J. Math. Biol. (2012) 65:237–262
DOI 10.1007/s00285-011-0460-2

Mathematical Biology

The size of epidemics in populations with heterogeneous susceptibility

Guy Katriel

Received: 22 September 2010 / Revised: 8 July 2011 / Published online: 10 August 2011
© Springer-Verlag 2011

Abstract We formulate and study a general epidemic model allowing for an arbitrary distribution of susceptibility in the population. We derive the final-size equation which determines the attack rate of the epidemic, somewhat generalizing previous work. Our main aim is to use this equation to investigate how properties of the susceptibility distribution affect the attack rate. Defining an ordering among susceptibility distributions in terms of their Laplace transforms, we show that a susceptibility distribution dominates another in this ordering if and only if the corresponding attack rates are ordered for every value of the reproductive number R_0. This result is used to prove a sharp universal upper bound for the attack rate valid for any susceptibility distribution, in terms of R_0 alone, and a sharp lower bound in terms of R_0 and the coefficient of variation of the susceptibility distribution. We apply some of these results to study two issues of epidemiological interest in a population with heterogeneous susceptibility: (1) the effect of vaccination of a fraction of the population with a partially effective vaccine, (2) the effect of an epidemic of a pathogen inducing partial immunity on the possibility and size of a future epidemic. In the latter case, we prove a surprising '50% law': if infection by a pathogen induces a partial immunity reducing susceptibility by less than 50%, then, whatever the value of $R_0 > 1$ before the first epidemic, a second epidemic will occur, while if susceptibility is reduced by more than 50%, then a second epidemic will only occur if R_0 is larger than a certain critical value greater than 1.

Keywords Epidemics · Heterogeneous susceptibility

Mathematics Subject Classification (2000) 92D30 · 92D25 · 92D40

G. Katriel (✉)
Biomathematics Unit, Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel
e-mail: haggaika@yahoo.com

1 Introduction

In the basic formulations of infectious disease dynamics, individuals are described as either susceptible or immune to being infected. This, however, is a simplified description, since individuals vary in their degree of susceptibility, that is in their probability of being infected under identical conditions, due to genetic factors and to previous encounters with antigenically similar pathogens (Bellamy 2004; Craig and Scherf 2003; Frank 2002). It is therefore important to understand how the heterogeneity of susceptibility can influence the epidemic dynamics. In this work we study an epidemic model allowing a general distribution of susceptibility among individuals. We explore how properties of this distribution determine the size of the epidemic generated by the model.

Several authors have previously formulated and studied epidemiological models with heterogenous susceptibility, either in terms of a finite number of different susceptibility classes (Andersson and Britton 1998; Ball 1985; Bonzi et al. 2010; Gart 1972; Hyman and Li 2005; Rodrigues et al. 2009; Scalia-Tomba 1986) or as a continuous distribution of susceptibility (Coutinho et al. 1999; Diekmann and Heesterbeek 2000; Dwyer et al. 1997, 2000; Novozhilov 2008). Below, as we describe our results, we will mention some of the results obtained in these works, and their relations with the present investigation.

In Sect. 2 we describe the general epidemic model with heterogeneous susceptibility. In this model, the distribution of susceptibility in the population is described by a probability measure μ on the positive real axis. Another general aspect of the model is that it allows an arbitrary generation-time distribution.

In Sect. 3 we derive the final-size equation corresponding to the model, an algebraic equation whose solution is the attack rate of the epidemic, that is the fraction of the population infected throughout the epidemic. The nonlinearity in this equation is given by the Laplace transform of the susceptibility distribution. In the case of the basic SIR model (exponentially distributed generation-time) an equivalent equation was already derived by Ball (1985) for a finite number of susceptibility classes, and by Novozhilov (2008) for general susceptibility distributions, using a general formalism for studying heterogeneous populations developed by Karev (2005). The final-size equation can also be derived from general results of Diekmann and Heesterbeek (2000) (see Chap. 6). We note that the fact that the final-size equation that we derive is identical with that previously derived for the case of exponentially distributed generation time is not surprising, in view of known general results which show that the total size of epidemics is independent of the generation-time distribution (Lefévre and Picard 1995). Our derivation of the final-size equation is therefore only a slight generalization of previous results, and the main concern of this paper is to employ this final-size formula in order to derive some general results on the relations between properties of the susceptibility distribution and the attack rate.

In Sect. 4 we define a partial order among susceptibility distributions μ, $\tilde{\mu}$, closely related to the Laplace order (Shaked and Shanthikumar 2007), and prove that one distribution dominates another with respect to this partial order if and only if the attack rate corresponding to this distribution is equal to or larger than that corresponding to the other distribution, for any value of the reproduction number R_0. We also show

that a susceptibility distribution with larger coefficient of variation will have smaller attack rate when R_0 is sufficiently close to 1, but not in general.

In Sects. 5,6 we use the result of Sect. 4 to derive universal upper and lower bounds for the attack rates, that is bounds depending only on basic quantitative characteristics of the susceptibility distribution, and not on its detailed form. In Sect. 5 we show that for any given reproductive number R_0, the largest attack rate is obtained for the case of homogeneous susceptibility. In the case of the basic SIR model with a finite number of susceptibility classes, this upper bound can already be found in Ball (1985). We also show, using a result of Shaked and Shanthikumar (2007) that if the susceptibility distribution has a monotonically decreasing density, a stronger bound can be obtained.

In Sect. 6 we first show that a lower bound on the attack rate in terms of the R_0 alone does *not* exist, or in other words that for a given R_0, one can find susceptibility distributions with arbitrarily small attack rates. We then show that a lower bound can be found in terms of R_0 *and* the coefficient of variation of the susceptibility distribution. Taken together, our upper and lower bounds say that the attack rate Z corresponding to an arbitrary susceptibility distribution μ satisfies

$$\frac{Z^*}{1+c_\mu^2} \leq Z \leq Z^*,$$

where Z^* is the attack rate corresponding to the case of homogeneous susceptibility with the same reproductive number, and c_μ is the coefficient of variation of μ. These bounds are both sharp in the sense that they are attained for appropriate susceptibility distributions.

Sections 7, 8 contain two applications of some of the previous results to questions of epidemiological interest. In Sect. 7 we study the effect of vaccination on a population with an arbitrary susceptibility distribution. We consider leaky vaccination, which reduces the susceptibility by a certain factor, and study how this factor, as well as the percentage of the population vaccinated and the susceptibility distribution before vaccination, affect the size of the epidemic and the possibility of preventing an epidemic by achieving herd immunity.

In Sect. 8 we study recurring epidemics. We assume that a population undergoes an epidemic which affords those who get infected partial immunity to the pathogen, which reduces their susceptibility by a certain factor. We then study whether another epidemic is possible, and what its size will be. A surprising result that we derive here is what we have called the '50% law', which says that, for any initial susceptibility distribution, if the reduction in susceptibility after infection is less than 50% then a second epidemic will always occur, while if it is greater than 50% then a second epidemic will only occur if the reproductive number R_0 associated with the first epidemic is higher than a certain critical value.

This paper deals with the case of epidemics. Several other studies model heterogeneous susceptibility when a continuous replenishment of susceptibles due to births or loss of immunity leads to an endemic equilibrium (Bonzi et al. 2010; Hyman and Li 2005; Reluga et al. 2008; Veliov 2005; White and Medley 1998).

Heterogeneity in susceptibility is only one of various types of heterogeneity that have been considered in the literature of mathematical epidemiology. In particular,

in the study of sexually transmitted diseases an important factor is the heterogeneity in the number of contacts that different individuals make (Andreasen 2011; May et al. 1988; Pastor-Satorras and Vespignani 2001). The model describing this situation and its analysis has some significant similarities with the heterogneous susceptibility model. However there are also essential differences. For example, while in the case of a heterogeneous distribution of the number of contacts the reproductive number is larger than that of the corresponding homogeneous population with the same mean number of contacts, in the case of heterogeneous susceptibility the reproductive number is the same for all susceptibility distribution with the same mean. The reason for the difference between the two cases is that heterogeneous number of contacts implies that the individuals who have a high number of contacts are both more susceptible *and* more infective than others, while in the case considered here the more susceptible individuals are not more infective than others when they are infected.

2 The general epidemic model with heterogeneous susceptibility

In this section we formulate an age-of-infection model with heterogeneous susceptibility, which generalizes the general epidemic model for a homogeneous population (Brauer 2008; Diekmann and Heesterbeek 2000; Rass and Radcliffe 2003). We note that the advantage of age-of-infection models is that they take account of the fact that the generation-time distribution is not necessarily exponential, as is assumed by the simple differential-equation SIR model.

We characterize individuals, with respect to a certain pathogen, by their level of *infectivity* $\lambda \geq 0$ and their level of *susceptibility* $\sigma \geq 0$. We also characterize the pathogen by the generation-time distribution whose density is $P(\tau) \in L^1[0, \infty)$ with

$$\int_0^\infty P(\tau)d\tau = 1.$$

The operational meaning of these concepts is as follows: an infected individual whose infectivity is λ and whose age-of-infection is τ (that is someone who was infected τ time-units ago), placed in a population with homogenous susceptibility level σ, will infect, on the average, $\lambda\sigma P(\tau)$ individuals per unit time.

