1 **Title:** Novel compartmental models of infectious disease transmission

2

3 **Authors:** Scott Greenhalgh[a]*, Carly Rozins[b]

4

5 **Affiliations**:

6 [a] Siena College, Department of Mathematics, Loudonville, NY, 12211, USA.

7 email: sgreenhalgh@siena.edu

8 [b] University of California Berkeley, Department of Integrative Biology, Berkeley, CA, 94720, USA.

9 email: crozins@berkeley.edu

10

11 * Corresponding author

12

13

14 **Author contributions:**

15 SG conceptualized the project. CR and SG conducted formal analysis, wrote, reviewed, and edit the

16 project manuscript.

17

18

Abstract

Many methodologies in disease modeling are invaluable in the evaluation of health interventions. Of these methodologies, one of most fundamental is compartmental modeling. Compartmental models have many different forms with one of the most general characterizations occurring from the description of disease dynamics with nonlinear Volterra integral equations. Despite this generality, the vast majority of disease modellers prefer the special case where nonlinear Volterra integral equations reduce to systems of differential equations through the traditional assumptions that 1) the infectiousness of a disease corresponds to incidence, and 2) the duration of infection follows either an exponential or Erlang distribution. However, these assumptions are not the only ones that simplify nonlinear Volterra integral equations in such a way. In what follows, we illustrate a biologically more accurate description of the total infectivity of a disease that reduces systems of nonlinear Volterra integral equations to a class of novel compartmental models, as described by systems of differential equations. We demonstrate the consistency of these novel compartmental models to their traditional counterparts when the duration of infection follows either an exponential or Erlang distribution, and provide a novel compartmental model for a Pearson distributed duration of infection. Significant outcomes of our work include a compartmental model that captures any Erlang distributed duration of infection with only 3 differential equations, instead of the typical inflated model sizes, and a compartmental models that capture any mean, standard deviation, skewness, and kurtosis of the duration of infection distribution with only 4 differential equations.

Keywords: Differential equations, integral equations, infectious disease models, compartmental models, disease infectivity, survival analysis, infectious period

42 **Author summary:**

43 Compartmental models are a powerful tool for predicting disease outbreaks, and evaluating public
44 health policies and intervention effectiveness. However, such models typically have an inability to
45 account for many of the biological features of a disease. For instance, the assumptions placed on the
46 duration of infection required by most compartmental models are due to mathematical convenience,
47 and are known to massively effect model behavior and quality of predictions. Our work illustrates a
48 simple solution to these erroneous assumptions by proposing a new simplification of the general model
49 proposed by Kermack and McKendrick. In doing so, we obtain a new class of compartmental models
50 with many of the features that make traditional compartmental the go-to disease model for the vast
51 majority the epidemiological modeling community, such as their formulations as systems of differential
52 equations, while adding the ability to more accurately account for effects of variability in an individual's
53 duration of infection. As such, our work may be viewed as the starting point for multiple research
54 avenues, as it opens up a new class of compartmental model for investigation under the contexts of
55 mathematics, public health, and evolutionary biology.

56

57

58

59

60

61

62

1. Introduction

The compartmental model of Kermack and McKendrick [1–3] is arguably one of the greatest development in disease modeling. The formulation of this model, in its original form as a system of nonlinear Volterra integral equations [4], provides a general characterization of the transmission cycle between susceptible individuals and a disease that propagates throughout an environment [5]. Despite this generality, the vast majority of disease modellers prefer differential equation compartmental models [6]. While this particular formulation of compartmental models has distinct advantages, such as the non-requirement of specialist knowledge to implement and well-developed numerical methods [7], they are in fact a special case of the aforementioned system of nonlinear Volterra integral equations [8]. Specifically, one obtains differential equation compartmental models from the nonlinear Volterra integral equations by imposing only two traditional assumptions: 1) the infectiousness of a disease corresponds to disease incidence, and 2) the duration of infection follows either an exponential or an Erlang distribution. Unfortunately, the vast majority of diseases do not have infectious periods that follow these distributions [9–14], and force fitting such a distributional structure is known to have a massive effect on the behavior and quality of model predictions [15]. Regardless of these issues, the compartmental models obtained by the two traditional assumptions have undergone many extensions. A few noteworthy examples include modifying the force of infection to account for the saturation of infection in a population [16,17], behavioral characteristics [18–21], modification of the recovery rate to better capture a disease's infectious period [22], disease burnout [23], and the inclusion of additional disease stages [8]. Furthermore, the applications of these models have also grown considerably from just predicting a disease's trajectory. Today, these traditional compartmental models are used to evaluate the health benefits and cost-effectiveness of public health policies and disease interventions [24,25], gauge the potential for disease virulence evolution [26], predict dominant influenza strains [27], investigate the complexities of disease co-infection [28], among many others. However, despite this

growth in theory and application, the generalization of the very foundational assumptions that simplifies systems of nonlinear Volterra integral equations to differential equation compartmental models remains largely undeveloped.

