

# Searching for Superspreaders: Identifying Epidemic Patterns Associated with Superspreading Events in Stochastic Models



**Christina J. Edholm, Blessing O. Emerenini, Anarina L. Murillo,  
Omar Saucedo, Nika Shakiba, Xueying Wang, Linda J. S. Allen,  
and Angela Peace**

**Abstract** The importance of host transmissibility in disease emergence has been demonstrated in historical and recent pandemics that involve infectious individuals, known as superspreaders, who are capable of transmitting the infection to a large number of susceptible individuals. To investigate the impact of superspreaders on epidemic dynamics, we formulate deterministic and stochastic models that incorporate differences in superspreaders versus nonsuperspreaders. In particular, continuous-time Markov chain models are used to investigate epidemic features associated with the presence of superspreaders in a population. We parameterize the models for two case studies, Middle East respiratory syndrome (MERS) and Ebola. Through mathematical analysis and numerical simulations, we find that the probability of outbreaks increases and time to outbreaks decreases as the prevalence of superspreaders increases in the population. In particular, as disease outbreaks

---

C. J. Edholm (✉)

Department of Mathematics, University of Tennessee, Knoxville, TN, USA

e-mail: [cedholm@utk.edu](mailto:cedholm@utk.edu)

B. O. Emerenini

Department of Physics, Ryerson University, Toronto, ON, Canada

A. L. Murillo

Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL, USA

O. Saucedo

Mathematical Biosciences Institute, The Ohio State University, Columbus, OH, USA

N. Shakiba

Institute of Biomaterials and Biomedical Engineering (IBBME), University of Toronto, Toronto, ON, Canada

X. Wang

Department of Mathematics and Statistics, Washington State University, Pullman, WA, USA

L. J. S. Allen · A. Peace

Department of Mathematics and Statistics, Texas Tech University, Lubbock, TX, USA

occur more rapidly and more frequently when initiated by superspreaders, our results emphasize the need for expeditious public health interventions.

**Keywords** Superspreader · Host heterogeneity · Middle East respiratory syndrome · Ebola · Deterministic model · Stochastic model

## 1 Background

The prevalence of most infectious diseases is often assumed to emerge from person-to-person interactions among a population of individuals who are considered homogeneous with respect to contact, transmission, and recovery behavior. However, it is more realistic to assume that diseases spread among a heterogeneous population. Host heterogeneity may be due to physiological, behavioral, or immunological differences [33]. Behavioral differences may also be related to the environmental setting [33]. For example, some individuals are at a higher risk for spreading the disease due to increased contact with susceptible persons or longer length of infection. This has been observed in the spread of sexually transmitted and vector-borne diseases, where high-risk individuals are characterized by the “20/80” rule, in which 20% of the infected individuals are responsible for 80% of the disease transmission [21, 41]. The 2002–2003 SARS epidemic highlighted the role of superspreaders (SS), defined as people who infect a large number of individuals, in comparison to nonsuperspreaders (NS) who transmit the disease to few or none [6, 15, 30, 40]. However, the exact characteristics of SS and their impact on disease dynamics are difficult to define. Lloyd et al. studied the effects of heterogeneity in infectiousness and then found that the proportion of SS contributed to high levels of heterogeneity for several infectious diseases (e.g., SARS, measles, influenza, rubella, smallpox, Ebola, and other diseases) [26]. Currently, there are no well-known methods for identifying SS in the population or control efforts to reduce the disease transmission at the individual or population levels based on SS. We consider two infectious diseases, Ebola and Middle East respiratory syndrome (MERS), that are associated with certain cultural and health behaviors for which contact patterns may be traceable. Focusing on these two epidemic cases, we will provide insight into disease patterns associated with superspreading events.

Ebola virus was first discovered in 1976 in Africa, in the country now named the Democratic Republic of the Congo, near the Ebola river. Ebola virus can persist in the environment through animal-to-animal transmission, e.g., bats can transmit the virus to apes, monkeys, antelopes, and other animals. The virus can also be transmitted to humans through contact with infected animals in the environment during hunting, meat preparation, or from an animal bite. Infection can be transmitted to other humans through contact with bodily fluids, such as blood, secretions, and organs of sick or diseased individuals, or with contaminated objects, such as bedding and clothes. According to the World Health Organization (WHO), the 2014–2016 Ebola outbreak in West Africa had the most cases and deaths of any

Ebola outbreak to date [38]. This spread might have been increased due to infected health-care workers' close contact with susceptible individuals. Additionally, burial ceremonies may increase contact with infectious deceased bodies that contain the virus. The incubation period, defined as the time of infection to onset of symptoms, ranges from 2 to 21 days [38]. Individuals can recover from Ebola; however, mortality rates range from 25 to 90%. In 2016, the WHO announced that the first vaccine trial implemented in Guinea was 100% effective [17, 37]. The recent preventive measures announced by the Centers for Disease Control and Prevention (CDC) include: reducing contacts with infected animals or bodily fluids of infected individuals, isolating infected and deceased individuals, early detection of infected individuals, and maintaining a clean environment [8].

MERS was first identified in 2012 from an outbreak that occurred in Saudi Arabia [40]. The source of infection was identified as dromedary camels. However, most cases are not due to camel-to-human infections. MERS outbreaks among humans arise from human-to-human interactions, where many cases occur in healthcare settings with poor health prevention and control practices. In 2015, an outbreak of MERS in South Korea was driven by three SS, initiated with one SS contracting MERS during international travel. The first SS was responsible for 29 secondary infections through various clinical visits. Two subsequently infected individuals were responsible for 106 tertiary infections [39, 40]. Individuals infected with MERS can be asymptomatic, while others may experience the following symptoms: fever, coughs, shortness of breath, diarrhea, and pneumonia. Nearly, 35% of MERS cases resulted in death. While no vaccine or treatments are available, individuals are advised to maintain good hygiene when coming into contact with animals, particularly camels, such as washing hands and avoiding contact with sick animals. Additional prevention strategies include consuming thoroughly cooked and prepared animal products [39].

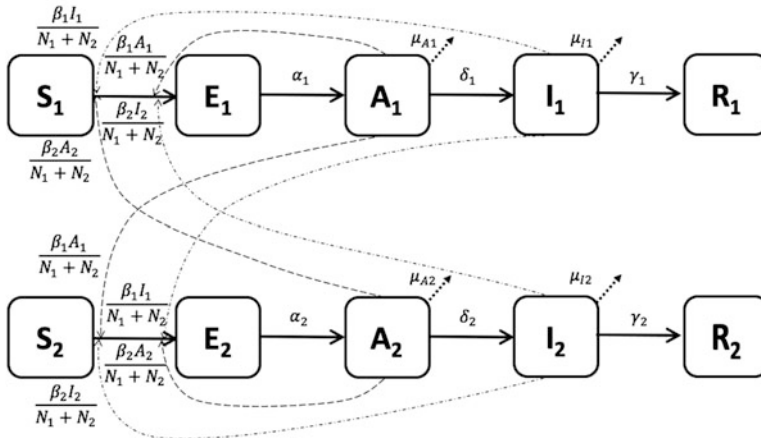
Mathematical models formulated for recent outbreaks of MERS and Ebola have applied the compartmental setting with various disease stages such as susceptible, exposed, infectious, and recovered (SEIR) or performed statistical analyses to identify important parameters in spread of the disease ([11, 23, 24] MERS and [4, 10, 16, 22] Ebola). Additional classes for asymptomatic, hospitalized, or isolated individuals were also included in MERS models [11, 23]. Time-dependent transmission parameters accounted for superspreading events (e.g., [22–24]). Superspreading events have also been investigated with multitype branching processes by including individual heterogeneity in offspring generating functions [18, 26]. All of these models have contributed to a better understanding of the role of superspreaders in disease outbreaks. Our models incorporate the compartmental framework and apply stochastic simulations with theory from branching processes to further elucidate the role of superspreaders in disease dynamics.

In this investigation, we develop a mathematical modeling framework that incorporates the heterogeneity of hosts through differences in transmission rates to assess the role of SS in disease spread at the population level. Specifically, we aim to study the disease dynamics in a heterogeneous population consisting of SS and NS individuals, and develop a deterministic model based on ordinary

differential equations (ODEs) which is expanded to a stochastic model that is implemented as a continuous-time Markov chain (CTMC) system and approximated by a multitype branching process [1, 2]. We incorporate estimated parameter values from published data of prior MERS and Ebola epidemics into our models. Next, we compute the basic reproduction number for the ODE model, and perform sensitivity analysis using Latin hypercube sampling and partial rank correlation. By varying the initial size of SS and model parameters of the CTMC model, we derive and verify analytical estimates obtained using multitype branching process approximations with model simulations to predict the probability of an epidemic outbreak. In further numerical simulations of the CTMC model, we compute sample paths, probability of outbreak, number of deaths, time to outbreak, time to peak infection, and peak number of infectious individuals. Our analyses and numerical simulations reveal how SS influence the dynamics of epidemic outbreaks, which may provide useful insight for public health interventions.

## 2 Deterministic Model

We formulate a simple modeling framework for host heterogeneity due to differences in individuals that account for either SS or NS. In particular, SS or NS may differ in transmission, transitions between disease stages, deaths, recovery, or population size. The SS and NS mix homogeneously, such as in a hospital setting (MERS) or at a large gathering such as a funeral (Ebola). Our basic modeling framework is a system of ODEs with five disease stages for SS and NS as described by the compartmental diagram in Fig. 1 and by the differential equations in (2.1),



**Fig. 1** Flow diagram for the ODE model with classes  $S_i$ ,  $E_i$ ,  $A_i$ ,  $I_i$ , and  $R_i$ , where  $i = 1$  represents NS and  $i = 2$  represents SS. The solid lines denote transitions between classes. Meanwhile, the dashed curves are transmission and death rates in the model, where the transmission of infection from classes  $E_i$  and  $A_i$  to the susceptible class results in a transition from  $S_i$  to  $E_i$

**Table 1** Description of the variables used in the ODE model (Fig. 1) and in the CTMC model

Variable	Description
$S_i$	Number of NS/SS Susceptible individuals
$E_i$	Number of NS/SS Exposed individuals
$A_i$	Number of NS/SS Asymptomatic individuals
$I_i$	Number of NS/SS Infected individuals
$R_i$	Number of NS/SS Recovered individuals

Subscripts are  $i = 1$  for NS and  $i = 2$  for SS

where  $i = 1$  is NS and  $i = 2$  is SS. The description of the model variables are summarized in Table 1. Such types of models have been used in metapopulation settings and are referred to as multigroup models (e.g., [25, 34]).

$$\begin{aligned}
 \frac{dS_i}{dt} &= -\frac{S_i}{N_1 + N_2} (\beta_1(I_1 + A_1) + \beta_2(I_2 + A_2)) \\
 \frac{dE_i}{dt} &= \frac{S_i}{N_1 + N_2} (\beta_1(I_1 + A_1) + \beta_2(I_2 + A_2)) - \alpha_i E_i \\
 \frac{dA_i}{dt} &= \alpha_i E_i - \delta_i A_i - \mu_{A_i} A_i \\
 \frac{dI_i}{dt} &= \delta_i A_i - \mu_{I_i} I_i - \gamma_i I_i \\
 \frac{dR_i}{dt} &= \gamma_i I_i
 \end{aligned} \tag{2.1}$$

For the ODE model, we assume that the disease duration is short and, therefore, we do not include birth or natural death rates. In addition, we make the simplifying assumption that NS cannot become SS and vice versa. We make this assumption due to the short duration of the epidemic period and the fact that no control measures are applied (which could change the transmission patterns). In the model, the number of susceptible NS and SS are denoted by  $S_1$  and  $S_2$ , respectively. Susceptible individuals transition into their respective exposed classes,  $E_1$  and  $E_2$ , at a rate of

$$\frac{\beta_1(I_1 + A_1) + \beta_2(I_2 + A_2)}{N_1 + N_2},$$

where  $\beta_1$  is the transmission rate of the NS asymptomatic  $A_1$  and infective  $I_1$  classes with  $N_1$  as the total number of NS and similarly for SS variables. The total number of individuals is  $N = N_1 + N_2$ , where  $N_i = S_i + E_i + A_i + I_i + R_i$ , for  $i = 1, 2$ . From the exposed class, individuals transition to the asymptomatic class,  $A_1$  and  $A_2$ , at a rate of  $\alpha_1$  or  $\alpha_2$ . In the asymptomatic class, there is disease-induced mortality with rates  $\mu_{A_1}$  or  $\mu_{A_2}$ , respectively. Asymptomatic individuals do not display symptoms but are infectious. Individuals transition into the infective class at a rate of  $\delta_1$  or  $\delta_2$ , where the disease-induced mortality rates are  $\mu_{I_1}$  or  $\mu_{I_2}$ .