The population is characterized by a susceptibility distribution which measures the fractions of the population with different susceptibility levels. The susceptibility distribution is given by a Borel probability measure μ with support

$$\Omega = supp(\mu) \subset [0, \infty),$$

so that for any interval $[\sigma_1, \sigma_2]$, $\mu([\sigma_1, \sigma_2])$ is the fraction of the population with susceptibility in this interval, before the beginning of the epidemic. This general formulation allows us to include in the same framework both the case of discrete susceptibility susceptibility classes, in which case μ is a measure with a discrete support

$$d\mu(\sigma) = \sum_{n=1}^{N} p_k \delta(\sigma - \sigma_k) d\sigma, \quad p_k > 0, \quad \sum_{n=1}^{N} p_k = 1, \tag{1}$$

(N might be finite or $N = \infty$), and the case of continuously-distributed susceptibility of the form $d\mu = \rho(\sigma)d\sigma$, where $\rho \in L^1[0, \infty)$, or mixtures of the two types of distributions.

For $\sigma \in \Omega$, $S(\sigma, t)$ denotes the fraction of the population with susceptibility σ which remains susceptible (that is, has not been infected yet) at time t. Thus the fraction of the population with susceptibility in the interval $[\sigma_1, \sigma_2]$ at time t is given by $\int_{\sigma_1}^{\sigma_2} S(\sigma, t)d\mu(\sigma)$. $i(\sigma, t)$ denotes the incidence rate among individuals with susceptibility σ, that is the fraction of these individuals who are infected per unit time, at time t. The total incidence rate is given by

$$i(t) = \int_{\Omega} i(\sigma, t) d\mu(\sigma). \tag{2}$$

The age-of-infection model with heterogeneous susceptibility is thus

$$S_t(\sigma, t) = -i(\sigma, t), \quad \sigma \in \Omega \tag{3}$$

$$i(\sigma, t) = \lambda \sigma S(\sigma, t) \int_0^{\infty} P(\tau) i(t - \tau) d\tau, \quad \sigma \in \Omega. \tag{4}$$

Equation (3) is an accounting equation which states that the rate of reduction of the susceptible population with susceptibility level σ is equal to the incidence rate in that class. Equation (4) calculates this incidence rate by noting that at time t the fraction of the population whose age-of-infection in the interval $[\tau - d\tau, \tau]$ is $i(t - \tau)d\tau$, so their contribution to the force of infection $\lambda P(\tau) i(t - \tau) d\tau$, and this quantity is integrated over all values of the age-of-infection τ to obtain the total force of infection.

We note that when the population has a homogeneous susceptibility, that is when $d\mu(\sigma) = \delta(\sigma - \sigma_0)d\sigma$, the above model reduces to the standard age-of-infection model (Brauer 2008; Diekmann and Heesterbeek 2000; Rass and Radcliffe 2003).

Substituting (4) into (2) we have

$$i(t) = \lambda \int_{\Omega} \sigma S(\sigma, t) d\mu(\sigma) \cdot \int_0^{\infty} P(\tau) i(t - \tau) d\tau, \tag{5}$$

and from (3) and (4) we have

$$S_t(\sigma, t) = -\lambda \sigma S(\sigma, t) \int_0^{\infty} P(\tau) i(t - \tau) d\tau, \quad \sigma \in \Omega. \tag{6}$$

Equations (5) to (6) are a more convenient form of the model, since only the total incidence rate $i(t)$ appears in them.

Calculating the number of cases infected by a single individual with infectivity λ in a population with susceptibility structure μ (that is ignoring the depletion of susceptibles), we get the reproductive number

$$R_0 = \lambda \int_\Omega \sigma d\mu(\sigma) \cdot \int_0^\infty P(\tau) d\tau = \lambda \bar{\sigma}_\mu,$$

where

$$\bar{\sigma}_\mu = \int_\Omega \sigma d\mu(\sigma)$$

is the mean susceptibility corresponding to μ. It is useful to re-parameterize the model in terms of R_0 instead of λ, re-writing Eqs. (5),(6) as

$$i(t) = R_0 \frac{1}{\bar{\sigma}_\mu} \int_\Omega \sigma S(\sigma, t) d\mu(\sigma) \cdot \int_0^\infty P(\tau) i(t-\tau) d\tau, \tag{7}$$

$$S_t(\sigma, t) = -R_0 \frac{1}{\bar{\sigma}_\mu} \sigma S(\sigma, t) \int_0^\infty P(\tau) i(t-\tau) d\tau, \quad \sigma \in \Omega. \tag{8}$$

3 Attack rate of an epidemic

We now re-write the model (7)–(8) in another form, from which the final size of the epidemic generated by the model will easily follow. For this we will need to introduce an initial condition for the fraction of susceptibles at $t = -\infty$,

$$S(\sigma, -\infty) = \lim_{t \to -\infty} S(\sigma, t) = 1, \quad \sigma \in \Omega. \tag{9}$$

We are thus assuming that, in each susceptibility class, all individuals are susceptible before the epidemic. This represents no loss of generality, since if $0 \in \Omega$ we have a class of individuals with susceptibility 0, so that the model does take into account those individuals who are completely immune to infection.

Writing (8) as

$$[\log(S(\sigma, t)]_t = -R_0 \frac{\sigma}{\bar{\sigma}_\mu} \int_0^\infty P(\tau) i(t-\tau) d\tau, \quad \sigma \in \Omega, \tag{10}$$

and integrating from $-\infty$ to t, using (9), we obtain

$$S(\sigma, t) = e^{-R_0 \frac{\sigma}{\bar{\sigma}_\mu} \int_{-\infty}^{t} \int_0^\infty P(\tau) i(s-\tau) d\tau ds}, \quad \sigma \in \Omega. \tag{11}$$

Substituting (11) into (7) we have

$$\begin{aligned} i(t) &= R_0 \frac{1}{\bar{\sigma}_\mu} \int_\Omega \sigma e^{-R_0 \frac{\sigma}{\bar{\sigma}_\mu} \int_{-\infty}^{t} \int_0^\infty P(\tau) i(s-\tau) d\tau ds} d\mu(\sigma) \cdot \int_0^\infty P(\tau) i(t-\tau) d\tau \\ &= -\frac{d}{dt} \int_\Omega e^{-R_0 \frac{\sigma}{\bar{\sigma}_\mu} \int_{-\infty}^{t} \int_0^\infty P(\tau) i(s-\tau) d\tau ds} d\mu(\sigma). \end{aligned} \tag{12}$$

We define $C(t)$ to be the cumulative fraction of the population infected up to time t, given by

$$C(t) \doteq \int_{-\infty}^{t} i(s) ds. \tag{13}$$

Integrating both sides of (12), taking into account the fact that $C(-\infty) = 0$, we have

$$C(t) = 1 - \int_\Omega e^{-R_0 \frac{\sigma}{\bar{\sigma}_\mu} \int_{-\infty}^{t} \int_0^\infty P(\tau) i(s-\tau) d\tau ds} d\mu(\sigma). \tag{14}$$

As a simplification of notation, we introduce the function F_μ defined by

$$F_\mu(x) \doteq \int_\Omega e^{-\frac{\sigma}{\bar{\sigma}_\mu} x} d\mu(\sigma). \tag{15}$$

Noting also that

$$\int_{-\infty}^{t} \int_0^\infty P(\tau) i(s-\tau) d\tau ds = \int_0^\infty P(\tau) \int_{-\infty}^{t-\tau} i(s) ds d\tau = \int_0^\infty P(\tau) C(t-\tau) d\tau, \tag{16}$$

we see that (14) can be written as

$$C(t) = 1 - F_\mu \left(R_0 \int_0^\infty P(\tau) C(t-\tau) d\tau \right). \tag{17}$$

We note that

$$F_\mu(x) = \mathcal{L}(\mu)\left(\frac{x}{\bar{\sigma}_\mu}\right), \tag{18}$$

where $\mathcal{L}(\mu)$ is the Laplace transform of the measure μ, so F_μ is simply a scaling of the Laplace transform of the susceptibility distribution.

Equation (17) is equivalent to the original model (7), (8). Indeed, given a solution $C(t)$ of (17), we have $i(t) = C'(t)$ and using (11) and (16)

$$S(\sigma, t) = e^{-R_0 \frac{\sigma}{\sigma_\mu} \int_0^\infty P(\tau) C(t-\tau) d\tau}. \tag{19}$$

We can now derive an equation for the attack rate of the epidemic

$$Z_\mu(R_0) \doteq \lim_{t\to\infty} C(t). \tag{20}$$

Since $C(t)$ is monotonically increasing and bounded from above by 1, this limit exists. By (20) and the Lebesgue monotone convergence theorem we get

$$\lim_{t\to\infty} \int_0^\infty P(\tau) C(t-\tau) d\tau = Z_\mu(R_0) \int_0^\infty P(\tau) d\tau = Z_\mu(R_0). \tag{21}$$

Therefore, taking the limit $t \to \infty$ on both sides of (17), we obtain the final size equation.

Theorem 1 *The attack rate $Z_\mu(R_0)$ of (7)–(8) satisfies the equation*

$$Z = 1 - F_\mu(R_0 Z). \tag{22}$$

We note that, apart from notation, (22) is identical to the final-size equation obtained by Novozhilov 2008. Since $F_\mu(0) = 1$, we see that $Z = 0$ is always a solution of (22), but we seek a nontrivial solution.

We can write the final-size equation in another way, which is convenient for several purposes. We define $G_\mu : [0, \infty) \to \mathbb{R}$ by

$$G_\mu(x) \doteq \frac{1 - F_\mu(x)}{x}, \tag{23}$$

(in a moment we will see that G_μ is well-defined also at 0) so that (22) is rewritten as

$$R_0 G_\mu(R_0 Z) = 1. \tag{24}$$

We shall use the two equivalent forms (22) and (24) of the final-size equation alternately, according to convenience. One advantage of the form (24) is that the trivial solution $Z = 0$ of (22) is eliminated.

