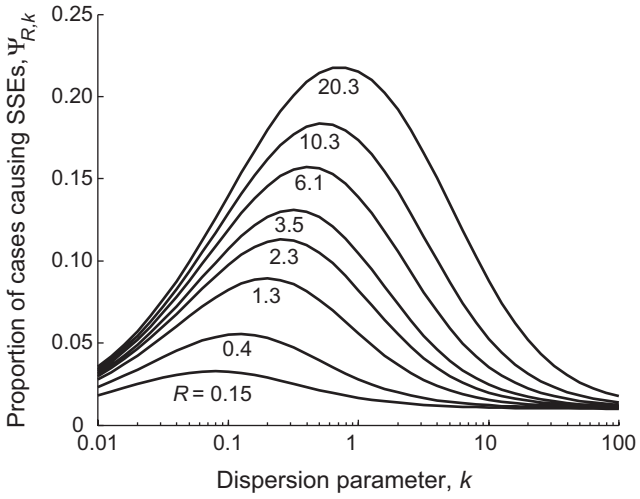
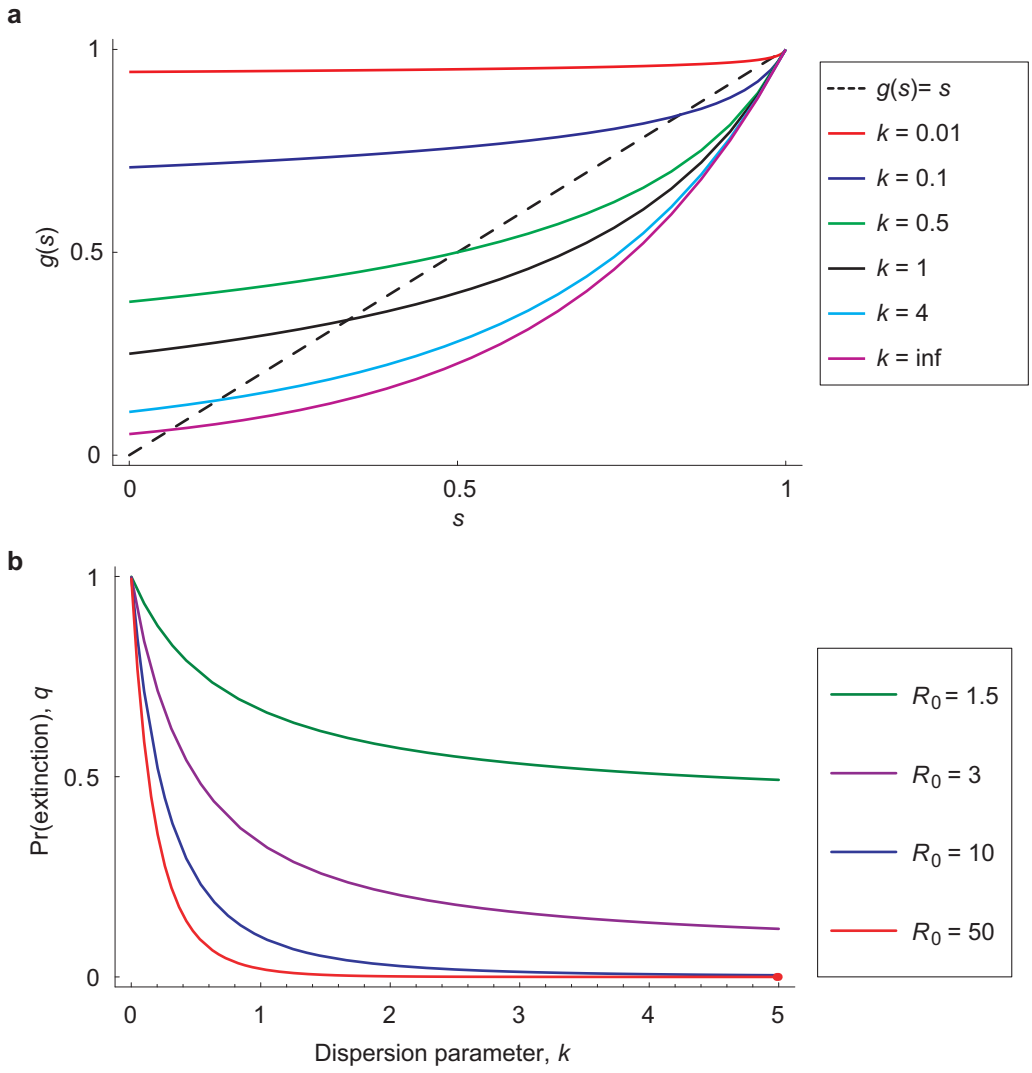
Supplementary Figures