**Table 2** Description of the parameters used in the ODE model (Fig. 1) and in the CTMC model

	Parameter	MERS		Ebola	
		Baseline	Range	Baseline	Range
$\beta_1$	Transmission rate NS	0.06 [31]	(0.04, 0.08)	0.128 [10]	(0.04, 0.38)
$\beta_2$	Transmission rate SS	0.6931**	(0.4, 0.8)	1.0150**	(0.5, 1.2)
$\alpha_i^{-1}$	Latent period	6.3 [12]	(2–8)	10 [10]	(7, 14)
$\delta_i^{-1}$	Duration of asymp. stage	0.4 [12]	(0.1, 2)	$10^{-4}$ [10]	$(9.9, 100) \times 10^{-5}$
$\mu_{ji}$	Disease-induced death rate	0.08 [7]	(0.02, 0.14)	0.09 [9]	(0.075, 0.125)
$\gamma_i$	Recovery rate	0.075 [19]	(0.05, 0.1)	0.05 [31, 35]	(0.04, 0.1)

Subscripts are  $i = 1, 2$  and  $j = I, A$ . Baseline values are used in all simulations. Range of values is used for the parameter sensitivity analysis (Sect. 4). We assume  $\alpha_1 = \alpha_2$ ,  $\delta_1 = \delta_2$ , and  $\mu_{A1} = \mu_{A2} = \mu_{I1} = \mu_{I2}$ . Estimated parameters are marked by \*\*

In the infective classes, individuals display symptoms and are infectious. Lastly, individuals can transition into the recovered class at a rate of  $\gamma_1$  or  $\gamma_2$ . Due to the short duration of the epidemic, we assume that the recovered individuals are immune for the duration of the outbreak. The definition and values of parameters are summarized in Table 2. For Ebola parameter values, we used the outbreak in Sierra Leon in 2014 [7, 12, 19, 31] and for the MERS 2015 outbreak in South Korea [9, 10, 31, 35]. Note that these parameter values are taken from a single outbreak of MERS and Ebola, which means that they vary from other outbreaks and may present some constraints when asserting conclusions for outbreaks of the same infectious disease [14, 20]. However, the parameter values used from the two outbreaks provide an excellent baseline for our model simulations.

## 2.1 Basic Reproduction Number

We compute the reproduction number for the ODE system (2.1) using the next-generation matrix [34]. The basic reproduction number,  $\mathcal{R}_0$ , is defined as the number of secondary cases produced by the introduction of a single infected individual into a fully susceptible population. If  $\mathcal{R}_0 > 1$ , an outbreak occurs in the ODE model. We start by defining two matrices,  $F$  and  $V$ , where the  $F$  matrix represents the newly infected rates in the system, and  $V$  represents the remaining rates in the infected compartments, Eqs. (1) and (2) respectively. The matrix  $F - V$  is the Jacobian matrix of the infected compartments evaluated at the disease-free equilibrium (DFE), where  $\bar{S}_i = N_i(0)$  and  $E_i = A_i = I_i = R_i = 0$ ,  $i = 1, 2$ . We find the spectral radius of the matrix  $FV^{-1}$  (Appendix 1), which equals the basic reproduction number,

$$\mathcal{R}_0 = \underbrace{\frac{\beta_1 \frac{N_1}{N} (\gamma_1 + \delta_1 + \mu_{I1})}{(\delta_1 + \mu_{A1})(\gamma_1 + \mu_{I1})}}_{NS} + \underbrace{\frac{\beta_2 \frac{N_2}{N} (\gamma_2 + \delta_2 + \mu_{I2})}{(\delta_2 + \mu_{A2})(\gamma_2 + \mu_{I2})}}_{SS}. \quad (2.2)$$

The basic reproduction number has the form typical of a multigroup/stage progression model [34]. It is the sum of two basic reproduction numbers, one for each group, NS when  $i = 1$  and SS when  $i = 2$ . In particular,  $\mathcal{R}_0 = \sum_{i=1}^2 \mathcal{R}_0^i$ , where

$$\mathcal{R}_0^i = \frac{\beta_i N_i / N}{\delta_i + \mu_{A_i}} + \frac{\beta_i (N_i / N) \delta_i}{(\delta_i + \mu_{A_i})(\gamma_i + \mu_{I_i})}, \quad i = 1, 2.$$

The two terms in the preceding expression represent new infections resulting from either the asymptomatic stage  $A_i$  or from the infectious stage  $I_i$ ,  $i = 1, 2$ . In addition for group  $i$ , the term  $\beta_i(N_i/N)$  is the number of successful transmissions from an individual in stage  $A_i$  (first term) or from an individual in stage  $I_i$  (second term) that result in exposed individuals. The term  $1/(\delta_i + \mu_{A_i})$  is the average length of the asymptomatic stage while  $1/(\gamma_i + \mu_{I_i})$  is the average length of the infectious stage, and  $\delta_i/(\delta_i + \mu_{A_i})$  is the probability of transitioning from  $A_i$  to  $I_i$ . For parameter values in Table 2 and for equal proportion of SS and NS,  $N_1/N = 0.5 = N_2/N$ , the basic reproduction number for MERS is  $\mathcal{R}_0 = 2.36$  and for Ebola it is  $\mathcal{R}_0 = 3.75$ .

### 3 Markov Chain Model

If the number of hosts/pathogens is sufficiently small, an ODE model is not appropriate. To that end, we utilize a continuous-time Markov chain (CTMC) model, which is continuous in time and discrete in the state space, to study the variability at the initiation of an outbreak, in time to outbreak, and in the peak level of infection. For simplicity, we use the same notation for the state variables as in the ODE model. In particular, time  $t \in [0, \infty)$  and the states are discrete random variables, e.g.,  $S_i, E_i, A_i, I_i, R_i \in \{0, 1, 2, \dots\}$ . The Markov property implies that the future states of the stochastic process only depend on the current states. In particular, there is an exponential waiting time between events.

To formulate a CTMC, it is necessary to define the infinitesimal transition probabilities corresponding to each change (event) in the state variables. The CTMC model consists of 12 distinct events, six events for each of the groups, NS and SS. The changes and the corresponding infinitesimal transition rates are summarized in Table 3.

#### 3.1 Branching Process Approximation

The theory of multitype (Galton–Watson) branching processes has a long history (e.g., [2, 13, 36] and references therein). It has been used to approximate the dynamics of the CTMC model near the DFE and the stochastic threshold for a disease outbreak [1–3, 36]. In fact, the stochastic threshold (i.e., probability of a disease outbreak) is directly related to the basic reproduction number as defined in

**Table 3** State transitions and rates for the CTMC model with Poisson rates  $r_i \Delta t + o(\Delta t)$ 

Event	Description ( $i = 1, 2$ )	Transition	Rate, $r_i$
1, 2	Infection of $S_i$	$S_i \rightarrow S_i - 1$ $E_i \rightarrow E_i + 1$	$\frac{S_i}{N} \sum_{k=1}^2 \beta_k (A_k + I_k)$
3, 4	Transition to $A_i$	$E_i \rightarrow E_i - 1$ $A_i \rightarrow A_i + 1$	$\alpha_i E_i$
5, 6	Transition to $I_i$	$A_i \rightarrow A_i - 1$ $I_i \rightarrow I_i + 1$	$\delta_i A_i$
7, 8	Death of $A_i$	$A_i \rightarrow A_i - 1$	$\mu_{A_i} A_i$
9, 10	Death of $I_i$	$I_i \rightarrow I_i - 1$	$\mu_{I_i} I_i$
11, 12	Recovery of $I_i$	$I_i \rightarrow I_i - 1$ $R_i \rightarrow R_i + 1$	$\gamma_i I_i$

the corresponding deterministic model (2.1) (see [3, 36]). More specifically, if the basic reproduction is less than unity, then disease extinction occurs with probability one. In this case, the branching process is called subcritical. However, if the basic reproduction number is greater than unity, the probability of disease extinction is less than one (probability of outbreak is greater than zero) and the process is referred to as supercritical.

In what follows, we will apply a multitype branching process approximation of the CTMC model at the DFE to estimate disease extinction probability. First, let us define the offspring probability generating function (pgf) for the exposed, asymptomatic, and infectious individuals in NS and SS. Let  $X = (X_1, \dots, X_6) := (E_1, A_1, I_1, E_2, A_2, I_2)$  be a vector of integer-valued random variables and  $\delta_{ij}$  denote the Kronecker delta (i.e.,  $\delta_{ij} = 1$  if  $i = j$  and zero otherwise). In general, the offspring pgf for type  $i$  given  $X_j(0) = \delta_{ij}$  is a function from  $[0, 1]^6$  to  $[0, 1]$ , and it takes the form:

$$f_i(x_1, \dots, x_6) = \sum_{k_1=0}^{\infty} \dots \sum_{k_6=0}^{\infty} P_i(k_1, k_2, \dots, k_6) x_1^{k_1} \dots x_6^{k_6}. \quad (3.1)$$

Here,  $P_i(k_1, k_2, \dots, k_6)$  is the probability that the individual of type  $i$  gives ‘‘birth’’ to  $k_j$  individuals of type  $j$  for  $j = 1, 2, \dots, 6$ . In particular, the pgfs  $f_i : [0, 1]^6 \rightarrow [0, 1]$  ( $i = 1, \dots, 6$ ) are given by:

$$\begin{aligned} f_1(x_1, x_2, \dots, x_6) &= x_2, \\ f_2(x_1, x_2, \dots, x_6) &= \frac{\frac{N_1}{N} \beta_1 x_1 x_2 + \frac{N_2}{N} \beta_1 x_2 x_4 + \delta_1 x_3 + \mu_{A1}}{\beta_1 + \delta_1 + \mu_{A1}}, \\ f_3(x_1, x_2, \dots, x_6) &= \frac{\frac{N_1}{N} \beta_1 x_1 x_3 + \frac{N_2}{N} \beta_1 x_3 x_4 + \mu_{I1} + \gamma_1}{\beta_1 + \mu_{I1} + \gamma_1}, \\ f_4(x_1, x_2, \dots, x_6) &= x_5 \end{aligned}$$



$$f_5(x_1, x_2, \dots, x_6) = \frac{\frac{N_1}{N}\beta_2 x_1 x_5 + \frac{N_2}{N}\beta_2 x_5 x_4 + \delta_2 x_6 + \mu_{A2}}{\beta_2 + \delta_2 + \mu_{A2}},$$

$$f_6(x_1, x_2, \dots, x_6) = \frac{\frac{N_1}{N}\beta_2 x_1 x_6 + \frac{N_2}{N}\beta_2 x_6 x_4 + \mu_{I_2} + \gamma_2}{\beta_2 + \mu_{I_2} + \gamma_2}.$$

According to the theory of multitype branching processes [5, 13], the fixed points of the offspring pgfs give an estimate of the disease extinction probability. Let  $(q_1, q_2, q_3, q_4, q_5, q_6)$  be the minimal fixed points of pgfs; that is,  $f_i(q_1, \dots, q_6) = q_i$  for  $i = 1, \dots, 6$ . Then, an estimate of the extinction probability given  $X(0) = (a_1, e_1, i_1, a_2, e_2, i_2)$  is

$$P_{ext} = \lim_{t \rightarrow \infty} P(X(t) = \mathbf{0}) = q_1^{a_1} q_2^{e_1} q_3^{i_1} q_4^{a_2} q_5^{e_2} q_6^{i_2},$$

and hence the probability of an outbreak is

$$P_{out} = 1 - P_{ext} = 1 - q_1^{a_1} q_2^{e_1} q_3^{i_1} q_4^{a_2} q_5^{e_2} q_6^{i_2}.$$

However, due to the simplicity of  $f_1$  and  $f_4$  (no deaths during stage  $E_i$ ), the pgfs can be simplified. That is,  $x_1 = x_2$  and  $x_4 = x_5$ . Therefore, we only solve for  $q_2, q_3, q_5,$  and  $q_6$ .