We can derive another natural representation of the function G_μ. Let us define $\Phi_\mu : [0, \infty) \to [0, 1]$ by

$$\Phi_\mu(\sigma) \doteq \mu([\sigma, \infty)). \tag{25}$$

Thus $\Phi_\mu(\sigma)$ is the fraction of the population whose susceptibility is greater than or equal to σ at the beginning of the epidemic. Φ_μ is monotonically decreasing, continuous from the left, and satisfies $\Phi_\mu(0) = 1, \lim_{\sigma\to+\infty}\Phi_\mu(\sigma) = 0$. Employing integration by parts (using the Riemann–Stieltjes integral notation) we have

$$-\int_\Omega e^{-x\sigma}d\mu(\sigma) = \int_0^\infty e^{-x\sigma}d\Phi_\mu(\sigma) = e^{-x\sigma}\Phi_\mu(\sigma)\Big|_{\sigma=0}^{\sigma=\infty} - \int_0^\infty \Phi_\mu(\sigma)d(e^{-x\sigma})$$
$$= -1 + x\int_0^\infty \Phi_\mu(\sigma)e^{-x\sigma}d\sigma.$$

We therefore have

$$G_\mu(x) = \frac{1}{\bar\sigma_\mu}\mathcal{L}(\Phi_\mu)\Big(\frac{x}{\bar\sigma_\mu}\Big). \tag{26}$$

The following properties of the function G_μ will be used repeatedly in what follows:

Lemma 1 (i) $G_\mu(0) = \lim_{x\to 0+} G_\mu(x) = 1$.

(ii) $G'_\mu(0) = -\frac{c_\mu^2+1}{2}$, *where* c_μ *is the coefficient of variation (c.v.) of the susceptibility distribution, defined as its standard deviation divided by its mean*

$$c_\mu \doteq \frac{\sqrt{\overline{\sigma^2}_\mu - \bar\sigma_\mu^2}}{\bar\sigma_\mu}.$$

(iii) G_μ *is decreasing and convex.*

(iv) $G_\mu(x) > 0$ *for all* $x \geq 0$ *and* $\lim_{x\to\infty} G_\mu(x) = 0$.

Proof (i) Since $F_\mu(0) = 1$ we have, from (23),

$$\lim_{x\to 0} G_\mu(x) = -F'_\mu(0).$$

By differentiating (15) one sees that

$$F'_\mu(0) = -\frac{1}{\bar\sigma_\mu}\int_\Omega \sigma d\mu(\sigma) = -1, \tag{27}$$

and the result follows.

(ii) Using L'Hôpital's rule, we have

$$\lim_{x\to 0} G'_\mu(x) = \lim_{x\to 0}\frac{F_\mu(x) - xF'_\mu(x) - 1}{x^2} = -\frac{1}{2}\lim_{x\to 0} F''_\mu(x) = -\frac{1}{2}F''_\mu(0).$$

The result then follows from the fact that

$$F''_\mu(0) = \frac{1}{\bar{\sigma}_\mu^2} \int_\Omega \sigma^2 d\mu(\sigma) = \frac{\overline{\sigma^2}_\mu}{\bar{\sigma}_\mu^2} = c_\mu^2 + 1. \tag{28}$$

(iii) follows from (26) and the fact that the Laplace transform of a non-negative function is decreasing and convex.
(iv) follows from (23) and the fact that $0 < F_\mu(x) \leq 1$.

From the above properties of G_μ we immediately have the threshold result

Theorem 2 *Equation* (24) (*hence also* (22)) *has a solution* $Z \in (0, 1)$ *if and only if* $R_0 > 1$, *and this solution is then unique.*

We thus see that the threshold condition for the triggering of an epidemic depends on the susceptibility distribution μ only through the mean susceptibility $\bar{\sigma}_\mu$. As for the size of the epidemic, we see from (22) that it depends on the susceptibility distribution through its entire Laplace transform.

Since G_μ is decreasing , the inverse function G_μ^{-1} is well defined on $(0, 1]$, and by (24) we have the following explicit representation for $Z_\mu(R_0)$:

$$Z_\mu(R_0) = \begin{cases} 0 & R_0 \leq 1 \\ \frac{1}{R_0} G_\mu^{-1}(\frac{1}{R_0}) & R_0 > 1 \end{cases}. \tag{29}$$

In addition to the total attack rate $Z_\mu(R_0)$, it is also of interest to consider, for $\sigma \in \Omega$, the attack rate $Z_\mu(R_0, \sigma)$ among individuals with susceptibility σ. This is equal to

$$Z_\mu(R_0, \sigma) = 1 - \lim_{t\to\infty} S(\sigma, t),$$

and using (19) and (21) we get

$$Z_\mu(R_0, \sigma) = 1 - e^{-R_0 Z_\mu(R_0)\frac{\sigma}{\bar{\sigma}_\mu}}, \quad \sigma \in \Omega. \tag{30}$$

As expected, the attack rate among more susceptible individuals is higher.

We now apply the above results to some explicit examples of susceptibility distributions.

In the particular case of a homogeneous population, that is when $d\mu(\sigma) = \delta(\sigma - \bar{\sigma})d\sigma$, we have $F_\mu(x) = e^{-x}$, so (22) reduces to

$$Z = 1 - e^{-R_0 Z}, \tag{31}$$

the familiar final size equation for the basic SIR model, and many of its variants (Ma and Earn 2006). For future use we shall denote the solution of (31) by $Z^*(R_0)$.

More generally, in the case of a finite number of susceptibility classes (1) we have

$$F_\mu(x) = \sum_{k=1}^n p_k e^{-\frac{\sigma_k}{\bar{\sigma}_\mu}x}, \quad \bar{\sigma}_\mu = \sum_{k=1}^n p_k \sigma_k,$$

so the final-size equation is

$$Z = 1 - \sum_{k=1}^{n} p_k e^{-\frac{\sigma_k}{\bar{\sigma}_\mu} R_0 Z}, \tag{32}$$

already given by Ball (1985).

As another interesting example, already considered in Novozhilov (2008), is the case in which the susceptibility of the population is Gamma-distributed, with mean $\bar{\sigma}$ and shape parameter κ, that is

$$d\mu(\sigma) = \frac{1}{\Gamma(\kappa)} \left(\frac{\kappa}{\bar{\sigma}}\right)^{\kappa} \sigma^{\kappa-1} e^{-\frac{\kappa}{\bar{\sigma}}\sigma} d\sigma. \tag{33}$$

Then

$$F_\mu(x) = \left(1 + \frac{1}{\kappa}x\right)^{-\kappa}, \quad G_\mu(x) = \frac{1}{x}\left[1 - \left(1 + \frac{1}{\kappa}x\right)^{-\kappa}\right], \tag{34}$$

and the final size equation is thus

$$Z = 1 - \left(1 + \frac{R_0}{\kappa} Z\right)^{-\kappa}. \tag{35}$$

In general G_μ cannot be inverted in terms of simple functions, but in the particular case $\kappa = 1$, that is when susceptibility is exponentially distributed $d\mu = \frac{1}{\bar{\sigma}} e^{-\frac{1}{\bar{\sigma}}\sigma} d\sigma$, we have $G_\mu(x) = \frac{1}{x+1}$, so $G_\mu^{-1}(y) = \frac{1}{y} - 1$ and we get a very simple expression for the attack rate

$$Z_\mu(R_0) = 1 - \frac{1}{R_0}. \tag{36}$$

4 Ordering of susceptibility distributions and attack rates

Application of the final-size formula yields a sufficient criterion for the attack rate corresponding to one instance of the model (7)–(8), defined by a susceptibility measure μ and by R_0, to be larger than the attack rate corresponding to another instance, defined by $\tilde{\mu}$ and $\tilde{R}_0$.

Lemma 2 *Let $\mu, \tilde{\mu}$ be initial susceptibility distributions for the model (7)–(8), let $R_0, \tilde{R}_0 \geq 0$ and assume that*

$$F_\mu(R_0 z) \leq F_{\tilde{\mu}}(\tilde{R}_0 z), \quad \forall z \in (0, 1) \tag{37}$$

Then the corresponding attack rates satisfy $Z_{\tilde{\mu}}(\tilde{R}_0) \leq Z_\mu(R_0)$.

Proof By (37) we have, for all $z \in (0, 1)$,

$$G_\mu(R_0 z) = \frac{1 - F_\mu(R_0 z)}{R_0 z} \geq \frac{1 - F_{\tilde{\mu}}(\tilde{R}_0 z)}{R_0 z} = \frac{\tilde{R}_0}{R_0} \frac{1 - F_{\tilde{\mu}}(\tilde{R}_0 z)}{\tilde{R}_0 z} = \frac{\tilde{R}_0}{R_0} G_{\tilde{\mu}}(\tilde{R}_0 z),$$

Thus, setting $z = Z_\mu(R_0)$ and using the fact that $Z_\mu(R_0)$ satisfies (24),

$$G_{\tilde{\mu}}(\tilde{R}_0 Z_\mu(R_0)) \leq \frac{R_0}{\tilde{R}_0} G_\mu(R_0 Z_\mu(R_0)) = \frac{1}{\tilde{R}_0}.$$

Since $G_{\tilde{\mu}}$ is monotonically decreasing (Lemma 1 (iii)) we get, using also the final-size formula (29) applied to $\tilde{\mu}$,

$$\tilde{R}_0 Z_\mu(R_0) \geq G_{\tilde{\mu}}^{-1}\left(\frac{1}{\tilde{R}_0}\right) = \tilde{R}_0 Z_{\tilde{\mu}}(\tilde{R}_0),$$

so we have $Z_{\tilde{\mu}}(\tilde{R}_0) \leq Z_\mu(R_0)$.

