An event-based model of superspreading in epidemics

Alex James¹, Jonathan W. Pitchford^{2,*} and Michael J. Plank¹

¹*Department of Mathematics and Statistics, University of Canterbury, Christchurch 8140, New Zealand*

²*Department of Biology and York Centre for Complex Systems Analysis, University of York, PO Box 373, York YO10 5YW, UK*

Many recent disease outbreaks (e.g. SARS, foot-and-mouth disease) exhibit superspreading, where relatively few individuals cause a large number of secondary cases. Epidemic models have previously treated this as a demographic phenomenon where each individual has an infectivity allocated at random from some distribution. Here, it is shown that superspreading can also be regarded as being caused by environmental variability, where superspreading events (SSEs) occur as a stochastic consequence of the complex network of interactions made by individuals. This interpretation based on SSEs is compared with data and its efficacy in evaluating epidemic control strategies is discussed.

Keywords: superspreading; epidemiology; branching process; environmental stochasticity

1. INTRODUCTION

Superspreading, loosely defined as a situation where relatively few individuals cause a large number of secondary infections while a majority of infected individuals cause few (or no) infections, is an important phenomenon in human and animal epidemiology; understanding its causes and consequences is a priority for epidemiological research (Galvani & May 2005; Matthews & Woolhouse 2005). Progress has been made by modelling superspreading as a demographic phenomenon, allowing variation in infectivity between individuals. This may be interpreted as variation in quantities such as pathogen shedding rate (Matthews *et al.* 2006) or total lifetime infectiousness (Lloyd-Smith *et al.* 2005). Lloyd-Smith *et al.* (2005) combine contact tracing data with a stochastic model based on individual variability to implicate superspreading in human diseases including SARS, Ebola haemorrhagic fever (HF), measles and smallpox. Their results indicate that the traditional metric R_0 (the mean number of infections caused by an infected individual in a fully susceptible population) is an inadequate indicator of whether or not an epidemic will be triggered; repeated field observations of low R_0 may lead to serious underestimation of epidemic potential.

In §2, a simple model is proposed whereby every individual in the population has the potential of superspreading (SS), with a superspreading event (SSE) being defined as a single individual causing a potentially large number of infections in a short time. This allows the development of a stochastic model with parameters which may both aid understanding and help to formulate and evaluate improved and practical control strategies (§3). In common with any simple model of a complex process, this event-driven description is an idealization. The ‘reality’ of superspreading lies somewhere between the event-oriented model proposed here and the individual-oriented description of Lloyd-Smith *et al.* (2005); these issues are explored more comprehensively in §4.

2. A SIMPLE MECHANISTIC SUPERSPREADING MODEL

Discrete-time branching processes (Diekmann & Heesterbeek 2000) and the general class of susceptible–infected–removed (SIR) dynamic models (Murray 1989; Anderson & May 1991) are fundamental tools in epidemiology (Lloyd-Smith *et al.* 2005). In a branching process, each infected individual infects a random number of susceptibles in each time-step (generation) and the epidemic spreads generation-by-generation. Individuals infect independently, following some probabilistic offspring distribution (defined as the number of secondary cases Z caused by an infectious individual). Successful monitoring, prevention and control of epidemics therefore depend on understanding this offspring distribution. The simplest branching process model (BP1) ignores variation between individuals; each individual infects as a Poisson process, $Z \sim \text{Poisson}(R_0)$. A slightly more sophisticated model (BP2) allows some heterogeneity among the individuals by introducing a random variable ν , the expected number of infections caused by an infected individual. BP2 assumes that ν is exponentially distributed with mean R_0 , so that $Z \sim \text{Geometric}(R_0)$. Mechanistically, this model can be interpreted as individuals being infectious for an exponentially distributed time, t_i , motivated by a constant SIR removal rate. The model in Lloyd-Smith *et al.* (2005), herein labelled as LS, generalizes this notion by allowing ν to be gamma distributed, yielding a negative binomial distribution for Z which is argued to more effectively describe epidemics where SS is a key feature.

