Mathematical Biosciences 277 (2016) 89–107

Contents lists available at ScienceDirect

Mathematical Biosciences

journal homepage: www.elsevier.com/locate/mbs

# Mathematical models of Ebola—Consequences of underlying assumptions

Zhilan Feng[a,*], Yiqiang Zheng[a], Nancy Hernandez-Ceron[b], Henry Zhao[a], John W. Glasser[c], Andrew N. Hill[c]

[a] Department of Mathematics, Purdue University, West Lafayette, IN 47907, USA
[b] University of Michigan, Ann Arbor, MI 48109, USA
[c] Centers for Disease Control and Prevention, Atlanta, GA 30333, USA

ARTICLE INFO

*Article history:*
Received 8 December 2015
Revised 16 March 2016
Accepted 9 April 2016
Available online 26 April 2016

*Keywords:*
Mathematical models
Ebola
Arbitrarily distributed disease stage
Exponential waiting time
Model assumptions

ABSTRACT

Mathematical models have been used to study Ebola disease transmission dynamics and control for the recent epidemics in West Africa. Many of the models used in these studies are based on the model of Legrand et al. (2007), and most failed to accurately project the outbreak's course (Butler, 2014). Although there could be many reasons for this, including incomplete and unreliable data on Ebola epidemiology and lack of empirical data on how disease-control measures quantitatively affect Ebola transmission, we examine the underlying assumptions of the Legrand model, and provide alternate formulations that are simpler and provide additional information regarding the epidemiology of Ebola during an outbreak. We developed three models with different assumptions about disease stage durations, one of which simplifies to the Legrand model while the others have more realistic distributions. Control and basic reproduction numbers for all three models are derived and shown to provide threshold conditions for outbreak control and prevention.



## 1. Introduction

Mathematical models have been very helpful in evaluating and identifying alternative strategies for infectious disease control and prevention. However, for the recent epidemics of Ebola in West Africa, the success of mathematical models has been very limited. As pointed out in Butler [2], "mathematical models have failed to accurately project the outbreak's course". Although various reasons may explain why "on-the-ground data contradict the projections of published models", including incomplete and unreliable data on Ebola epidemiology (especially in the hardest-hit areas) and lack of empirical data on how disease-control measures quantitatively affect Ebola transmission, it is important to examine the appropriateness of assumptions made in the models on which the projections are based. This is the objective of the current paper. There have been various modeling approaches, including deterministic and stochastic models, or relatively simple models consisting of ordinary differential equations (ODEs) and more complicated agent-based models, among others. Many of the ODE models are variations of the model studied by Legrand et al. [8], to which we refer as the Legrand model. It has been pointed out that some of the assumptions made in the Legrand model may not have clear justifications (see, for example, Rivers et al. [11]). Thus, it is important to examine the critical assumptions made in this model and better understand their possible impact on model outcomes.

It often happens that, when a model is formulated, certain assumptions are made without consideration of their consequences. One of the most common assumptions made in ODE models is the exponential waiting time in disease stages. That is, the survival probability is described by a negative exponential function. For example, if the model assumes that an infected individual will recover at a constant per-capita rate $\gamma$, then it implicitly assumes that the infectious period is exponentially distributed, and the probability that an individual is still infectious $s > 0$ units of time since onset is given by

$$P_I(s) = e^{-\gamma s}.$$

That is, if $X_I$ denotes the random variable for the waiting time in the infectious class $I$ before exiting, then

$$\mathbb{P}[X_I > s] = P_I(s) = e^{-\gamma s}.$$

In this case, the average waiting time before recovery (or the mean infectious period) is given by

$$\mathbb{E}[X_I] = \int_0^\infty P_I(s)ds = \frac{1}{\gamma}.$$

* Corresponding author. Tel.: +1 765 494 1901; fax: +1 765 494 0548.
*E-mail address:* fengz@purdue.edu (Z. Feng).