We will now concentrate on the case in which the reproductive numbers of the two instances of the model are equal, $R_0 = \tilde{R}_0$, so that we wish to compare the functions $Z_\mu(R_0)$, $Z_{\tilde{\mu}}(R_0)$. This brings us naturally to the definition of an ordering relation for probability measures on $[0, \infty)$.

Definition 1 Given two probability measures μ, $\tilde{\mu}$ on $[0, \infty)$, we say that μ dominates $\tilde{\mu}$, denoted $\tilde{\mu} \preccurlyeq \mu$, if $F_\mu(x) \leq F_{\tilde{\mu}}(x)$ for all $x \geq 0$.

The relation $\preccurlyeq$ is a *partial* order, so that for many pairs of distributions we have neither $\tilde{\mu} \preccurlyeq \mu$ nor $\mu \preccurlyeq \tilde{\mu}$. This relation is closely related to the Laplace order (Shaked and Shanthikumar 2007), which is defined by: $\tilde{\mu} \leq_{\mathcal{L}} \mu$ iff $\mathcal{L}(\mu)(x) \leq \mathcal{L}(\tilde{\mu})(x)$ for all $x \geq 0$. Indeed if two distributions have the same mean, then the above relation $\preccurlyeq$ is equivalent to the Laplace order. We note that the Laplace order has recently appeared in a quite different epidemiological application (Yan and Feng 2010), where it is shown that the Laplace order over latent and infectious period distributions can be used to rank the effectiveness of control measures like quarantine and isolation.

From Lemma 2 we have that if $\tilde{\mu} \preccurlyeq \mu$, the corresponding attack rates satisfy $Z_{\tilde{\mu}}(R_0) \leq Z_\mu(R_0)$, for all values of R_0. The following theorem says that the converse to this statement is also true:

Theorem 3 *Let $\mu, \tilde{\mu}$ be susceptibility distributions. Then the following statements are equivalent*

(i) $\tilde{\mu} \preccurlyeq \mu$.
(ii) *For any value of R_0, the attack rates corresponding to $\mu, \tilde{\mu}$ satisfy $Z_{\tilde{\mu}}(R_0) \leq Z_\mu(R_0)$.*

Proof The implication (i) $\Rightarrow$ (ii) follows from Lemma 2. To prove (ii) $\Rightarrow$ (i), we assume that (i) does not hold, and will prove that (ii) does not hold. Thus, assume that

there exist some value $x^* > 0$ for which

$$F_\mu(x^*) > F_{\tilde{\mu}}(x^*). \tag{38}$$

Defining R_0 by

$$R_0 = \frac{1}{G_\mu(x^*)},$$

we have $x^* = G_\mu^{-1}(\frac{1}{R_0})$, so that, using (29)

$$x^* = R_0 Z_\mu(R_0). \tag{39}$$

Using (38) we have

$$G_{\tilde{\mu}}(x^*) = \frac{1 - F_{\tilde{\mu}}(x^*)}{x^*} > \frac{1 - F_\mu(x^*)}{x^*} = G_\mu(x^*) = \frac{1}{R_0}$$

and since G_μ is decreasing this implies, using also the final-size formula (29) applied to $\tilde{\mu}$,

$$x^* < G_{\tilde{\mu}}^{-1}\left(\frac{1}{R_0}\right) = R_0 Z_{\tilde{\mu}}(R_0), \tag{40}$$

and by (39) and (40) we get $Z_\mu(R_0) < Z_{\tilde{\mu}}(R_0)$, so (ii) does not hold.

Thus, given two susceptibility distributions, if they are ordered, that is either $\tilde{\mu} \preccurlyeq \mu$ or $\mu \preccurlyeq \tilde{\mu}$, then the attack rate corresponding to one of the distributions is equal to or larger than that corresponding to the other, for any value of R_0. On the other hand, if the two distributions are not ordered, that is neither $\tilde{\mu} \preccurlyeq \mu$ nor $\mu \preccurlyeq \tilde{\mu}$ holds, then the above theorem tells us that there will be some values of R_0 for which the attack rate corresponding to μ will be larger, and other values of R_0 for which the attack rate corresponding to $\tilde{\mu}$ will be larger.

If $\mu, \tilde{\mu}$ are two susceptibility distributions with $\tilde{\mu} \preccurlyeq \mu$, we have, since $F_{\tilde{\mu}}(x) \geq F_\mu(x)$, $F_\mu(0) = F_{\tilde{\mu}}(0) = 1$, $F'_\mu(0) = F'_{\tilde{\mu}}(0) = -1$ that $F''_{\tilde{\mu}}(0) \geq F''_\mu(0)$. By (28) we get

Lemma 3 *If $\tilde{\mu} \preccurlyeq \mu$ then $c_\mu \leq c_{\tilde{\mu}}$.*

Thus, a distribution dominating another has a lower coefficient of variation (c.v.). The converse is, of course, not true: $c_\mu \leq c_{\tilde{\mu}}$ does not imply $\tilde{\mu} \preccurlyeq \mu$, so a lower c.v. of the susceptibility distribution does *not*, in general, imply higher attack rate (see example below), although this can be true for particular families of distributions: in the case of Gamma-distributed susceptibility (see (33)), we have $c_\mu^2 + 1 = F''_\mu(0) = 1 + \kappa^{-1}$, so for two members $\mu, \tilde{\mu}$ of this family, with shape parameters $\kappa, \tilde{\kappa}$, we have $c_\mu \leq c_{\tilde{\mu}}$ iff $\kappa \geq \tilde{\kappa}$, which implies that $F_\mu(x) = (1 + \kappa^{-1}x)^{-\kappa} \leq (1 + \tilde{\kappa}^{-1}x)^{-\tilde{\kappa}} = F_{\tilde{\mu}}(x)$, so

$\tilde{\mu} \preccurlyeq \mu$, and in particular the attack rate corresponding to μ is smaller, a fact already noted by Novozhilov (2008).

We now show that a weakened form of the converse statement discussed above *is* valid, namely ordering of the c.v.'s does imply ordering of the attack rate when the reproductive number R_0 is sufficiently close to the epidemic threshold $R_0 = 1$. This will follow from

Lemma 4 *For any susceptibility distribution μ, as $R_0 \to 1+$,*

$$Z_\mu(R_0) = \frac{2}{c_\mu^2 + 1}(R_0 - 1) + O((R_0 - 1)^2).$$

Proof Differentiating (29) we have, for $R_0 \geq 1$,

$$Z'_\mu(R_0) = -\frac{1}{R_0^2} G_\mu^{-1}\Big(\frac{1}{R_0}\Big) - \frac{1}{R_0^3} \frac{1}{G'_\mu(G_\mu^{-1}(\frac{1}{R_0}))}.$$

Substituting $R_0 = 1$ and using the fact that $G_\mu(0) = 1$, $G'_\mu(0) = -\frac{c_\mu^2+1}{2}$ (lemma 1(i), (ii)) we obtain

$$Z'_\mu(1) = -\frac{1}{G'_\mu(0)} = \frac{2}{c_\mu^2 + 1},$$

and the result follows by the Taylor approximation.

Thus, if we have two susceptibility distributions, for R_0 sufficiently close to the threshold 1, the susceptibility distribution with larger c.v. will lead to a smaller attack rate. As stressed above, this does not hold for general R_0.

As a simple explicit example, consider the finitely supported distributions

$$d\mu(\sigma) = \frac{1}{4}\delta\Big(\sigma - \frac{1}{4}\Big) + \frac{3}{4}\delta(\sigma - 5), \quad d\tilde{\mu}(\sigma) = \frac{3}{4}\delta\Big(\sigma - 1\Big) + \frac{1}{4}\delta(\sigma - 10). \quad (41)$$

We have $\bar{\sigma}_\mu = \frac{61}{16}$, $\bar{\sigma}_{\tilde{\mu}} = \frac{13}{4}$, so

$$F_\mu(x) = \frac{1}{4}e^{-\frac{4}{61}x} + \frac{3}{4}e^{-\frac{80}{61}x}, \quad F_{\tilde{\mu}}(x) = \frac{3}{4}e^{-\frac{4}{13}x} + \frac{1}{4}e^{-\frac{40}{13}x}.$$

Since $F_\mu(0) = F_{\tilde{\mu}}(0) = 1$, $F'_\mu(0) = F'_{\tilde{\mu}}(0) = -1$ (as always) and $F''_\mu(0) = 1.29 < 2.44 = F''_{\tilde{\mu}}(0)$, we have that $F_\mu(x) < F_{\tilde{\mu}}(x)$ for x sufficiently small. On the other hand, for x large $F_{\tilde{\mu}}(x) \approx \frac{3}{4}e^{-\frac{4}{13}x} < \frac{1}{4}e^{-\frac{4}{61}x} \approx F_\mu(x)$. Thus $\mu, \tilde{\mu}$ are not ordered according to the partial order $\preccurlyeq$. Plotting the attack rates corresponding to the two distributions as functions of R_0 (Fig. 1), computed by solving (22) numerically, we see that for low values of R_0 the attack rate corresponding to μ is larger while for higher values of R_0 the attack rate corresponding to $\tilde{\mu}$ is larger.

