



## A Reed–Frost Model Taking into Account Uncertainties in the Diagnostic of the Infection

RENÉE X. DE MENEZES\*

Department of Medical Statistics,  
Leiden University Medical Centre,  
Postbus 9604,  
2300 RC Leiden,  
The Netherlands

*E-mail:* r.x.menezes@lumc.nl

NELI R. S. ORTEGA AND EDUARDO MASSAD

Discipline of Medical Informatics,  
School of Medicine,  
University of São Paulo,  
Brazil

In this paper, we model the epidemic course of a pathogen infection within a semi-closed group which generates clinical signals which do not necessarily permit its ready and certain identification. Typical examples of such a pathogen are influenza-type viruses. We allow for time-varying infectivity levels among individuals, and model the probability of infection per contact as a function of the clinical signals. In order to accomplish this, we introduce a modified chain-binomial Reed–Frost model. We obtain an expression for the basic reproduction ratio and determine conditions which guarantee that the epidemic does not survive in the long-term. These conditions being functions of the signal's distribution, they can be used to design and evaluate interventions, such as treatment protocols.

© 2003 Society for Mathematical Biology. Published by Elsevier Ltd. All rights reserved.

### 1. INTRODUCTION

The Reed–Frost model was proposed by Reed and Frost in a series of lectures held at Johns Hopkins University (Abbey, 1952). It is a particular case of a chain-binomial model, in which it is assumed that each infected individual infects susceptible individuals independently, and that individuals are under the same contact rate with each other. If we represent by  $p$  the probability of a contact between a susceptible and an infected individual resulting in a new case, we have that, at time  $t$ , the probability that a susceptible individual does become infected,  $c_t$ , is

---

\* Author to whom correspondence should be addressed.

equal to the probability of at least one infectious contact, i.e.,

$$c_t = 1 - (1 - p)^{I_t}, \quad t > 0, \quad (1)$$

where  $I_t$  is equal to the number of infected individuals at time  $t$ . Time  $t$  is assumed to be a discrete variable, and an individual's classification, as either susceptible, infected or resistant, can only change when time changes from  $t$  to  $t + 1$ . Other assumptions are that the probability of an infectious contact is fixed along the epidemic course and is the same, for all individuals.

The Reed–Frost model (1) can be used to describe the spread of any infectious disease affecting closed, uniformly-mixed groups. The group has a constant and small size  $N$ , is homogeneous, both from the susceptibility and infectivity viewpoints [see Bailey (1975)], with individual members spending a significant and constant part of the day in close contact. From the infection viewpoint, the infectious period is assumed to be short compared to the incubation period which, in its turn, is taken as constant.

The Reed–Frost model assumes that individuals are classified according to their disease status: susceptible and infected and, in some cases, also resistant or immune. No error involved in the classification process, such as a truly infected individual being classified as susceptible, is considered in the model. For a great number of infectious diseases, however, such a diagnostic test is neither readily nor easily available: examples are influenza and several other viral and bacterial infections. The corresponding diagnostic process involves uncertainty, and is based upon a set of clinical characteristics, often subjective, which we call *signals*. It is then important to consider, in the epidemic model, the uncertainty involved in the classification process.

The homogeneity assumption is unlikely to hold in real epidemics, especially in large groups [see, for example, Becker (1979)]. In certain cases, the assumption of time-invariant susceptibility/infectivity levels does not hold either. Each individual may have a varying susceptibility level to infections, depending on physical and psychological factors, even within a short-lasting epidemic. Infectivity levels may also vary according to similar factors. Indeed, the capacity of an infected individual to produce an infectious contact may depend upon the set of signals developed. Some signals, such as sneezing and coughing in influenza-type infections, may increase the probability that a contact be infectious, while others, such as fever, may decrease it, by making the host less prone to contacts.

In this paper, we shall consider studies involving small groups, within which both homogeneous mixing and homogeneous susceptibility still hold. We consider the clinical signals involved in the classification process in the study of the epidemic course. These clinical signals may include symptoms, results from laboratorial and physical exams. We assume that, after being infected, no resistance is gained and the individual becomes susceptible again.

In Section 2 we model an individual's infectivity as a function of the signals, therefore allowing for time-dependent, heterogeneous infectivity. We consider this

model in the context of both prospective and retrospective studies. For each study, we obtain expressions for the epidemic basic reproduction ratio in [Section 3](#), and for its probability function in [Section 4](#). In [Section 5](#) conditions are given under which the proposed model is reduced to the classic Reed–Frost model. Finally, in [Sections 6](#) and [7](#) we illustrate how the results can be applied to study long-term disease establishment and impact of intervention, respectively.

## 2. A REED–FROST MODEL WITH UNCERTAINTIES

**2.1. Overview.** Clinical signals are recorded and are taken into account in the epidemic course via a signal summary, both as part of the classification process and to define the probability of an infectious contact. It is assumed that, the higher the signal summary, the higher the probability that a contact be infectious. The probability that an individual has at least one infectious contact, which is the core of the Reed–Frost model, is then computed taking into account the heterogeneous infectivity in the group.

The model can include both signals linked to an increased infectiousness and signals linked to a decreased infectiousness. Both types of signals enter the signal summary, affecting it in opposite directions. Distinct signals can have different weights in the summary, reflecting the impact they are believed to have on both the classification process and on the infectious contact probability.

A probability distribution is assigned to the signal summary, conditioning on the previous probability of at least one infectious contact. This distribution is a mixture of the one given the individual is infected, with the one given the patient is susceptible. The classification is seen as a probabilistic step conditioned on the signal summary. The probability of an infectious contact is taken as a deterministic, polynomial function of the signal summary.