In what follows, we propose new assumptions to simplify systems of nonlinear Volterra integral equations to systems of differential equations. The biological motivation for these new assumptions stem from the idea that a disease's average duration of infection changes throughout an epidemic, whereas the average infectious period of a disease remains constant. Consequently, we extend current models from solely tracking the disease incidence to tracking the number of person-days of infected individuals. To do this, we assume 1) the total infectiousness of a disease corresponds to the product of disease incidence with a time-varying average duration of infection and 2) the duration of infection is distributed according to a non-homogeneous analog to the exponential distribution. Under these assumptions, we derive a novel class of differential equation compartmental models, provide model equilibria, and disease reproductive numbers [29].

Essential in the development of this new class of models is the use of survival analysis. Specifically the development of our novel class of models requires the hazard function and the mean residual waiting-time of a distribution [30,31], which is used to describe the time-varying average duration of infection. Therefore, we briefly outline some of the fundamental properties of these functions, in addition to their relationship to one another. We then demonstrate the consistency of our new class of models to traditional models when the duration of infection follows an exponential distribution. Next, we consider our model with a duration of infection that follows an Erlang distribution, and illustrate how this choice of distribution has a representation as either an ODE system of 3 equations, regardless of the Erlang distribution parameters, or as an ODE system that features a chain of infected equations, as is typical from the linear chain trick [9,32,33]. Finally, we consider a duration of infection that is Pearson

distributed. In choosing the Pearson distribution, we develop a model that is capable of accounting for any possible mean, standard deviation, skewness, and kurtosis of the duration of infection. Thereby, we provide a simple approach for measuring how altering the infectivity profile of a disease, as described by the first four statistical moments, influences a diseases trajectory.

2. Methods

In what follows, we develop a novel class of ODE compartmental models to describe the progression of a disease throughout a population. To obtain such models, we impose new assumptions on the notion of infectivity used in the integral equations of Kermack and McKendrick. Before reducing the integral equations of Kermack and McKendrick to their differential equation counterparts, we briefly highlight the formal relationship in our assumptions in the context of survival analysis. Finally, we present the class of novel compartmental models, their equilibria, and basic reproductive number.

2.1 Traditional compartmental models. To begin, considered the general compartmental model of Kermack and McKendrick [1–3,5]. For this general compartmental model, we consider the number of susceptible individuals to be denoted as S, and the total infectivity of the disease (at time t) to be denoted as $\phi(t)$. We define $\phi(t)$ as the sum of the products of the number of individuals at a particular age of infection that remain infectious with their mean infectivity for that particular infection age. We also define ϕ_0 as the total infectivity of the individuals initial infected with the disease at the start of the epidemic. In addition, we characterize the fraction of infected individuals remaining infected at time t, who where initially infected at time x, with the duration of infection distribution, as described by the survival function $P(t,x)$. Furthermore, we also define the mean infectivity of an infected individual $t-x$ units of time after the start of the infection as $\pi(t-x)$, where $0 \leq \pi(t-x) \leq 1$, and the mean infectivity of individuals in the population at time t who where initially infected at time x as

$$A(t,x) = \pi(t-x)P(t,x).$$

For simplicity, we assume that $\pi(t-x) = 1$, and so the progression of an epidemic throughout a population can be described with the integral equations,

$$S(t) = S_0 - \int_0^t \lambda(x)S(x)dx,$$
$$\phi(t) = \phi_0(t) + \int_0^t \lambda(x)S(x)P(t,x)dx. \quad (1)$$

Here λ is the force of infection, which we assume to be

$$\lambda(t) = \beta\frac{\phi(t)}{N}, \quad (1b)$$

where β is the average number of contacts individuals in a population make per unit of time.