From: Superspreading and the impact of individual variation on disease emergence

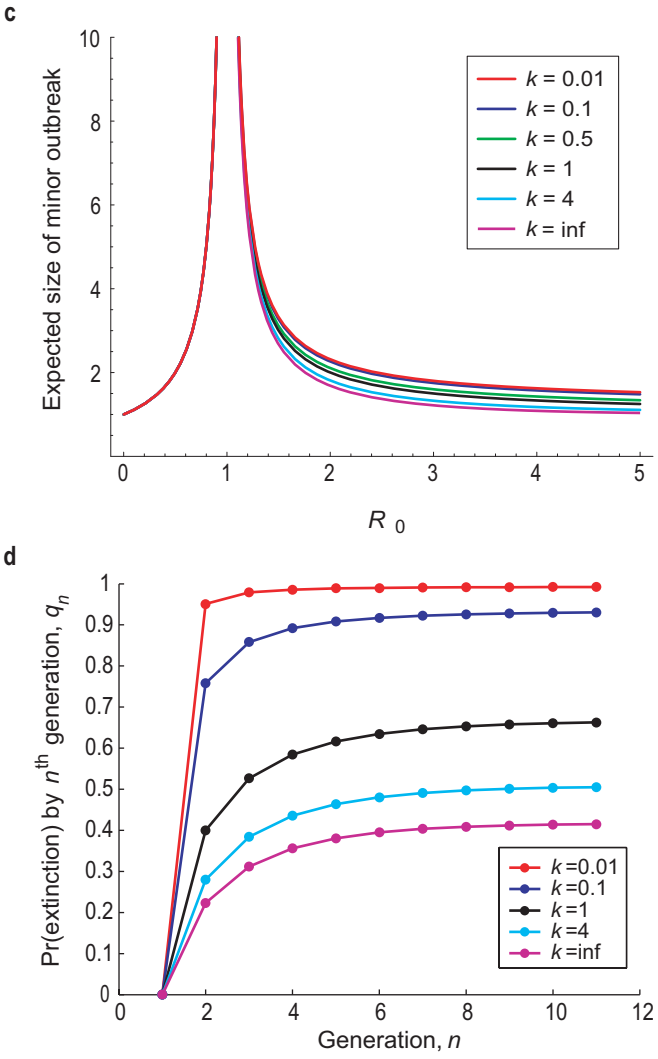
J.O. Lloyd-Smith, S.J. Schreiber, P.E. Kopp, W.M. Getz



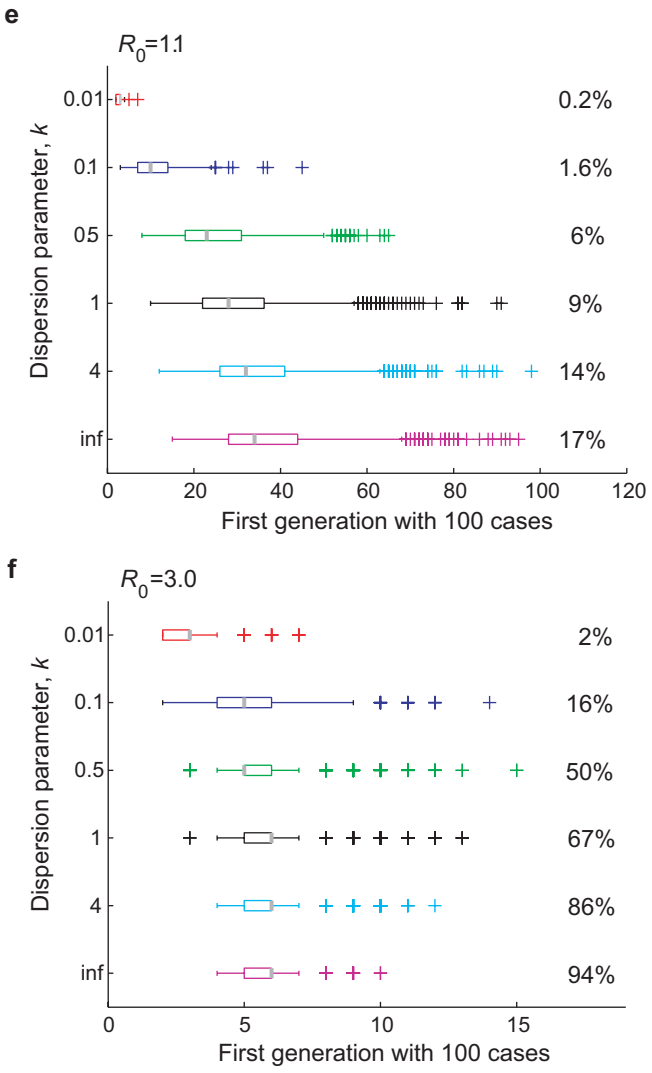
Supplementary Fig 1. Prediction of SSE frequency.
The expected proportion of infectious cases causing 99th-percentile SSEs ($\Psi_{R,k}$) for outbreaks with Z~NegB(R,k), plotted versus k. Each curve shows the relationship for a particular value of the effective reproductive number, R. The values of R plotted were selected such that $\Pr(Z \leq Z^{(99)} | Z \sim \text{Poisson}(R)) = 0.01$. See Supplementary Notes for details of the calculation.