**2.2. Definitions.** In this formulation, we take a susceptible individual as reference, and construct a generalized Reed–Frost model. Hereafter, upper case letters such as  $(S_{ij,t}, D_{i,t}, G_{i,t}, P_{ij,t}, C_{i,t})$  represent random variables, while lower case letters  $(s_{ij,t}, d_{i,t}, g_{i,t}, p_{ij,t}, c_{i,t})$  represent values the corresponding variables may assume. Greek letters  $(\alpha, \beta, \eta)$  represent possibly unobservable quantities.

We first assume that, at time  $t$ , each individual  $i$  has a true health status represented by  $\eta_{i,t}$ , which takes value 1 if the individual is infected at  $t$ , and 0 if the individual is susceptible. Thus, the number of individuals infected at  $t$  is given by

$$I_t = \sum_{i=1}^N \eta_{i,t}. \quad (2)$$

Each individual has one or more clinical signals, which can be summarized by one variable  $D_{i,t}$ , taking values between 0 and 1. At time  $t$ , the probability  $P_{il,t}$  that a contact between a susceptible individual  $i$  and an infected individual  $l$  results

in a new case is a function of the signals of the infected individual only,  $D_{l,t}$ , as a consequence of the susceptibility homogeneity assumption. We assume in particular that this function can be written as a polynomial of degree  $M$ . That is,

$$P_{il,t} \equiv P_{l,t} = \sum_{j=1}^M \phi_j D_{l,t}^j, \quad (3)$$

where  $0 \leq \phi_j \leq 1$  and  $\sum_j \phi_j = 1$ , that is,  $P_{l,t}$  is a convex combination of  $D_{l,t}^j$ , guaranteeing that  $P_{l,t} \in [0, 1]$  for all  $i, t$ . Then the probability that a susceptible individual has, at time  $t$ , at least one infectious contact defines our Reed–Frost model as

$$C_t = 1 - \prod_{l=1}^N (1 - P_{l,t})^{\eta_{l,t}}. \quad (4)$$

We can interpret  $C_t$  here as the probability that an individual be infected at time  $t + 1$ , as in the classic Reed–Frost model, and we write  $C_t = P\{\eta_{i,t+1} = 1\}$ .

In some instances  $\eta_{i,t}$  is unknown, so individuals have to be diagnosed as either infected or susceptible. This consists of a classification procedure which takes into account the clinical signals or, for simplicity, the signals summary  $D_{i,t}$ , and is defined outside the model, probably by specialists. Let  $G_{i,t} = 1$  indicate that the individual  $i$  is diagnosed as infected at  $t$ , and  $G_{i,t} = 0$  indicate that the individual is diagnosed as susceptible. The number of individuals diagnosed as infected at time  $t$  is an estimator of the number of infected individuals at  $t$ , and is given by

$$\hat{I}_t = \sum_{i=1}^N G_{i,t}. \quad (5)$$

The probability that a contact between a susceptible individual  $i$  and an infected individual  $l$  results in a new case is defined as before by (3) and, in this case, (4) is estimated as

$$\hat{C}_t = 1 - \prod_{l=1}^N (1 - P_{l,t})^{G_{l,t}}. \quad (6)$$

Thus,  $\hat{C}_t$  here is the estimated probability that an individual be infected at time  $t + 1$ .

**2.3. Two kinds of study.** This approach can be used in two related contexts. One is that of a retrospective study, in which patients' health status are observable and modelled as random variables. The objective of such a study is typically to estimate the parameters of the signals' distributions, and it involves relations (2)–(4).

Once these parameter estimates are available, the approach can be used in a prospective study, where the true  $\{\eta_{i,t}\}$  are not known, due to either time or cost

constraints. Such a study could include as an objective evaluating the function  $P(\cdot)$ , and it involves expressions (3), (5) and (6).

This consists of recording patients' clinical signals over a certain period of time, and then estimating their true health status using the model and the classification process.

**2.4. Probabilistic structure for retrospective study.** Each infectious disease produces clinical signals with varying degrees of severity, which depend upon both pathogenic and individual variability. Susceptible individuals may also present some of these signals, for reasons other than infection by the pathogen considered, but it is expected that they do so with a lower severity than if they were infected.

The true health status  $\eta_{i,t}$  is a binary variable. For  $t > 1$  and given all the epidemic information up to  $t - 1$ ,  $P\{\eta_{i,t} = 1\}$  is equal to the probability of having at least one infectious contact at  $t - 1$ ,  $C_{t-1}$ , which is the same for all individuals due to the homogeneous mixing assumption. For  $t = 1$ , we define  $P\{\eta_{i,1} = 1\} \equiv \theta_1$  as the *a priori* probability that any individual is infected at the epidemic onset, which must be evaluated via populational measurements (e.g., the estimated prevalence of the pathogen).

For the pathogen under study, represent by  $X_I$  the clinical summary for any infected individual, and by  $X_S$  the clinical summary for any susceptible individual. Given an individual's health status  $\eta$ ,  $X_I$ ,  $X_S$  are random variables intrinsically linked to the pathogen, their distributions remaining unaffected by the epidemic progress. Here we shall assume that they take a value within the interval  $[0, 1]$  with a distribution within the beta family, as follows:

$$X_I \sim \text{Beta}(\alpha_I, \beta_I), \quad X_S \sim \text{Beta}(\alpha_S, \beta_S).$$

We define  $\mu_I = E(X_I)$  and  $\mu_S = E(X_S)$ . Note that all moments  $E(X_I^k)$ ,  $E(X_S^k)$  of  $X_I$ ,  $X_S$  are finite, for all  $k = 1, 2, \dots$  (see appendix for their expressions). We also define  $\Delta = X_I - X_S$  and  $\delta \equiv E(\Delta) = \mu_I - \mu_S$ .