The model formulated below exploits the basic framework of branching processes, but allows ‘individual variation’ (*sensu* Galvani & May 2005; Lloyd-Smith *et al.* 2005) to arise as a consequence of each individual infecting others according to realizations of the same underlying stochastic process. The resulting mechanistic formulation allows analytical and numerical exploration of epidemics where SSEs are the primary cause of observed variation in the infectivity of individuals.

* Author for correspondence (jwp5@york.ac.uk).

Table 1. Maximum log-likelihood estimates of CP parameters r , ρ and λ . (N is the total number of infected individuals in the dataset; R_0 is the average number of secondary cases per infected individual. AIC_c^i is the modified Akaike information criterion for the best-fit parameter values for model i ; K is the number of parameters associated with each model. The ' α (% SSE)' column indicates the percentage of infections that are caused by SSEs resulting in two or more cases according to the fitted parameters. Abbreviations: before control (b.c.); after control (a.c.); FMD, foot-and-mouth disease.)

outbreak	N	R_0	r	ρ	λ	α (% SSE)	AIC_c^{CP} $K=3$	AIC_c^{BP1} $K=1$	AIC_c^{BP2} $K=1$	AIC_c^{LS} $K=2$	references
SARS Beijing 2003 b.c.	34	1.88	0.41	0.12	12.0	78	110	281	129	97	Shen <i>et al.</i> (2004)
SARS Beijing 2003 a.c.	43	0.28	0.00	0.02	12.0	100	20	97	60	22	Shen <i>et al.</i> (2004)
SARS Singapore 2003 b.c.	57	1.63	0.48	0.05	21.9	70	151	411	201	160	Leo <i>et al.</i> (2003)
SARS Singapore 2003 a.c.	114	0.68	0.22	0.03	17.2	67	187	491	260	178	Leo <i>et al.</i> (2003)
FMD UK 2001	78	0.99	0.33	0.06	11.2	67	181	380	217	169	Matthews & Woolhouse (2005)
Measles MA 1984	27	0.89	0.29	0.38	1.59	32	77	79	73	75	Nkowane <i>et al.</i> (1987)
Pneumonic plague 1907–1933	74	1.32	0.13	1.26	0.94	22	240	252	237	239	Gani & Leach (2004)
Ebola HF Uganda 2000	13	1.50	0.00	3.18	0.48	8	51	45	47	48	Francesconi <i>et al.</i> (2003)
Hantavirus Argentina 1996	20	0.70	0.48	0.12	1.81	17	53	49	48	51	Wells <i>et al.</i> (1997)

Explicitly, suppose infectious individuals again are all independent, and that each individual is infectious for a deterministic time t_I (without loss of generality, henceforth assume $t_I=1$). Infections occur in one of two ways: 'normal' infections, modelled as a Poisson process with intensity r as in BP1, and rare SSEs which occur as a Poisson process with intensity ρ . In each SSE, the number of infections is Poisson distributed with mean λ . The number of offspring in this case is described by a cumulative process (CP; Cox 1962) with probability generating function

$$G(s) = \exp\{r(s-1) + \rho(\exp(\lambda(s-1)) - 1)\}, \quad (2.1)$$

and mean and variance

$$E(Z) = (r + \rho\lambda), \quad \text{Var}(Z) = (r + \rho\lambda + \rho\lambda^2).$$

To provide a fair comparison with alternative models with mean offspring size, R_0 , one simply imposes

$$(r + \rho\lambda) = R_0. \quad (2.2)$$

Table 1 shows maximum log-likelihood estimates of the parameters in equations (2.1) and (2.2) for several disease outbreaks. Results are shown for CP, BP1, BP2 and LS models, allowing direct comparisons to be made. Discussion of the alternative SS model (LS, based on variation between individuals) is reserved for §4. Under the CP model, the importance of SS in a given epidemic is indicated by the proportion α of infections which are caused by SSEs

$$\alpha = \frac{\rho\lambda}{R_0} (1 - e^{-\lambda(1+\lambda)}).$$

This definition does not include *all* events from the Poisson process of intensity ρ , only those resulting in at least two secondary cases. This reflects the fact that, in practice, one could not distinguish a SSE causing a single infection from a single infection arising from the non-SS Poisson process r (this is a necessary consequence of the simplicity of the CP model, and causes no inconsistencies on a practical or mathematical level).