The expectation matrix  $M = (m_{ij})$  can be shown to be directly related to the basic reproduction number [3] with  $m_{ij} = \frac{\partial f_j}{\partial x_i} |_{X=1}$ . We include this calculation in Appendix 2. It is known that the spectral radius of  $M$ , denoted as  $\rho(M)$ , determines whether the disease extinction probability is equal to or less than the unity [2, 3, 13]. Specifically, if  $\rho(M) < 1$ ,  $q_1 = \dots = q_6 = 1$ , then the extinction probability is one; if  $\rho(M) > 1$ , then there exists a unique fixed point  $(q_1, \dots, q_6) \in (0, 1)^6$ , and hence the extinction probability is strictly less than one. By the Threshold Theorem of reference [3], it follows that the spectral radius of the matrix  $M$  is strictly less than one if and only if the basic reproduction number is strictly less than one. Analogous statements hold whenever the spectral radius of  $M$  is equal to one or is strictly greater than one.

## 4 Parameter Sensitivity Analysis

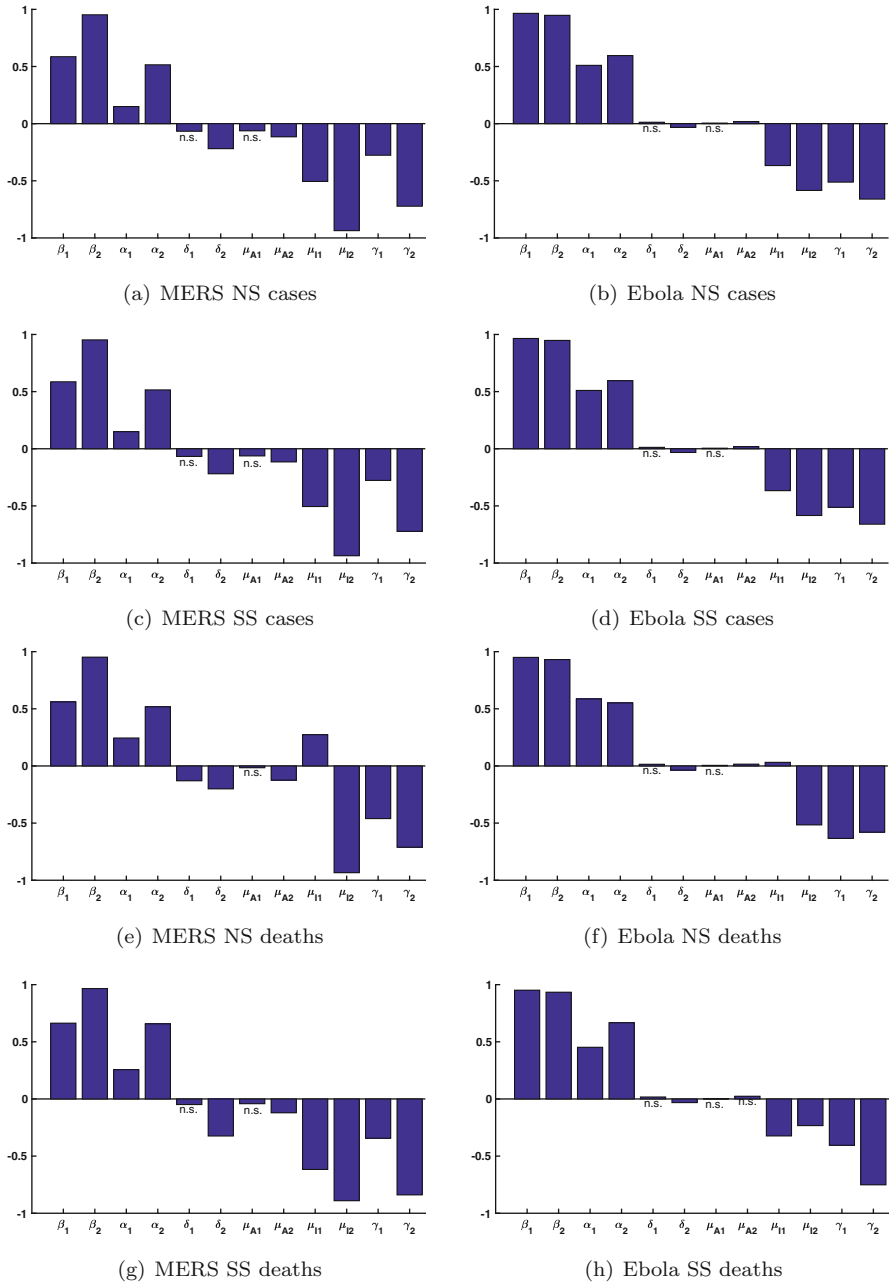
We perform a sensitivity analysis on the parameters ranges given in Table 2 for the ODE models for MERS and Ebola using a uniform distribution for the values. Latin hypercube sampling (LHS), first developed by McKay et al. [29], with the statistical sensitivity measure partial rank correlation coefficient (PRCC), performs a sensitivity analysis that explores a defined parameter space of the model. The parameter space considered is defined by the parameter intervals depicted in Table 2. Rather than simply exploring one parameter at a time with

other parameters held fixed at baseline values, the LHS/PRCC sensitivity analysis method globally explores multidimensional parameter space. LHS is a stratified Monte Carlo sampling without replacement technique that allows an unbiased estimate of the average model output with limited samples. The PRCC sensitivity analysis technique works well for parameters that have a nonlinear and monotonic relationship with the output measure. PRCC shows how the output measure is influenced by changes in a specific parameter value when the linear effects of other parameter values are removed. The PRCC values were calculated as Spearman (rank) partial correlations using the *partialcorr* function in MATLAB 2016. Their significances, uncorrelated  $p$ -values, were also determined. The PRCC values vary between  $-1$  and  $1$ , where negative values indicate that the parameter is inversely proportional to the output measure. Following Marino et al. [27], we performed a  $z$ -test on transformed PRCC values to rank significant model parameters in terms of relative sensitivity. According to the  $z$ -test, parameters with larger magnitude PRCC values had a stronger effect on the output measures.

We start by verifying the monotonicity of the output measures. Monotonicity was observed for all parameters except  $\mu_{I2}$  with total SS deaths, which exhibited two monotonic ranges  $[0.02, 0.0278]$  and  $[0.0278, 0.14]$ . For non-monotonic trends, alternative methods based on decomposition of model output variances such as eFAST (extended Fourier Amplitude Sensitivity Test) can be used instead of PRCC [27]; however, since all other parameters were monotonic, we use PRCC and just consider the two monotonic ranges of  $\mu_{I2}$  separately. PRCC analysis of these two ranges produces similar results. For an analysis of the monotonicity, refer to Appendix 3. Once meeting the monotonicity requirements, we proceed to utilize LHS with PRCC for both MERS and Ebola parameters. For each disease, we calculate the PRCC for the following output measures: total NS cases, total SS cases, total NS deaths, and total SS deaths. The number of total cases refers to the total number of transmission events where susceptible individuals become exposed (latently infected) individuals. For the outputs of NS/SS cases, the PRCC results were similar in both Ebola and MERS. According to the PRCC values, the  $\beta_2$  and  $\mu_{I2}$  are significant in the model for MERS. Meanwhile, in the Ebola model, both transmission parameters are significant in the model (see Fig. 2). Note that  $\beta_2$  is calculated from  $\mathcal{R}_0$ , which we will vary later in simulations.

## 5 CTMC Analysis

For the CTMC model, we numerically simulate sample paths to compute the probability of an outbreak, number of deaths, time to outbreak, time to peak infection, and peak number of infectious individuals. For sample paths and probability of outbreak, we compare our results with the deterministic model. In the remainder of this analysis, we assume that the initial total population size is  $N(0) = 2000$ . Reference to infected individuals will imply the variables  $I_1$  and  $I_2$ , unless stated otherwise. For example, peak number of infectious individuals refers to the maximum value of



**Fig. 2** PRCC values for output measures (a)–(b) number of NS cases, (c)–(d) number of SS cases, (e)–(f) number of NS deaths, and (g)–(h) number of SS deaths with  $\mu_{I2}$  range [0.02, 0.0278]. The number of total cases refers to the total number of transmission events where susceptible individuals become exposed (latently infected) individuals. *P*-values that are greater than 0.05 are labeled as not significant (n.s.)

$I_1 + I_2$  and the time to peak infection refers to the time  $t$  at which this maximum occurs. However, an outbreak means that the total number in classes  $E_i$ ,  $A_i$ , and  $I_i$  for both NS and SS has reached at least 50, i.e.,  $\sum(E_i + A_i + I_i) \geq 50$ . In addition, we note that for the CTMC model, an outcome measure (e.g., peak values, time to peak, and number of deaths) is defined by a corresponding probability distribution and a sample path yields one outcome from the distribution.

## 5.1 Sample Paths

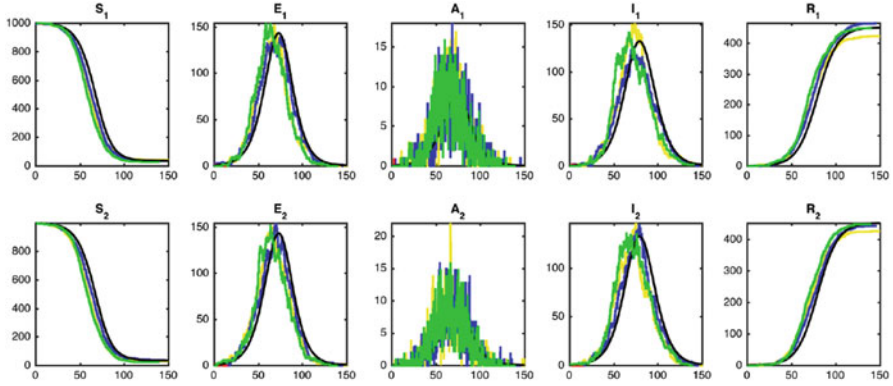
An example of the sample paths resulting from our CTMC model is shown in Fig. 3, for both MERS and Ebola cases. These sample paths are generally well aligned with the population average response that is captured by our ODE model (shown with a black line). However, the sample paths of the CTMC model illustrate the potential variability in timing of the peak level of infection and the peak number of infectious individuals. Note that some sample paths are not shown because in those simulations the disease becomes extinct. Also, note that the  $A$  class is not shown for Ebola (Fig. 3b) given that the asymptomatic stage is extremely short for this disease.