http://dx.doi.org/10.1016/j.mbs.2016.04.002
0025-5564/

**Table 1**
Definition of symbols commonly used in all models including arbitrary stage distributions.

| Symbol | Definition |
|---|---|
| $T_P$, $T_L$, $T_M$ | Random variables for the waiting times in $I$ before moving to $R$, $H$, $D$, respectively |
| $X_I$ | Random variable for the overall time spent in the $I$ compartment |
| $X_H$ | Random variable for the overall time spent in the $H$ compartment |
| $P_i(s)$ | Probability that a living individual remains infectious $s$ units of time since onset for Models I, II, III when $i = 1, 2, 3$, respectively. That is, $\mathbb{P}[T_{P_i} > s] = P_i(s)$ |
| $L_i(s)$ | Probability of a living individual not being hospitalized $s$ units of time since onset for Models I, II when $i = 1, 2$, respectively. That is, $\mathbb{P}[T_{L_i} > s] = L_i(s)$ |
| $M_1(s)$ | Probability of surviving the disease $s$ units of time since onset for Model I. That is, $\mathbb{P}[T_{M_1} > s] = M_1(s)$ |
| $Q_3(s)$ | Probability of not having recovered $s$ units of time after being hospitalized for Model III |
| $\mathbb{E}[T_P]$ | Mean duration from onset to recovery (absent intervention or death) |
| $\mathbb{E}[T_L]$ | Mean duration from onset to hospitalization (given hospitalized and not dead) |
| $\mathbb{E}[T_M]$ | Mean duration between onset and death (absent intervention or recovery) |
| $\mathbb{E}[X_I]$ | Mean duration in the $I$ compartment (hospitalization and death included) |
| $\mathbb{E}[X_H]$ | Mean duration in the $H$ compartment (death included) |
| $\mathcal{D}_{HR}$ | Mean duration from hospitalization to recovery |
| $\mathcal{D}_{HD}$ | Mean duration from hospitalization to death |
| $\gamma_{IR}$ | $= 1/\mathbb{E}[T_P]$, |
| $\gamma_{IH}$ | $= 1/\mathbb{E}[T_L]$, |
| $\gamma_{ID}$ | $= 1/\mathbb{E}[T_M]$, |
| $\omega_{HR}$ | $= 1/\mathcal{D}_{HR}$, per-capita rate of transition from $H$ to $R$ if the transition is exponential |
| $\omega_{HD}$ | $= 1/\mathcal{D}_{HD}$, per-capita rate of transition from $H$ to $D$ if the transition is exponential |
| $p$ | Proportion hospitalized (dependent on control effort) |
| $f$ | Probability of death (with or without hospitalization) |
| $\gamma$ | $= 1/\mathbb{E}[X_I]$, per-capita rate of exiting $I$ if $X_I$ is exponential |

Because of the memoryless property of exponential distributions, this leads to ODE models that are easy to analyze. However, when isolation of infectious individuals is considered as a control strategy, models with exponentially distributed infectious stages can lead to misleading or even incorrect evaluations of effectiveness (see, e.g., Feng et al. [4]). Similar results hold for discrete-time models, in which the analogue of the exponential distribution is geometric [5,6]. These findings demonstrate some of the drawbacks of ODEs models with exponentially-distributed disease stages.

It is not clearly specified in the Legrand model (see model (A.1) in Appendix A) what underlying assumptions have been made regarding the distributions of waiting times for epidemiological processes such as recovery (transition from $I$ to $R$), hospitalization (transition from $I$ to $H$), and death (transition from $I$ to $D$). For ease of reference, we refer to these three transitions as IR, IH, and ID, respectively. In addition, the two possible transitions for hospitalized individuals, recovery or death, are denoted by HR and HD. Let $T_P$, $T_L$ and $T_M$ denote random variables for the waiting times associated with IR, IH and ID, and let the associated survival functions be denoted by $P(t)$, $L(t)$ and $M(t)$, respectively. The mean durations of these transitions are respectively $\mathbb{E}[T_P]$, $\mathbb{E}[T_L]$, $\mathbb{E}[T_D]$. Similarly, let $\mathcal{D}_{HR}$ and $\mathcal{D}_{HD}$ denote the mean durations from hospitalization to recovery or death, respectively. For ease of comparison between models presented in this paper, we list in Table 1 some of the quantities that play common roles and have clear biological meaning in these models. Several of these quantities should have values that are independent of model assumptions, including the mean duration (absent intervention) from onset to recovery $\mathbb{E}[T_P]$, the probability of hospitalization $p$, and the probability of death $f$.