The data in table 1 indicate two distinct classes of epidemic. Parameters fitted to the SARS epidemics and the UK foot-and-mouth disease (FMD) epidemics indicate that SS is a vital ingredient. The proportion of infections occurring as a result of SSEs, α , is relatively large and the average number of secondary infections occurring as a result of a single SSE, λ , is also large (between 11 and 22). (Note that λ is an average; a single SSE can result in many more secondary infections; see electronic supplementary material of Lloyd-Smith *et al.* (2005) for examples from contact tracing data.) Table 1 also shows the values of the modified Akaike information criterion (AIC_c) for each of the four models (Burnham & Anderson 2002). This is defined in terms of the maximized log-likelihood $\ln(L)$ as follows

$$AIC_c = -2 \ln(L) + 2K + \frac{2K(K+1)}{N-K-1},$$

where K is the number of fitted parameters in the model ($K=3$ for CP; $K=2$ for LS; and $K=1$ for BP1 and BP2). The AIC_c is used to compare how well the different models fit the data and penalizes additional parameters: lower the value of the AIC_c , better the model represents the available information. The values of AIC_c in table 1

show that, for the SARS and FMD outbreaks, the SS models (CP and LS) provide a much better representation of the data than do the BP1 and BP2 models. For most of these datasets, there is little difference in the AIC_c between the CP and LS models, indicating that neither of these models is clearly favoured over the other.

In contrast to SARS and FMD, the fitted CP parameters for the other four epidemics in table 1 provide no clear evidence of the SS phenomenon. In these cases, α is small and, more conclusively, λ is small (less than 2), i.e. according to the CP model more than 40% of the so-called SSEs caused zero or one secondary infections. The fact that $\rho > r$ in the fitted parameters, suggesting that SSEs are more frequent than ordinary infections, is therefore misleading. In such cases, it could be argued that SS, modelled without regard to demographic variability, is not a significant feature of the epidemic and the CP model is not the most appropriate tool to use. The AIC_c values show that the CP model provides a worse representation of the data than do the basic BP1 (which is a special case of the CP model with $r=R_0$ and $\rho=0$) and BP2 models.

The SARS data used in table 1 are split into cases before and after stringent control measures were instigated (Leo *et al.* 2003; Shen *et al.* 2004). For the Beijing outbreak, the 'before control' cases are those from generations 1 and 2 of the underlying branching process, and the 'after control' from generations 3 and 4. For the Singapore outbreak, the 'before control' cases are those from generations 1 to 3 and the 'after control' from generations 4 to 7 (Lloyd-Smith *et al.* 2005). Interestingly, comparing the 'before control' and 'after control' data suggests that, although control measures reduced R_0 and brought an end to the epidemics, they did not significantly reduce the proportion α of infections caused by SSEs. In fact, for the Beijing outbreak, α increased to 100% after initiation of control measures: a single infected individual caused all of the 12 secondary cases.

The observation that a small proportion of the population contributes disproportionately to the number of infections is not a new one and has been previously formulated as the '20/80 rule' (Woolhouse *et al.* 1997). This states that the most infectious 20% of individuals (i.e. those causing the greatest number of secondary cases) are responsible for at least 80% of all infections. An analysis of the outbreak data for the diseases in table 1 in the framework of the 20/80 rule is revealing. Figure 1 shows the cumulative percentage of infections caused (i.e. the contribution to R_0) by the most infectious individuals. For the diseases identified in table 1 as having a major SS component (SARS and FMD), the most infectious 20% of individuals caused more than 80% of all infections. In fact, in all these epidemics, 80% of all infections were attributable to the most infectious 2–10% of individuals. For the other disease outbreaks, the most infectious 20% of individuals caused less than 80% of all infections (approx. 70% for measles and 60% for Ebola HF, hantavirus and plague).