The observed clinical summary for individual  $i$  at time  $t$ ,  $D_{i,t}$ , is equal to  $X_I$  if the individual is infected and it is equal to  $X_S$  if the individual is susceptible. So, we can write

$$D_{i,t} = X_I \eta_{i,t} + X_S (1 - \eta_{i,t}) = X_S + \Delta \eta_{i,t}, \quad (7)$$

for all  $i = 1, \dots, N$  and all  $t \geq 1$ . For  $t > 1$ , the conditional mean of  $D_{i,t}$ , given  $C_{t-1}$ , is given by

$$E(D_{i,t} \mid C_{t-1}) = \mu_S + \delta C_{t-1}. \quad (8)$$

Similarly for  $t = 1$ ,

$$E(D_{i,t} \mid \theta_1) = \mu_S + \delta \theta_1. \quad (9)$$

In other words, the expected value of the signal summary at time  $t$  is a convex combination of the mean signals for infected and susceptible individuals, based upon the probability of being truly infected.

Given  $C_{t-1}$ , the number of infected,  $I_t$ , defined by (2), is a sum of conditionally independent binomial variables, all with the same probability of success:  $C_{t-1}$  for  $t > 1$ , and  $\theta_1$  for  $t = 1$ . Thus,  $I_t$  is in this case a binomial random variable with probability  $C_{t-1}$  and sample size  $N$ .

Since the probability of an infectious contact  $P_{l,t}$  is a deterministic function of  $D_{l,t}$ , it is a constant, when  $D_{l,t}$  is given. Unconditionally, a probability distribution is effectively assigned to each probability of an infectious contact,  $P_{l,t}$ , for each individual  $l$  at each generation  $t$ ,  $l = 1, \dots, N$ ,  $t > 0$ . The probability of at least one infectious contact,  $C_t$ , does not have a well-known probabilistic distribution, but its conditional expected value, given the clinical summaries  $\{D_{i,t}\}$ , can be computed.

In the above, we have preferred to perform calculations in terms of  $\mu_S$ ,  $\delta$  instead of  $\mu_S$ ,  $\mu_I$ . This separates the contribution of the signal's distribution, which can be treated as unaffected by the disease spread, from the probability that each individual is infected. Therefore, it incorporates treatment effects naturally as a reduction of the difference between mean signals,  $\delta = \mu_I - \mu_S$ .

**2.5. Probabilistic structure for prospective study.** In a prospective study context, most of the structure introduced in Section 2.4 applies, but the diagnostic uncertainty must be included.

When the  $\{\eta_{i,t}\}$  are unknown, patients must be classified as either infected or susceptible. The classification is represented by  $G_{i,t}$  and defined as a simple process: given the clinical summaries  $\{D_{i,t}\}$ , each individual is independently classified as infected with probability:

$$P\{G_{i,t} = 1 \mid D_{i,t}\} = D_{i,t}. \quad (10)$$

We can interpret the outlined probabilistic structure for  $D_{i,t}$ ,  $G_{i,t}$  as defining a conditional binomial distribution for  $G_{i,t}$ , given the probability of success  $D_{i,t}$  which itself has a conditional beta distribution, given the health status  $\eta_{i,t}$ . As a consequence, we have that  $E(G_{i,t} \mid D_{i,t}) = D_{i,t}$ .

The conditional probability that an individual is classified as infected in generation  $t$ , given the probability of at least one infectious contact in the previous generation  $C_{t-1}$ , is given by:

$$\begin{aligned} P\{G_{i,t} = 1 \mid C_{t-1}\} &= E[P\{G_{i,t} = 1 \mid D_{i,t}\} \mid C_{t-1}] \\ &= E[D_{i,t} \mid C_{t-1}], \quad t > 1, \end{aligned} \quad (11)$$

where either (8) or (9) can be used to re-express this as a function of the signals.

The number of patients diagnosed as infected,  $\hat{I}_t$ , defined by (5), is again a sum of conditionally independent binomial variables, given  $\{D_{i,t}\}$ , but since each one of these has a different probability of success, the distribution of  $I_t$  is not the usual binomial.

### 3. THE BASIC REPRODUCTION RATIO

**3.1. Definition.** We shall now study the basic reproduction ratio,  $R_0$ . It is defined as the number of secondary infections resulting from a single case in an entirely susceptible group, during its infectious period (Anderson and May, 1991). In our context,  $R_0 = I_2$ , given that  $I_1 = 1$ . Note that this definition of  $R_0$  is coherent with the Diekmann *et al.* (1990) definition of the next generation operator.

**3.2. Expected value in retrospective study.** In this case, we have

$$E(R_0) = E(I_2 \mid I_1 = 1) = \sum_{j=1}^N E(\eta_{j2} \mid I_1 = 1).$$

By conditioning on  $C_1$ , we get  $E(\eta_{j2} \mid I_1 = 1) = E[E(\eta_{j2} \mid C_1) \mid I_1 = 1] = E(C_1 \mid I_1 = 1)$ , for all  $j = 1, \dots, N$ . This means that

$$E(R_0) = NE(C_1 \mid I_1 = 1). \quad (12)$$

The same result is obtained by considering that, given  $C_1$ ,  $I_2$  has binomial distribution with mean  $NC_1$ . Using the definition of  $I_1$ , we can re-write  $E(C_1 \mid I_1 = 1)$  as

$$\sum_{j=1}^N E\{C_1 \mid \eta_{j1} = 1, I_1 = 1\} P\{\eta_{j1} = 1 \mid I_1 = 1\}. \quad (13)$$