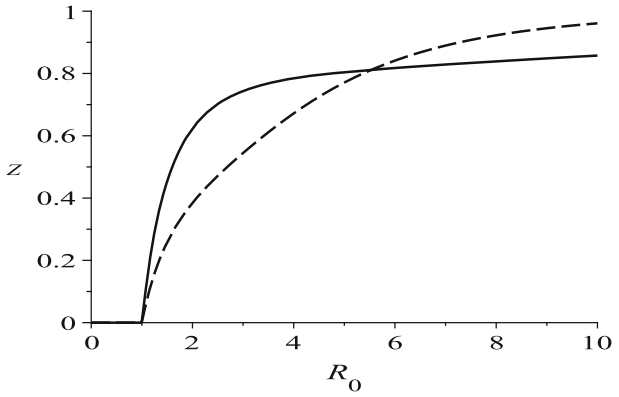

Fig. 1 Attack rates as a function of R_0 for the susceptibility distributions given by (41). *Full line* corresponds to μ, *dashed line* to $\tilde{\mu}$

5 Universal upper bounds for the attack rate

We now apply Theorem 3 to obtain an upper bound for the attack rate in terms of R_0 alone, valid for any susceptibility distribution μ. The following theorem is a generalization of the result obtained by Ball 1985 for a finite number of susceptibility classes.

Theorem 4 *For any susceptibility distribution μ, assuming $R_0 > 1$, the attack rate $Z_\mu(R_0)$ is bounded from above by*

$$Z_\mu(R_0) \leq Z^*(R_0), \tag{42}$$

where $Z^*(R_0)$ is the attack rate for a homogeneous population, given as the solution of (31)*.*

Proof Applying Jensen's inequality with the convex function $f(\sigma) = e^{-x\frac{\sigma}{\bar{\sigma}_\mu}}$, we obtain

$$F_\mu(x) = \int_\Omega f(\sigma) d\mu(\sigma) \geq f\left(\int_\Omega \sigma d\mu(\sigma)\right) = e^{-x\frac{1}{\bar{\sigma}_\mu}\int_\Omega \sigma d\mu(\sigma)} = e^{-x} = F_{\hat{\mu}}(x),$$

where $d\hat{\mu}(\sigma) = \delta(\sigma - \bar{\sigma}_\mu)d\sigma$, so that $\mu \preccurlyeq \hat{\mu}$, and the result follows from Theorem 3.

Since $R_0 = \lambda\bar{\sigma}$, the above theorem implies that, for given infectivity λ, among the susceptibility distributions with the same mean, the largest attack rate is obtained for the distribution concentrated at this mean, that is for the distribution with the least variation. We may thus say that heterogeneity of the susceptibility leads to decrease of the attack rate, relative to the homogeneous case with the same reproductive number. However, as we stressed and demonstrated in the previous section, this does not mean

that susceptibility distributions with more variation, as measured by the coefficient of variation, necessarily have a lower attack rate.

Since the upper bound given by Theorem 4 is attained for the homogeneous case, it is sharp and cannot be improved unless we somehow restrict the class of susceptibility distributions. We now derive a stronger upper bound for the class of susceptibility distributions which have a monotonically decreasing density, that is $d\mu = \rho(\sigma)d\sigma$ with $\rho \in L^1[0,\infty)$ a monotonically decreasing function. This depends on the following result from Shaked and Shanthikumar (2007), Theorem 3.A.46 (a).

Lemma 5 *Let μ be a probability measure on $[0,\infty)$ with a monotonically decreasing density and mean $\bar{\sigma}$. Let $\tilde{\mu}$ be the uniform measure on the interval $[0, 2\bar{\sigma}]$. Then, for any convex function $\phi : \mathbb{R} \to \mathbb{R}$:*

$$\int_0^\infty \phi(\sigma)d\mu(\sigma) \le \int_0^\infty \phi(\sigma)d\tilde{\mu}(\sigma) = \frac{1}{2\bar{\sigma}}\int_0^{2\bar{\sigma}} \phi(\sigma)d\sigma.$$

In particular, if we take $\phi(\sigma) = e^{-x\frac{\sigma}{\bar{\sigma}}}$ we conclude that $F_\mu(x) \ge F_{\tilde{\mu}}(x)$, so that $\mu \preccurlyeq \tilde{\mu}$. Noting that

$$F_{\tilde{\mu}}(x) = \frac{1}{2\bar{\sigma}}\int_0^{2\bar{\sigma}} e^{-x\frac{\sigma}{\bar{\sigma}}}d\sigma = \frac{1-e^{-2x}}{2x},$$

and applying Theorem 3, we therefore have

Theorem 5 *If the susceptibility distribution μ has a monotonically decreasing density, then for all $R_0 > 1$ we have $Z_\mu(R_0) \le Z_{\tilde{\mu}}(R_0)$, where $Z_{\tilde{\mu}}(R_0)$ is the solution of the equation*

$$Z = 1 - \frac{1}{2R_0 Z}(1 - e^{-2R_0 Z}).$$

In Fig. 2 we plot the upper bound on the attack rate for general susceptibility distributions, and the upper bound for the case of distributions with a monotonically decreasing density, which is of course lower.

6 Universal lower bounds for the attack rate

A natural question, in view of Theorem 4, is whether it is possible to find a bound from *below* for the attack rate, in terms of R_0, independent of the susceptibility distribution. The answer is negative: for given R_0, one can find susceptibility distributions for which the corresponding attack rate is arbitrarily small. To see this, let us consider a population divided into two subpopulations: a fraction α has uniform susceptibility σ_1, and the rest are totally immune (susceptibility 0). The mean susceptibility is then given

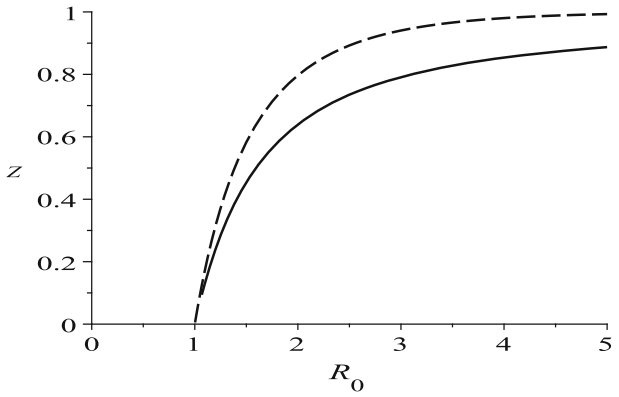

Fig. 2 Upper bound for the attack rate as a function of R_0 for arbitrary susceptibility distributions (*dashed line*) and for susceptibility distributions with a monotonically decreasing density

by $\alpha\sigma_1$, and we now fix the mean to be a given number $\bar{\sigma}$ by setting $\sigma_1 = \frac{\bar{\sigma}}{\alpha}$, so that the susceptibility distribution is:

$$d\breve{\mu}(\sigma) = \left[(1-\alpha)\delta(\sigma) + \alpha\delta\left(\sigma - \frac{\bar{\sigma}}{\alpha}\right)\right]d\sigma, \tag{43}$$

and we have

$$F_{\breve{\mu}}(x) = 1 - \alpha + \alpha e^{-\frac{1}{\alpha}x}.$$

If we take $\alpha \to 0$, we see that $F_{\breve{\mu}} \to 1$ uniformly, so that the final size equation (22) becomes $Z = 0$. Therefore, by taking α sufficiently small, we can get an attack rate which is arbitrarily close to 0. Thus the number of infectives, beyond those initially infected, can be made arbitrarily small for a given R_0. The intuitive reason is obvious—in this example there is a very small population of very highly susceptible population, with the rest of the population totally immune, so the epidemic can affect only the small highly susceptible population.

We will now show, however, that a lower bound for the attack rate *can* be obtained if we restrict the coefficient of variation c_μ of the susceptibility distribution.

Note first that for the specific example (43), using (28) we have

$$c_{\breve{\mu}}^2 + 1 = F''_{\breve{\mu}}(0) = \frac{1}{\alpha},$$

so that fixing the coefficient of variation to be $c_{\breve{\mu}} = c$ determines the value of α to be $\alpha = \frac{1}{c^2+1}$, hence $\breve{\mu}$ becomes

$$d\breve{\mu}(\sigma) = \frac{c^2}{1+c^2}\delta(\sigma)d\sigma + \frac{1}{1+c^2}\delta\left(\sigma - (1+c^2)\bar{\sigma}\right)d\sigma. \tag{44}$$

It turns out that the distribution given by (44) is dominated by *any* distribution μ with the same coefficient of variation $c_\mu = c$. This is the content of the following result from Shaked and Shanthikumar (2007), see Theorem 5.A.21:

Lemma 6 *Let μ be any probability measure on $[0, \infty)$, with finite first and second moments, and let c_μ denote its coefficient of variation. Let $\check{\mu}$ be the probability measure defined by (44), with $\bar{\sigma} = \bar{\sigma}_\mu$ and $c = c_\mu$. Then $\mathcal{L}(\mu)(x) \leq \mathcal{L}(\check{\mu})(x)$ for all $x \geq 0$.*

Therefore, with $\mu, \check{\mu}$ as above we have, using (18), that $\check{\mu} \preceq \mu$, so by Theorem 3 we have $Z_{\check{\mu}}(R_0) \leq Z_\mu(R_0)$ for all R_0. The final size equation (22) for $Z_{\check{\mu}}(R_0)$ is

$$\check{Z} = (1 + c^2)^{-1}[1 - e^{-(1+c^2)R_0\check{Z}}]. \tag{45}$$

Note that if we set $Z' = (1 + c^2)\check{Z}$, then (45) becomes $Z' = 1 - e^{-R_0 Z'}$, which is the same as (31). We therefore have $Z' = Z^*(R_0)$, so

$$\check{Z} = \frac{1}{1 + c^2} Z^*(R_0).$$

We have thus obtained

Theorem 6 *For any susceptibility distribution μ, the corresponding attack is bounded from below by*

$$Z_\mu(R_0) \geq \frac{1}{1 + c_\mu^2} Z^*(R_0), \tag{46}$$

where c_μ is the coefficient of variation of μ and $Z^*(R_0)$ is the attack rate for a homogeneous population, given as the solution of (31).