This suggests a useful rule-of-thumb for when it is appropriate to employ a model explicitly including SS within the CP context. Datasets complying with the 20/80 rule (i.e. with a curve that passes to the left of the point (20, 80) on the graph in figure 1) are the good candidates for fitting with the CP model. If the dataset has a curve

Figure 1. The cumulative percentage of secondary cases caused by the most infectious individuals for a number of disease outbreaks. The vertical line shows the percentage of cases caused by the most infectious 20% of individuals; the horizontal line shows the percentage of individuals responsible for 80% of all infections. For the SARS and FMD outbreaks, the most infectious 20% of individuals caused more than 80% of all infections; for the other disease outbreaks, the most infectious 20% of individuals caused less than 80% of all infections. Abbreviations: before control (b.c.); after control (a.c.).

that passes to the right of the point (20, 80), then a simple model such as BP1 or BP2 is more appropriate.

3. CONSEQUENCES FOR EPIDEMICS AND CONTROL

(a) Probability of ultimate extinction

One aim of disease control, when $R_0 > 1$, is to increase the probability that an outbreak will eventually cease, rather than continuing indefinitely (Lloyd-Smith *et al.* 2005). (When $R_0 \leq 1$, the epidemic becomes extinct with probability 1.) This is quantified using the probability η of ultimate extinction (PUE), which is the smallest positive root of $G(\eta) = \eta$. For CP models, η satisfies

$$\eta = \exp\{r(\eta - 1) + \rho(\exp(\lambda(\eta - 1)) - 1)\}. \quad (3.1)$$

For comparison, let η_0 be the PUE for the equivalent BP1 model with the same value of R_0 :

$$\eta_0 = \exp(R_0(\eta_0 - 1)).$$

Simple algebra shows that, for any CP model containing SS (i.e. with $\rho\lambda = R_0 - r > 0$),

$$G(\eta_0) = \eta_0 \exp\{\rho(\exp(\lambda(\eta_0 - 1)) - 1 - \lambda(\eta_0 - 1))\} > \eta_0.$$

Since $G(s)$ has strictly positive first and second derivatives on $[0, 1]$, one can assert that SS always increases the PUE (relative to the equivalent BP1 model). This is intuitively reasonable and in agreement with the model of Lloyd-Smith *et al.* (2005): SS models, on an average, die out more often than the equivalent BP1 model with the same value of R_0 .

For a given R_0 , the PUE for a range of r , ρ and λ is computed by solving $G(\eta) = \eta$ numerically. Figure 2a plots the PUE for varying proportions of SS from $(r/R_0) = 0$ (only SS) to $(r/R_0) = 1$ (no SS), for a range of values of R_0 and fixed $\lambda = 5$. In each case, the larger the proportion of infections attributable to SSEs, the larger the PUE; SS epidemics are

Figure 2. The probability η of ultimate extinction (PUE) for different levels of superspreading: $(r/R_0)=1$ corresponds to simple Poisson infections (BP1); $(r/R_0)=0$ corresponds to spread purely by SSEs. (a) PUE against r/R_0 for fixed R_0 and $\lambda=5$. (b) PUE against r/R_0 and λ for fixed $R_0=2$.

much more likely to die out than non-SS epidemics. Figure 2b plots the PUE against (r/R_0) and λ . For a fixed level of superspreading (i.e. fixed r/R_0), increasing λ increases the PUE. Thus, an infection characterized by rare (low ρ) but potentially severe (high λ) SSEs has a higher extinction probability than one characterized by frequent, mild SSEs. Again this is intuitively reasonable as the former is a more severe example of SS.

(b) Observations versus expectations

The expectation and variance of an outbreak trajectory for CP are:

$$E(X_n) = R_0^n, \quad (3.2)$$

$$\text{Var}(X_n) = \frac{(R_0 + \rho\lambda^2)R_0^{n-1}(R_0 - 1)}{R_0 - 1} \quad R_0 \neq 1, \quad (3.3)$$

$$\text{Var}(X_n) = n(R_0 + \rho\lambda^2) \quad R_0 = 1, \quad (3.4)$$

where X_n is the number of infected individuals in generation n . The expected trajectory is independent of SSEs, depending only on R_0 . However, the variance increases with the frequency and severity of SSEs. Thus, SS results in outbreaks that are more likely to become extinct, but those epidemics that do not become extinct will be significantly more severe than any ‘average’ outbreak. This is in agreement with the results of Lloyd-Smith *et al.* (2005).