## 5.2 Probability of Outbreak

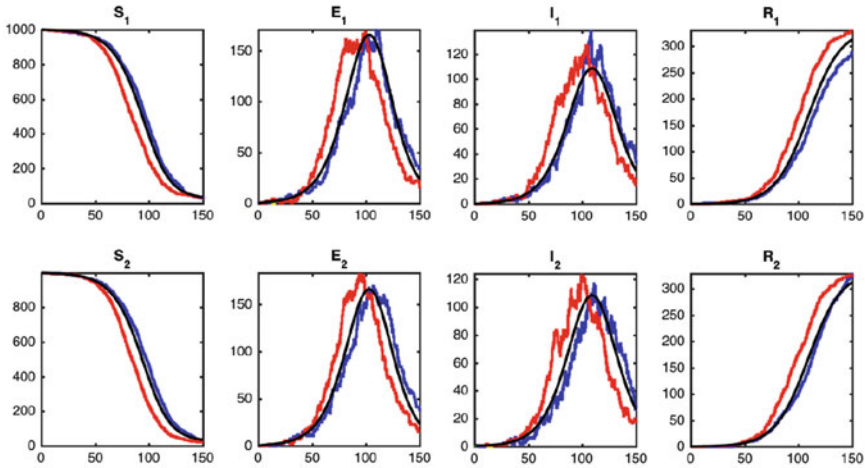
Next, in order to do a comprehensive comparison of the stochastic simulation and ODE model results, we probe the relationship between two model parameters—the value of  $\mathcal{R}_0$  (Fig. 4) as well as the fraction of the susceptible population that are in the SS class (Fig. 4) and a key model output: the probability of outbreak. Probability of outbreak is defined by monitoring the number of people in the  $E$ ,  $A$ , and  $I$  classes and an outbreak is declared when the cumulative size of these compartments reaches the threshold value of 50. Although the value of 50 appears relatively large, it is reasonable given that we are counting the cumulative number in all three classes for a relatively large population size of 2000. For these simulations, we vary  $\beta_2$  given the significant effect of this parameter on the model outputs as confirmed by the LHS analysis.

We note a negative correlation between the proportion of SS in the S class and the probability of outbreak (Fig. 4) and attribute this to the fact that the value of  $\beta_2$  is varied in order to maintain a constant value of  $\mathcal{R}_0$  (MERS,  $\mathcal{R}_0 = 2.5$  and Ebola,  $\mathcal{R}_0 = 2.39$ ). In other words, as the fraction of SS susceptible individuals is increased, the value of  $\beta_2$  decreases and results in a reduction in the probability of outbreak ( $1 - q_6$ ). Results in Fig. 4 are shown only for  $q_3$  and  $q_6$  since these outputs are similar to  $q_1$  and  $q_4$ , respectively. We also note that  $q_1 = q_2$  and  $q_4 = q_5$  and therefore exclude those plots as well.

As expected, the probability of an outbreak is dependent on the initial fraction of the population that is infected, with an increasing chance of an outbreak



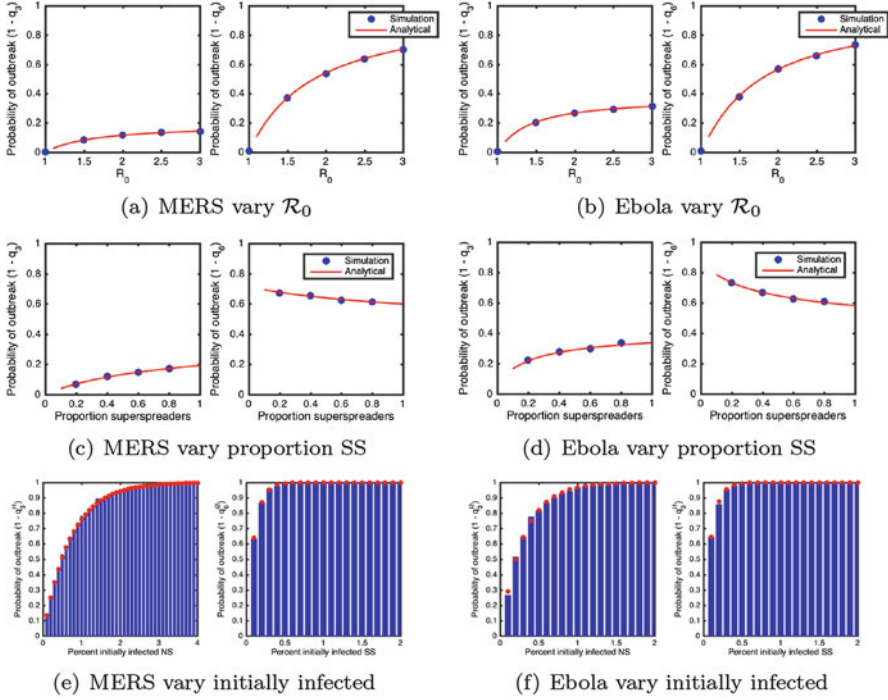
(a) MERS epidemic dynamics



(b) Ebola epidemic dynamics

**Fig. 3** Sample paths of the (a) MERS and (b) Ebola epidemics. Four sample paths (in color) of the CTMC system are shown and overlaid on the deterministic ODE model (in black). In (a) only three sample paths and in (b) only two sample paths are visible on the graphs, as one sample path in (a) and two sample paths in (b) did not result in an outbreak. The visible sample paths illustrate the potential variability in timing and epidemic size for a total population size of  $N_1 = 1000$  and  $N_2 = 1000$ , with one initial infected SS individual ( $I_2(0) = 1$ ) and all NS individuals are susceptible,  $I_1(0) = 0$

as the number of initially infected individuals increases (Fig. 4). Furthermore, the probability of outbreak is significantly enhanced when the initially infected population is composed of SS rather than NS individuals. We also find a strong agreement between the probability of outbreak predicted by stochastic simulations of the CTMC model and the associated branching process approximations for all of these analyses (Fig. 4a–f).

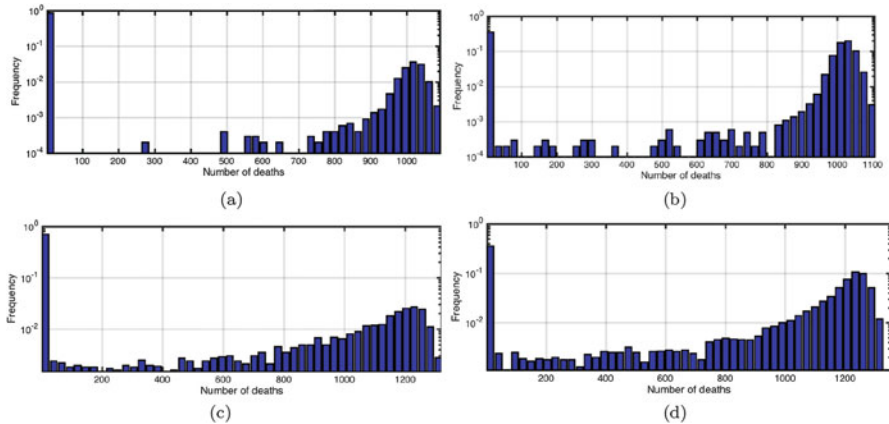


**Fig. 4** Probability of an outbreak as a function of  $\mathcal{R}_0$  for MERS in (a) and Ebola in (b), the Proportion SS Susceptible Individuals for MERS in (c) and Ebola in (d), and the Number Initially Infected for MERS in (e) and Ebola in (f). The stochastic simulations (red) and analytical calculations (blue dots) are overlaid for MERS in (a), (c), and (e), and Ebola in (b), (d), and (f). The total population size is  $N_1 = N_2 = 1000$  and the probability of an outbreak  $(1 - q_3)$ , after introducing one infected NS,  $I_1(0) = 1$ , is shown for MERS and Ebola on the left in (a)–(f). Similarly, the probability of an outbreak  $(1 - q_6)$ , after introducing one infected SS,  $I_2(0) = 1$ , is shown for MERS and Ebola on the right in (a)–(f)

### 5.3 Number of Deaths

Utilizing our stochastic model of MERS and Ebola dynamics within a population of individuals, we next sought to investigate whether the presence of SS individuals within the population could be reflected in key metrics that capture the severity of disease outbreak: the number of deaths, time to disease outbreak, probability of outbreak, time to peak number of infections, and the peak number of infectious individuals.

We first assess the impact of SS individuals on the number of deaths that accumulate over a 150-day time frame following disease initiation. We observe a modest increase in the frequency of deaths as the size of the susceptible SS class of individuals is increased from 5 to 50% of the total population for both MERS and Ebola disease simulations (not shown). We note a higher frequency of epidemics

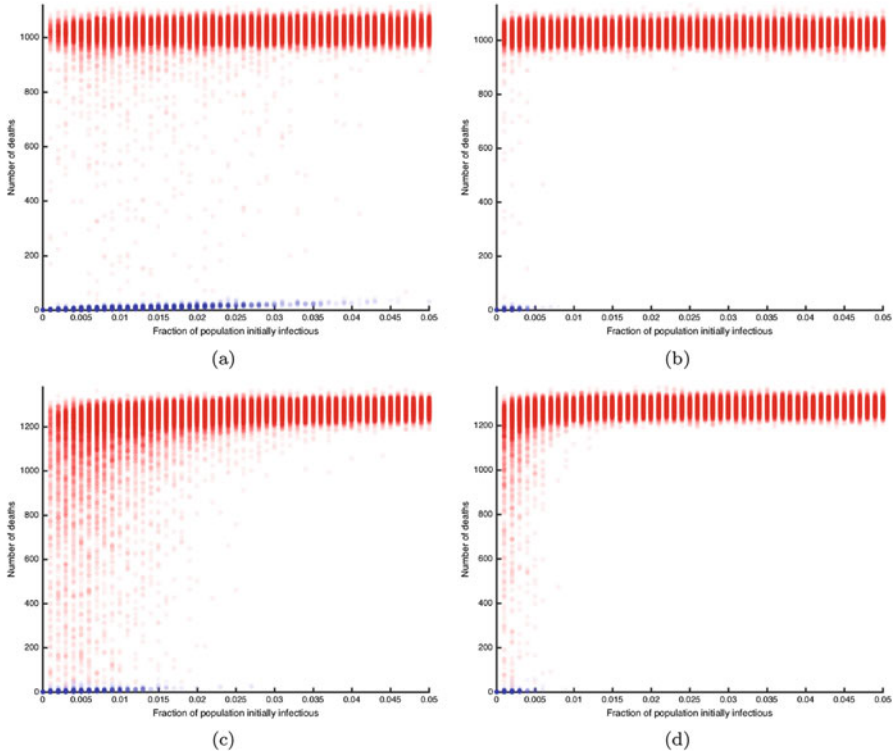


**Fig. 5** Total number of deaths when initial infected is varied. Histograms of the number of deaths by day 150 calculated from 10,000 sample paths, with a total initial population size of  $N_1 = N_2 = 1000$ . The initial number of infected is varied in each panel. In MERS simulations, one NS is introduced  $I_1(0) = 1$  and  $I_2(0) = 0$  in (a) and one SS is initiated in  $I_1(0) = 0$  and  $I_2(0) = 1$  in (b). Similarly, in (c) one NS is introduced and in (d) one SS is introduced. The distributions are bimodal

with lower numbers of deaths when the fraction of SS individuals in the susceptible fraction is lower (not shown).

For all subsequent simulations, we initialize the population consisting of 1000 SS and 1000 NS susceptible individuals. Most notably, there is a ten-fold increase in the frequency of deaths expected when the initial infected individual (for both MERS and Ebola) is an SS rather than an NS (Fig. 5). The statistical significance of the difference between NS- and SS-initiated epidemics is confirmed with a Kolmogorov–Smirnov test ( $p < 0.001$ ) [28]. It is clear that the distributions are bimodal. This is due to the fact that there may be only a minor outbreak (with probability  $q_3$  or  $q_6$ ) with none or a few deaths or a major outbreak (with probability  $1 - q_3$  or  $1 - q_6$ ) with a significant number of deaths.