As $c_\mu \to 0$, the lower bound (46) approaches the upper bound (42). In other words, (46) gives quantitative expression to the fact that low heterogeneity leads to an attack rate which is close to the maximal attack rate possible, which is the attack rate corresponding to the homogeneous case.

To conclude this discussion, we consider a question that arises naturally at this point. We have seen that the attack rate can be bounded from above given the reproductive number alone, while to bound the attack rate from below we also need information on the c.v. of the susceptibility distribution. Our question is whether, given knowledge of the c.v., one can improve the upper bound given in terms R_0 alone, and get an upper bound which is smaller. We now show that the answer is negative, that is, given any values $R_0 > 1, c > 0$ we can find a susceptibility distribution μ with $c_\mu = c$, for which the attack rate $Z_\mu(R_0)$ is as close as we wish to $Z^*(R_0)$, the attack rate in the homogeneous case. To construct such an example, we consider distributions which are supported on two points

$$d\mu(\sigma) = \alpha\delta(\sigma - \sigma_1) + (1 - \alpha)\delta(\sigma - \sigma_2)$$

We fix the first two moments of μ:

$$\alpha\sigma_1 + (1-\alpha)\sigma_2 = \bar{\sigma}, \quad \alpha\sigma_1^2 + (1-\alpha)\sigma_2^2 = \overline{\sigma^2}.$$

Solving this pair of equations for α, σ_2 we have

$$\alpha = \frac{\bar{\sigma}^2 - \overline{\sigma^2}}{2\sigma_1\bar{\sigma} - \sigma_1^2 - \overline{\sigma^2}}, \quad \sigma_2 = \frac{\overline{\sigma^2} - \sigma_1\bar{\sigma}}{\bar{\sigma} - \sigma_1},$$

hence

$$F_\mu(x) = \frac{\bar{\sigma}^2 - \overline{\sigma^2}}{2\sigma_1\bar{\sigma} - \sigma_1^2 - \overline{\sigma^2}} e^{-\frac{\sigma_1}{\bar{\sigma}}x} + \frac{2\sigma_1\bar{\sigma} - \sigma_1^2 - \bar{\sigma}^2}{2\sigma_1\bar{\sigma} - \sigma_1^2 - \overline{\sigma^2}} e^{-\frac{\overline{\sigma^2} - \sigma_1\bar{\sigma}}{\bar{\sigma} - \sigma_1}\frac{1}{\bar{\sigma}}x}$$

Taking the limit $\sigma_1 \to \bar{\sigma}$, we get $\alpha \to 1, \sigma_2 \to \infty$, and therefore $F_\mu(x) \to e^{-x}$ uniformly. Therefore in such a limit the final-size equation (22) becomes arbitrarily close to the final-size equation for the homogeneous model, so that the solution, which gives the attack rate, is arbitrarily close to the attack rate of the homogeneous model, despite the fact that the c.v. of the susceptibility distribution is fixed. Thus knowledge of the c.v. of the susceptibility distribution does not enable to improve the upper bound for the attack rate given by $Z^*(R_0)$.

7 Vaccination in a population with heterogeneous susceptibility

We now apply some of the previous results to consider the effect of vaccination. We assume that the susceptibility distribution of the unvaccinated population is μ, and the reproductive number before vaccination is R_0. A vaccine may be completely effective, in which case any individual who is vaccinated will have susceptibility 0, or leaky, meaning that it only reduces the susceptibility of those vaccinated by a given factor $0 < r < 1$ (Halloran et al. 2009). If we vaccinate the entire population with such a leaky vaccine, the susceptibility distribution will be the scaled distribution μ_r, defined by

$$\mu_r([a,b]) = \mu([r^{-1}a, r^{-1}b]), \tag{47}$$

that is, the fraction of the population which will have susceptibility in the interval $[a,b]$ after vaccination is the fraction of the population that had susceptibility in the interval $[r^{-1}a, r^{-1}b]$ prior to vaccination.

More generally, if we have a limited quantity of vaccine, sufficient to vaccinate a fraction v of the population, the susceptibility distribution after vaccination will be the mixture of distributions

$$\mu_{r,v} = v\mu_r + (1-v)\mu.$$

The mean susceptibility after vaccination is

$$\bar{\sigma}_{\mu_{r,v}} = v\bar{\sigma}_{\mu_r} + (1-v)\bar{\sigma}_\mu = [1 - v(1-r)]\bar{\sigma}_\mu,$$

where $\bar{\sigma}_\mu$ is the mean susceptibility before vaccination, so that the post-vaccination reproduction number is

$$R_0^{r,v} = \lambda\bar{\sigma}_{\mu_{r,v}} = R_0[1 - v(1-r)],$$

and by Theorem 2, the condition for achieving herd immunity and preventing an epidemic is $R_0^{r,v} \le 1$, that is

$$v(1-r) \ge 1 - \frac{1}{R_0}. \tag{48}$$

With Φ defined by (25), we have, by (47),

$$\Phi_{\mu_{r,v}}(\sigma) = v\Phi_{\mu_r}(\sigma) + (1-v)\Phi_\mu(\sigma) = v\Phi_\mu(r^{-1}\sigma) + (1-v)\Phi_\mu(\sigma),$$

so, using (26),

$$\mathcal{L}(\Phi_{\mu_{r,v}})(x) = vr\mathcal{L}(\Phi_\mu)(rx) + (1-v)\mathcal{L}(\Phi_\mu)(x),$$

$$G_{\mu_{r,v}}(x) = \frac{1}{1-v(1-r)}\Big[vrG_\mu\Big(\frac{rx}{1-v(1-r)}\Big) + (1-v)G_\mu\Big(\frac{x}{1-v(1-r)}\Big)\Big].$$

In case that (48) does not hold, so that vaccination is not sufficient for preventing an epidemic, we thus have

Proposition 1 *For any susceptibility distribution μ, assuming the reproduction number before vaccination is R_0 and $v(1-r) < 1 - \frac{1}{R_0}$, the attack rate*

$$Z_\mu(r, v, R_0) = Z_{\mu_{r,v}}(R_0^{r,v})$$

after vaccination of a fraction v of the population which reduces susceptibility by a fraction r is the positive solution of the equation

$$R_0\Big[vrG_\mu(rR_0Z) + (1-v)G_\mu(R_0Z)\Big] = 1. \tag{49}$$

While the threshold condition (48) for preventing an epidemic depends only on the reproductive number R_0 before vaccination and on the product $v(1-r)$, the dependence of the attack rate on r, v, expressed by (49), is not only through the product $v(1-r)$, so that different combinations of v, r with the same value of $v(1-r)$, which thus lead to the same reduction of the reproduction number, can lead to very different attack rates.

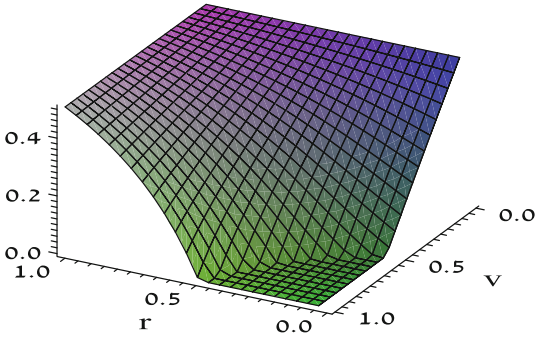

Fig. 3 Attack rate as a function of the fraction of population vaccinated v and reduction in susceptibility of vaccinated individuals r, for a population with exponentially distributed susceptibility and $R_0 = 2$

We now consider the example in which the susceptibility distribution before vaccination is Gamma-distributed, see (33), in which case G_μ is given by (34), so (49) is

$$Z = 1 - v\Big(1 + \frac{1}{\kappa} r R_0 Z\Big)^{-\kappa} - (1 - v)\Big(1 + \frac{1}{\kappa} R_0 Z\Big)^{-\kappa}.$$

In particular, when $\kappa = 1$ so that susceptibility is exponentially distributed $d\mu(\sigma) = \frac{1}{\bar{\sigma}} e^{-\frac{1}{\bar{\sigma}}\sigma} d\sigma$, the above equation, after simplification and division by Z to eliminate the trivial solution, gives a quadratic equation, whose solution in $(0, 1)$ is

$$Z_\mu(r, v, R_0) = \frac{1}{2R_0 r}\Big[(R_0 - 1)r - 1 + \Big([(R_0 - 1)r + 1]^2 - 4vR_0 r(1 - r)\Big)^{\frac{1}{2}}\Big]. \tag{50}$$

The attack rate as a function of r, v, for $R_0 = 2$, is plotted in Fig. 3.