This phenomenon is quantified by considering the average observed epidemic, i.e. the average trajectory of those which do not become extinct. The probability of extinction by generation n is given by the n th iteration of the generating function evaluated at 0,

$$p_n = G^n(0).$$

The expected observed trajectory is

$$E(\hat{X}_n) = \frac{E(X_n)}{1 - p_n}.$$

When $R_0 > 1$, outbreaks doomed to extinction are likely to become extinct quickly, so it is reasonable to approximate $p_n \approx \eta$ for all but the first few generations n (figure 3a), allowing one to write the expected observed trajectory as

$$E(\hat{X}_n) \approx \frac{R_0^n}{1 - \eta}.$$

After several generations, any observed trajectory is likely to be amplified by a factor of approximately $1/(1 - \eta)$ compared with the naive R_0^n prediction, and the above

results on the PUE (η) show that SS exacerbates the effect significantly. The effect is confirmed in figure 3b, showing the observed and expected trajectory for different proportions of superspreading for a constant R_0 ($R_0 = 1.1$ and $\lambda = 3$ in all cases).

The larger the proportion of infections from SSEs, the more severe the observed outbreaks. The effect is large, confirming the importance of superspreading in understanding epidemics and indicating that, if the aim of control is to decrease the severity of successful outbreaks, as opposed to increasing the probability of outbreak extinction, then decreasing the proportion of infections from SSEs is paramount.

(c) Epidemic control

Historically, the primary aim of control strategies was to reduce R_0 . The SS phenomenon adds a further complication and reveals inadequacies in this approach (Lloyd-Smith *et al.* 2005). Where SS is attributable to individual variation, it can lead to improved control strategies provided potentially superspreading individuals can be identified (Lloyd-Smith *et al.* 2005; Matthews *et al.* 2006). The approach based on SSEs, summarized by the CP model, allows further insight without the need to target particular individuals.

The fitted parameters for SARS and FMD in table 1 show that infections caused by non-SSEs can be relatively insignificant ($r \leq 0.5$ in all cases, even when $R_0 > 1$). This indicates an opportunity to develop targeted control policies based on reducing the frequency (ρ) or severity (λ) of SSEs. In terms of human epidemics, a reduction in SSE frequency ρ corresponds to reducing the frequency of large gatherings of people, for example, by temporarily reducing the working/school week from 6 to 3 days. A reduction in SSE severity λ could similarly be achieved by reducing the maximum number of people gathering together, for example, by segregating the work force and/or the physical environment and staggering break times, or encouraging local small-scale meetings rather than large assemblies. However, SSEs are by no means confined to the school and workplace (Lloyd-Smith *et al.* 2005); so halving the working week, for example, would not correspond to the halving of ρ . While it is unrealistic to imagine all potential for SSEs being eradicated, or all non-SS infections being entirely prevented, the CP

Figure 3. (a) Probability of extinction (p_n) against generation number (n) for different levels of superspreading. (b) Expected trajectory $E(X)$ and average observed trajectories for different levels of superspreading. $R_0 = 1.1$ and $\lambda = 3$.

framework allows the effect of such imperfect measures to be estimated.

The FMD data analysed here use properties (farms) as the unit of infection rather than individual animals (Keeling *et al.* 2001). Therefore, a SSE might be precipitated by gatherings of animals from large numbers of properties such as markets, and one could similarly evaluate the efficacy of reducing the frequency (proportional to ρ) and size (proportional to λ) of markets. If SS is known to be significant and economics demand that a constant number of animals be processed per unit time, so that $\rho\lambda$ remains constant, reference to equation (2.2) and figure 2b shows that holding more markets of a smaller size will diminish the impact of SS compared with the alternative of holding fewer larger markets. (Note, however, that large markets are also argued to offer logical foci for targeted interventions, where such interventions exist (Woolhouse *et al.* 2005)).

The implications for control are necessarily speculative given the simplicity of the CP model. While the above temporary control measures would hardly be popular, and would not be compatible with thriving economic and social activity in the medium- and long-term, they do offer avenues for epidemic control without the need for potentially contentious discrimination between individuals. In reality, the best choice of control strategy must strike a balance between limiting the potential for SSEs in general, and targeting particular individuals with a high propensity for SS. Where it is available, information regarding variability of infectiousness between individuals (i.e. the ingredients of demographic variability models discussed in §4) will also be central to effective epidemic control.

