Host behaviour driven by awareness of infection risk amplifies the chance of superspreading events

Kris V Parag[1,2,*] and Robin N Thompson[3-5]

[1]MRC Centre for Global Infectious Disease Analysis, Imperial College London, London, UK.
[2]NIHR HPRU in Behavioural Science and Evaluation, University of Bristol, Bristol, UK.
[3]Mathematics Institute, University of Warwick, Coventry, UK.
[4]Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research, University of Warwick, Coventry, UK.
[5]Mathematical Institute, University of Oxford, Oxford, UK.

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

Abstract

We demonstrate that heterogeneity within host populations in the perceived risks associated with infection can amplify the chance that superspreading occurs during the crucial, early stages of an epidemic. Under our behavioural model, individuals who are less concerned about the risks of infection are more likely to be infected and to attend larger sized events. For directly communicable diseases, such as COVID-19, this leads to infections being introduced at rates above the population prevalence to events that are most conducive to superspreading, for fixed overall transmission levels. We develop a computational framework for evaluating within-event risks and derive a small-scale reproduction number to measure how introductions and transmission heterogeneities determine the number of infections resulting from an event of given duration. This generalises earlier frameworks and both quantifies and clarifies how event-scale dynamics depend on population-level characteristics. As event duration and size increase, our reproduction number converges to the widely used basic reproduction number, R_0. We show that even moderate levels of heterogeneity in the perceived risks associated with infection can substantially increase the risk that disproportionately large numbers of infections are generated at larger events, as compared to assuming homogeneous perceptions of risk across the host population. Behavioural dynamics, which remain understudied in the context of epidemiological modelling, are essential to consider when assessing the risk posed by an invading pathogen in the early stages of an infectious disease epidemic.

Keywords: infectious diseases; behavioural models; risk awareness; reproduction numbers; importations; superspreading events.

Introduction

The prediction and prevention of superspreading events, which are characterised by primary infected individuals generating disproportionately large numbers of secondary infections [1], is a central challenge in infectious disease epidemiology. For acute, directly communicable diseases such as COVID-19 and Ebola virus disease, superspreading is a critical transmission driver that leads to less frequent but more explosive outbreaks than we might expect under epidemiological models that do not account for superspreading [2]. During early or emergent stages of a potential epidemic, when there are limited immunity levels in the host population and transmission dynamics are inherently stochastic, superspreading events have been found responsible for spurring both the initial growth and eventual persistence of epidemics and for limiting the effectiveness of non-pharmaceutical interventions [1,3–5].

Consequently, identifying the main factors that underly the risk of superspreading is crucial for effective disease management [4]. Many of these factors are known, with heterogeneities in (i) host characteristics (e.g., susceptibility, infectiousness and contact patterns), (ii) pathogen biology (e.g., transmission routes and viral loads), (iii) environmental effects (e.g., ventilation and gathering size) and (iv) host behaviours (e.g., social customs and intervention adherence) all contributing to the risk of superspreading [3,4,6–8]. However, incorporating these factors in parsimonious modelling frameworks can be difficult because the mechanisms linking them to superspreading are still not fully understood. This is particularly the case for factors (iii) and (iv), with recurrent calls emphasising the need to study the relationships among behavioural, environmental and epidemiological dynamics [9–11]. Here we explore how a key feature of host behaviour relates to superspreading.

We consider how heterogeneity in perceptions of the risk associated with infection across a host population impact heterogeneity in the transmission of new infections. Risk awareness is a documented phenomenon in which individuals adopt self-protective behaviours in response to perceived risks of acquiring infection and its associated dangers. The implications of risk awareness can include precautionary reductions in mobility that improve intervention efficacy, changes to the amplitude and duration of outbreaks and, in some instances, complete suppression of onward transmission [11–15]. However, the interplay between risk awareness and superspreading risk has been unstudied, with previous research focussing on pathogen and host, instead of behavioural, characteristics. We study this interplay under a simple but plausible hypothesis – that risk-aware hosts are more likely

to avoid events of larger size, due to the perception of a heightened infection risk at those events [15].

This awareness mechanism implies, for a fixed overall transmission level, that larger events (e.g., concerts or sport matches) are more likely to be attended by individuals who are less risk averse. During the early stages of an epidemic, when there is limited infection-induced immunity in the host population, these individuals are also more likely to already be infected. We posit that this coupling between behaviour and environment (i.e., risk and event size) may amplify the chances of superspreading occurring at larger events, which have the capacity to support excessive numbers of infections. To test this hypothesis, we develop a framework to model the number of infections y generated at an event of size n, given that x initially infected individuals attend that event. This yields a small-scale reproduction number that extends recent approaches [16–18] to understanding within-event transmission in three directions.

First, we explicitly model the transmission-reducing effects of both finite numbers of imports (x) and susceptible individuals ($n - x$). As event size and duration grow, these finite size effects become less important and our small-scale reproduction number converges to the widely used basic reproduction number, R_0. Second, we embed heterogeneity in transmission at the event within our small-scale reproduction number by allowing for variations in secondary infections using the dispersion parameter, k. This essentially provides a within-event version of the seminal model of superspreading applied in [1,5,19] and includes the broad influence of factors (i)-(ii) described above. Third, we account for how x changes (stochastically) with n. This depends on the prevalence of infection in the wider population, the extent to which larger events are likely to feature disproportionately more importations due to risk awareness (which we denote by $\epsilon(n)$), and considers factors (iii)-(iv).

The functional dependence of ϵ on n serves as a parsimonious model of risk awareness and allows us to assess how host behaviour shapes the risk of superspreading. We explore our central hypothesis by comparing the relative and combined impact of $\epsilon(n)$ and k on the tail probability of observing a disproportionately large value of y at an event. We demonstrate, for a fixed overall transmission or import rate, that risk awareness can substantially amplify the chance of superspreading at a large event, compared to the scenario in which all individuals attending the large event are assumed to have a similar perceptions of infection risk. This pattern holds regardless of k and, in some instances, we find the increase in

superspreading risk due to risk-aware behaviour outweighs that from inherent transmission heterogeneity.

Methods

Event reproduction numbers including import risk and transmission heterogeneity

We develop a framework for quantifying the risk of acquiring infection at an event (e.g., a party or concert), parameterised by a small-scale (within-event) reproduction number and the number of infections likely to occur at that event. An event is defined as a short-term grouping of n people and we allow $0 \leq x \leq n$ of the individuals attending that event to be infectious. We say there are x introductions or imported infections at this event and $n - x$ susceptible hosts initially. We assume no prior immunity in the population and let $\mathbf{P}(y|n)$ be the probability of $0 \leq y \leq n - x$ new infections being generated at that event. This depends on $\mathbf{P}(y|x,n)$, the probability of y infections occurring given x infectious individuals initially (for events of size n) and the prior probability of those x imports occurring, $\mathbf{P}(x|n)$, as on the left of **Eq. (1)**.

$$\mathbf{P}(y|n) = \sum_{x=0}^{n} \mathbf{P}(y|x,n)\mathbf{P}(x|n)\,, \qquad R(x) \stackrel{\text{def}}{=} \frac{1}{x}\sum_{y=0}^{n-x} y\mathbf{P}(y|x,n). \qquad (1)$$

We define a reproduction number for this event, $R(x) \stackrel{\text{def}}{=} x^{-1}\mathbf{E}[y|x,n]$, as on the right of **Eq. (1)** by expanding the expectation $\mathbf{E}[.]$. Here $R(x)$ measures the expected number of new infections generated by each import when there are x imports in total. The expected infections generated by all x imports attending the event is $\mathbf{E}[y|x,n]$.