We next explore the relationship between the number of deaths and the number of initially infected individuals. For each fraction of the population initially infected, there are 1000 points, one point from each of the 1000 sample paths, representing the total number of deaths over a 150-day time period. As expected, we find that as the number of initially infected NS individuals increases, the expected number of deaths increases as well (Fig. 6). Interestingly, we find a threshold response as the fraction of initially infected SS individuals increases. As the fraction of initially infected SS individuals increases beyond 0.005 for MERS (Fig. 6b) and 0.0075 for Ebola (Fig. 6d), we find that the simulation always gives rise to an outbreak, resulting in a maximal number of around 1000 deaths over a 150-day simulated period. We also note that there is a decrease in the variability in the number of deaths when the outbreak is initiated by an SS rather than an NS, which contributes to this threshold response. The seemingly binary response in the number of deaths resulting from a



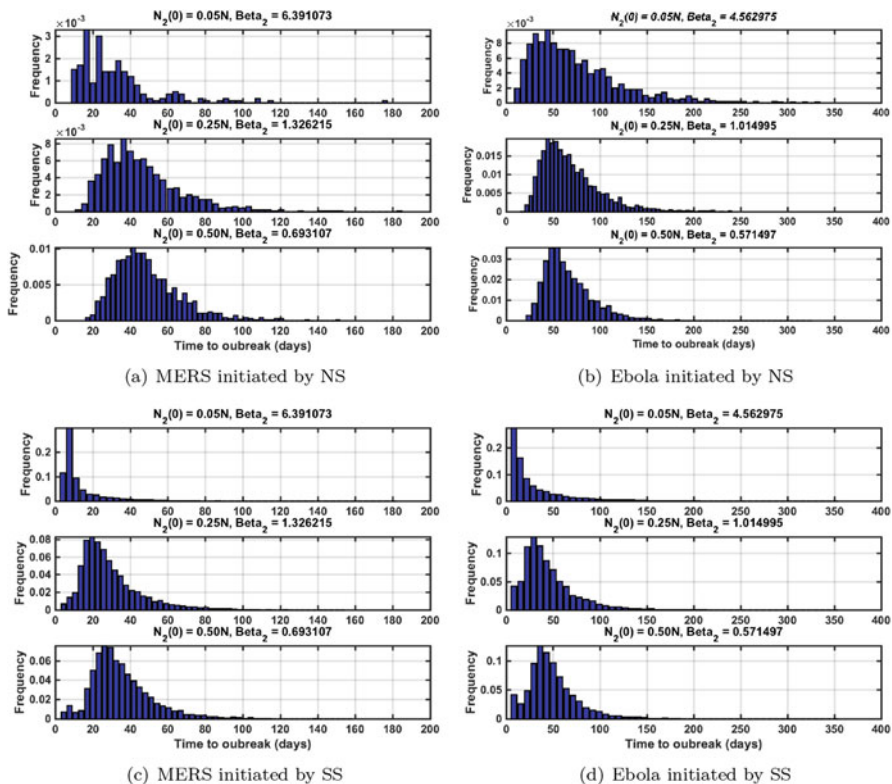
**Fig. 6** Number of deaths as a function of fraction of initially infected. Scatterplot of the number of deaths over a 150-day period calculated from 1000 sample paths with a total initial population  $N_1 = N_2 = 1000$ . The fraction of the population is initially increasing for NS (a) and SS (b) for MERS and NS (c) and SS (d) for Ebola. Outbreaks (red) and non-outbreak cases (blue) are shown

MERS or Ebola epidemic initiated by infected SS individuals who only contribute to 0.5–0.75% of the starting population is a good indication that by tracking the number of deaths in an epidemic, the presence of an SS may be predicted. Thus, while the observation that an outbreak has occurred does not necessarily suggest the existence of SS individuals in the population, the severity of the outbreak in terms of lives lost may be more suggestive of the presence of an SS, especially when the number of known initial infections is low.

## 5.4 Time to Outbreak

Similarly, we find that the time to outbreak—where an outbreak is defined as 50 or more people in all of the  $E$ ,  $A$ , and  $I$  classes—is reduced when the initial

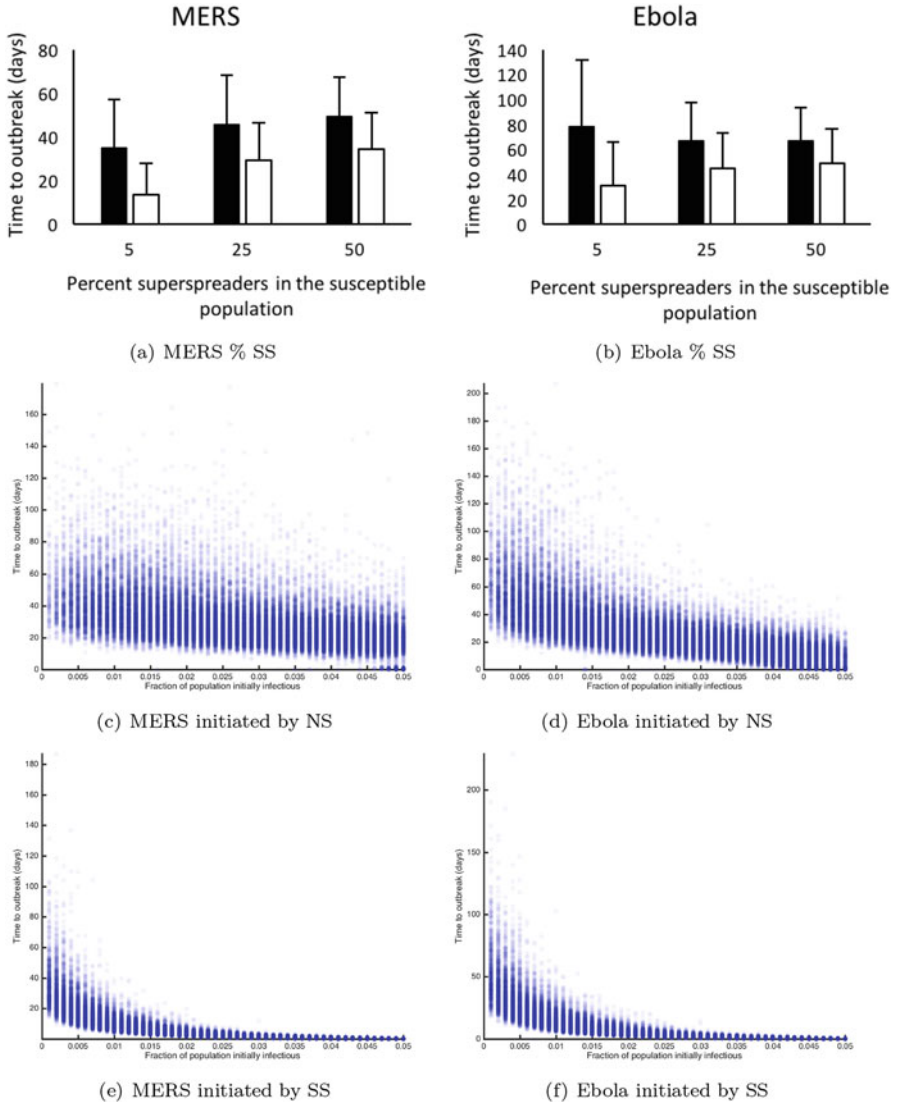




**Fig. 7** Time to outbreak for MERS (a) and (c), and Ebola (d) and (b). Calculations for 10,000 sample paths with an increasing fraction of SS in the susceptible populations are shown. The epidemic is initiated with a single infected (a)–(b) NS individual and (c)–(d) SS individual ( $I_i(0) = 1$ ), and  $E_i(0) = 0$  and  $R_i(0) = 0$ . The initial population is  $N = 2000$  individuals where  $\mathcal{R}_0 = 2.5$  for MERS is held constant. Simulations (a) and (c) are run for three cases:  $N_2(0) = 0.05N$  and  $\beta_2 = 6.391$  (top),  $N_2(0) = 0.25N$  and  $\beta_2 = 1.326$  (middle), and  $N_2(0) = 0.50N$  and  $\beta_2 = 0.693$  (bottom).  $\mathcal{R}_0 = 2.5$  for Ebola is held constant. Simulations (b) and (d) are run for three cases:  $N_2(0) = 0.05N$  and  $\beta_2 = 4.562$  (top),  $N_2(0) = 0.25N$  and  $\beta_2 = 1.014$  (middle), and  $N_2(0) = 0.50N$  and  $\beta_2 = 0.571$

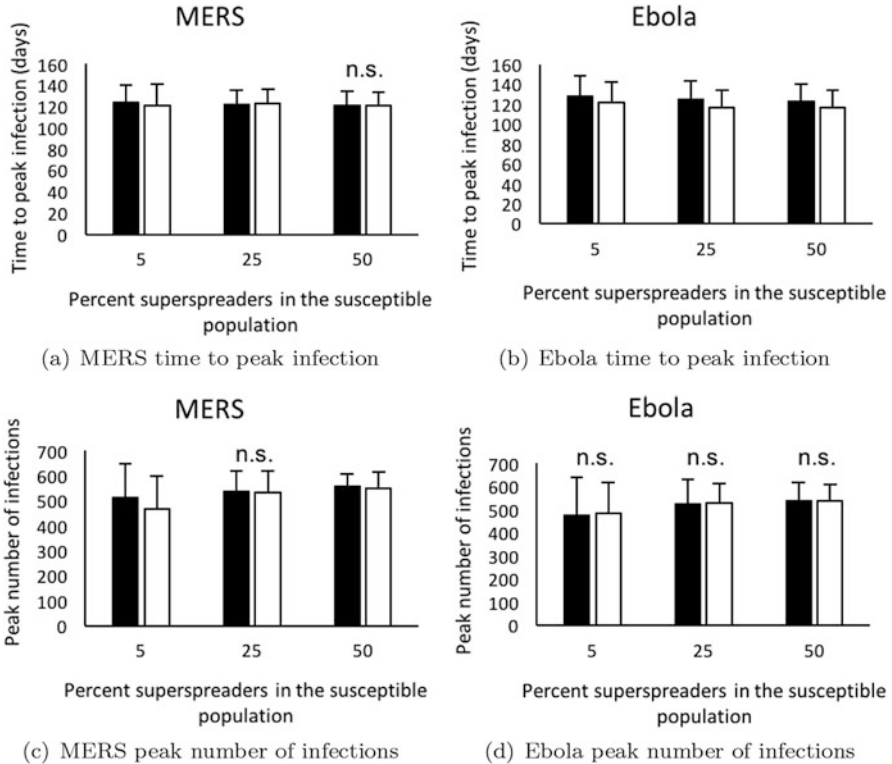
infected individual in a simulated MERS or Ebola disease situation is an SS rather than an NS (Fig. 7a, c). We confirmed that this reduction is, indeed, statistically significant (Fig. 8a–b). These results also illustrate that as the fraction of susceptible SS increases the time to outbreak increases as well, which we attribute to the fact that  $\beta_2$  values decrease (detailed in Figs. 7 and 8). In Fig. 7, each distribution is based on 10,000 sample paths, whereas in Fig. 8, for each fraction initially infected, the time points are based on 1000 sample paths.

We also find a clear separation between the time to outbreak of an epidemic initiated by a fraction of SS versus NS infected individuals. Mean differences were



**Fig. 8** Time to outbreak as a function of percent of SS and fraction of initially infected. In (a)–(b), time to outbreak (mean  $\pm$  SD) for 500 sample paths initiated with a single infected NS (black) versus SS (white) individual. In (c)–(f), time to outbreak for 1000 sample paths with an increasing fraction of initially infected

significantly distinct for each percentage in (Fig. 8a–b),  $p < 0.001$ . In fact, if a MERS or Ebola outbreak is initiated by 1.5% or more of the initial population size and these individuals are SS, then the time to outbreak is predicted to be no more than 20 days where an outbreak is defined as 2.5% of the population becoming infected (Fig. 8c–f).