8 Recurring epidemics

In this section we consider the following situation: a population undergoes an epidemic with a pathogen that, upon recovery, induces a complete short-term protection (say for a few months), so that individuals do not get infected more than once during the epidemic. At a later point in time (perhaps a year later), the pathogen re-enters the population and a new epidemic occurs. By this time the individuals who had been infected in the previous epidemic have only a partial protection to being infected, reducing their susceptibility by a factor r. This can be due either to waning immunity or to antigenic evolution of the pathogen. The individuals who had not been infected during the first epidemic are, of course, just as susceptible to the new epidemic as the were to the previous one. We want to study the degree of protection that the previous encounter with the pathogen affords the population, and answer questions like:

– What are the conditions on the susceptibility distribution μ, the reproductive number R_0 for the first epidemic, and the reduction of susceptibility r, so that a second epidemic will be prevented?

– In case a second epidemic does occur, how does its size depend on the above factors?

The essential difference between the situation considered here and that of vaccination studied in the previous section is that in the case of vaccination of a randomly chosen fraction of the population we vaccinate an equal fraction of people with low susceptibility and with high susceptibility. In the case of partial immunity afforded by exposure to the pathogen during the first epidemic, individuals with a higher susceptibility will have been infected at a higher rate so that a larger proportion of them will have some protection. Let us mention the paper of Bansal and Meyers (2008) which studies closely related questions under different heterogeneity assumptions.

Let μ be the susceptibility distribution of the population before the first epidemic, and R_0 the reproductive number, so that $Z_\mu(R_0)$ is the attack rate of the first epidemic. We are going to calculate the susceptibility distribution $\tilde{\mu}$ and the reproductive number $\tilde{R}_0$ for the second epidemic.

Since, of the individuals with susceptibility σ, a fraction $e^{-R_0 Z_\mu(R_0)\frac{\sigma}{\bar{\sigma}_\mu}}$ were not infected during the first epidemic (see (30)) so that their susceptibility remains the same, while a fraction $1-e^{-R_0 Z_\mu(R_0)\frac{\sigma}{\bar{\sigma}_\mu}}$ were infected so that their susceptibility is reduced by a factor of r, we get that the susceptibility distribution following the first epidemic is given by

$$\tilde{\mu}([a,b]) = \int_a^b e^{-R_0 Z_\mu(R_0)\frac{\sigma}{\bar{\sigma}_\mu}} d\mu(\sigma) + \int_{r^{-1}a}^{r^{-1}b} (1-e^{-R_0 Z_\mu(R_0)\frac{\sigma}{\bar{\sigma}_\mu}}) d\mu(\sigma). \tag{51}$$

We therefore have

$$\begin{aligned}
\mathcal{L}(\Phi_{\tilde{\mu}})(x) &= \int_0^\infty e^{-x\sigma}\Big[\int_\sigma^\infty e^{-R_0 Z_\mu(R_0)\frac{\sigma'}{\bar{\sigma}_\mu}} d\mu(\sigma') + \int_{r^{-1}\sigma}^\infty (1-e^{-R_0 Z_\mu(R_0)\frac{\sigma'}{\bar{\sigma}_\mu}}) d\mu(\sigma')\Big] d\sigma \\
&= \int_0^\infty e^{-R_0 Z_\mu(R_0)\frac{\sigma'}{\bar{\sigma}_\mu}} \int_0^{\sigma'} e^{-x\sigma} d\sigma d\mu(\sigma') + \int_0^\infty (1-e^{-R_0 Z_\mu(R_0)\frac{\sigma'}{\bar{\sigma}_\mu}}) \int_0^{r\sigma'} e^{-x\sigma} d\sigma d\mu(\sigma') \\
&= \frac{1}{x}\int_0^\infty e^{-R_0 Z_\mu(R_0)\frac{\sigma'}{\bar{\sigma}_\mu}} [1-e^{-x\sigma'}] d\mu(\sigma') + \frac{1}{x}\int_0^\infty (1-e^{-R_0 Z_\mu(R_0)\frac{\sigma'}{\bar{\sigma}_\mu}})[1-e^{-rx\sigma'}] d\mu(\sigma') \\
&= \frac{1}{x}\Big[1 - F_\mu\Big(\bar{\sigma}_\mu x + R_0 Z_\mu(R_0)\Big) - F_\mu(r\bar{\sigma}_\mu x) + F_\mu\Big(r\bar{\sigma}_\mu x + R_0 Z_\mu(R_0)\Big)\Big],
\end{aligned} \tag{52}$$

hence, by (26),

$$G_{\tilde{\mu}}(x) = \frac{1}{x}\Big[1 - F_\mu\Big(\frac{\bar{\sigma}_\mu}{\bar{\sigma}_{\tilde{\mu}}} x + R_0 Z_\mu(R_0)\Big) - F_\mu\Big(r\frac{\bar{\sigma}_\mu}{\bar{\sigma}_{\tilde{\mu}}} x\Big) + F_\mu\Big(r\frac{\bar{\sigma}_\mu}{\bar{\sigma}_{\tilde{\mu}}} x + R_0 Z_\mu(R_0)\Big)\Big]. \tag{53}$$

Taking the limit $x \to 0$ in (53) and using the fact that $G_{\tilde{\mu}}(0) = 1$ (lemma 1(i)) and (27), we get

$$1 = \lim_{x\to 0} \frac{1}{x}\Big[1 - F_\mu\Big(\bar{\sigma}_\mu x + R_0 Z_\mu(R_0)\Big) - F_\mu(r\bar{\sigma}_\mu x) + F_\mu\Big(r\bar{\sigma}_\mu x + R_0 Z_\mu(R_0)\Big)\Big]$$
$$= [r - (1-r)F'_\mu(R_0 Z_\mu(R_0))]\frac{\bar{\sigma}_\mu}{\bar{\sigma}_{\tilde{\mu}}},$$

hence

$$\bar{\sigma}_{\tilde{\mu}} = [r - (1-r)F'_\mu(R_0 Z_\mu(R_0))]\bar{\sigma}_\mu. \tag{54}$$

The reproductive number for the second epidemic is therefore, assuming the infectivity λ remains the same,

$$\tilde{R}_0 = \lambda\bar{\sigma}_{\tilde{\mu}} = \lambda\bar{\sigma}_\mu[r - (1-r)F'_\mu(R_0 Z_\mu(R_0))] = R_0[r - (1-r)F'_\mu(R_0 Z_\mu(R_0))]. \tag{55}$$

The threshold condition for a second epidemic is $\tilde{R}_0 > 1$, or, equivalently

$$r > r_0(R_0) \doteq 1 - \frac{1 - R_0^{-1}}{1 + F'_\mu(R_0 Z_\mu(R_0))}. \tag{56}$$

Thus a second epidemic will be prevented if the protection afforded by previous infection is sufficiently strong, that is if $r \le r_0(R_0)$. Assuming that $R_0 > 1$, i.e. that a first epidemic occurred, we have $r_0(R_0) < 1$, so that there is always a range $r \in (r_0(R_0), 1)$ of values for which the partial immunity generated by the first epidemic is too weak to prevent a second epidemic.

The relation $r = r_0(R_0)$ defines a curve in the (R_0, r) plane, such that a second epidemic occurs when (R_0, r) is above this curve. Let us now calculate the curve $r = r_0(R_0)$ explicitly for two specific susceptibility distributions. In the case of a homogeneous population with susceptibility $\bar{\sigma}$, $d\mu(\sigma) = \delta(\sigma - \bar{\sigma})d\sigma$, we have $F_\mu(x) = e^{-x}$, so $Z_\mu(R_0) = Z^*(R_0)$ given as the solution of (31), so

$$F'_\mu(R_0 Z_\mu(R_0)) = -e^{-R_0 Z^*(R_0)} = Z^*(R_0) - 1$$

and

$$r_0(R_0) = 1 - \frac{1 - R_0^{-1}}{Z^*(R_0)}.$$

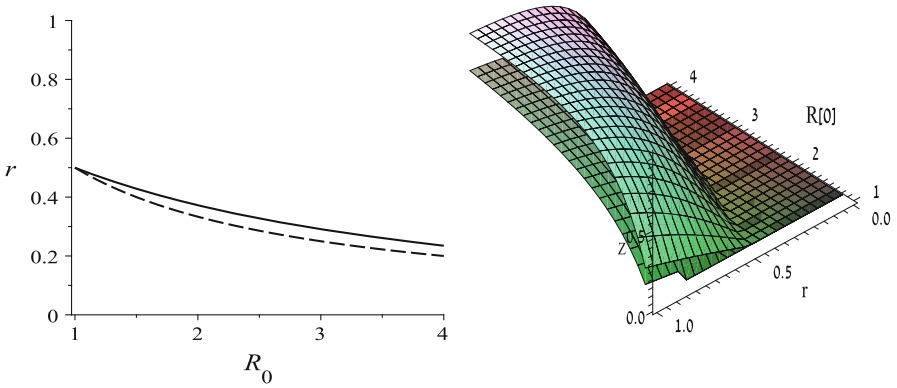

Fig. 4 *Left*. The curves $r = r_0(R_0)$ for a homogeneous population and for a population with exponentially distributed susceptibility (*dashed*). *Right*. Attack rate for a second epidemic as a function of reproduction number R_0 of the first epidemic and reduction of susceptibility r following infection. The higher surface corresponds to homogeneous susceptibility, and the lower surface corresponds to exponentially distributed susceptibility

