, for example viral loads or symptom severity, and infectiousness or transmissibility, may also make identifying and curbing specific spreading events easier [34], [35]. However, even if these identification problems are overcome, policy resistance can nullify the benefits of targeted strategies [36].

Despite these points, our analysis of spreading events and our framework have several practical implications. First, we showed that the benefits of curbing super-spreading diminish with incidence. If targeted control is more costly than population-wide surveillance measures, this can support policy switches as the epidemic wanes. Second, we exposed the inherent risk of control measures that neglect sub-spreading. This knowledge can evidence more risk-averse approaches to end-of-epidemic declarations, in line with works such as [14]. Third, key formulae within our framework (Eq. (2) – Eq. (8) and Eq. (10)) are valid for arbitrary reproduction number distributions. Consequently, we can test the influence of other types of heterogeneity (e.g. demographic or spatial effects) on epidemic lifetimes and even incorporate empirical effective reproduction number distributions when available.

An important consequence of our analyses emerges from the mathematical analogue between control and case ascertainment [18]. Uniform, super-spreading and sub-spreading control are equivalent to constant, size-inverse and size-biased reporting. Size-biased reporting protocols are common in practice and generally include any sampling scheme that observes infectious events with likelihoods that grow with the number of secondary cases produced by that event [24]. We investigated a specific size-biased scheme, similar to that in [18], where sub-spreading events are unobserved. This leads (by analogy to our sub-spreading control results) to a drastic underestimation of the variance of possible elimination times exhibited by heterogeneous epidemics.

Previously it was shown that the mean end-of-epidemic declaration time is biased by any type of under-reporting [15]. As a result, practical surveillance schemes, which will almost surely feature some degree of size-biased

reporting, are likely to promote simultaneously biased and overconfident end-of-epidemic declarations. These effects emphasise why sustained high quality epidemic monitoring is crucial for guiding the relaxation of interventions or the reopening of economies during the waning phases of outbreaks. In the absence of such monitoring our results recommend a more cautious approach to designating the end of an outbreak.

While the results we have presented expose the complexities and biases in adjudging epidemic fade-out, there are some limitations to our analyses. We have assumed that the control effort or reporting rate of ρ can be realised with equal ease for both the uniform and selective strategies and that gamma or truncated gamma reproduction number distributions are sensible. These are necessary and standard assumptions for comparing strategies and sensibly summarising heterogeneity [2], [7] but their validity may depend on the disease being considered and the properties (e.g. contact networks) of the area under study. Scale may also matter. For example, across larger regions population-wide measures may more easily achieve a given ρ than more targeted ones.

Moreover, we have not accounted for how interventions (e.g. lockdowns or quarantines) may alter the characteristics of serial interval and offspring distributions [37] and hence the adequacy of the renewal models we have employed or the realistic effectiveness of the strategies we have examined. While lack of data precludes improvement here, we note that our framework for investigating epidemic fade-out (a) is sufficiently general to handle empirical (and non-gamma) reproduction number distributions if such knowledge becomes available and (b) applies to the waning phases of the epidemic, after most drastic changes would likely have already occurred.

Our framework provides a general toolkit for testing hypotheses about targeted controls or case ascertainment schemes and measuring their influence on end-of-epidemic declaration times. It can also be easily extended to include additional factors such as imported cases or to investigate other types of heterogeneity (e.g. age-based reproduction numbers) [15]. As the current COVID-19 pandemic underscores, much still remains unknown about the relative merits of elimination approaches, e.g. “zero COVID” strategies [11], and the influence of heterogeneity [4]. Improved understanding of epidemic dynamics can only aid preparedness and decision-making. We hope that our framework, which exposed unexpected consequences of understudied spreading events [19], can contribute towards this goal and help inform safe intervention relaxation and end-of-epidemic declaration strategies.

Funding

This work is jointly funded under grant reference MR/R015600/1 by the UK Medical Research Council (MRC) and the UK Department for International Development (DFID) under the MRC/DFID Concordat agreement and is also part of the EDCTP2 programme supported by the European Union.

References

[1] Nishiura H, Chowell G. The Effective Reproduction Number as a Prelude to Statistical Estimation of Time-Dependent Epidemic Trends. In: Mathematical and statistical estimation approaches in epidemiology. Springer; 2009. p. 103–21.

[2] Lloyd-Smith J, Schreiber S, Kopp P, et al. Superspreading and the effect of individual variation on disease emergence. Nature. 2005;438(17):355–9.

[3] Parag K, Cowling B, Donnelly C. Deciphering early-warning signals of the elimination and resurgence potential of SARS-CoV-2 from limited data at multiple scales. medRxiv. 2021;2020.11.23.20236968.

[4] Susswein Z, Bansal S. Characterizing superspreading of SARS-CoV-2 : from mechanism to measurement. medRxiv. 2020;2020.12.08.20246082.