**Fig. 9** Time to peak infection and peak number of infections over percent of SS. (a)–(b) compare time to outbreak (mean  $\pm$  SD), where (c) and (d) compare the number of peak infections initiated with a single infected NS (black) versus SS (white) individual for MERS in (a) and (c), and Ebola in (b) and (d). The initial population is  $N = 2000$ . The fraction of susceptible SS is increased for comparison. All results were statistically significant when  $p < 0.05$  from t-test (two-tailed). Comparisons that are not statistically significant were denoted n.s.

### 5.5 Time to Peak Infection and Peak Number of Infectious Individuals

Given that the time to outbreak shows a significant difference between epidemics initiated by SS versus NS individuals, we next asked whether SS-initiated epidemics will also reach peak infection in a shorter time. To investigate this, we calculated mean ( $\pm$  SD) of time to peak infection (in days) for MERS and Ebola, where the percent of SS varied in the susceptible population (see Fig. 9a–b). Mean differences between the introduction of 1 infected NS (black) compared to 1 infected SS (white) were assessed separately as the percent of SS varied (e.g., 5%, 25%, and 50%) using t-tests where statistical significance was accepted when  $p < 0.05$ . For MERS, time to peak infection was slightly significantly lower for SS when 5%

and 25% of the susceptible population was SS, but not significant when 50% of the population was SS (Fig. 9a). For Ebola, time to peak infection was significantly lower for SS regardless of changes in the percent of SS in the susceptible population (Fig. 9b). Hence, while the differences in mean time to peak infection between SS and NS-initiated epidemics are only modestly different, we find their difference to be statistically significant (Fig. 9a–b). Thus, this confirmed that epidemics initiated by infected SS individuals reaches its peak value more quickly.

We repeated the same analysis to assess mean differences in the peak number of infections. Surprisingly, we did not find a significant difference between the peak number of infections for epidemics initiated from a single infected NS versus SS individual for Ebola (Fig. 9d). However, significant differences were observed when 5% or 50% of the susceptible population was SS for MERS.

## 6 Discussion

In this investigation, we capture the dynamics of MERS and Ebola epidemics by applying both deterministic and stochastic modeling strategies. To investigate the role of SS on the epidemic dynamics and to compare our results, we keep the  $\mathcal{R}_0$  constant for both MERS and Ebola while varying  $\beta_2$ , the transmission rate of SS. Parameter sensitivity analysis, using Latin hypercube sampling and partial rank correlation coefficient, shows that  $\beta_2$  has a significant effect on all the output measures (Fig. 2).

From Fig. 4, we can conclude that the stochastic model simulations agree with the branching process analytical results. As the value of  $\mathcal{R}_0$  increases, we observe that the probability of an outbreak increases for both diseases. This result is expected since more individuals in the population are infected. The probability of an outbreak is greater for Ebola than MERS, which is due to the transmission parameters for Ebola being larger than MERS. Furthermore, these results show that if the outbreak is initiated by an SS, then the probability of an outbreak is significantly higher. Additionally, fewer SS individuals than NS individuals are sufficient to cause an outbreak irrespective of the disease (MERS or Ebola).

As an outbreak initiated by SS has a greater probability of occurrence and peaks earlier than with NS, the accumulated number of deaths is more severe in an epidemic initiated with the same proportion of SS than NS (Figs. 5 and 6). Disease severity (number of deaths) for both MERS and Ebola occurs earlier with SS than NS. Our findings agree with prior epidemiological studies on superspreading events [15, 32, 40]. For example, the 2003 outbreak of the respiratory infection SARS in Beijing found that SS had higher mortality rates, higher attack rates, and greater number of contacts in comparison to NS [40]. From a public health perspective, as SS events will be observed more frequently, intervention/prevention methods must have rapid response to reduce disease severity. For example, Wong et al. [40] suggested that several community-based efforts could have been made to reduce the number of MERS and Ebola cases in Guinea and Sierra Leone, such as tracking

contacts, earlier diagnosis, treatment strategies, and community education. Effective responses to control superspreading events and reduce disease transmission in MERS and Ebola outbreaks included: “early discovery, diagnosis, intervention, and quarantine of confirmed cases.” [40]. Other epidemics that are more likely to occur in hospital settings, e.g., SARS, could be controlled through hospital administrative strategies, such as reducing contact between the infected patient and healthcare workers, visitors, or other patients whose immune system may be comprised due to other infections [32]. Thus, a rapid response is needed to reduce disease severity of SS events.

Evident in Figs. 7 and 8, when an outbreak is initiated by an SS rather than an NS, the time to outbreak is shorter and has less variability. Therefore, if the number of disease cases rises rapidly, there may be SS in the community. In this scenario, healthcare managers should search for potential SS. Similar results apply for time to peak infection, Fig. 9. If peak infection occurs quickly, it is more likely that there is an SS in the population.

Interestingly, varying the percentage of SS in the population has little influence on the peak number of infections (Fig. 9c, d). This is likely due to the fact that the  $\mathcal{R}_0$  values are held constant.

## 7 Future Work

We have formulated, analyzed, and numerically simulated deterministic and stochastic epidemic models that include heterogeneity in transmission for NS and SS. We applied our models to emerging and re-emerging infectious diseases, MERS and Ebola, where the models were parameterized with data from the literature but with a fixed initial population size of 2000. There are a number of extensions and generalizations that we will consider in the future work. We assumed homogeneous mixing and only two types of classifications of individuals (NS/SS) for the entire population. Generalizing this model to include heterogeneous mixing and spatial components are key features that can provide insight on how a superspreaders can be classified. In our model, we considered inter-host variability, which naturally leads to constructing a model with intra-host variability utilizing stochastic differential equations or other types of models. In addition, variability of the pathogen on epidemic dynamics can be explored. Additionally, we will validate our models’ findings against time series data, test our models’ abilities to detect the presence of SS, and interpret the results for public health implementation. Finding answers to these problems will lead to our ultimate goal of constructing novel ways to quantify, characterize, and identify an SS during the initiation of an outbreak.

**Acknowledgements** The work described in this chapter was initiated during the Association for Women in Mathematics collaborative workshop Women Advancing Mathematical Biology (WAMB) hosted by the Mathematical Biosciences Institute (MBI) at Ohio State University in April 2017. Funding for the workshop was provided by MBI, NSF ADVANCE “Career Advancement for

Women Through Research-Focused Networks” (NSF-HRD 1500481), Society for Mathematical Biology, and Microsoft Research. We give special thanks to the WAMB organizers: Ami Radunskaya, Rebecca Segal, and Blerita Shtylla. Anarina L. Murillo acknowledges that this work has been supported in part by the grant T32DK062710 from the National Institute of Diabetes and Digestive and Kidney Diseases and grant T32HL072757 from the National Heart, Lung, and Blood Institute. Omar Saucedo acknowledges that this research has been supported in part by the MBI and the grant NSF-DMS 1440386. Nika Shakiba is the recipient of the NSERC Vanier Canada Graduate Scholarship. This work was partially supported by a grant from the Simons Foundation (#317047 to Xueying Wang). In addition, we thank Texas Tech University for hosting our second WAMB group meeting and the Paul Whitfield Horn Professorship of Linda JS Allen for providing financial support. We thank the two anonymous reviewers for their helpful suggestions on the original manuscript.

## Appendix 1: Basic Reproduction Number Calculation

In the calculations below, we denote  $N_i(0)$  as  $N_i$  and  $N(0) = N_1(0) + N_2(0)$  as  $N$ :

$$F = \begin{pmatrix} 0 & \beta_1 \frac{N_1}{N} & \beta_1 \frac{N_1}{N} & 0 & \beta_2 \frac{N_1}{N} & \beta_2 \frac{N_1}{N} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \beta_1 \frac{N_2}{N} & \beta_1 \frac{N_2}{N} & 0 & \beta_2 \frac{N_2}{N} & \beta_2 \frac{N_2}{N} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad (1)$$

and

$$V = \begin{pmatrix} \alpha_1 & 0 & 0 & 0 & 0 & 0 \\ -\alpha_1 & \delta_1 + \mu_{A_1} & 0 & 0 & 0 & 0 \\ 0 & -\delta_1 & \mu_{I_1} + \gamma_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \alpha_2 & 0 & 0 \\ 0 & 0 & 0 & -\alpha_2 & \delta_2 + \mu_{A_2} & 0 \\ 0 & 0 & 0 & 0 & -\delta_2 & \mu_{I_2} + \gamma_2 \end{pmatrix}. \quad (2)$$

The next-generation matrix is

$$FV^{-1} = \begin{pmatrix} \xi_1 N_1 & \xi_1 N_1 & \frac{\beta_1 \frac{N_1}{N}}{\gamma_1 + \mu_{I_1}} & \xi_2 N_1 & \xi_2 N_1 & \frac{\beta_2 \frac{N_1}{N}}{\gamma_2 + \mu_{I_2}} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \xi_1 N_2 & \xi_1 N_2 & \frac{\beta_1 \frac{N_2}{N}}{\gamma_1 + \mu_{I_1}} & \xi_2 N_2 & \xi_2 N_2 & \frac{\beta_2 \frac{N_2}{N}}{\gamma_2 + \mu_{I_2}} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

where

$$\xi_i = \frac{\beta_i \frac{1}{N} (\gamma_i + \delta_i + \mu_{I_i})}{(\mu_{A_i} + \delta_i) (\gamma_i + \mu_{I_i})}, \quad i = 1, 2.$$

## Appendix 2: Expectation Matrix

For each pgf, we take partial derivatives with respect to  $x_1, \dots, x_6$  (Jacobian matrix of the functions  $f_1, \dots, f_6$ ), then evaluate at  $x_1, \dots, x_6 = 1$  and take the transpose of the matrix. The result is the expectation matrix,  $M$ , where Eq. (3) reduces to (4). In addition, we create the matrix  $W$ , Eq. (5), a diagonal matrix.

$$M = \begin{pmatrix} 0 & \frac{\frac{N_1}{N} \beta_1 x_2}{\beta_1 + \delta_1 + \mu_{A1}} & \frac{\frac{N_1}{N} \beta_1 x_3}{\beta_1 + \gamma_1 + \mu_{I1}} & 0 & \frac{\frac{N_1}{N} \beta_2 x_5}{\beta_2 + \delta_2 + \mu_{A2}} & \frac{\frac{N_1}{N} \beta_2 x_6}{\beta_2 + \mu_{I2} + \gamma_2} \\ 1 & \frac{\frac{N_1}{N} \beta_1 x_1 + \frac{N_2}{N} \beta_1 x_4}{\beta_1 + \delta_1 + \mu_{A1}} & 0 & 0 & 0 & 0 \\ 0 & \frac{\delta_1}{\beta_1 + \delta_1 + \mu_{A1}} & \frac{\frac{N_1}{N} \beta_1 x_1 + \frac{N_2}{N} \beta_1 x_4}{\beta_1 + \gamma_1 + \mu_{I1}} & 0 & 0 & 0 \\ 0 & \frac{\frac{N_2}{N} \beta_1 x_2}{\beta_1 + \delta_1 + \mu_{A1}} & \frac{\frac{N_2}{N} \beta_1 x_3}{\beta_1 + \gamma_1 + \mu_{I1}} & 0 & \frac{\frac{N_2}{N} \beta_2 x_5}{\beta_2 + \delta_2 + \mu_{A2}} & \frac{\frac{N_2}{N} \beta_2 x_6}{\beta_2 + \mu_{I2} + \gamma_2} \\ 0 & 0 & 0 & 1 & \frac{\frac{N_1}{N} \beta_2 x_1 + \frac{N_2}{N} \beta_2 x_4}{\beta_2 + \delta_2 + \mu_{A2}} & 0 \\ 0 & 0 & 0 & 0 & \frac{\delta_2}{\beta_2 + \delta_2 + \mu_{A2}} & \frac{\frac{N_1}{N} \beta_2 x_1 + \frac{N_2}{N} \beta_2 x_4}{\beta_2 + \mu_{I2} + \gamma_2} \end{pmatrix} \Big|_{x_i=1} \quad (3)$$