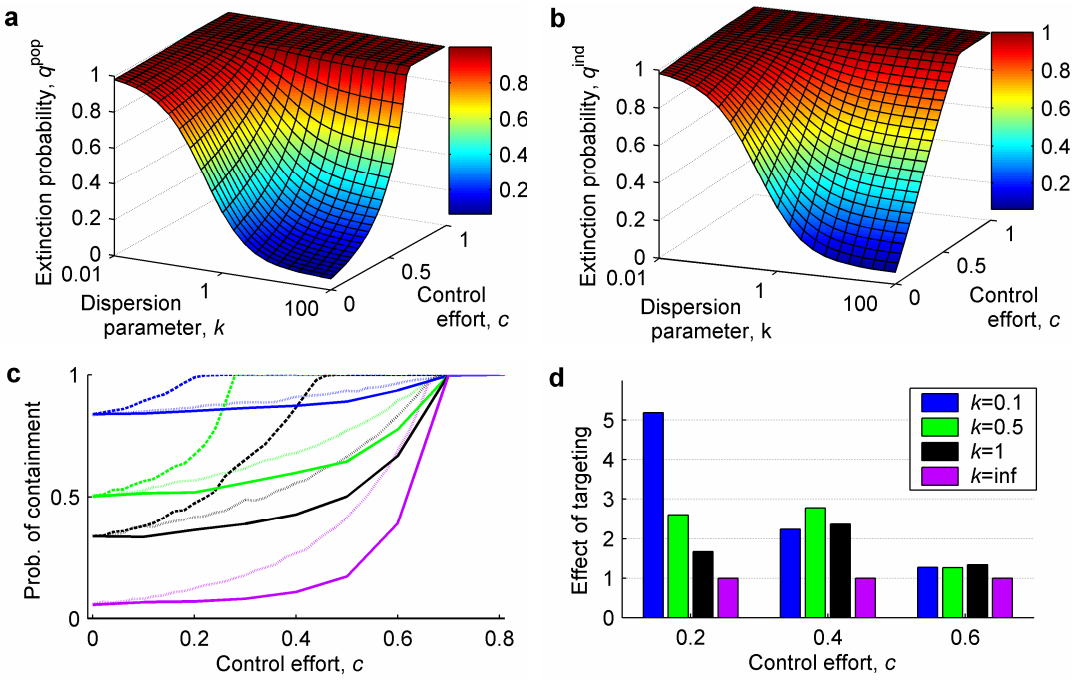
Supplementary Fig 2. Branching process results for Z~NegB(R_0,k).
(a) The probability generating function of the negative binomial distribution, plotted for R_0=3 and different dispersion parameters k. The y-intercept of the pgf equals p_0, the probability that an infected individual will infect nobody, and is a major factor in the rising probability of extinction as k decreases. The extinction probability q is determined by the point of intersection of the pgf with a line of slope 1 (dashed) through the origin. (b) The probability of stochastic extinction given introduction of a single infected individual, q, rises to 1 as $k \to 0$ for any value of R_0.