4. DISCUSSION

The CP model presented in this paper uses a mechanistic description of infection, based on meaningful parameters, to characterize the probabilistic course of epidemics with a significant SS component. An *ad hoc* criterion to define such an epidemic is that it obeys the 20/80 rule, i.e. the most infectious 20% of individuals cause at least 80% of all the cases. The model demonstrates that SS is not necessarily a demographic phenomenon caused by possibly unidentifiable population heterogeneities, but may also be observed in populations where every

individual has the potential to cause rare, but severe SSEs. Analysis shows that SS can dramatically influence the probability of ultimate extinction and the variability of epidemic trajectories.

In contrast to the event-based CP model, variation in the infectiousness of individuals can be introduced demographically via a random variable ν , the expected number of infections caused by an infected individual (Lloyd-Smith *et al.* 2005). The LS model is particularly appealing because it includes the BP1 and the BP2 offspring distributions as special cases, and it very naturally allows superspreading (it is ‘heavy-tailed’). The parameter ν encompasses ‘all variation in infectious histories of individuals, including properties of the host and pathogen and environmental circumstances’ (Lloyd-Smith *et al.* 2005). The generality of this definition is a strength in that it allows flexibility, particularly where the underlying mechanism of the infection process is unclear. However, this generality may cause problems in devising effective control strategies; the most effective control strategies target potential super-spreaders before they infect, posing practical problems of identifying such individuals (see Lloyd-Smith *et al.* (2005) for further discussion).

One notable difference is that, in the LS model, the extinction probability tends to 1 as the infection dynamic becomes dominated by SS (fig. 2b of electronic supplementary material of Lloyd-Smith *et al.* 2005), whereas the CP model predicts that extinction probabilities increase, but remain less than 1, even when infections are spread solely by SSEs (figure 2a). The reason for this difference is that the LS model is based on a heavy-tailed distribution, which becomes highly over-dispersed for high levels of SS. The extinction probabilities for the CP model also tend to 1 as the severity of the SSEs (λ) becomes very large (figure 2b), but this is due to a ‘top-heavy’ distribution rather than to a heavy-tailed one. Another difference is that the CP model has the potential for a bimodal offspring (infection) distribution: there is a Poisson-like distribution at the left-hand end and a possible second mode dominated by SSEs. This contrasts with the heavy-tailed unimodal distribution of LS. Figure 4 shows two SARS offspring distributions, showing the original data together with the best fit of the CP and LS models. The data do not allow definite conclusions regarding bimodality to be drawn,

Figure 4. Offspring distributions for the SARS outbreaks before control: (a) Singapore, 2003; (b) Beijing, 2003. $P(Z)$ is the probability of an infected individual causing Z secondary infections during his/her infectious lifetime, according to data (bars) and the CP and LS models (points). The logarithmic scale highlights the occurrence of rare SSEs (which are all counts of 1 or 2).

especially given that the CP model has three parameters compared with two in LS, but the possibility cannot be ruled out at this stage.

Despite approaching the problem of SS from different directions, the CP and LS models show a good deal of agreement about its epidemiological consequences. Both predict that increasing the degree of SS increases the probability that any given outbreak will go extinct and, conversely, that any outbreak which is successful (from the point of view of the disease) will potentially be much more severe than would naively be anticipated from the value of R_0 . The two models characterize SS in different ways, thus lending themselves to different epidemiological scenarios and different approaches to control. Indeed, the modelling of superspreading as events driven by a stochastic environment and as heterogeneities between the individuals both represent idealized infection dynamics. Reality lies somewhere in between, where heterogeneous individuals infect according to stochastic events driven by a random environment of social (or other) interactions. The CP model efficiently describes epidemics driven largely by rare events drawn from similar underlying distributions, whereas the LS model is more appropriate for populations with an overdispersed distribution of individual infectiousness. An attempt to marry these two approaches by modelling both individual heterogeneity and rare SSEs is an important challenge for the future.