At first glance (A.1) (see Appendix A), it may seem that the transitions IR, IH and ID are assumed to be independent in the Legrand model and the waiting times are all exponentially distributed with mean durations $1/\gamma_{IR}$, $1/\gamma_{IH}$ and $1/\gamma_{ID}$, respectively. If so, however, the mean overall rate of exiting the $I$ compartment would not be given by $\Delta$ in (A.5) (see Table 2 and Appendix A). In fact, if we denote the survival functions by $P(t) = e^{-\gamma_{IR}t}$, $L(t) = e^{-\gamma_{IH}t}$ and $M(t) = e^{-\gamma_{ID}t}$, respectively, then the overall waiting time in $I$ is

$$\begin{aligned}\mathbb{P}[\min\{T_P, T_L, T_M\} > t] &= \mathbb{P}[\{T_P > t\} \cap \{T_L > t\} \cap \{T_M > t\}] \\ &= \mathbb{P}[T_P > t]\mathbb{P}[T_L > t]\mathbb{P}[T_M > t].\end{aligned}$$

Thus, the mean overall time spent in the $I$ compartment is

$$\mathbb{E}[\min\{T_P, T_L, T_M\}] = \int_0^\infty P(t)L(t)M(t)dt = \frac{1}{\gamma_{IR} + \gamma_{IH} + \gamma_{ID}}.$$

It follows that the overall rate of exiting $I$ is $\gamma_{IR} + \gamma_{IH} + \gamma_{ID}$, which is not a weighted average, unlike $\Delta$ in (A.5) for the Legrand model.

To fully understand the underlying assumptions used in the Legrand model, we develop three models based on general distributions for the waiting times of key transitions and consider various assumptions about their relationships. We demonstrate that, when specific stage distributions are considered, one of these general models reduces to the Legrand model (A.1). The models that we develop in this paper under arbitrarily distributed disease stages consist of integro-differential equations. It has been asserted that, when the arbitrary distributions are replaced by Gamma distributions, the so-called "linear chain trick" can be applied to reduce the integral equations to ODEs (MacDonald [10], Hethcote and Tudor [7], Lloyd [9]). However, apart from a proof of the linear chain trick in a simpler setting by Smith [13, Section 7.1], a rigorous derivation of this fact for more complex epidemic models, such as the one in this paper, is lacking. In this paper, we provide a derivation (see Section 3).

We also provide a simpler (but equivalent) formulation of the Legrand model. In particular, we define the overall waiting time in the $I$ class to be the weighted combination of waiting times $\mathbb{E}[T_P]$, $\mathbb{E}[T_L]$ and $\mathbb{E}[T_M]$ of the following form:

$$\mathbb{E}[X_I] = p\mathbb{E}[T_H] + (1-p)f\mathbb{E}[T_M] + (1-p)(1-f)\mathbb{E}[T_P], \tag{1}$$

where $p$ and $f$ denote the probability of hospitalization and case-fatality, respectively. Using this assumption, as shown in Section 2, we obtain a much simpler formulation of the Legrand model (A.1). This facilitates identification of its underlying assumptions, as illustrated in Sections 3 and 4.

Our paper is organized as follows. In Section 2, we present an equivalent Legrand model with a simpler formulation. Section 3 is devoted to the derivation of three models with arbitrarily distributed disease stages under various assumptions about the relationships between the overall waiting time in the $I$ class and