Traditionally, to reduce (1) to a system of differential equations requires that 1) the duration of infection follows the exponential distribution,

$$P(t,x) = e^{-\gamma(t-x)},$$

where γ is the recovery rate, and 2) that the infectiousness of a disease corresponds directly to the number of infected individuals, $\phi = I$. Combining these assumptions, along with an additional compartment to track recovered individuals, R, transforms system (1) into

$$S(t) = S_0 - \int_0^t \beta\frac{I(x)}{N}S(x)dx, \quad (2)$$

$$I(t) = I_0 e^{-\gamma t} + \int_0^t \beta \frac{I(x)}{N} S(x) e^{-\gamma(t-x)} dx,$$

$$R(t) = \left(R_0 + I_0 - I_0 e^{-\gamma t}\right) + \int_0^t \beta \frac{I(x)}{N} S(x)\left(1 - e^{-\gamma(t-x)}\right) dx.$$

An important feature of system (2) is that it conserves the total population:

$$S(t) + I(t) + R(t) = S_0 + I_0 + R_0 = N.$$

Differentiating system (2) with respect to t, and substituting the integral equation for $I(t)$ for the remaining integrals yields the classic SIR system:

$$S'(t) = -\beta \frac{I(t)}{N} S(t),$$

$$I'(t) = \beta \frac{I(t)}{N} S(t) - \gamma I(t), \tag{2b}$$

$$R'(t) = \gamma I(t).$$

Alternatively, (1) reduces to a system of differential equations when 1) the duration of infection is the survival function of the Erlang distribution,

$$P(t,x) = P(t-x) = \sum_{j=0}^{k-1} \frac{(\gamma(t-x))^j}{j!} e^{-\gamma(t-x)},$$

where k is a shape parameter that determines the total number of infection stages and $\frac{1}{\gamma}$ is the average duration spent in each stage, and 2) the total infectivity of the disease corresponds to k identical stages (in terms of the average duration spent in each stage) of infected individuals,

$$\phi = I = \sum_{j=1}^{k} I_j.$$

Combining these assumptions, along with an additional compartment to track recovered individuals, R, transforms system (1) into

$$S(t) = S_0 - \int_0^t \beta \frac{I(x)}{N} S(x) dx,$$

$$I(t) = I_0 \sum_{j=0}^{k-1} \frac{(\gamma t)^j}{j!} e^{-\gamma t} + \int_0^t \beta \frac{I(x)}{N} S(x) \sum_{j=0}^{k-1} \frac{(\gamma(t-x))^j}{j!} e^{-\gamma(t-x)} dx, \tag{3}$$

$$R(t) = \left(R_0 + I_0 - I_0 \sum_{j=0}^{k-1} \frac{(\gamma t)^j}{j!} e^{-\gamma t} \right) + \int_0^t \beta \frac{I(x)}{N} S(x) \left(1 - \sum_{j=0}^{k-1} \frac{(\gamma(t-x))^j}{j!} e^{-\gamma(t-x)} \right) dx.$$

Equivalently, if the linear chain trick is applied, we have that

$$I_j(t) = I_{0,j} \frac{(\gamma t)^{j-1}}{(j-1)!} e^{-\gamma t} + \int_0^t \beta \frac{I(x)}{N} S(x) \frac{(\gamma(t-x))^{j-1}}{(j-1)!} e^{-\gamma(t-x)} dx,$$

where

$$I(t) = \sum_j I_j(t).$$

An important feature of system (3) is that it conserves the total population:

$$S(t) + I(t) + R(t) = S_0 + \sum_j I_{0,j} + R_0 = N.$$

Differentiating system (3) with respect to t, applying the 'linear chain trick', and substituting $\gamma I_j(t)$ as needed yields the classic SI^{k}R system:

$$S^{'}(t) = -\beta\frac{I(t)}{N}S(t),$$

$$I_1^{'}(t) = \beta\frac{\sum_{j=1}^{k} I_j(t)}{N}S(t) - \gamma I_1(t), \tag{3b}$$

$$I_j^{'}(t) = \gamma I_{j-1}(t) - \gamma I_j(t),$$

$$R^{'}(t) = \gamma I_j(t).$$

To obtain our novel differential equation compartmental models, we generalize the assumptions used to formulate system (2). We first assume the infectiousness of a disease corresponds to the product of the number of infected individuals with their average duration of infectiousness at time t:

$$\phi(t) = I(t)m(t). \tag{4}$$

Here $m(t)$ is a mean residual waiting-time [31]. Motivation for choosing $m(t)$ over the typical (constant) average waiting-time stems from the fact that the composition of infected individuals $I(t)$ includes individuals from different initial infectiontimes, and that individuals infected at different times are not likely to remain infectious for the same time period. By including $m(t)$ in (4), our notion of infectivity is able to differentiatie between similar quantities of infected individuals that may (or may not) contribute differently to the spread of an epidemic.