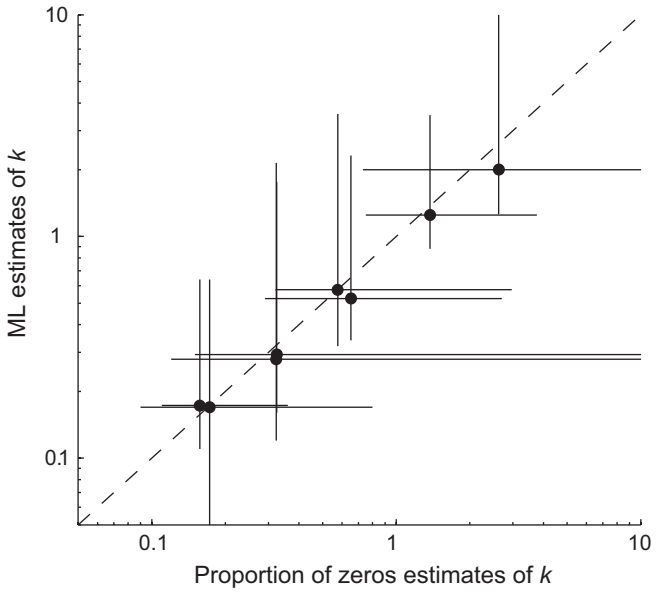
Supplementary Fig 2. Branching process results for Z~NegB(R_0,k) (cont).
(c) Expected size of a minor outbreak (i.e. an outbreak that dies out spontaneously) versus R_0. Curves for all k values are identical for R_0<1. (d) The probability of stochastic extinction by the n^{th} generation of transmission, q_n, for R_0=3 and a range of k. Interestingly, for the third and subsequent generations, the k=1 case has the highest continuing probability of extinction.



Supplementary Fig 2. Branching process results for Z~NegB(R_0,k) (cont).
(e) Growth rate of simulated outbreaks with R_0=1.1 and one initial case, conditional on non-extinction. Boxes show interquartile range (IQR) and median (in grey) of the first disease generation with 100 cases; whiskers show most extreme values within 1.5×IQR of the boxes, and crosses show outliers. Percentages show the proportion of 10,000 simulated outbreaks that reached the 100-case threshold (i.e. roughly 1-q). (f) Growth rate of simulated outbreaks with R_0=3. Both (e) and (f) are exactly analogous to Fig 2c except for different values of R_0.



Supplementary Figure 3. Impact of control measures.
(a) Probability of stochastic extinction for diseases with different degrees of individual variation, k, under population-wide control policies where the infectiousness of all individuals is reduced by a factor c. (b) Probability of stochastic extinction under individual-specific control policies where a randomly-selected proportion c of infectious individuals have their infectiousness reduced to zero. In (a) and (b), outbreaks had R_0=3 and began with a single infectious case, and control was assumed to be present from the outset. The difference between (b) and (a) is shown in Fig. 3a in the main text. (c) Effect of random versus targeted control measures. The plot is exactly analogous to Fig. 3c in the main text, except that in the targeted control scenario individuals in the top 20% of infectiousness are ten-fold more likely to be controlled than those in the bottom 80% (rather than four-fold more likely as in Fig. 3c), so 71% of control effort is focused in the top 20% of cases (rather than 50% in Fig. 3c). The probability of outbreak containment (i.e. never reaching 100 cases) is shown for four diseases with R_0=3 and k=0.1 (blue), k=0.5 (green), k=1 (black), or $k\to\infty$ (purple). Control policies are population-wide (solid lines), random individual-specific (dotted lines), or targeted individual-specific (dashed lines). (d) The factor by which targeting increased the impact of control on preventing a major outbreak relative to random individual-specific control, for the simulations shown in (c). For $k\to\infty$, targeting has no effect so this factor is 1, and dotted and dashed lines overlay one another in (c). Results in (c) and (d) are the mean of 10,000 simulations, with control beginning in the second generation of cases.



Supplementary Fig 4. Estimation of the negative binomial dispersion parameter *k* from full datasets and from mean and proportion of zeroes.

Each point corresponds to an outbreak for which we have full information on *Z*, so we are able to estimate $\hat{k}_{mle}$ and the corresponding bias-corrected bootstrap 90% confidence interval. For the same dataset, we then discarded all information except the mean and proportion of zeros and estimated $\hat{k}_{pz}$ and Anscombe's large-sample confidence interval (method 4 in Section 2.2.4 of the Supplementary Notes).