[5] De Serres G, Gay N, Farrington P. Epidemiology of Transmissible Diseases after Elimination. Am J Epidemiol. 2000;151(11).

[6] Woolhouse M, Dye C, Etard J, et al. Heterogeneities in the transmission of infectious agents: Implications for the design of control programs. PNAS. 1997;94:338–42.

[7] Garske T, Rhodes C. The effect of superspreading on epidemic outbreak size distributions. J Theor Biol. 2008;253:228–37.

[8] Frieden T, Lee C. Identifying and Interrupting Superspreading Events— Implications for Control of Severe Acute Respiratory Syndrome Coronavirus 2. Emerg Infect Dis. 2020;26(6):1059–66.

[9] Britton T, House T, Loyd A, et al. Five challenges for stochastic epidemic models involving global transmission. Epidemics. 2015;10:54–7.

[10] Parag K. Improved estimation of time-varying reproduction numbers at low case incidence and between epidemic waves. medRxiv. 2020;2020.09.14.20194589.

[11] Baker M, Wilson N, Blakely T. Elimination may be the optimal response strategy for COVID-19 and other emerging pandemic diseases. BMJ. 2020;371:m4907.

[12] Lee H, Nishiura H. Sexual transmission and the probability of an end of the Ebola virus disease epidemic. J Theor Biol. 2019;471:1–12.

[13] Nishiura H, Miyamatsu Y, Mizumoto K. Objective determination of end of MERS outbreak, South Korea. Emerg Infect Dis. 2016;22:146–8.

[14] Djaafara B, Imai N, Hamblion E, et al. A quantitative framework to define the end of an outbreak: application to Ebola Virus Disease. Am J Epidemiol. 2020;p. kwaa212.

[15] Parag K, Donnelly C, Jha R, et al. An exact method for quantifying the reliability of end-of-epidemic declarations in real time. PLOS Comput Biol. 2020;16(11):e1008478.

[16] WHO. WHO recommended criteria for declaring the end of the Ebola virus disease outbreak; 2020. Available from: https://www.who.int/who-documents-detail/who-recommended-criteria-for-declaring-the-end-of-the-ebola-virus-disease-outbreak.

[17] Lloyd-Smith J, Cross P, Briggs C, et al. Should we expect population thresholds for wildlife disease? Trends Ecol Evol. 2005;20(9):511–19.

[18] Lloyd-Smith J. Maximum Likelihood Estimation of the Negative Binomial Dispersion Parameter for Highly Overdispersed Data, with Applications to Infectious Diseases. PLOS One. 2007;2:e180.

[19] Lawyer G. Understanding the influence of all nodes in a network. Sci Rep. 2015;5(8665).

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

[21] Fraser C, Cummings D, Klinkenberg D, et al. Influenza Transmission in Households During the 1918 Pandemic. Am J Epidemiol. 2011;174(5):505–14.

[22] Cori A, Ferguson N, Fraser C, et al. A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics. Am J Epidemiol. 2013;178(9):1505–12.

[23] Parag K, Donnelly C. Adaptive Estimation for Epidemic Renewal and Phylogenetic Skyline Models. Syst Biol. 2020;69(6):1163–79.

[24] Karlis D, Xekalaki E. Mixed Poisson Distributions. Intern Statist Rev. 2005;73(1):35–58.

[25] Churcher T, Cohen J, Ntshalintshali N, et al. Measuring the path toward malaria elimination. Science. 2014;344(6189):1230–32.

[26] Lau M, Dalziel B, Funk S, et al. Spatial and temporal dynamics of superspreading events in the 2014-2015 West Africa Ebola epidemic. PNAS. 2017;114(9):2337–42.

[27] Parag K, Donnelly C. Using information theory to optimise epidemic models for real-time prediction and estimation. PLOS Comput Biol. 2020;16(7):e1007990.

[28] Azmon A, Faes C, Hens N. On the estimation of the reproduction number based on misreported epidemic data. Stats Med. 2014;33:1176–92.

[29] Nishiura H. Methods to determine the end of an infectious disease epidemic: a short review. In: Chowell G, Hyman J, editors. Mathematical and statistical modeling for emerging and re-emerging infectious diseases; 2016. p. 291–301.

[30] Yan P, Chowell G. Quantitative Methods for Investigating Infectious Disease Outbreaks. vol. 70 of Texts in Applied Mathematics. Cham, Switzerland: Springer; 2019.

[31] Wasserman L. All of Statistics: A Concise Course in Statistical Inference,. Berlin, Germany: Springer-Verlag; 2003.

[32] Xekalaki E. On the distribution theory of over-dispersion. J Statistic Distrib Applic. 2014;1(19).

[33] Jombart T, Frost S, Nouvellet P, et al. outbreaks: A Collection of Disease Outbreak Data; 2019 [cited 31 Jul 2020]. Available from: https://github.com/reconhub/outbreaks.