In addition to assumption (4), we also assume that the duration of infection distribution corresponds to a non-homogeneous analog of the exponential distribution, namely the survival function given by

$$P(t,x) = P(T > t|T > x) = \frac{P(T > t)}{P(T > x)} = \frac{e^{-\int_0^t \eta(z)dz}}{e^{-\int_0^x \eta(z)dz}} = e^{-\int_x^t \eta(z)dz}, \quad (5)$$

where $t \geq x$, and T is a random variable that denotes the sojourn time in the infectious state.

2.2 The hazard function and mean residual waiting-time. Due to the importance of assumptions (4) and (5), we now provide a brief overview of the relationship between the hazard function, the mean residual waiting-time, and the survival function [30]. Consider a random variable T characterized by the survival function (5), which represents the probability of remaining infectious t units after becoming infectious at time x. The associated hazard function to T is given by

$$\eta(t) = -\frac{1}{P(t,x)}\frac{dP}{dt}, \quad (6)$$

where $P(t,x)$ is the survival function (for remaining infected) for individuals initially infected at time x.

Similarly, the mean residual waiting-time associated with T is also determined from (5) [30], and defined as

$$m(t) = \mathbb{E}[T - t|T > t] = \frac{1}{P(t,x)}\int_t^\infty P(z,x)dz. \quad (7)$$

From the mean residual waiting-time (7), it is also possible to uniquely determine the hazard function (6) through the relation [30],

$$\eta(t) = \frac{m'(t) + 1}{m(t)}. \quad (8)$$

An important feature of the mean residual waiting-time (7) is that it is initially equivalent to the average duration of infectiousness, as

$$\lim_{t\to 0} m(t) = \lim_{t\to x} \frac{1}{P(t,x)} \int_{x}^{\infty} P(z,x)dz = \mu. \tag{9}$$

189 It follows from (7) and (8), and through the application of l'Hôpitals rule, that

$$\lim_{t\to\infty} \frac{m'(t)+1}{\eta(t)} = \lim_{t\to\infty} \frac{1}{P(t,x)} \int_{t}^{\infty} P(z,x)dz = \lim_{t\to\infty} -\frac{P(t,x)}{\frac{dP}{dt}} = \lim_{t\to\infty} \frac{1}{\eta(t)}. \tag{10}$$

190 This implies if $\eta(t) > 0 \; \forall t$ that

$$\lim_{t\to\infty} m'(t) = 0. \tag{11}$$

191 In addition to determining $\eta(t)$ from $m(t)$, the receprical relation is also possible. For convenience we

192 restrict (5) to the family of distributions [30] that satisfy

$$g'(t)\frac{dP}{dt} + g(t)\frac{d^2P}{dt^2} = (\mu - t)\frac{dP}{dt}. \tag{12}$$

193 Note condition (12) ensures that both $m(t)$ and $\eta(t)$ exist and are finite [30]. Integrating (12) over $[t,\infty)$,

194 and using the property that $\lim_{t\to\infty} P(t,x) = 0$, we obtain

$$-g(t)\frac{dP}{dt} = -\mu P(t,x) - \int_{t}^{\infty} t\frac{dP}{dt}dt. \tag{13}$$