The branching process model analysed here is a simple representation of a complex phenomenon, but nevertheless its message is important for more realistic simulation models. The choice of a deterministic infectious period imposes a natural generational time-step on any resulting branching process, but this need not be restrictive. Infectious individuals need not 'die' after each time-step, but could simply be included as an offspring in the next generation with a certain probability, resulting in a discrete version of the BP2 exponential infective period model. Such possibilities are not considered here, partly because one must then address the issue of implausibly long infectious periods. Since SS offspring distributions can arise without indefinitely long infectious periods, it seems unnecessary to introduce them. However, the roles of a finite population

of susceptibles, and any explicit spatial structure, require further (possibly disease-specific) modelling outside the scope of this communication.

The authors are grateful to J. Lloyd-Smith and an anonymous referee for their very valuable comments and constructive criticisms. The authors also thank J. Lloyd-Smith for providing datasets for the SARS Singapore (before control), SARS Beijing (before control), plague, Ebola HF and hantavirus outbreaks.

REFERENCES

- Anderson, R. M. & May, R. M. 1991 *Infectious diseases of humans: dynamics and control*. Oxford, UK: Oxford University Press.
- Burnham, K. P. & Anderson, D. R. 2002 *Model selection and multimodel inference: a practical information-theoretic approach*. New York, NY: Springer.
- Cox, D. R. 1962 *Renewal theory*. London, UK: Methuen and Co.
- Diekmann, O. & Heesterbeek, J. A. 2000 *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*. Chichester, UK: Wiley.
- Francesconi, P. *et al.* 2003 Ebola Hemorrhagic fever transmission and risk factors of contacts, Uganda. *Emerg. Infect. Dis.* **9**, 1430–1437.
- Galvani, A. P. & May, R. M. 2005 Dimensions of superspreading. *Nature* **438**, 293–295. (doi:10.1038/438293a)
- Gani, R. & Leach, S. 2004 Epidemiological determinants for modeling pneumonic plague outbreaks. *Emerg. Infect. Dis.* **10**, 608–614.
- Keeling, M. J. *et al.* 2001 Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science* **294**, 813–817. (doi:10.1126/science.1065973)
- Leo, Y. S. *et al.* 2003 Severe acute respiratory syndrome—Singapore, 2003. *Morbid. Mortal. Weekly Rep.* **52**, 405–411.
- Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E. & Getz, W. M. 2005 Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355–359. (doi:10.1038/nature04153)
- Matthews, L. & Woolhouse, M. E. J. 2005 New approaches to quantifying the spread of infection. *Nat. Rev. Microbiol.* **3**, 529–537. (doi:10.1038/nrmicro1178)

- Matthews, L. *et al.* 2006 Heterogeneous shedding of *Escherichia coli* 0157 in cattle and implications for control. *Proc. Natl Acad. Sci. USA* **103**, 547–552. (doi:10.1073/pnas.0503776103)
- Murray, J. D. 1989 *Mathematical biology*. Berlin, Germany: Springer.
- Nkowane, B. M., Bart, S. W., Orenstein, W. A. & Baltier, M. 1987 Measles outbreak in a vaccinated school population: epidemiology, chains of transmission and the role of vaccine failures. *Am. J. Public Health* **77**, 434–438.
- Shen, Z., Ning, F., Zhou, W., He, X., Lin, C., Chin, D., Zhu, Z. & Schuchatt, A. 2004 Superspreading SARS events, Beijing 2003. *Emerg. Infect. Dis.* **10**, 256–260.
- Wells, R. M., Estani, S. S., Yadon, Z. E., Enria, D., Padula, P., Pini, N., Mills, J. N., Peters, C. J. & Segura, E. S. 1997 An unusual hantavirus outbreak in Southern Argentina: person-to-person transmission? *Emerg. Infect. Dis.* **3**, 171–174.
- Woolhouse, M. E. J. *et al.* 1997 Heterogeneities in the transition of infectious agents: implications for the design of control programs. *Proc. Natl Acad. Sci. USA* **94**, 338–342. (doi:10.1073/pnas.94.1.338)
- Woolhouse, M. E. J., Shaw, D. J., Matthews, L., Liu, W.-C. & Mellor, M. R. 2005 Epidemiological implications of the contact network structure for cattle farms and the 20–80 rule. *Biol. Lett.* **1**, 350–352. (doi:10.1098/rsbl.2005.0331)