[34] Edwards A, Ausiello D, Salzman J, et al. Exhaled aerosol increases with COVID-19 infection, age, and obesity. PNAS. 2021;118(8):e2021830118.

[35] Goyal A, Reeves D, Cardozo-Ojeda E, et al. Viral load and contact heterogeneity predict SARS-CoV-2 transmission and super-spreading events. eLife. 2021;10:e63537.

[36] Wang A, Bauch C, Bhattacharyya S, et al. Statistical physics of vaccination. Phys Rep. 2016;664:1–113.

[37] Ali S, Wang L, Lau E, et al. Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions. Science. 2020;369(6507):1106–9.

APPENDIX

A. Declaration time histograms for SARS and EVD

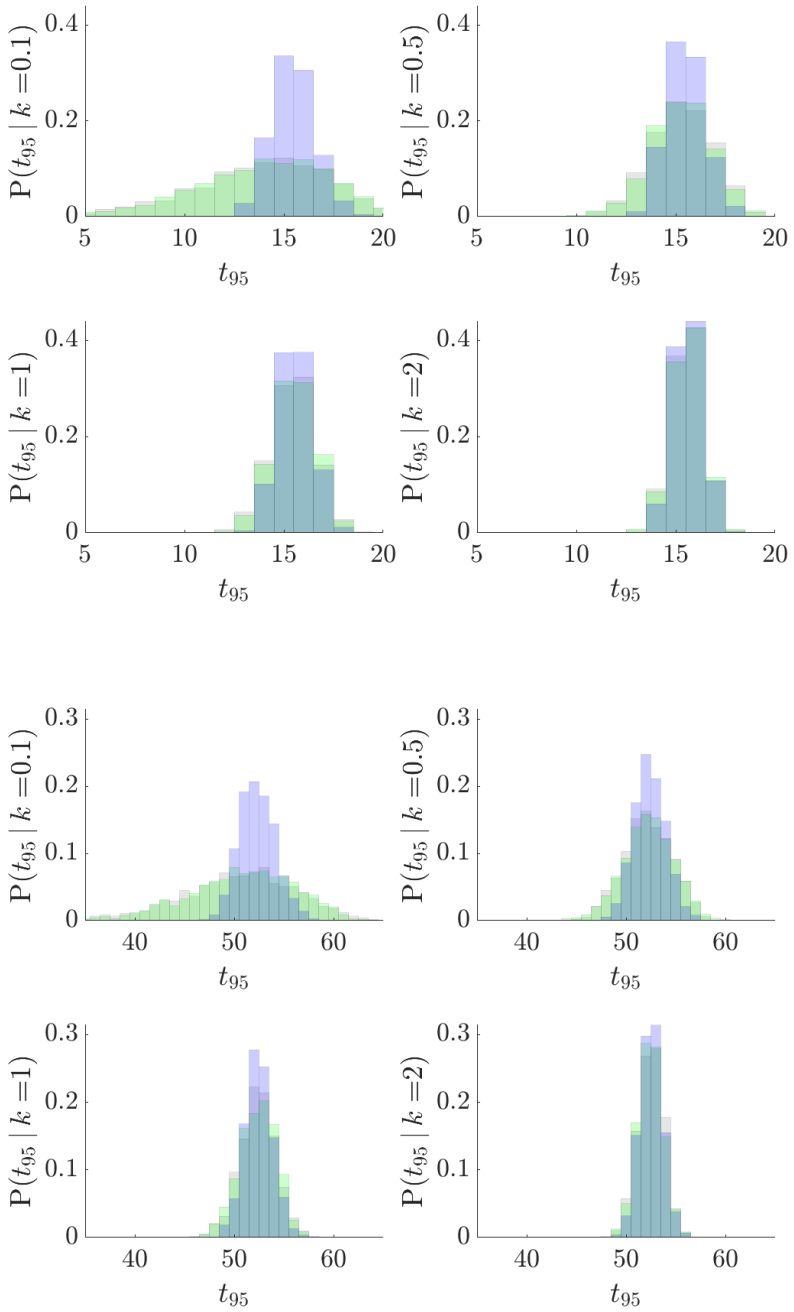

Fig. A.1: **Declaration times for different conditioned incidence curves.** We simulate 2000 future trajectories of zero case-days, conditioned on the incidence data of the SARS epidemic in Hong Kong in 2003 (top) [22] and the simulated incidence from for EVD (bottom) [33]. We generate trajectories by sampling from the R^{∞}_{s+1} distributions for uniform (green), super-spreading (red) and sub-spreading (blue) control measures and plot corresponding 95% declaration times (t_{95}). The uniform and super-spreading results largely overlap. All times are relative to that of the last observed case and assume $\rho\mu = 0.5$. As in the main text, we see that sub-spreading control is most effective at reducing the variability among possible t_{95} and hence increases the reliability of any end-of-epidemic declaration.