195 Noting that $\int_t^\infty t\frac{dP}{dt}dt = \mathbb{E}[T|T > t]$, definitions (6) and (7), and dividing through by $P(t,x)$ we obtain

$$\mathbb{E}[T|T > t] = \mu + g(t)\eta(t). \tag{14}$$

196 Thus, subtracting t from both sides we have that

$$m(t) = \mu - t + g(t)\eta(t). \tag{15}$$

2.3 Novel compartmental models. Before presenting the novel compartmental model we briefly contextualize its differences in comparison to traditional compartmental models. The traditional compartmental models of (2b) and (3b) track the changes in the number of persons suscpetible to, infected with, and recovered from infection for a disease circulating in a population. The premise of our novel compartmental model is to instead consider tracking the change in the number of person-days (i.e. the number of people in a particular state multipled by a duration of time) susceptible to, infected with, and recovered from infection for a disease circulating in a population. Therefore, to appropriately account for the temporal aspect of person-days we incorporate $m(t)$ into each compartment and the total population. Thus, we have that the person-days of the susceptible, infected, and recovered individuals in a population are governed by,

$$\begin{aligned} S(t)m(t) &= Nm(t) - (I_0 + R_0)\mu - \int_0^t \lambda(x)S(x)dx, \\ I(t)m(t) &= I_0P(t,0)\mu + \int_0^t \lambda(x)S(x)P(t,x)dx, \\ R(t)m(t) &= (R_0 + I_0 - I_0P(t,0))\mu + \int_0^t \lambda(x)S(x)(1 - P(t,x))dx, \end{aligned} \tag{16}$$

where $m(0) = \mu$ by (9), and $P(t,x)$ is given by (5). Note, the term $Nm(t) - (I_0 + R_0)\mu$ accounts for changes in the time-varying reference frame for the total person-days of those susceptible to infection in the population.

Adding the equations of system (16) together, it follows that

$$(S(t) + I(t) + R(t))m(t) = Nm(t) - (N - S_0 - I_0 - R_0)\mu.$$

Given $N = S_0 + I_0 + R_0$ and $m(t) \neq 0$ for all t, we have that

$$S(t) + I(t) + R(t) = N.$$

Imposing (4-7), taking the time derivative of system (16), and applying Leibniz rule for the derivatives of integrals as needed, we obtain

$$\begin{aligned}
&S'(t)m(t) + S(t)m'(t) = Nm'(t) - \beta\frac{I(t)}{N}m(t)S(t),\\
&I'(t)m(t) + I(t)m'(t) = -\eta(t)I_0P(t,0)\mu - \eta(t)\int_0^t \beta\frac{I(x)}{N}m(x)S(x)P(t,x)dx + \beta\frac{I(t)}{N}m(t)S(t),\\
&R'(t)m(t) + R(t)m'(t) = \eta(t)I_0P(t,0)\mu + \eta(t)\int_0^t \beta\frac{I(x)}{N}m(x)S(x)P(t,x)dx,
\end{aligned} \tag{18}$$

where $P(t,x)$ is given by (5), and $\eta(t)$ is given by (8).

Substituting (16) into (1) to eliminate the integrals, and isolating for each of $S'(t)$, $I'(t)$, and $R'(t)$, we obtain:

$$\begin{aligned}
&S'(t) + \frac{m'(t)}{m(t)}S(t) = N\frac{m'(t)}{m(t)} - \beta\frac{I(t)}{N}S(t),\\
&I'(t) + \frac{m'(t)}{m(t)}I(t) = \beta\frac{I(t)}{N}S(t) - \left(\frac{m'(t)+1}{m(t)}\right)I(t),\\
&R'(t) + \frac{m'(t)}{m(t)}R(t) = \left(\frac{m'(t)+1}{m(t)}\right)I(t).
\end{aligned} \tag{19}$$

2.4 Equilibria, the basic reproductive number $\mathcal{R}_0$, and the effective reproductive number $\mathcal{R}_e$. The compartmental model (19) potentially has several equilibria. As is standard, (19) possess the standard disease free equilibrium:

$$S^* = N,$$

$$I^* = 0, \tag{20}$$

$$R^* = 0.$$

222 In addition, an equilibrium where the infection was exhausted, leaving susceptible individuals that

223 escaped infection, and recovered individuals:

$$\hat{S} = S_\infty,$$

$$\hat{I} = 0, \tag{21}$$

$$\hat{R} = R_\infty.$$

224 Turning our attention to the reproductive numbers of the disease, the basic reproductive number

225 obtained directly from the survival function is

226 $$\mathcal{R}_0 = \beta \int_0^\infty P(t,0)dt = \beta\mu.$$

227 To estimate the basic reproductive number using the next-generation method, we have that $\mathcal{F} = \frac{\beta}{N}S$ and