All individuals are equally likely to be the first one infected, so  $P\{\eta_{j1} = 1 \mid I_1 = 1\} = 1/N$ . Moreover, given that  $\eta_{j1} = 1$  and all other  $\eta_{k1}$  are equal to zero, we have from (4) that  $C_1 = P(D_{j1}) = P(X_I)$ , since this individual is infected. Therefore, we can re-write (13) as

$$\frac{1}{N} \sum_{j=1}^N E\{P(X_I)\} = E\{P(X_I)\},$$

and thus

$$E(R_0) = NE\{P(X_I)\} = N \sum_{k=1}^M \phi_k E(X_I^k). \quad (14)$$

In the simple case where  $P(D) \equiv D$ , we get

$$E(R_0) = N\mu_I = N(\mu_S + \delta). \quad (15)$$

**3.3. Expected  $R_0$  in prospective study.** In a prospective study, a diagnostic is estimated with a certain margin of error, and similar arguments to the ones used in Section 3.2 can be used to derive an expression for the expected value of  $R_0$ . Indeed, we now have

$$E(R_0) = E(\hat{I}_2 \mid I_1 = 1) = \sum_{j=1}^N E(G_{j2} \mid I_1 = 1),$$

where it is obvious that we must be sure about the first recorded case being indeed an infection. By conditioning on  $D_{j2}$ , we get  $E(G_{j2} \mid I_1 = 1) = E(D_{j2} \mid I_1 = 1)$ , for all  $j = 1, \dots, N$ . Now conditioning on  $C_1$ , we get  $E(D_{j2} \mid I_1 = 1) = \mu_S + \delta E(C_1 \mid I_1 = 1)$ , for all  $j$ . This means that

$$E(R_0) = N\{\mu_S + \delta E(C_1 \mid I_1 = 1)\}. \quad (16)$$

In Section 3.2 we saw that  $E(C_1 \mid I_1 = 1) = E\{P(X_I)\}$ , and the same still holds here, as conditioning on having one infected individual at  $t = 1$  there is no diagnostic uncertainty at  $t = 1$ . Thus, we obtain

$$E(R_0) = N\{\mu_S + \delta E\{P(X_I)\}\}. \quad (17)$$

In particular, when  $P(D) \equiv D$ , we get

$$E(R_0) = N\{\mu_S + \delta(\mu_S + \delta)\}. \quad (18)$$

#### 4. PROBABILITY FUNCTION FOR $R_0$

**4.1. Motivation.** Our main objective with the evaluation of  $R_0$  and related functions is to define criteria yielding clues as to the long-term disease establishment. While the expected value gives some clues, it is also important to evaluate how likely the value of  $R_0$  is to spread around it. The traditional statistical approach is to construct a confidence interval for  $R_0$ , and check if the value contains those leading to long-term disease establishment, in this case any value greater than, or equal to, 1. In our problem, however, this is of limited use:  $R_0$  being a random variable assuming only nonnegative integer values  $0, 1, 2, \dots$ , all of these values have positive probability mass and, thus, any confidence interval is likely to include the value  $R_0 = 1$  at least. A more useful measurement seems to be the probability that  $R_0 \geq 1$ . While taking uncertainty into account, this is perhaps more useful: if, under certain conditions, it is known that the probability that  $R_0 \geq 1$  is about 20%, then only 1 in 5 independent initial cases are likely to propagate the disease.

**4.2. Retrospective study.** In order to evaluate the uncertainty around the  $R_0$  estimate, we must evaluate its probability distribution. We start with the

retrospective study. We have:

$$\begin{aligned} P\{R_0 = z\} &= P\{I_2 = z \mid I_1 = 1\} \\ &= E(P\{I_2 = z \mid C_1\} \mid I_1 = 1) \\ &= E(n_z C_1^z (1 - C_1)^{N-z} \mid I_1 = 1), \end{aligned}$$

where we have used the fact that, given  $C_1$ ,  $I_2$  has a binomial distribution, with  $n_z$  representing the binomial coefficient equal to  $N!/[z!(N-z)!]$ . If we then condition on the true health status  $\{\eta_{j1}\}$ , we can write

$$\begin{aligned} E(C_1^z (1 - C_1)^{N-z} \mid I_1 = 1) &= \sum_{j=1}^N E\{C_1^z (1 - C_1)^{N-z} \mid \eta_{j1} = 1, I_1 = 1\} \frac{1}{N} \\ &= E\{P(X_I)^z (1 - P(X_I))^{N-z}\}. \end{aligned}$$

In particular for  $z = 0$ ,

$$P\{R_0 = 0\} = E\{(1 - P(X_I))^N\}. \quad (19)$$

**4.3. Prospective study.** Using conditional probability properties, we can write

$$\begin{aligned} P\{R_0 = z\} &= P\{\hat{I}_2 = z \mid I_1 = 1\} = P\left\{\sum_{j=1}^N G_{j2} = z \mid I_1 = 1\right\} \\ &= E\left[P\left\{\sum_{j=1}^N G_{j2} = z \mid \{D_{i2}\}\right\} \mid I_1 = 1\right], \end{aligned} \quad (20)$$

where the fact that, given  $\{D_{i2}\}$ , the random variables  $\{G_{i2}\}$  are conditionally independent of  $I_1 = 1$  is used. Since each  $G_{i2}$  is a binary random variable,  $\sum_{j=1}^N G_{j2} = z$  occurs whenever exactly  $z$  of the  $\{G_{i2}\}$  are equal to 1. Define  $J_z$  as a set of exactly  $z$  indices, ranging from 1 to  $N$ . That is,  $J_z$  is a subset of exactly  $z$  elements of the discrete set  $\{1, 2, \dots, N\}$ . Note that there exist  $n_z = N!/[z!(N-z)!]$  such subsets. Let these be represented by  $J_{z,1}, J_{z,2}, \dots, J_{z,n_z}$ . Given  $\{D_{i2}\}$ , the conditional probability that only the variables within the subset  $\{G_{i2}, i \in J_{z,l}\}$  are equal to 1 is equal to