$$= \begin{pmatrix} 0 & \frac{\frac{N_1}{N} \beta_1}{\beta_1 + \delta_1 + \mu_{A1}} & \frac{\frac{N_1}{N} \beta_1}{\beta_1 + \gamma_1 + \mu_{I1}} & 0 & \frac{\frac{N_1}{N} \beta_2 x_5}{\beta_2 + \delta_2 + \mu_{A2}} & \frac{\frac{N_1}{N} \beta_2}{\beta_2 + \mu_{I2} + \gamma_2} \\ 1 & \frac{\beta_1}{\beta_1 + \delta_1 + \mu_{A1}} & 0 & 0 & 0 & 0 \\ 0 & \frac{\delta_1}{\beta_1 + \delta_1 + \mu_{A1}} & \frac{\beta_1}{\beta_1 + \gamma_1 + \mu_{I1}} & 0 & 0 & 0 \\ 0 & \frac{\frac{N_2}{N} \beta_1}{\beta_1 + \delta_1 + \mu_{A1}} & \frac{\frac{N_2}{N} \beta_1}{\beta_1 + \gamma_1 + \mu_{I1}} & 0 & \frac{\frac{N_2}{N} \beta_2}{\beta_2 + \delta_2 + \mu_{A2}} & \frac{\frac{N_2}{N} \beta_2}{\beta_2 + \mu_{I2} + \gamma_2} \\ 0 & 0 & 0 & 1 & \frac{\beta_2}{\beta_2 + \delta_2 + \mu_{A2}} & 0 \\ 0 & 0 & 0 & 0 & \frac{\delta_2}{\beta_2 + \delta_2 + \mu_{A2}} & \frac{\beta_2}{\beta_2 + \mu_{I2} + \gamma_2} \end{pmatrix}, \quad (4)$$

and

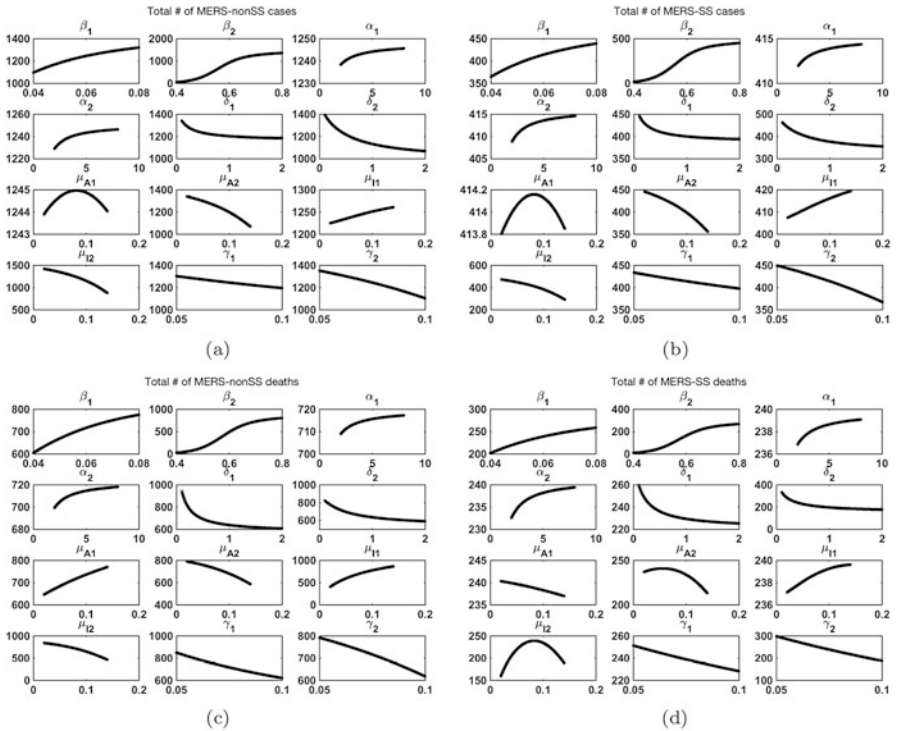
$$W = \begin{pmatrix} \alpha_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \beta_1 + \delta_1 + \mu_{A1} & 0 & 0 & 0 & 0 \\ 0 & 0 & \beta_1 + \mu_{I1} + \gamma_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \alpha_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \beta_2 + \delta_2 + \mu_{A2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \beta_2 + \mu_{I2} + \gamma_2 \end{pmatrix}. \quad (5)$$

Finally, we check that  $(M - I)W = F - V$ , where  $(M - I)W$  is displayed in Eq. (6),

$$\begin{pmatrix} -\alpha_1 & \frac{N_1}{N}\beta_1 & \frac{N_1}{N}\beta_1 & 0 & \frac{N_1}{N}\beta_2 & \frac{N_1}{N}\beta_2 \\ \alpha_1 & -\delta_1 - \mu_{A1} & 0 & 0 & 0 & 0 \\ 0 & \delta_1 & -\gamma_1 - \mu_{I1} & 0 & 0 & 0 \\ 0 & \frac{N_2}{N}\beta_1 & \frac{N_2}{N}\beta_1 & -\alpha_2 & \frac{N_2}{N}\beta_2 & \frac{N_2}{N}\beta_2 \\ 0 & 0 & 0 & \alpha_2 & -\delta_2 - \mu_{A2} & 0 \\ 0 & 0 & 0 & 0 & \delta_2 & -\mu_{I2} - \gamma_2 \end{pmatrix}. \quad (6)$$

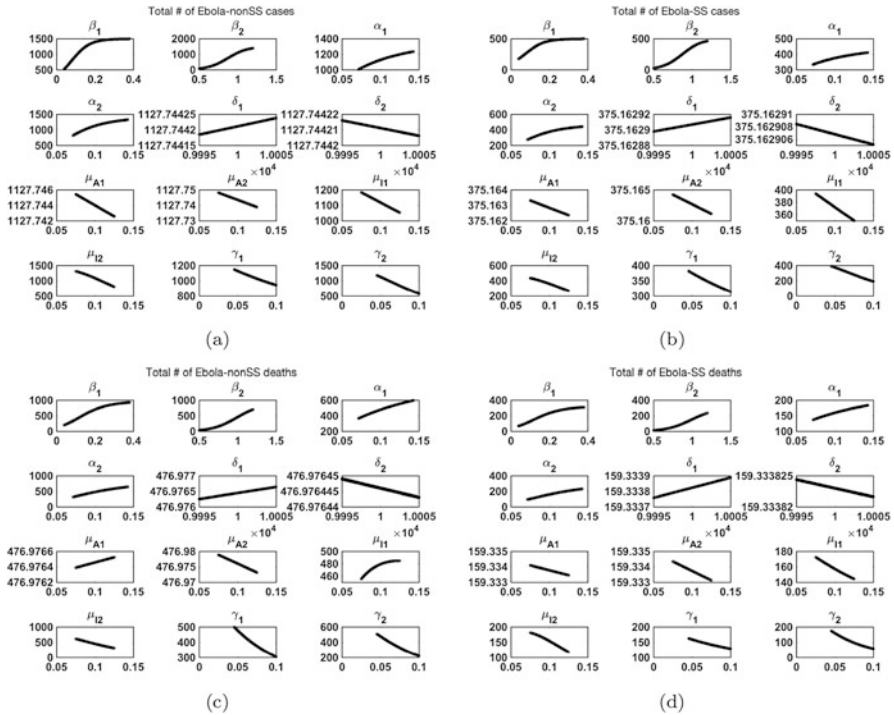
### Appendix 3: Latin Hypercube and Partial Rank Correlation Coefficient

For the MERS parameters, the graph of the output measure with  $\mu_{I2}$  parameter had a concave curvature implying that another sensitivity analysis maybe implemented. However, the range of the output measure is small enough that we can ignore the monotonicity. The remaining graphs for MERS (Fig. 10) and all the graphs for Ebola (Fig. 11) are all monotonic which means that we can trust the sensitivity analysis and proceed to the PRCC analysis (Fig. 12, Tables 4, 5 and 6).

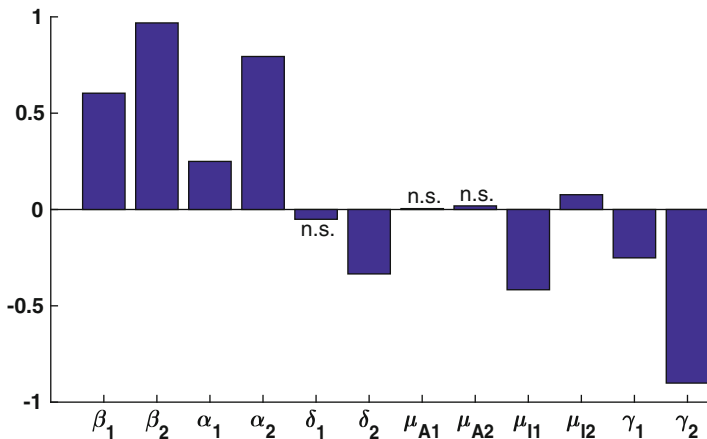


**Fig. 10** Monotonicity plots for MERS. Total nonSS and SS cases are shown in (a) and (b), respectively. Total nonSS and SS deaths are shown in (c) and (d), respectively





**Fig. 11** Monotonicity plots for Ebola. Total nonSS and SS cases are shown in (a) and (b), respectively. Total nonSS and SS deaths are in (c) and (d), respectively



**Fig. 12** PRCC values of MERS for output measures number of SS deaths with  $\mu_{I2}$  range [0.0278, 0.14].  $P$ -values that are greater than 0.05 are labeled as not significant (n.s.)

**Table 4** LHS MERS A

Param	NS Cases		NS Deaths		SS Cases		SS Deaths	
	PRCC	<i>P</i> -value	PRCC	<i>P</i> -value	PRCC	<i>P</i> -value	PRCC	<i>P</i> -value
$\beta_1$	0.687	5.300E-139	0.505	4.030E-65	0.687	5.140E-139	0.604	2.790E-99
$\beta_2$	0.979	0.000E+00	0.953	0.000E+00	0.979	0.000E+00	0.969	0.000E+00
$\alpha_1$	0.299	7.950E-22	0.311	1.430E-23	0.299	7.920E-22	0.250	1.620E-15
$\alpha_2$	0.797	5.410E-218	0.656	1.460E-122	0.797	5.350E-218	0.794	8.200E-216
$\delta_1$	-0.075	1.855E-02	-0.109	5.711E-04	-0.075	1.859E-02	-0.051	1.099E-01
$\delta_2$	0.110	1.826E-03	-0.069	3.036E-02	-0.099	1.830E-03	-0.334	2.860E-27
$\mu_{A1}$	-0.015	6.419E-01	0.059	6.433E-02	-0.015	6.425E-01	0.005	8.874E-01
$\mu_{A2}$	-0.199	2.850E-10	-0.145	4.540E-06	-0.199	2.860E-10	0.019	5.562E-01
$\mu_{I1}$	-0.427	3.990E-45	0.593	6.370E-95	-0.427	3.940E-45	-0.417	8.000E-43
$\mu_{I2}$	-0.325	8.370E-26	-0.195	6.410E-10	-0.325	8.340E-26	0.077	1.573E-02
$\gamma_1$	-0.359	1.890E-31	-0.477	2.030E-57	-0.359	1.870E-31	-0.251	1.100E-15
$\gamma_2$	-0.903	0.000E+00	-0.784	7.600E-207	-0.903	0.000E+00	-0.901	0.000E+00