228 $\mathcal{V}^{-1} = \frac{m(t)}{m'(t)+1}$. Note, the motivation for $\mathcal{V}^{-1} = \frac{m(t)}{m'(t)+1}$ instead of $\mathcal{V}^{-1} = \frac{m(t)}{2m'(t)+1}$ arises from the $\frac{m'(t)}{m(t)}I$

229 (t) accounting for a change in average duration of infectivity, instead of the transfer of infected

230 individuals to the recovered state. It follows that

$$\widehat{\mathcal{R}_0} = \rho\left(\mathcal{F}|_{DFE} \cdot \mathcal{V}^{-1}|_{DFE}\right) = \beta m^*, \tag{22}$$

231 where

232 $$m^* \in \left[\lim_{t\to\infty} \frac{m(t)}{m'(t)+1}, \lim_{t\to 0} \frac{m(t)}{m'(t)+1},\right]$$

233 Finally, we consider the effective reproductive number, in the face of time variability of $m(t)$, $m'(t)$, and

234 $S(t)$, to be

$$\mathcal{R}_e(t) = \frac{\beta m(t)}{m'(t)+1}\frac{S(t)}{N}. \tag{23}$$

235

236 **3. Special cases of the duration of infection distribution**

237 We now consider the duration of infection distributions used in traditional differential equation

238 compartmental models, namely the exponential distribution, and the Erlang distribution. In addition, we

239 illustrate a compartmental model that accounts for any mean, standard deviation, skewness, and excess

240 kurtosis, by assuming that the duration of infection is Pearson distributed.

241 **3.1 The exponential distribution.** If the duration of infection is exponentially distributed, then the

242 associated survival function at the onset of the epidemic is $P(t,x) = e^{-\gamma(t-x)}$. Under this assumption,

243 we have that [30],

$$\eta(t) = -\frac{1}{P(t,x)}\frac{dP}{dt} = \gamma, \tag{24}$$

244 and

$$g(t) = \frac{1}{\gamma}t. \tag{25}$$

245 Solving (8) under the assumption of (25) yields,

$$m(t) = \frac{1}{\gamma}. \tag{26}$$

246 Substituting (26) into (19), we arrive at the traditional form of differential equation compartmental

247 models:

$$\begin{aligned} S'(t) &= -\beta \frac{I(t)}{N} S(t), \\ I'(t) &= \beta \frac{I(t)}{N} S(t) - \gamma I(t), \\ R'(t) &= \gamma I(t). \end{aligned} \tag{27}$$

248 Finally, because $m(t) = \frac{1}{\gamma} = \mu$, the basic reproductive numbers are

$$\mathcal{R}_0 = \frac{\beta}{\gamma} \text{ and } \widehat{\mathcal{R}_0} = \frac{\beta}{\gamma} \tag{28}$$

249

250 **3.5 The Erlang distribution.** If the duration of infection is Erlang distributed, then the survival function is

$$P(t,x) = \frac{P(T > t)}{P(T > x)} = \left(\sum_{j=0}^{k-1} \frac{(\gamma x)^j}{j!} e^{-\gamma x} \right)^{-1} \sum_{j=0}^{k-1} \frac{(\gamma t)^j}{j!} e^{-\gamma t}. \tag{29}$$

251 Note, this formulation of an Erlang distributed random variable differs from that used in the derivation

252 of traditional compartmental models, as in general $P(t,x) \neq P(t-x)$.

253 From (6) and (29), it follows that

$$\eta(t) = \left(\sum_{j=0}^{k-1} \frac{(\gamma t)^j}{j!} e^{-\gamma t} \right)^{-1} \frac{\gamma(\gamma t)^{k-1}}{(k-1)!} e^{-\gamma t} = \left(\sum_{j=0}^{k-1} \frac{(\gamma t)^j}{j!} \right)^{-1} \frac{\gamma(\gamma t)^{k-1}}{(k-1)!}. \tag{30}$$

254 We also have that [30]

$$g(t) = \frac{1}{\gamma}t\,. \tag{31}$$

255 Substituting (31) in (8), we obtain the mean residual waiting-time:

$$m(t) = \frac{k}{\gamma} - t + \frac{1}{\gamma}t\left(\sum_{j=0}^{k-1}\frac{(\gamma t)^j}{j!}\right)^{-1}\frac{\gamma(\gamma t)^{k-1}}{(k-1)!}. \tag{32}$$

256 Using (32) and its derivative in system (19), we obtain a compartmental model of only 3 equations that

257 features a duration of infection that is Erlang distributed, regardless of the value of k.