$$\begin{aligned} &\left[\prod_{j \in J_{z,i}} P\{G_{j2} = 1 \mid D_{j2}\}\right] \left[\prod_{j \notin J_{z,i}} P\{G_{j2} = 0 \mid D_{j2}\}\right] \\ &= \left[\prod_{j \in J_{z,i}} D_{j2}\right] \left[\prod_{j \notin J_{z,i}} \{1 - D_{j2}\}\right]. \end{aligned} \quad (21)$$

When taking the conditional expectation given  $C_1$ , we use the fact that the  $\{D_{i2}\}$  are conditionally independent and we get that (21) is

$$\left[ \prod_{j \in J_{z,i}} E(D_{j2} | C_1) \right] \left[ \prod_{j \notin J_{z,i}} E(1 - D_{j2} | C_1) \right],$$

which means that (20) becomes

$$P\{R_0 = z\} = E[n_z(\mu_S + \delta C_1)^z (1 - \mu_S - \delta C_1)^{N-z} | I_1 = 1]. \quad (22)$$

Note that we have hereby shown that, given  $C_1$ ,  $R_0$  has conditional binomial distribution with mean  $N(\mu_S + \delta C_1)$ . For  $z = 0$  the right-hand side of (22) becomes  $E[(1 - \mu_S - \delta C_1)^N | I_1 = 1]$  and, given that only one individual is infected,  $C_1 = P(X_I)$  as before. Thus, we can write

$$P\{R_0 = 0\} = E\{[1 - \mu_S - \delta P(X_I)]^N | I_1 = 1\} \quad (23)$$

$$= \sum_{k=0}^N \frac{N!}{k!(N-k)!} (-1)^k E\{[\mu_S + \delta P(X_I)]^k | I_1 = 1\}. \quad (24)$$

Now we use

$$\mu_S + \delta P(X_I) = (\mu_S + \delta) \left\{ 1 - \frac{\delta[1 - P(X_I)]}{\mu_S + \delta} \right\},$$

to re-express (24) as

$$\sum_{k=0}^N \sum_{l=0}^k \frac{N!}{k!(N-k)!} \frac{k!}{l!(k-l)!} (-1)^{k+l} \delta^l (\mu_S + \delta)^{k-l} E\{[1 - P(X_I)]^l | I_1 = 1\}.$$

Thus, we get

$$P\{R_0 = 0\} = \sum_{k=0}^N \sum_{l=0}^k \sum_{m=0}^l \frac{N!(-1)^{k+l+m} \delta^l (\mu_S + \delta)^{k-l}}{(N-k)!(k-l)!m!(l-m)!} E\{[P(X_I)]^m\}. \quad (25)$$

**4.4. Evaluating uncertainty.** The probability  $P\{R_0 \geq 1\}$  can be easily evaluated from the formulae obtained in Sections 4.2 and 4.3, since  $P\{R_0 \geq 1\} = 1 - P\{R_0 = 0\}$ . Here we shall assume that  $P(D) \equiv D$  for simplicity, and evaluate this probability in both retrospective and prospective studies, along with  $E(R_0)$ , for a range of values of  $\mu_I, \delta$ , assuming  $N = 5$ . There is no need to evaluate the functions for other values of  $N$  as, from a theoretical viewpoint,  $N$  is just a scale factor. In practice, for larger values of  $N$  the function defining the probability of

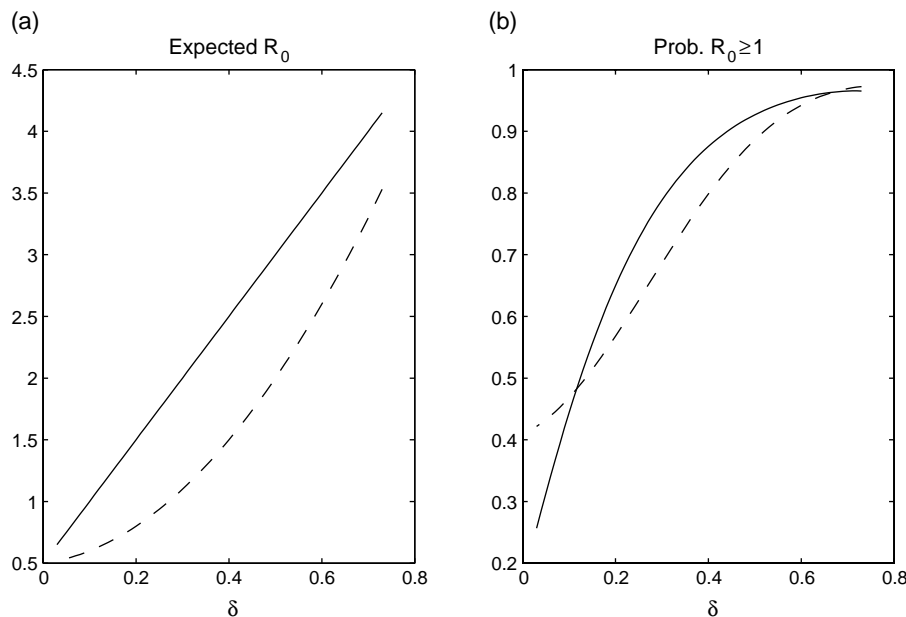


Figure 1. Functions for retrospective study (solid line) and prospective study (dashed line) for several values of  $\delta$ . (a) (left): expected  $R_0$ . (b) (right): probability that  $R_0 \geq 1$ .

an infective contact given the signal,  $P(D)$ , is likely to be other than the identity, reflecting lower contact rates between individuals.