A central idea of this study is the importance of $\mathbf{P}(x|n)$ and its dependence on event size n. Earlier work assumed that $\mathbf{P}(x|n)$ depends on the prevalence of infection in the population [18] but neglected how heterogeneities in human behaviour may affect the number of imported cases at a given event of size n. To our knowledge, alternative models for $\mathbf{P}(x|n)$ informed by human behaviour and the impact of this behaviour on infections generated at the event have not been explored. Our event or small-scale reproduction number also generalises prior research by including the effects of finite x and n. The original event reproduction number [17] considers a single imported case and relates to our $R(1)$, which

we later show upper bounds $R(x)$. By extending the definition of the event reproduction number, we model the influence of $\mathbf{P}(x|n)$ on the risk of acquiring infection at any event directly. As we develop below, $R(x)$ also embeds heterogeneity in transmission from host characteristics or pathogen biology [1], and is explicitly related to the population-level basic reproduction number, R_0 [20].

To convert **Eq. (1)** into a computable form we draw on characteristics of both the event and disease. We denote the (frequency-dependent) transmission rate as β and the expected duration of an individual infection as d, so that $R_0 = \beta d$. We then consider an event that lasts for time τ, which is assumed to be substantially shorter than d, so that infectiousness outlasts the event and at most one generation of infection is possible at the event. If there is one infected individual at the start of the event, then the probability that any given susceptible host gets infected is the secondary attack rate (SAR), $p = 1 - e^{-\frac{\beta\tau}{n}}$, assuming exponentially distributed times to infection. When there are s susceptible individuals, then $\mathbf{E}[y|1, n] = sp$. While this assumes that all susceptible individuals are exposed to all infectious ones, realistic contact networks can be modelled as in [20] by modifying s to be subset of susceptible hosts likely to be exposed to each infection (this connects network and random mixing models).

We generalise this approach in three main directions. First, we model the effect of variability in the number of imported infections. If there are x imports to the event, then the SAR becomes $p = 1 - e^{-\left(\frac{\tau}{d}\right)\left(\frac{x}{n}\right)R_0}$ with $\beta = \frac{R_0}{d}$. Since there are initially $n - x$ susceptible individuals, the expected number of infections generated at the event is $\mathbf{E}[y|x, n] = (n - x)p$. The leads to the event reproduction number $R(x)$ in **Eq. (2)** below. Note that $R(0) = R(n) = 0$.

$$R(x) = \frac{n-x}{x}\left(1 - e^{-\left(\frac{\tau}{d}\right)\left(\frac{x}{n}\right)R_0}\right) \approx \left(1 - \frac{x}{n}\right)\left(\frac{\tau}{d}\right)R_0 + \mathbf{O}\left(\frac{x}{n}\right), \quad (2)$$

This formulation has interesting limiting behaviour at various x. As the number of susceptibles grows in excess of imports i.e., $\frac{n}{x}$ increases, the $\mathbf{O}\left(\frac{x}{n}\right)$ terms in the Taylor series approximation of $R(x)$ in **Eq. (2)** become negligible. As n becomes large, we find $R(x) \to \left(\frac{\tau}{d}\right)R_0$. If the event lasts for the duration of infectiousness ($\tau = d$), then $R(x) \to R_0$. This convergence makes sense since our formulation is equivalent to a finite or small-scale version of random mixing.

Second, we expand this model to include realistic heterogeneity due to host characteristics or pathogen biology. It is unlikely that every infectious individual has the same transmissibility and we expect substantial variations in the numbers of infections generated by each infected individual [1,22]. We therefore allow R_0 to have some distribution from which every import is randomly sampled and let R_0^j indicate the sample for the jth of the x imports at the event. Accordingly, the heterogeneous version of $R(x)$ and its expected infections in **Eq. (3)**.

$$R(x) = \frac{n-x}{x}\left(1 - e^{-\left(\frac{\tau}{d}\right)\left(\frac{1}{n}\right)\sum_{j=1}^{x} R_0^j}\right), \qquad \mathbf{E}[y|x,n] = xR(x). \quad (3)$$

We compute the mean of $R(x)$ across the transmission heterogeneity for x infectious imports in **Eq. (4)**, with $\mathbf{E}_{\text{het}}$ indicating expectation about the distributions of the R_0^j and $\mathbf{M}_b(a)$ as the moment generating function about b evaluated at a. As the transmissibility of the x imported infections are independently sampled, $\mathbf{M}_{\sum_{j=1}^{x} R_0^j}(a) = \prod_{j=1}^{x} \mathbf{M}_{R_0^j}(a)$. This reduces to $\mathbf{M}_{R_0}(a)^x$ if samples are identically distributed. The expected number of infections under this model as a function of x is $\mathbf{E}_{\text{het}}\left[\mathbf{E}[y|x,n]\right] = x\mathbf{E}_{\text{het}}[R(x)]$ with $\mathbf{E}_{\text{het}}[R(x)]$ from **Eq. (4)**.

$$\mathbf{E}_{\text{het}}[R(x)] = \frac{n-x}{x}\left(1 - \mathbf{M}_{\sum_{j=1}^{x} R_0^j}\left(\left(\frac{\tau}{d}\right)\left(\frac{1}{n}\right)\right)\right). \quad (4)$$

Following [22], we evaluate the variance around $R(x)$ as $\mathbf{V}_{\text{het}}[R(x)]$ with $a_n = \left(\frac{\tau}{d}\right)\left(\frac{1}{n}\right)$ in **Eq. (5)**. This involves expanding $\mathbf{E}_{\text{het}}[R(x)^2] - \mathbf{E}_{\text{het}}[R(x)]^2$ and applying properties of $\mathbf{M}_b(a)$. The variance on the expected number of infections is $\mathbf{V}_{\text{het}}\left[\mathbf{E}[y|x,n]\right] = x^2\mathbf{V}_{\text{het}}[R(x)]$. All of these statistics remain valid for any model of transmission heterogeneity but we will derive analytic relations under the most widely used model of [1] in the subsequent section.

$$\mathbf{V}_{\text{het}}[R(x)] = \left(\frac{n-x}{x}\right)^2\left(\mathbf{M}_{\sum_{j=1}^{x} R_0^j}(-2a_n) - \mathbf{M}_{\sum_{j=1}^{x} R_0^j}(-a_n)^2\right). \quad (5)$$

Third, we examine how the likelihood of finding that x infectious individuals have attended the event impacts the above quantities. This involves evaluating how $\mathbf{P}(x|n)$ weights the formulae in **Eqs. (3)-(5)**. This weighting may be random, depend on behavioural dynamics as we posit in the next section (i.e., risk awareness) or be assigned using other rules. We

propose that a more informative measure of the risk of acquiring infection from an event of size n and duration τ is the import-weighted event reproduction number R_{imp} and its associated expected number of infections $\mathbf{E}_{\text{imp}}[y|n]$. We introduce these measures in **Eq. (6)** below.

$$R_{\text{imp}} = \sum_{x=0}^{n} R(x)\mathbf{P}(x|n), \qquad \mathbf{E}_{\text{imp}}[y|n] = \sum_{x=0}^{n} xR(x)\mathbf{P}(x|n). \quad (6)$$

The quantities in **Eq. (6)** are random variables that vary with the samples from the distribution controlling transmission heterogeneity. Accordingly, we can evaluate statistics $\mathbf{E}_{\text{het}}\left[R_{\text{imp}}\right]$ and $\mathbf{E}_{\text{het}}\left[\mathbf{E}_{\text{imp}}[y|n]\right]$ by summing and weighting $\mathbf{E}_{\text{het}}[R(x)]$ and $x\mathbf{E}_{\text{het}}[R(x)]$ respectively using $\mathbf{P}(x|n)$. Similarly, we get $\mathbf{V}_{\text{het}}\left[R_{\text{imp}}\right]$ and $\mathbf{V}_{\text{het}}\left[\mathbf{E}_{\text{imp}}[y|n]\right]$ from $\mathbf{V}_{\text{het}}[R(x)]$ and $x^2\mathbf{V}_{\text{het}}[R(x)]$ but with squared weightings $\mathbf{P}(x|n)^2$. These all follow from the properties of expectations and variances applied to a linear weighted sum with independent terms.