In the case of a population with exponentially distributed susceptibility $d\mu = \frac{1}{\bar{\sigma}} e^{-\frac{1}{\bar{\sigma}}\sigma} d\sigma$, we have $F_\mu(x) = (1+x)^{-1}$, $Z_\mu(R_0) = 1 - \frac{1}{R_0}$ (see (36)), and therefore

$$F'_\mu(R_0 Z_\mu(R_0)) = -(1 + R_0 Z_\mu(R_0))^{-2} = -R_0^{-2},$$
$$r_0(R_0) = 1 - \frac{1 - R_0^{-1}}{1 - R_0^{-2}} = \frac{1}{R_0 + 1}.$$

The curves $r = r_0(R_0)$ for the two cases that were calculated above are plotted in Fig. 4 (left). Recall that the region above each curve is the region for which a second epidemic will occur for the corresponding susceptibility distribution. An interesting fact we notice when looking at these two curves is that for both of them $r_0(1) = \frac{1}{2}$. It turns out that this is a general fact:

Theorem 7 *For any susceptibility distribution μ the corresponding curve $r = r_0(R_0)$ satisfies $r_0(1) = \frac{1}{2}$.*

The meaning of this result, thinking now of r as fixed and varying R_0, is that if $r \geq \frac{1}{2}$, that is if immunity following infection reduces susceptibility by less than 50%, then for *any* value of $R_0 > 1$ a second epidemic will occur, while if previous infection reduces susceptibility by more than 50% ($r < \frac{1}{2}$) then then an epidemic will *not* occur if R_0 is smaller then a critical value given by $r_0^{-1}(r)$ (see Fig. 4 left). We call this surprising result the 50%-law.

Proof By the definition of $r_0(R_0)$, we need to show that

$$\lim_{R_0 \to 1+} \frac{1 - R_0^{-1}}{1 + F'_\mu(R_0 Z_\mu(R_0))} = \frac{1}{2}. \tag{57}$$

Since $Z_\mu(1) = 0$, we have $F'_\mu(Z_\mu(1)) = F'_\mu(0) = -1$, and we see that both the numerator and the denominator in (57) vanish when $R_0 = 1$. Therefore to evaluate the limit we use L'Hôpital's rule,

$$\lim_{R_0\to 1+} \frac{1 - R_0^{-1}}{1 + F'_\mu(R_0 Z_\mu(R_0))} = \lim_{R_0\to 1+} \frac{\frac{d}{dR_0}[1 - R_0^{-1}]}{\frac{d}{dR_0}[1 + F'_\mu(R_0 Z_\mu(R_0))]}$$
$$= \lim_{R_0\to 1+} \frac{R_0^{-2}}{F''_\mu(R_0 Z_\mu(R_0))[Z_\mu(R_0) + R_0 Z'_\mu(R_0)]} = \frac{1}{F''_\mu(0) Z'_\mu(1)}.$$

We now recall (28), which tells us that $F''_\mu(0) = c_\mu^2 + 1$, and Lemma 4, which tells us that $Z'_\mu(1) = \frac{2}{c_\mu^2+1}$, so we get (57).

Let us now assume that $r > r_0(R_0)$, that is $\tilde{R}_0 > 1$, so that a second epidemic occurs, and denote the attack rate of the second epidemic by $\tilde{Z} = Z_{\tilde{\mu}}(\tilde{R}_0)$. Computing the final-size equation (24), using (53) and the fact that $\frac{\bar{\sigma}_\mu}{\bar{\sigma}_{\tilde{\mu}}} = \frac{R_0}{\tilde{R}_0}$, we have

Proposition 2 *In case $r > r_0(R_0)$, the attack rate $\tilde{Z}$ of the second epidemic is given as the positive solution of*

$$\tilde{Z} = 1 - F_\mu\Big(R_0(\tilde{Z} + Z_\mu(R_0))\Big) + F_\mu\Big(R_0 r\tilde{Z}\Big) - F_\mu\Big(R_0(r\tilde{Z} + Z_\mu(R_0))\Big).$$

For a homogeneous population we get the equation

$$\tilde{Z} = 1 - e^{-R_0(\tilde{Z}+Z^*(R_0))} - e^{-rR_0\tilde{Z}} + e^{-R_0(r\tilde{Z}+Z^*(R_0))},$$

and using the fact that $Z^*(R_0) = 1 - e^{-R_0 Z^*(R_0)}$ this can be written as

$$\tilde{Z} = 1 - (1 - Z^*(R_0))e^{-R_0\tilde{Z}} - Z^*(R_0)e^{-rR_0\tilde{Z}}. \tag{58}$$

As another example, in the case of an exponential susceptibility distribution we have

$$\tilde{Z} = 1 - (1 + R_0(Z_\mu(R_0) + \tilde{Z}))^{-1} - (1 + R_0 r\tilde{Z})^{-1} + (1 + R_0(Z_\mu(R_0) + r\tilde{Z}))^{-1},$$

and using the fact that $Z_\mu(R_0) = 1 - \frac{1}{R_0}$ we get

$$\tilde{Z} = 1 - (R_0(1 + \tilde{Z}))^{-1} - (1 + R_0 r\tilde{Z})^{-1} + (R_0(1 + r\tilde{Z}))^{-1}, \tag{59}$$

which can be written as a third order polynomial equation for $\tilde{Z}$.

In Fig. 4 (right) we plot the attack rate as a function of R_0 and r for the two cases considered above, by solving Eqs. (58) and (59) numerically.

Acknowledgments The author acknowledges support of EU-FP7 Grant Epiwork.

References

Andersson H, Britton T (1998) Heterogeneity in epidemic models and its effect on the spread of infection. J Appl Probab 35:651–661

Andreasen V (2011) The final size of an epidemic and its relation to the basic reproduction number. Bull Math Biol (Online First)

Ball F (1985) Deterministic and stochastic epidemics with several kinds of susceptibles. Adv Appl Probab 17:1–22

Bansal S, Meyers LA (2008) The impact of past epidemics on future disease dynamics. Preprint, arxiv:0910.2008v

Bellamy R (ed) (2004) Susceptibility to infectious diseases: the importance of host genetics. Cambridge University Press, Cambridge

Bonzi B, Fall AA, Iggidr A, Sallet G (2010) Stability of differential susceptibility and infectivity models, epidemic models. J Math Biol [Epub ahead of print]

Brauer F (2008) Age-of-infection and the final size relation. Math Biosci Eng 5(2008):681–690

Coutinho FAB, Massad E, Lopez LF, Burattini MN, Struchiner CJ, Azevedo-Neto RS (1999) Modelling heterogeneities in individual frailties in epidemic models. Math Comput Model 30:97–115

Craig A, Scherf A (2003) Antigenic variation. Academic Press, Amsterdam

Diekmann O, Heesterbeek JAP (2000) Mathematical epidemiology of infectious diseases. Wiley, New York

Dwyer G, Elkinton JS, Buonaccorsi JP (1997) Host heterogeneity in susceptibility and disease dynamics: tests of a mathematical model. Am Nat 150:685–707

Dwyer G, Dushoff J, Elkinton JS, Levin SA (2000) Pathogen-driven outbreaks in forest defoliators revisited: building models from experimental data. Am Nat 156:105–120

Frank SA (2002) Immunology and evolution of infectious diseases. Princeton University Press, Princeton

Gart J (1972) The statistical analysis of chain-binomial epidemic models with several kinds of susceptibles. Biometrics 28:921–930

Halloran ME, Longini IM, Struchiner CJ (2009) Design and analysis of vaccine studies. Springer, New York

Hyman JM, Li J (2005) Differential susceptibility epidemic models. J Math Biol 50:626–644

Karev GP (2005) Dynamics of heterogeneous populations and communities and evolution of distributions. Discrete Contin Dyn Sys (Suppl):487–496

Lefévre C, Picard P (1995) Collective epidemic processes: a general modelling approach to the final outcome of SIR infectious diseases. In: Mollison J (ed) Epidemic models: their structure and relation to data

Ma J, Earn DJD (2006) Generality of the final size formula for an epidemic of a newly invading infectious disease. Bull Math Biol 68:679–702

May RM, Anderson RM, Irwin ME (1988) The transmission dynamics of human immunodeficiency virus (HIV). Philos Trans R Soc Lond B 321:565–607

Novozhilov AS (2008) On the spread of epidemics in a closed heterogeneous population. Math Biosci 215:177–185

Pastor-Satorras R, Vespignani A (2001) Epidemic spreading in scale free networks. Phys Rev Lett 86:3200–3203

Rass L, Radcliffe J (2003) Spatial deterministic epidemics. American Mathematical Society, Providence

Reluga TC, Medlock J, Perelson AS (2008) Backward bifurcation and multiple equilibria in epidemic models with structured immunity. J Theor Biol 252:155–165

Rodrigues P, Margheri A, Rebelo C, Gomes MGM (2009) Heterogeneity in susceptibility to infection can explain high reinfection rates. J Theor Biol 259:280–290

Scalia-Tomba G (1986) Final-size distribution of the multitype Reed–Frost process. J Appl Probab 23:563–584

Shaked M, Shanthikumar JG (2007) Stochastic orders. Springer, New York

Veliov VM (2005) On the effect of population heterogeneity on dynamics of epidemic diseases. J Math Biol 51:124–143

White LJ, Medley GF (1998) Microparasite population dynamics and continuous immunity. Proc R Soc Lond B 265:1977–1983

Yan P, Feng Z (2010) Variability order of the latent and the infectious periods in a deterministic SEIR epidemic model and evaluation of control effectiveness. Math Biosci 224:43–52