For the retrospective study, we compute  $E(R_0)$  and  $P\{R_0 = 0\}$  using (14) and (19), respectively, whilst for the prospective study we use (17) and (23). In Fig. 1(a) we can see  $E(R_0)$  as a function of  $\delta$  in both kinds of study, assuming the variances for both  $X_S$ ,  $X_I$  are fixed as equal to 0.06,  $\mu_S = 0.1$  and  $\mu_I$  varies from 0.13 to 0.83 by 0.1. First, we note that the expected value in the retrospective study is an upper bound for the one in the prospective study. This suggests that the uncertainty involved in the diagnostic process implies an underestimation of  $E(R_0)$ .

We can also note from Fig. 1(a) that the two quantities are similar for values of  $\delta = \mu_I - \mu_S$  in the extremes of the range considered, differing more for intermediate values. This was expected: first, when  $\mu_I \rightarrow 1$ ,  $\mu_S + \delta\mu_I$  tends to  $\mu_S + \delta = \mu_I$  and, thus  $E(R_0)$  in the prospective study (17) tends to  $N\mu_I$ , which is  $E(R_0)$  in the retrospective study, (14). On the other hand, when  $\mu_I \rightarrow \mu_S$ , we also have  $\delta \rightarrow 0$  and  $\mu_S + \delta\mu_I \rightarrow \mu_S$ , and thus in both studies we have  $E(R_0) \rightarrow N\mu_S$ .

Note, however, that in spite of the two expectancies converging to the same value as  $\delta$  decreases, there is a *discordance* region where one of them is above 1, while the other is below.

In Fig. 1(b) we have the computed probabilities of  $R_0$  being greater than zero, as a function of  $\delta$ , for the same parameter values considered. Differently from the expected values, the probabilities for one study type are not consistently greater than those for the other study type. Perhaps the most important aspect highlighted

by this figure is the fact that  $P\{R_0 \geq 1\}$  may solve apparent discordances between the expected values from different study types. Indeed, for the points within the discordance region in Fig. 1(a), the computed probabilities lie between 0.5 and 0.7 for both study types indicating that, while long-term disease establishment is not entirely without doubt, the chance is far from negligible in both study types.

## 5. COMPARING WITH THE CLASSIC REED–FROST MODEL

The classic Reed–Frost model can be seen as a particular case of both Reed–Frost models introduced in Section 2. Let us consider the more general model for a prospective study; the result follows for the retrospective study model.

Suppose that  $\mu_S = 0$  and  $\mu_I = \delta = 1$ , so that the variances of  $X_I$ ,  $X_S$  are both equal to zero. This implies that  $D_{i,t} = X_I = \mu_I$  for all  $i, t$ . Then, from (10)  $P\{G_{i,t} = 1 \mid D_{i,t}\} = 1$  for an infected individual, and 0 otherwise, with no uncertainty, implying that  $\sum_l G_{l,t} = I_t$  in this case. Suppose also that  $M = 1$  in (3), meaning that  $P_{l,t} = \phi_1 \equiv P$  for all infected individuals, whilst  $P_{l,t} = 0$  for all susceptible individuals, for all  $t$ . Then expression (4) becomes:

$$C_t = 1 - (1 - P)^{\sum_l G_{l,t}} = 1 - (1 - P)^{I_t},$$

which is equation (1).

## 6. LONG-TERM DISEASE ESTABLISHMENT

**6.1. The proposed model.** Conditions under which the disease establishes itself in the group can be obtained by determining conditions under which  $E(R_0)$  is greater than, or equal to, 1. Assume for simplicity  $P(D) \equiv D$ . In a retrospective study, we use expression (14) to say that the disease establishes itself in the population whenever

$$\mu_S + \delta \geq \frac{1}{N}.$$

In a prospective study, we use (18) to say that the disease establishes itself in the population whenever

$$\mu_S(\delta + 1) + \delta^2 \geq \frac{1}{N}. \quad (26)$$

In this context, it is also interesting to consider  $P\{R_0 \geq 1\}$  as a stochastic way of evaluating how likely the disease is of establishing itself in the long term. For that, we simply compare the value obtained for  $P\{R_0 \geq 1\}$  to a pre-specified threshold; when the probability is below the threshold, the disease may be said to be unlikely to establish itself. This threshold may vary according to disease and context.

**6.2. The classic Reed–Frost model.** We fall into the classic Reed–Frost model when  $\mu_S = 0$  and  $\mu_I = \delta = 1$  and, in this case, from (26) the epidemic establishes itself whenever

$$E(C_1 | I_1 = 1) \geq \frac{1}{N} \quad \text{or} \quad P \geq \frac{1}{N}$$

which implies that, for large  $N$ , any epidemic with nonnegligible probability of an infectious contact,  $P$ , establishes itself in the population. It is important to point out, however, that this remark has a purely theoretical interest: in practice, when  $N$  is large the homogeneous mixing assumption rarely holds.

## 7. APPLICATION TO INTERVENTION DESIGN

An important practical application of the models presented here is that to studies aiming at intervention design. An intervention may involve simply a change in risk behaviour, thus changing  $P(D)$ , or it may involve administering treatment, which might affect both  $P(D)$  and infected individuals' signals summary distribution. The signals summary distribution for susceptible individuals, including the mean  $\mu_S$ , is assumed to not be affected, as it represents the populational distribution of the signals summary under study, due to causes other than the disease. For simplicity, we assume that  $P(D) = \phi D$ .