Statistical models for event reproduction numbers and importation dynamics

Having outlined measures of infection risk in **Eq. (6)**, we build into our framework some likely approaches for integrating transmission heterogeneities and import dynamics (including when those imported infections are risk-sensitive). This allows us to parsimoniously model traditional and behavioural drivers of superspreading, respectively. Additionally, we incorporate process stochasticity and provide a full Bayesian formulation for our framework. We start by including the seminal heterogeneity model of [1], which describes individual variations in transmissibility via a gamma distribution with dispersion k and mean R_0. We write this as $R_0^j \sim \mathbf{Gam}\left(k, \frac{R_0}{k}\right)$ with **Gam** as a shape-scale parameterised gamma distribution. Using scaling and summing properties of these gamma variables, we hence obtain $\sum_{j=1}^{x} R_0^j \sim \mathbf{Gam}\left(kx, \frac{R_0}{k}\right)$.

This assumes that samples of the basic reproduction number of individuals are independent and identically distributed and lets us analytically evaluate the moment generating function as $\mathbf{M}_{\sum_{j=1}^{x} R_0^j}(-a_n) = \left(1 + \frac{R_0 a_n}{k}\right)^{-kx}$. We substitute this into **Eqs. (3)-(5)** to precisely compute the mean and variance of the infections and event reproduction number conditional on the x introductions as detailed above. We can relax the assumption that the R_0^j are independent and identically distributed by instead sampling them from different distributions or by

applying alternative dispersion models [22]. The heterogeneous R_0^j constitute a major (and traditionally modelled) source of stochasticity that underpins the risk metrics we propose in **Eq. (6)**.

A less studied source of stochasticity is variability in the probability that infectious individuals attend the event. Previous work [18] has treated this deterministically, setting the probability or rate that an attending individual is infected as equal to the population prevalence ρ (or ρ adjusted by an exposure factor, if it is known that the event draws individuals who are less or more likely to be infected). This is modelled as $x \sim \mathbf{Bin}(n, \rho)$, with **Bin** indicating a binomial distribution. We generalise this under our behavioural hypothesis. We posit, for a fixed overall importation level, that this import probability increases with n. This models risk awareness, in which risk-averse individuals who are less likely to be infected avoid larger events, or equally the individuals attending larger sized events are less risk-averse and more likely to be infected. Risk awareness may also depend on event duration τ, but we do not explore this here.

We model event size bias using sorted Dirichlet weights. We consider m events, the jth of which has size n_j and importation rate $\epsilon(n_j)$. Sizes sequentially span all integers from $n_{\min}$ to $n_{\max}$ uniquely (i.e., $n_{\max} = n_{\min} + m - 1$) but we can relax this to include any distribution over chosen event sizes. We fix the overall transmission level i.e., the total rate of importation across all m events. This constrains $\sum_{j=1}^{m} n_j \epsilon(n_j) = \rho \sum_{j=1}^{m} n_j$, conserving the total infections introduced across all events so that the mean importation rate still equals the prevalence. The $\epsilon(n_j)$ values encode event size bias. We sample random weights from a symmetrical Dirichlet distribution $\{w_j\} \sim \mathbf{Dir}(\{r\})$ with r as a shape parameter and the set notation indicating that this distribution is across all m weights. Smaller r simulates a more skewed weighting.

We then sort the sampled $\{w_j\}$ over many runs so that they have the same rank as the n_j. The sorted weights divided by n_j produces $\epsilon(n_j)$, which varies across runs but has a fixed mean from the Dirichlet model. Also note that $\sum_{j=1}^{m} w_j = \rho$. The sorting ensures $\epsilon(n_j)$ increases with n_j and the choice of r controls the magnitude of this size biased or risk-aware effect. We can generalise this model by allowing r to also be size dependent (i.e., $r(n_j)$). While we assume a fixed overall transmission level as determined by our constraints to benchmark results, this may be violated if risk awareness itself changes the prevalence.

However, our framework still approximately models this if we use the updated prevalence in the constraint formulae.

Integrating the above models for heterogeneity and importation dynamics, we complete our algorithm for sampling import weighted distributions of event risk using **Eq. (1)**. We formulate this in **Eqs. (7)-(8)** with semi-colons discriminating between the probabilities we are evaluating from a distribution and the parameters specifying that distribution. For notational convenience, we use $S_x = \sum_{j=1}^{x} R_0^j$ for the heterogeneous samples and $\epsilon(n)$ to indicate general size bias.

$$\mathbf{P}(y|n,x) = \int_0^\infty \mathbf{Bin}\left(y; n-x, \left(1-e^{-\frac{\tau S_x}{d\,n}}\right)\right) \mathbf{Gam}\left(S_x;\ kx, \frac{R_0}{k}\right) dS_x. \quad (7)$$

$$\mathbf{P}(y|n) = \int_0^1 \sum_{x=0}^{n} \mathbf{P}(y|n,x)\ \mathbf{Bin}(x; n, \epsilon(n))\ \mathbf{P}(\epsilon(n)|n)\ d\epsilon(n). \quad (8)$$

We can use the probability distributions in **Eqs. (7)-(8)** together with the definitions of **Eq. (1)** to compute the measures of event risk that we propose in **Eq. (6)**. These marginalise over the distributions of importation rate and transmission heterogeneity, which are degenerate when $\epsilon(n)$ is constant for all n or all $R_0^j = R_0$, respectively. In the Results, we examine the properties of our computational framework and apply it to explore how behaviour affects superspreading. Our modelling framework is freely available at: https://github.com/kpzoo/smallscaleR.

Results

In the Methods, we developed a framework to assess the risk of acquiring infection at an event by deriving a small-scale reproduction number and the expected number of infections that will occur at the event. Both measures depend on the levels of heterogeneity in transmission and variability in the rate at which infectious individuals are likely to attend the event (i.e., imports). Here we examine the influence of these two key factors in determining outbreak dynamics.

Superspreading risk depends on importations and dispersion

Much research has investigated how heterogeneity in transmission can cause superspreading and hence increase the number of infections likely to result from a gathering or event [1,16]. Specifically, there has been study of how the dispersion parameter k

modulates the risk of superspreading events [19,22,23]. Generally, smaller values of $k < 1$ are predictive of larger transmission heterogeneity and superspreading risk. However, the influence of the number of importations x at an event of size n has received relatively little attention. We examine this by computing the statistics derived in **Eqs. (3)-(5)**, in which we defined the reproduction number $R(x)$ as a function of the imports and the resulting number of expected infections $\mathbf{E}[y|x,n]$.