**Table 5** LHS MERS B

Param	NS Cases		NS Deaths		SS Cases		SS Deaths	
	PRCC	<i>P</i> -value	PRCC	<i>P</i> -value	PRCC	<i>P</i> -value	PRCC	<i>P</i> -value
$\beta_1$	0.668	1.240E-128	0.638	3.980E-114	0.668	1.240E-128	0.663	1.960E-126
$\beta_2$	0.965	0.000E+00	0.963	0.000E+00	0.965	0.000E+00	0.966	0.000E+00
$\alpha_1$	0.260	1.090E-16	0.319	8.430E-25	0.260	1.090E-16	0.256	2.850E-16
$\alpha_2$	0.592	1.040E-94	0.586	2.110E-92	0.592	1.060E-94	0.657	2.230E-123
$\delta_1$	-0.060	5.720E-02	-0.118	2.010E-04	-0.060	5.740E-02	-0.049	1.226E-01
$\delta_2$	-0.278	4.870E-19	-0.272	2.970E-18	-0.278	4.820E-19	-0.324	1.210E-25
$\mu_{A1}$	-0.032	3.222E-01	0.025	4.306E-01	-0.032	3.223E-01	-0.042	1.899E-01
$\mu_{A2}$	-0.172	4.840E-08	-0.164	2.210E-07	-0.172	4.820E-08	-0.121	1.315E-04
$\mu_{I1}$	-0.613	2.970E-103	0.210	2.570E-11	-0.613	2.960E-103	-0.617	9.760E-105
$\mu_{I2}$	-0.613	2.970E-103	0.210	2.570E-11	-0.613	2.960E-103	-0.617	9.760E-105
$\gamma_1$	-0.353	1.800E-30	-0.519	2.600E-69	-0.353	1.790E-30	-0.343	9.950E-29
$\gamma_2$	-0.788	7.340E-210	-0.770	1.130E-194	-0.788	7.170E-210	-0.839	1.040E-263

**Table 6** LHS Ebola

Param	NS Cases		NS Deaths		SS Cases		SS Deaths	
	PRCC	P-value	PRCC	P-value	PRCC	P-value	PRCC	P-value
$\beta_1$	0.965	0.000E+00	0.950	0.000E+00	0.965	0.000E+00	0.952	0.000E+00
$\beta_2$	0.948	0.000E+00	0.931	0.000E+00	0.948	0.000E+00	0.934	0.000E+00
$\alpha_1$	0.510	0.000E+00	0.586	0.000E+00	0.510	0.000E+00	0.452	3.580E-250
$\alpha_2$	0.595	0.000E+00	0.551	0.000E+00	0.595	0.000E+00	0.667	0.000E+00
$\delta_1$	0.012	3.796E-01	0.015	2.938E-01	0.012	3.797E-01	0.017	2.212E-01
$\delta_2$	-0.033	2.102E-02	-0.036	1.034E-02	-0.033	2.102E-02	-0.032	2.360E-02
$\mu_{A1}$	0.005	7.200E-01	0.005	7.035E-01	0.005	7.197E-01	0.002	9.095E-01
$\mu_{A2}$	0.019	1.850E-01	0.016	2.620E-01	0.019	1.850E-01	0.024	9.168E-02
$\mu_{I1}$	-0.367	1.080E-158	0.032	2.520E-02	-0.367	1.080E-158	-0.324	5.050E-122
$\mu_{I2}$	-0.584	0.000E+00	-0.516	0.000E+00	-0.584	0.000E+00	-0.233	1.070E-62
$\gamma_1$	-0.512	0.000E+00	-0.634	0.000E+00	-0.512	0.000E+00	-0.405	1.930E-196
$\gamma_2$	-0.660	0.000E+00	-0.579	0.000E+00	-0.660	0.000E+00	-0.752	0.000E+00

## References

1. L.J.S. Allen, *Stochastic Population and Epidemic Models* (Springer, Cham, 2015)
2. L.J.S. Allen, G.E. Lahodny Jr., Extinction thresholds in deterministic and stochastic epidemic models. *J. Biol. Dyn.* **6**(2), 590–611 (2012)
3. L.J.S. Allen, P. van den Driessche, Relations between deterministic and stochastic thresholds for disease extinction in continuous-and discrete-time infectious disease models. *Math. Biosci.* **243**(1), 99–108 (2013)
4. C.L. Althaus, Estimating the reproduction number of Ebola virus (EBOV) during the 2014 outbreak in West Africa. *PLoS Curr.* **6** (2014). <https://doi.org/10.1371/currents.outbreaks.91afb5e0f279e7f29e7056095255b288>
5. K.B. Athreya, P.E. Ney, *Branching Processes* (Courier Corporation, New York, 2004)
6. C.T. Bauch, J.O. Lloyd-Smith, M.P. Coffee, A.P. Galvani, Dynamically modeling SARS and other newly emerging respiratory illnesses: past, present, and future. *Epidemiology* **16**(6), 791–801 (2005)
7. CDC, Mortality Risk Factors for Middle East Respiratory Syndrome Outbreak, South Korea, 2015
8. CDC, Ebola (Ebola Virus Disease) Prevention, 2015
9. CDC, 2014 Ebola Outbreak in West Africa-Case Counts, 2016
10. G. Chowell, H. Nishiura, Transmission dynamics and control of Ebola virus disease (EVD): a review. *BMC Med.* **12**(1), 196 (2014)
11. G. Chowell, S. Blumberg, L. Simonsen, M.A. Miller, C. Viboud, Synthesizing data and models for the spread of MERS-CoV, 2013: key role of index cases and hospital transmission. *Epidemics* **9**, 40–51 (2014)
12. B.J. Cowling, M. Park, V.J. Fang, P. Wu, G.M. Leung, J.T. Wu, Preliminary epidemiologic assessment of MERS-CoV outbreak in South Korea, May–June 2015. *Euro Surveillance: Bulletin European sur les Maladies Transmissibles Eur. Commun. Dis. Bull.* **20**(25), 7–13 (2015)
13. K.S. Dorman, J.S. Sinsheimer, K. Lange, In the garden of branching processes. *SIAM Rev.* **81**(46), 202–229 (2004)

14. S. Eifan, I. Nour, A. Hanif, A. Zamzam, S.M. ALJohani, A pandemic risk assessment of middle east respiratory syndrome coronavirus in Saudi Arabia. *Saudi J. Biol. Sci.* **24**(7), 1631–1638 (2017)
15. A.P. Galvani, R.M. May, Epidemiology: dimensions of superspreading. *Nature* **438**(7066), 293–295 (2005)
16. M.F.C. Gomes, A.P. y Piontti, L. Rossi, D. Chao, I. Longini, M.E. Halloran, A. Vespignani, Assessing the international spreading risk associated with the 2014 West African Ebola outbreak. *PLoS Curr.* **6** (2014). <https://doi.org/10.1371/currents.outbreaks.cd818f63d40e24aef769dda7df9e0da5>
17. A.M. Henao-Restrepo, I.M. Longini, M. Egger, N.E. Dean, W.J. Edmunds, A. Camacho, M.W. Carroll, M. Doumbia, B. Draguez, S. Duraffour, et al., Efficacy and effectiveness of an rVSV-vectored vaccine expressing Ebola surface glycoprotein: interim results from the Guinea ring vaccination cluster-randomised trial. *The Lancet* **386**(9996), 857–866 (2015)
18. A. James, J.W. Pitchford, M.J. Plank, An event-based model of superspreading in epidemics. *Proc. R. Soc. Lond. B: Biol. Sci.* **274**(1610), 741–747 (2007)
19. D. Jeong, C.H. Lee, Y. Choi, J. Kim, The daily computed weighted averaging basic reproduction number for MERS-CoV in South Korea. *Phys. A: Stat. Mech. Appl.* **451**, 190–197 (2016)
20. A. Khan, M. Naveed, M. Dur-e Ahmad, M. Imran, Estimating the basic reproductive ratio for Ebola outbreak in Liberia and Sierra Leone. *Infect. Dis. Poverty* **4**(13) (2015). <https://doi.org/10.1186/s40249-015-0043-3>
21. A. Lajmanovich, J.A. Yorke, A deterministic model for gonorrhea in a nonhomogeneous population. *Math. Biosci.* **28**(3), 221–236 (1976)
22. M.S.Y. Lau, B.D. Dalziel, S. Funk, A. McClelland, A. Tiffany, S. Riley, C.J.E. Metcalf, B.T. Grenfell, Spatial and temporal dynamics of superspreading events in the 2014–2015 West Africa Ebola epidemic. *Proc. Nat. Acad. Sci.*, **114**(9), 2337–2342 (2017)
23. J. Lee, G. Chowell, E. Jung, A dynamic compartmental model for the Middle East respiratory syndrome outbreak in the Republic of Korea: a retrospective analysis on control interventions and superspreading events. *J. Theor. Biol.* **408**, 118–126 (2016)
24. P.E. Lekone, B.F. Finkenstädt, Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study. *Biometrics* **62**(4), 1170–1177 (2006)
25. A.L. Lloyd, R.M. May, Spatial heterogeneity in epidemic models. *J. Theor. Biol.* **179**(1), 1–11 (1996)
26. J.O. Lloyd-Smith, S.J. Schreiber, P.E. Kopp, W.M. Getz, Superspreading and the effect of individual variation on disease emergence. *Nature* **438**(7066), 355 (2005)
27. S. Marino, I.B. Hogue, C.J. Ray, D.E. Kirschner, A methodology for performing global uncertainty and sensitivity analysis in systems biology. *J. Theor. Biol.* **254**(1), 178–196 (2008)
28. F.J. Massey Jr., The Kolmogorov-Smirnov test for goodness of fit. *J. Am. Stat. Assoc.* **46**(253), 68–78 (1951)
29. M.D. McKay, R.J. Beckman, W.J. Conover, Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **21**(2), 239–245 (1979)
30. R.N. Mohapatra, D. Porchia, Z. Shuai, Compartmental disease models with heterogeneous populations: a survey, in *Mathematical Analysis and Its Applications* (Springer, New Delhi, 2015), pp. 619–631
31. C.M. Rivers, E.T. Lofgren, M. Marathe, S. Eubank, B.L. Lewis, Modeling the impact of interventions on an epidemic of Ebola in Sierra Leone and Liberia (2014). arXiv preprint arXiv:1409.4607
32. Z. Shen, F. Ning, W. Zhou, X. He, C. Lin, D.P. Chin, Z. Zhu, A. Schuchat, Superspreading SARS events, Beijing, 2003. *Emerg. Infect. Dis.* **10**(2), 256 (2004)
33. R.A. Stein, Super-spreaders in infectious diseases. *Int. J. Infect. Dis.* **15**(8), e510–e513 (2011)

34. P. van den Driessche, J. Watmough, Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Math. Biosci.* **180**(1), 29–48 (2002)
35. G.E. Velásquez, O. Aibana, E.J. Ling, I. Diakite, E.Q. Mooring, M.B. Murray, Time from infection to disease and infectiousness for Ebola virus disease, a systematic review. *Clin. Infect. Dis.* (2015). <https://doi.org/10.1093/cid/civ531>
36. P. Whittle, The outcome of a stochastic epidemic—a note on Bailey’s paper. *Biometrika* **42**(1–2), 116–122 (1955)
37. WHO, Final trial results confirm Ebola vaccine provides high protection against disease (2016)
38. WHO, Ebola Virus Disease Fact Sheet (2017)
39. WHO, Middle East respiratory syndrome coronavirus (MERS-CoV) Fact Sheet (2017)
40. G. Wong, W. Liu, Y. Liu, B. Zhou, Y. Bi, G.F. Gao, MERS, SARS, and Ebola: the role of super-spreaders in infectious disease. *Cell Host Microbe* **18**(4), 398–401 (2015)
41. M.E.J. Woolhouse, C. Dye, J.-F. Etard, T. Smith, J.D. Charlwood, G.P. Garnett, P. Hagan, J.L.K. Hii, P.D. Ndhlovu, R.J. Quinnell, et al., Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *Proc. Nat. Acad. Sci.* **94**(1), 338–342 (1997)