First, let us consider the impact of a risk-behaviour reducing intervention, which can be represented by a change from  $P(D)$  to  $P^*(D) = \phi^* D$ , where  $\phi^* < \phi$ . In a retrospective study, the post-intervention  $R_0$  expected value is given by  $E(R_0^*) = NE[P^*(X_I)] = N\phi^*\mu_I$ , from (14). A desirable intervention yields  $E(R_0^*) < 1$ , which is guaranteed to hold if  $\phi^* < (N\mu_I)^{-1}$ .

In a similar way, such an intervention can be designed to guarantee that

$$1 - P\{R_0^* = 0\} < p_0, \quad (27)$$

where  $p_0$  is a pre-specified threshold. Indeed, using (19) we get that (27) is satisfied by designing the intervention so that

$$\sum_{k=1}^N \frac{N!}{k!(N-k)!} (-1)^k (\phi^*)^k E(X_I^k) > 1 - p_0$$

holds.

Similar conditions can be obtained to evaluate *a priori* the intervention impact in a prospective study. Indeed, using (17) we can conclude that the intervention will generate an expected basic reproduction ratio smaller than 1 whenever

$$[1 - \phi^* E(X_I^2)] \sum_{k=0}^{N-1} \frac{N!}{K!(N-k)!} (-1)^k (\phi^*)^k E(X_S^{2k}) > \frac{1}{\delta} \left( \mu_I - \frac{1}{N} \right).$$

No analytical expressions are available for the roots of this polynomial on  $\phi^*$  for general  $N$  and, thus, in practice this condition can be used mainly to check whether or not a specific value of  $\phi^*$  satisfies it.

The post-intervention probability of long-term disease establishment can be evaluated by replacing  $P(X_I)$  by  $P^*(X_I) = \phi^* X_I$  in (25), in the same way as before.

In the same vein, the impact of treatment affecting only the signals distribution can be evaluated prior to introduction via replacing  $\delta$  by  $\delta^*$  in expressions for  $E(R_0)$  and  $P\{R_0 \geq 0\}$ . For example, in a retrospective study such a treatment generates on average less than 1 new infected cases for each first case whenever

$$\delta^* < \frac{1}{N\phi} - \mu_S.$$

A treatment can also affect both  $\delta$  and  $P(D)$ . Expressing the treatment impact in the same way as before, the condition guaranteeing that on average less than 1 new infected cases are generated for each first case is

$$\phi^*(\mu_S + \delta^*) \leq \frac{1}{N}.$$

The treatment impact on the probability of long-term disease establishment, as well as in the prospective study case, can be evaluated in the same way.

The impact of a treatment affecting both the signals summary distribution and the probability of an infectious contact can be similarly evaluated, by combining the ideas above.

## 8. EXTENSIONS

Several variants of the proposed model can be obtained. Consider first the probability of an infectious contact, which is assumed to be a function of the signals. This function can have any polynomial form, and as such can potentially include any desired function: by assigning beta distributions to the individual signals, not only a flexible distribution family is used, but also one for which all moments are available, thus there is no limitation on the polynomial degree. The function of the signals can also be extended to allow some of the signals to yield an increase, and others a decrease, on the infection probability. Furthermore, it can be generalized to take the probability of an infectious contact as a probabilistic, rather than deterministic, function of the signals, or it can even consider frailty, or varying susceptibility levels.

Here we considered signals mainly as being disease symptoms, but in general these may include demographic variables, such as gender and age, as well as behavioural patterns. These may help in estimating model parameters, such as those involved in the signals distributions and in the classification process, as well as in yielding a better understanding between signals and the probability of an infectious contact.

Other variants of the proposed model can be obtained by considering more sophisticated classification procedures, which effectively suggests separating the clinical signals effect on different aspects of the epidemic: the one on the classification process may well be distinct from the one on the infectious contact probability.

In the above formulation it was assumed that the infectious period is of constant duration across individuals. If infectious period duration varies across individuals, but its duration can also be seen as a function of clinical signals, then different lags may be used for different individuals. The probability of an infectious contact is thus the product between the probability per time unit and the infectious period duration. The same conditional probability properties can be used as above to derive useful relations between the basic reproduction ratio and the signals' distribution parameters.

Some diseases are known to have an infectious period starting before clinical signals onset. This information, if known, can be used in the retrospective study model to yield better estimates for the signals distributions parameters, as well as for the probabilities of infectious contacts. In a prospective study context, however, when decisions must be made with regards to treatment, this information is less useful, as no signals exist to mark the infectious period onset.

## 9. DISCUSSION

The main idea behind our model is to use information available on signals to assess both the probability of an infectious contact and the diagnostic procedure. The basic reproduction ratio,  $R_0$ , is seen as the number of secondary infections caused over one generation, after the introduction of a single infected individual in an entirely susceptible population. This definition is coherent with another, introduced by Diekmann *et al.* (1990), based upon the next generation operator. We believe it is interpretable and suitable for our purposes. Within the context of the proposed model, it is a random variable and, as such, moments and a probability distribution function are available.

There have been several attempts to generalize the Reed–Frost model so as to consider a nonhomogeneous group, either from the susceptibility or from the infectivity viewpoint (Maia, 1952; Scalia-Tomba, 1986; Lefèvre and Picard, 1990; Picard and Lefèvre, 1991). In all these, the homogeneity assumption is relaxed by dividing the main group into subgroups, and considering that there is homogeneous mixing within each subgroup. Subgroups are closed and individuals remain within the same subgroup for the entire duration of the epidemics, which means that an individual's susceptibility and infectivity levels are taken as constant throughout the epidemic course.