We consider an event of size $n = 30$ over a range of dispersions $0.1 \leq k \leq 10$ with a large-scale limit (see **Eq. (2)**) of $\left(\frac{\tau}{d}\right)R_0 = 0.3$. We sample $R(x)$ and $\mathbf{E}[y|x,n]$ from heterogeneous gamma distributions describing the transmissibility of the sum of all imported infections (see Methods) and compute statistics from these samples using **Eqs. (4)-(5)**. We plot these results in **Fig *1***. We find several interesting points. First, $R(1)$ is an upper bound on $R(x)$. This single import scenario is related to the event reproduction number proposed in [17] and shows how these measures can overestimate the risk of acquiring infection at an event. If we assume, as is common in some branching process models, that all imports have a reproduction number of $R(1)$ instead of $R(x)$, then $\mathbf{E}[y|x,n]$ will be appreciably overestimated.

Second, increasing heterogeneity (decreasing k) increases the variance of our statistics but decreases our mean risk measures. We observe this as the inversion of the rank of the blue to red curves between the top and bottom panels in **Fig *1***. We show the ratio of variance to mean $\mathbf{VM}[.]$ to emphasise this inversion. Last, we see that the dependence of our statistics on the number of introductions is substantial and potentially as critical as the value of k for describing spread. This makes sense since, for a fixed event size, x controls the depletion of susceptible individuals and the sources of infection. However, these counteract causing the expected number of infections occurring at the event to peak at an intermediate x considerably above that at which $R(x)$ peaks. This underpins the importance of finite event size effects and signifies that a crucial factor controlling the risk of acquiring infection at an event of size n is the distribution over possible introductions to the event, $\mathbf{P}(x|n)$.

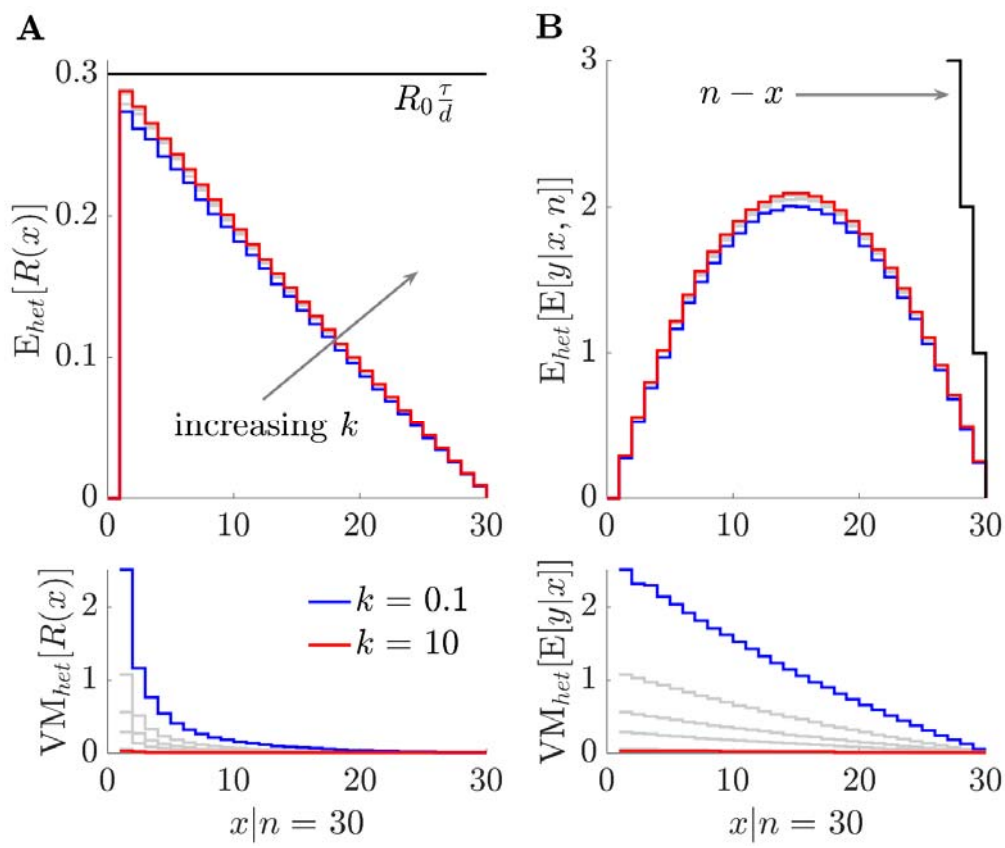

Fig 1: Risk statistics for an event with heterogeneous transmission. We plot the mean (**E**$[.]$, top subfigures) and variance to mean ratio (**VM**$[.]$, bottom subfigures) of the small-scale reproduction number $R(x)$ (panel A) and the mean count of new infections **E**$[y|x,n]$ (panel B) as a function of the number of imports x. We compute these via **Eqs. (3)-(5)** and compile results over 10^5 samples from heterogeneous offspring distributions (denoted with the het subscript) with dispersion parameter k ranging from 0.1 to 10 (increasing from blue to red with grey depicting intermediate values). For comparison, we show the large limit reproduction number $\left(\frac{\tau}{d}\right)R_0$ and the number of initial susceptible individuals at the event, $n-x$.

Population prevalence modulates the superspreading potential at events

Having observed the importance of the number of imports, x when assessing the transmission risk at events, we explore the influence of the distribution of introductions to the event $\boldsymbol{P}(x|n)$. Under a null model from [18], we define this as a binomial distribution with the probability of an import being equal to the prevalence of the infection in the wider population, ρ. We consider epidemics in their initial stages i.e., there is no vaccination- or infection-

acquired immunity, so ρ is small and there are $n - x$ susceptible individuals at the event. We use parameter settings from **Fig 1** but weight samples of small-scale reproduction numbers and mean numbers of imported infections using $\boldsymbol{P}(x|n)$, which is $\boldsymbol{Bin}(x; n, \rho)$ with ρ ranging from $0.01 - 0.1$ (1–10%). We compute histograms and statistics of these samples in **Fig 2**.

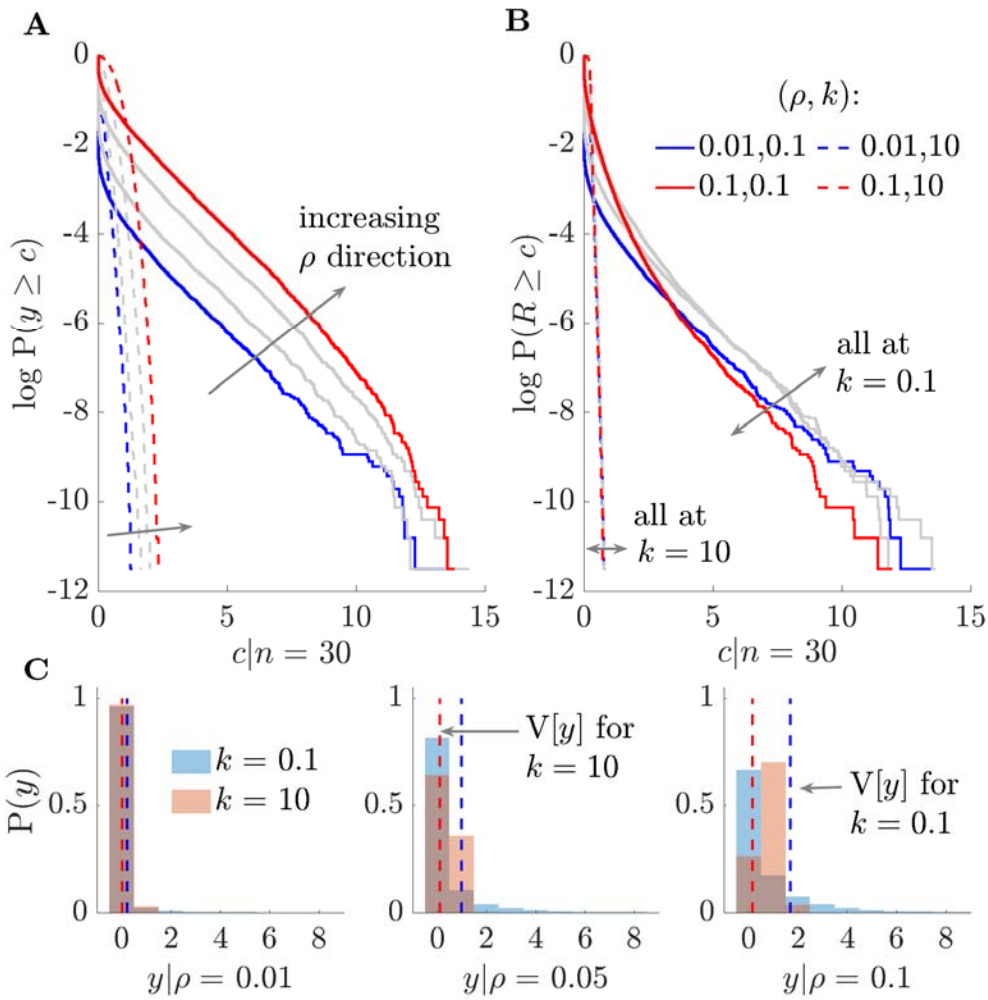

Fig 2: The importation rate magnifies the effects of heterogeneous transmission. We plot the log survival probabilities for the number of new infections y (panel A) and associated reproduction numbers R (panel B). We account for the probability of x imports (distributed as **Bin**$(x; n, \rho)$) at an event of size $n = 30$ with the population prevalence as ρ (increasing from blue to red with grey indicating intermediate values). Larger **P**$(y \geq c)$ signifies more realised heterogeneity (superspreading events occur when more infections are possible at the event), while larger **P**$(R \geq c)$ signifies more heterogeneity in transmissibility (which can engender superspreading). In panels A-B dashed curves are at $k = 10$ (spread is mostly

homogeneous) and solid curves at $k = 0.1$ (spread is heterogeneous). We compute these quantities from **Eqs. (3)-(6)**. Panel C shows histograms of 10^5 samples of y at some ρ values underpinning the results in panels A-B. Dotted lines show the variance of those samples $\mathbf{V}[y]$. We repeat this analysis at a larger event size of $n = 100$ in Supplement **Fig S1** for comparison.

We examine homogeneous ($k = 10$) and heterogeneous ($k = 0.1$) dispersion levels and plot the log survival probabilities of realised numbers of new infections y and associated small-scale reproduction numbers R in the top panels of **Fig *2*** for different values of ρ. We compute these probabilities using **Eq. (6)**. Larger values for these probabilities respectively indicate that superspreading is more likely (i.e., substantially more infections than $\mathbf{E}_{\text{het}}\left[\mathbf{E}_{\text{imp}}[y|n]\right]$ occur) and that imports have increased potential to cause superspreading (i.e., transmissibility above $\mathbf{E}_{\text{het}}[R_{\text{imp}}]$). The bottom panels of **Fig 2** display histograms of samples of the number of infections at the event for some of the values of ρ and show the variance of those samples (notated $\mathbf{V}[y]$ for convenience).

We find that increasing prevalence ranks the y survival curves for both k scenarios (panel A) (at a given threshold c probabilities increase with ρ) but has limited impact on the R survival curves (panel B). The increased risk of superspreading when the population prevalence is higher is confirmed by the histograms (panel C), which have thicker tail probabilities as ρ increases (even when k is large). The variances of the y values (dashed) also increase with ρ. In Supplement **Fig S1** we present an equivalent analysis for a larger sized event ($n = 100$) and obtain similar results. Consequently, the rate at which infections are introduced is crucial to assessing the chances of superspreading occurring at an event.

The risk of superspreading is a key determinant of whether cases of disease at the beginning of an outbreak will lead to a major epidemic, because larger infection clusters can propagate forward, snowballing into wider waves of infections. In standard models, the R survival curves correlate strongly with those of y [1,22]. However, the added variation we see in the y curves in **Fig 2** highlights that superspreading risk is above that expected from R alone and, further, that the chance of stochastic extinction is reduced ($\mathbf{P}(y = 0)$ falls with ρ). Understanding the interaction between the import rate (determined by the population prevalence, ρ) and finite size effects at the event is therefore essential to accurately inferring the risk of superspreading at an event and hence the chance of epidemic establishment.

Next we demonstrate that the risk of superspreading can further rise if attendance at an event depends on risk awareness.

Risk awareness controls importation rates and amplifies superspreading risk

We previously assumed that the importation rate into an event was small, constant and equal to the population infection prevalence ρ. However, this is unrealistic as event attendance will depend on individual preferences and data have found that individual perceptions of infection risk can regulate transmission dynamics [24,25]. Many models couple behavioural change to prevalence [10,13], and prevalence-elastic dynamics, in which self-protective behaviours vary with prevalence, have been observed. We hypothesise that, for a fixed prevalence of infection, heterogeneity in individual risk perception may mean that risk-averse individuals avoid larger events where they expect higher chances of becoming infected. Events with large numbers of attendees are then disproportionately likely to be attended by less risk-averse individuals, who have a higher chance of introducing infection to the event.

We explore this idea by altering the null model from the above section in which the probability that an event attendee is already infected is ρ. We propose a size-biased model where risk awareness adjusts the effective rate of importation based on the event size n. We realise this using sorted weights that assign a rate $\epsilon(n)$ that scales with n, but ensures the total number of imported infections to all events is conserved on average i.e., the overall transmission level is fixed. We consider a set of m events, the jth of which has size n_j. The weight w_j is generally assumed to increase with n_j but satisfies $\sum_{j=1}^{m} w_j = \rho$. The skew of the w_j i.e., strength of the size-bias, is controlled by the shape parameter r. We apply this model with differing weight strengths r using **Eqs.(7)-(8)** and under the parameter settings from **Fig 2**, to obtain **Fig 3**.

In **Fig 3**, we study weight choices characterising two risk-aware scenarios (green and red), in which the probability that each event attendee is an imported infection increases with the event size, a relatively risk-stable case (blue) and a null model (black, dashed) completely neglecting risk-awareness. We show corresponding importation rates in panel A of **Fig 3**. In panel B we demonstrate that size-biasing causes 1.6-5.6 times more infections at larger events, relative to the commonly used null model, for the shape parameters and overall transmission levels we consider here. This amplification outweighs the suppression of infections at smaller events and signifies that risk awareness can substantially alter infection dynamics. We also show the underlying mean small-scale reproduction numbers and

numbers of infections at all events (computed from **Eqs. (6)-(8)**) at differing within-event transmission heterogeneity levels (values of k). The impact of heterogeneity in transmissibility on these mean trends is minimal.