The Reed–Frost model proposed here handles varying infectivity levels by assuming that infectivity is determined by observable and quantifiable clinical signals, effectively assigning a probability distribution to each individual's infectious

contact probability. We do not associate individuals with fixed sub-groups, as this is of limited practical interest in small groups studies, our focus here. Moreover, individual infectiousness may vary with time, a possibility not included in other generalizations. Some of the possible applications involve disease spread within classrooms, hospital wards and work groups.

Extensions to larger group studies may involve subdivision into subgroups to ensure that homogeneous mixing within subgroups still holds, while allowing for infectiousness heterogeneity within subgroups as proposed here.

The inclusion of heterogeneity in individual susceptibility to infections gives a number of qualitative differences compared to ordinary methods. At present practical applications are not yet available, but the heterogeneity in frailty explains some unexpected results. For instance, in Coutinho *et al.* (1999) it is shown that large heterogeneity in individual susceptibility to infection results in a decreasing population force of infection with age. This is due to the fact that the population is subject to a heavy selection of highly susceptible individuals, the remaining being less and less susceptible with age.

Disease studies to which the proposed model can be applied include all fast-propagating infectious diseases, in the sense that the disease propagates at a faster rate than its diagnostic and control can be performed. Examples of such diseases are the influenza-types, such as SARS, and meningitis. Applications also include several kinds of confinement, such as the forced confinement of hospital wards, and the weather-imposed confinement of classrooms at winter time. The minimum required degree of confinement is such that, apart from the initial cases, all infections are acquired within the group. Thus, applications to diseases propagating within a classroom, for example, could only account for new cases within the same classroom.

Finally, we should make clear that our approach does not consist simply of generalizing the classical Reed–Frost model formulation. The inclusion of signals not only yields a more flexible model with heterogeneous susceptibilities, but also leads to conditions for effective intervention designs. Moreover, it incorporates naturally existing differences among individuals in order to make it applicable to real epidemic scenarios. For, on the one hand, there are several infections, like influenza, which, besides being transmitted among small groups of individuals, produce highly heterogeneous clinical pictures. On the other hand, the huge amount of genetic information provided by the emerging field of genomics (and proteomics) generates clinical information which may sharply distinguish individuals. These tailor-made diagnostic techniques make obvious the necessity of new tools to deal with heterogeneities.

#### ACKNOWLEDGEMENTS

The authors are grateful to Dr Clarisse V Machado and Dr Ana V A Mendes for many fruitful discussions in the preparation of this work, and for comments from

Dr Nico Nagelkerke and from both referees, which helped clarifying the text. This work was partially funded by PRONEX, CNPq and LIM01/HCFMUSP.

## APPENDIX

**Beta distribution.** For any random variable  $X$  with beta distribution with parameters  $(\alpha, \beta)$ , its probability density function is

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad 0 \leq x \leq 1, \alpha > 0, \beta > 0,$$

where  $\Gamma(x)$  is the gamma function, defined as

$$\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} du, \quad x > 0.$$

Note that  $\Gamma(x) = x\Gamma(x-1)$ , for all  $x > 1$ . The expected value of  $X^k$ , for any  $k > 0$ , integer, is

$$E(X^k) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha + \beta + k)} \times \frac{\Gamma(\alpha + k)}{\Gamma(\alpha)} = \frac{\prod_{m=0}^{k-1} (\alpha + m)}{\prod_{m=0}^{k-1} (\alpha + \beta + m)}. \quad (\text{A.1})$$

In particular,

$$\begin{aligned} E(X) &= \frac{\alpha}{\alpha + \beta}, \\ E(X^2) &= \frac{\alpha(\alpha + 1)}{(\alpha + \beta)(\alpha + \beta + 1)}, \\ \text{var}(X) &= \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}. \end{aligned}$$

## REFERENCES

- Abbey, H. (1952). An examination of the Reed–Frost theory of epidemics. *Human Biol.* **24**, 201–233.
- Anderson, R. M. and R. M. May (1991). *Infectious Diseases of Humans*, Oxford: Oxford University Press.
- Bailey, N. T. J. (1975). *The Mathematical Theory of Infectious Diseases and its Applications*, London: Charles Griffin & Co.
- Becker, N. G. (1979). *Analysis of Infectious Disease Data*, London: Chapman & Hall.

- Coutinho, F. A. B., E. Massad, L. F. Lopez, M. N. Burattini, C. J. Struchiner and R. S. Azevedo-Neto (1999). Heterogeneities in individual frailties in epidemic models: a theoretical framework. *Math. Comput. Modell.* **30**, 97–115.
- Diekmann, O., J. A. P. Heesterbeek and J. A. J. Metz (1990). On the definition and the computation of the basic reproduction ratio  $R_0$  in models for infectious diseases in heterogeneous populations. *J. Math. Biol.* **28**, 365–382.
- Lefèvre, C. and P. Picard (1990). A non-standard family of polynomials and the final size distribution of Reed–Frost epidemic processes. *Adv. Appl. Probability* **22**, 25–48.
- Maia, J. O. C. (1952). Some mathematical developments on the epidemic theory formulated by Reed and Frost. *Human Biol.* **24**, 167–200.
- Picard, P. and C. Lefèvre (1991). The dimension of Reed–Frost epidemic models with randomized susceptibility levels. *Math. Biosci.* **107**, 225–233.
- Scalia-Tomba, G. (1986). Asymptotic final size distribution of the multitype Reed–Frost process. *J. Appl. Probability* **23**, 563–584.

*Received 11 April 2003 and accepted 13 October 2003*