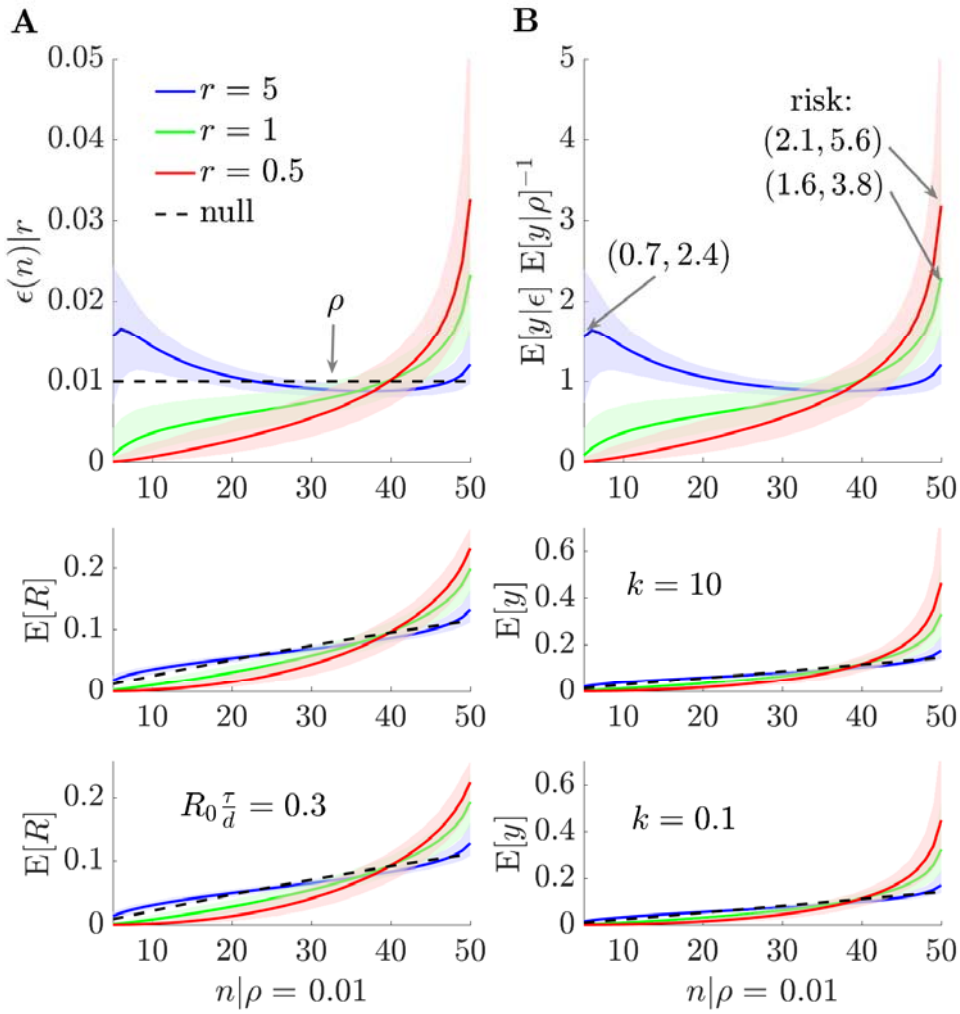

Fig 3*:* Event size bias substantially elevates the risk of infection. We compare the risk of acquiring infection at an event under models with size-biased introductions emerging from risk awareness to a null model with constant importation rate at the prevalence ρ. Panel A shows the size-biased rates $\epsilon(n)$, parametrised by r, for $m = 46$ events with sizes spanning $5{:}50$. Smaller r, decreasing from blue to green to red, indicates more skewed $\epsilon(n)$ distributions but conserves the overall transmission level (see text). Panel B illustrates the consequent risk of infections at an event $\mathbf{E}(y|\epsilon)$ relative to infections under the null model $\mathbf{E}(y|\rho)$ for dispersion $k = 0.1$. Panels C-D show mean reproduction numbers $\mathbf{E}[R]$ and infections $\mathbf{E}[y]$ across two dispersions. All subfigures plot medians with 95% credible

intervals and are computed from **Eqs. (6)-(8)**. These marginalise over 10^4 samples from both the distributions of transmission heterogeneity (controlled by k) and the importations (controlled by ρ and $\epsilon(n)$).

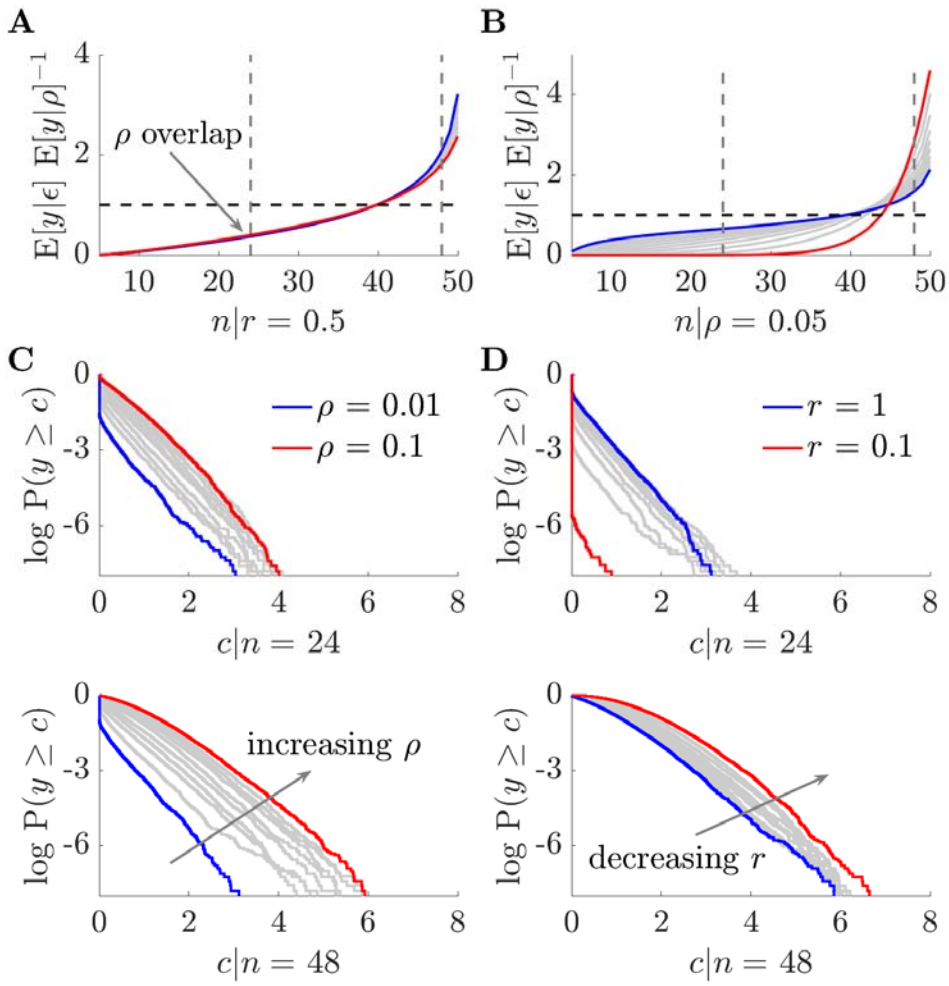

Fig 4: Superspreading risk increases with risk awareness. We repeat the analyses in **Fig *3*** but for varying prevalence rates ρ at a given risk-awareness strength $r = 0.5$ in panel A and for differing strengths at prevalence $\rho = 0.05$ in panel B. These show mean infections $\mathbf{E}(y|\epsilon)$ under risk-aware models relative to that from the null model $\mathbf{E}(y|\rho)$ (we plot only medians of distributions at dispersion $k = 0.1$). We demonstrate how risk awareness modulates the risk of superspreading at a medium and large sized event (dashed vertical lines in panels A-B) by exploring tail infection survival probabilities $\mathbf{P}(y \geq c)$ in panels C-D. See Supplement **Fig S2** and **Fig S3** for additional accompanying simulations and statistics.

We find that the variance in the number of infections and the small-scale reproduction numbers also rises with n (see Supplement **Fig S2** for variances and **VM** ratios of both quantities, which depend on k). These results (with **Fig 2**) suggest superspreading risk can be appreciably elevated if infection risk awareness affects host behaviour, as compared to standard models where the probability that each attendee is initially infected is completely determined by the prevalence ρ. We confirm this in **Fig 4**, where we illustrate how log survival or tail probabilities of infection ($\log \mathbf{P}(y \geq c)$) change with the risk awareness strength r and overall importation rate ρ. In panel A, we fix r and find that the relative risk of infections is mostly unchanged. This verifies that the skew of the size-bias induced by risk awareness is a key variable. In panel B, we illustrate how the relative risk at larger events rises as r falls, for fixed ρ. This causes the median number of expected infections to rise substantially at larger event sizes.

We choose two specific event sizes $n = (24, 48)$, which due to risk awareness have relative risks below and above one (dashed lines in A-B) and examine their tail probabilities in panels C-D. In C we find that, for both cases, superspreading risk rises with prevalence as tail probabilities at any threshold value c scale with ρ. In D we see that risk awareness at a given population infection prevalence can reduce the chances of superspreading at smaller events but considerably amplify the superspreading risk at larger events (seen as an inversion in the ranking of survival curves from blue to red). Associated variations in the statistics of new infections and small-scale reproduction numbers (see Supplement **Fig S3**) also support our conclusions and confirm that risk awareness is driving increased superspreading risk.

In our simulations, the risk of superspreading at large events was elevated by as much as 20 times due to host behaviour (roughly 3 natural log units in the tail probabilities of **Fig 4**). This could have critical ramifications, especially since larger events can support more infections and contribute disproportionately to the establishment of infection in the host population early in an epidemic. As a result, accurate characterisation of small-scale behavioural dynamics, in addition to estimation of both the wider-scale prevalence of infection in the population and transmission heterogeneities, are integral to correctly quantifying the risk of superspreading.

Discussion

Human behaviour is an important but understudied driver of the dynamics of infectious disease outbreaks. While it is known that variations in individual perception of the risks

associated with infection can affect important macroscopic properties of an epidemic, such as the shapes of disease incidence time series or patterns of spread [11,13,26], few studies have explored how human behaviour alters the chance that superspreading events occur. These events, which generate large numbers of infections, substantially influence the growth and persistence of epidemics, especially during early or emergent stages [3,5,19]. Here we aimed to resolve this gap by exploring the relationship between human behaviour and superspreading.

We developed a computational framework (**Eqs. (1)-(8)**) to model transmission at a local or small scale. Specifically, we considered transmission at events (e.g., weddings, parties, sports matches or concerts) where superspreading may arise and individual behaviour can impact pathogen dynamics. Our framework quantifies, under a standard random mixing assumption, how finite-size effects together with heterogeneities in both the transmissibility among hosts and the rate of introductions of infection to an event, contribute to the numbers of infections generated at that event. Our framework generalises several previous approaches [16–18] and allowed us to define a within-event (or small-scale) reproduction number. This measures how numbers of importations and individual-level variations impact transmissibility at events, and meaningfully links to population-level dynamics through its convergence to R_0 when the event size and duration scale asymptotically (see Methods).

Using our small-scale reproduction number, we found that earlier transmissibility metrics, whether derived from branching processes [1] or previous event-level approaches to [17], can overestimate transmission and the number of infections likely to occur at an event (**Fig *1***). This result holds for any model in which the limited number of available susceptible individuals to infect at an event is not accounted for, and is exacerbated when there are multiple imported cases (effectively reducing the number of susceptible individuals that each imported case can infect). Moreover, this finite-size effect highlighted that it is essential to measure the number of infections introduced to an event to accurately quantify superspreading risk (**Fig *2*** and Supplement **Fig S1**), which we found to strongly depend on both the number of imports and more conventionally evaluated heterogeneities (see Introduction) [3]. This insight also hints at one potential reason why behavioural dynamics may affect the risk of superspreading – if risk awareness alters the distribution of infections introduced to events then it may also modulate the chance of superspreading occurring at those events.

We explored this possibility using a parsimonious model of human behaviour. Specifically, we posited that variations in infection risk perceptions or awareness cause risk-aware individuals to be more likely to avoid larger events, which they believe present a higher

infection risk. Our framework allowed us to model this using a size-biased weighting on the infection introduction rate to an event that is a function of both the wider population prevalence and the size of the event. This draws on real-world observations that individuals adopt self-protective behaviours that are driven by infection prevalence and risk perception [10,15]. Across a number of model simulations, we found, for given event sizes and fixed overall transmission levels, that risk awareness amplifies the risk of superspreading at large events (**Fig *3*** and Supplement **Fig S2**) but leads to reduced transmission at smaller events.

Moreover, as either the population infection prevalence or strength of risk awareness increase, the superspreading risk elevates (**Fig 4** and Supplement **Fig S3**). This holds irrespective of the level of heterogeneity in transmissibility at the event (which describes conventional drivers of superspreading, including host characteristics and pathogen biology). Further, the mean, variance and probability of large numbers of infections at the event all support this trend. Since this amplification of within-event transmission occurs at those events with capacity to support larger numbers of infections (i.e., finite event size effects are less able to limit spread), this behavioural mechanism can have major consequences. This may be especially critical during the sensitive, initial stages of potential epidemics, where increased superspreading can drive growth and trigger progression from sporadic outbreaks into sustained waves of infection [3,5].

Although these results all indicate the importance of human behaviour in driving infectious disease outbreak dynamics, it should be noted that – like any mathematical modelling study – our analyses involved several assumptions. First, we assumed random mixing within events so that every susceptible individual interacts with every infectious individual with an equal probability. In reality, contact networks at events exist, and may differ with the size and type of event. Although modelling approaches exist to include contact structure in epidemiological models, for example incorporating both risk awareness and infection using multilayer networks [26,27], they can be complex and difficult to interpret or require high resolution data that are rarely available [9,24,28]. Note that our inclusion of transmission heterogeneities as in [1], together with our weighting of the risk of introductions based on event size, do model some features of real-world transmission networks while preserving interpretability.

Second, we assumed that event sizes and durations were pre-determined, and fixed the overall transmission level. However, these may all depend on risk awareness, which may itself reduce event sizes, durations, frequencies and hence the population prevalence of infection. Alternatively, if events are prevented, e.g., due to government guidance, then less risk-averse individuals may initiate their own gatherings, which may increase transmission (a

rebound effect). This feedback between behaviour and environment (risk and event properties) may shift the chances of superspreading [11]. Inclusion of these effects in our modelling framework is a target for further research. Finally, we did not attempt to model the influence of multiple data sources or surveillance biases on risk awareness. During initial epidemic stages and especially for novel pathogens, data may be sparse and erratic [28,29]. The perception of the risks associated with acquiring infection may therefore be affected by unreliable reports and considerable uncertainty about the true risk posed by the invading pathogen.

While our modelling approach is relatively simple, it provides clear evidence that behavioural dynamics can substantially amplify the risk of superspreading. Heterogeneity in infection risk perception within host populations translates into the potential for substantial transmission at large events early on in infectious disease epidemics. Because superspreading plays a pivotal role in epidemic growth and the risk of pathogen establishment, further study is required to uncover the mechanisms that underlie the coupling between behavioural and epidemiological dynamics. Accordingly, there is a well-defined need for high quality surveillance that includes behavioural data [9,24]. Surveys linking perceptions of infection risk with attendance at events [15] are essential to assess when risk awareness may be a key driver of superspreading. This is important to inform, design and effectively target public health interventions [4].

Bibliography

1. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. Nature. 2005;438: 355–359. doi:10.1038/nature04153
2. Heesterbeek H, Anderson RM, Andreasen V, Bansal S, De Angelis D, Dye C, et al. Modeling infectious disease dynamics in the complex landscape of global health. Science. 2015;347: aaa4339. doi:10.1126/science.aaa4339
3. Althouse BM, Wenger EA, Miller JC, Scarpino SV, Allard A, Hébert-Dufresne L, et al. Superspreading events in the transmission dynamics of SARS-CoV-2: Opportunities for interventions and control. PLoS Biol. 2020;18: e3000897. doi:10.1371/journal.pbio.3000897
4. Frieden TR, Lee CT. Identifying and Interrupting Superspreading Events-Implications for Control of Severe Acute Respiratory Syndrome Coronavirus 2. Emerging Infect Dis. 2020;26: 1059–1066. doi:10.3201/eid2606.200495
5. Lau MSY, Dalziel BD, Funk S, McClelland A, Tiffany A, Riley S, et al. Spatial and temporal dynamics of superspreading events in the 2014-2015 West Africa Ebola

epidemic. Proc Natl Acad Sci USA. 2017;114: 2337–2342. doi:10.1073/pnas.1614595114

6. Goyal A, Reeves DB, Cardozo-Ojeda EF, Schiffer JT, Mayer BT. Viral load and contact heterogeneity predict SARS-CoV-2 transmission and super-spreading events. Elife. 2021;10. doi:10.7554/eLife.63537
7. Sun K, Wang W, Gao L, Wang Y, Luo K, Ren L, et al. Transmission heterogeneities, kinetics, and controllability of SARS-CoV-2. Science. 2021;371. doi:10.1126/science.abe2424
8. Chen PZ, Bobrovitz N, Premji Z, Koopmans M, Fisman DN, Gu FX. Heterogeneity in transmissibility and shedding SARS-CoV-2 via droplets and aerosols. Elife. 2021;10. doi:10.7554/eLife.65774
9. Funk S, Bansal S, Bauch CT, Eames KTD, Edmunds WJ, Galvani AP, et al. Nine challenges in incorporating the dynamics of behaviour in infectious diseases models. Epidemics. 2015;10: 21–25. doi:10.1016/j.epidem.2014.09.005
10. Silk MJ, Carrignon S, Bentley RA, Fefferman NH. Improving pandemic mitigation policies across communities through coupled dynamics of risk perception and infection. Proc Biol Sci. 2021;288: 20210834. doi:10.1098/rspb.2021.0834
11. Weitz JS, Park SW, Eksin C, Dushoff J. Awareness-driven behavior changes can shift the shape of epidemics away from peaks and toward plateaus, shoulders, and oscillations. Proc Natl Acad Sci USA. 2020;117: 32764–32771. doi:10.1073/pnas.2009911117
12. Eksin C, Paarporn K, Weitz JS. Systematic biases in disease forecasting - The role of behavior change. Epidemics. 2019;27: 96–105. doi:10.1016/j.epidem.2019.02.004
13. Funk S, Gilad E, Watkins C, Jansen VAA. The spread of awareness and its impact on epidemic outbreaks. Proc Natl Acad Sci USA. 2009;106: 6872–6877. doi:10.1073/pnas.0810762106
14. Chan HF, Skali A, Savage DA, Stadelmann D, Torgler B. Risk attitudes and human mobility during the COVID-19 pandemic. Sci Rep. 2020;10: 19931. doi:10.1038/s41598-020-76763-2
15. Shiina A, Niitsu T, Kobori O, Idemoto K, Hashimoto T, Sasaki T, et al. Relationship between perception and anxiety about COVID-19 infection and risk behaviors for spreading infection: A national survey in Japan. Brain Behav Immun Health. 2020;6: 100101. doi:10.1016/j.bbih.2020.100101
16. Tupper P, Pai S, COVID Schools Canada, Colijn C. COVID-19 cluster size and transmission rates in schools from crowdsourced case reports. Elife. 2022;11. doi:10.7554/eLife.76174
17. Tupper P, Boury H, Yerlanov M, Colijn C. Event-specific interventions to minimize

COVID-19 transmission. Proc Natl Acad Sci USA. 2020;117: 32038–32045. doi:10.1073/pnas.2019324117

18. Champredon D, Fazil A, Ogden NH. Simple mathematical modelling approaches to assessing the transmission risk of SARS-CoV-2 at gatherings. Can Commun Dis Rep. 2021;47: 184–194. doi:10.14745/ccdr.v47i04a02
19. Garske T, Rhodes CJ. The effect of superspreading on epidemic outbreak size distributions. J Theor Biol. 2008;253: 228–237. doi:10.1016/j.jtbi.2008.02.038
20. Green DM, Kiss IZ, Kao RR. Parameterization of individual-based models: comparisons with deterministic mean-field models. J Theor Biol. 2006;239: 289–297. doi:10.1016/j.jtbi.2005.07.018
21. Sharker Y, Kenah E. Estimating and interpreting secondary attack risk: Binomial considered biased. PLoS Comput Biol. 2021;17: e1008601. doi:10.1371/journal.pcbi.1008601
22. Parag KV. Sub-spreading events limit the reliable elimination of heterogeneous epidemics. J R Soc Interface. 2021;18: 20210444. doi:10.1098/rsif.2021.0444
23. Edwards DA, Ausiello D, Salzman J, Devlin T, Langer R, Beddingfield BJ, et al. Exhaled aerosol increases with COVID-19 infection, age, and obesity. Proc Natl Acad Sci USA. 2021;118. doi:10.1073/pnas.2021830118
24. Verelst F, Willem L, Beutels P. Behavioural change models for infectious disease transmission: a systematic review (2010-2015). J R Soc Interface. 2016;13. doi:10.1098/rsif.2016.0820
25. Tan X, Li S, Wang C, Chen X, Wu X. Severe acute respiratory syndrome epidemic and change of people's health behavior in China. Health Educ Res. 2004;19: 576–580. doi:10.1093/her/cyg074
26. Granell C, Gómez S, Arenas A. Dynamical interplay between awareness and epidemic spreading in multiplex networks. Phys Rev Lett. 2013;111: 128701. doi:10.1103/PhysRevLett.111.128701
27. Bauch CT, Galvani AP. Epidemiology. Social factors in epidemiology. Science. 2013;342: 47–49. doi:10.1126/science.1244492
28. Parag KV, Donnelly CA, Zarebski AE. Quantifying the information in noisy epidemic curves. Nat Comput Sci. 2022;2: 584–594. doi:10.1038/s43588-022-00313-1
29. Yan P, Chowell G. Quantitative Methods for Investigating Infectious Disease Outbreaks. Cham, Switzerland: Springer; 2019.
30. West R, Michie S, Rubin GJ, Amlôt R. Applying principles of behaviour change to reduce SARS-CoV-2 transmission. Nat Hum Behav. 2020;4: 451–459. doi:10.1038/s41562-020-0887-9
31. Fenichel EP, Castillo-Chavez C, Ceddia MG, Chowell G, Parra PAG, Hickling GJ, et

al. Adaptive human behavior in epidemiological models. Proc Natl Acad Sci USA. 2011;108: 6306–6311. doi:10.1073/pnas.1011250108

Funding

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

Data availability statement

All data and code to reproduce the analyses and figures of this work and to compute formulae from the Methods are freely available (in MATLAB) at: https://github.com/kpzoo/smallscaleR.