258 We now apply 'linear chain trickery' [9,32,33] to the obtained compartmental model of only 3 equations

259 to further illustrate the effects of a duration of infection that is Erlang distributed.

260

261 Imposing the assumption of (30)-(32) on (22), we have that

$$m(t)I(t) = I_0\frac{k}{\gamma}\sum_{j=0}^{k-1}\frac{(\gamma t)^j}{j!}e^{-\gamma t} + \int_0^t \frac{\beta}{N}I(x)m(x)S(x)\left(\sum_{j=0}^{k-1}\frac{(\gamma x)^j}{j!}e^{-\gamma x}\right)^{-1}\left(\sum_{j=0}^{k-1}\frac{(\gamma t)^j}{j!}e^{-\gamma t}\right)dx. \tag{33}$$

262 Differentiating yields

$$m'(t)I(t) + m(t)I'(t) = I_0\frac{k}{\gamma}\left(\frac{-\gamma(\gamma t)^{k-1}}{(k-1)!}\right)e^{-\gamma t} +$$

$$\int_0^t \frac{\beta}{N}I(x)m(x)S(x)\left(\sum_{j=0}^{k-1}\frac{(\gamma x)^j}{j!}e^{-\gamma x}\right)^{-1}\left(\frac{-\gamma(\gamma t)^{k-1}}{(k-1)!}e^{-\gamma t}\right)dx + \frac{\beta}{N}I(t)m(t)S(t). \tag{34}$$

263 Mimicking the method of stages, we define

$$m_j(t)I_j(t) = I_0\frac{k}{\gamma}\left(\frac{(\gamma t)^{j-1}}{(j-1)!}\right)e^{-\gamma t} + \int_0^t \frac{\beta}{N}I(x)m(x)S(x)\left(\sum_{i=0}^{k-1}\frac{(\gamma x)^i}{i!}e^{-\gamma x}\right)^{-1}\left(\frac{(\gamma t)^{j-1}}{(j-1)!}e^{-\gamma t}\right)dx, \quad (35)$$

for $1 \leq j \leq k$.

Thereby, from (34) and (35), we have that

$$m'(t)I(t) + m(t)I'(t) = -\gamma m_k(t)I_k(t) + \frac{\beta}{N}I(t)m(t)S(t). \quad (36)$$

By differentiating (35), making the appropriate substitutions of $-\gamma m_j(t)I_j(t)$ and $\gamma m_{j-1}(t)I_{j-1}(t)$, and recognizing that $\left(\sum_{i=0}^{k-1}\frac{(\gamma t)^i}{i!}\right)^{-1}\left(\frac{(\gamma t)^{j-1}}{(j-1)!}\right) = \frac{(k-1)!\quad \eta(t)}{(j-1)!\gamma(\gamma t)^{k-j}}$, we have that

$$m_j'(t)I_j(t) + m_j(t)I_j'(t) = -\gamma m_j(t)I_j(t) + \gamma m_{j-1}(t)I_{j-1}(t) + \frac{\beta}{N}I(t)m(t)S(t)\frac{(k-1)!\quad \eta(t)}{(j-1)!\gamma(\gamma t)^{k-j}}. \quad (37)$$

for $1 < j \leq k$, and when $j = 1$,

$$m_1'(t)I_1(t) + m_1(t)I_1'(t) = -\gamma m_1(t)I_1(t) + \frac{\beta}{N}I(t)m(t)S(t)(k-1)!\frac{\eta(t)}{\gamma(\gamma t)^{k-1}}. \quad (38)$$

Summing over the index yields,

$$\sum_{j=1}^{k} m_j'(t)I_j(t) + m_j(t)I_j'(t) = -\gamma m_k(t)I_k(t) + \frac{\beta}{N}I(t)m(t)S(t)\eta(t)\sum_{j=0}^{k-1}\frac{(k-1)!\qquad 1}{j!\quad \gamma(\gamma t)^{k-j-1}}, \quad (39)$$

Noting that $\sum_{j=0}^{k-1}\frac{(k-1)!\qquad 1}{j!\quad \gamma(\gamma t)^{k-j-1}} = \frac{1}{\eta(t)}$, it follows from (36) that

$$m'(t)I(t) + m(t)I'(t) = \sum_{j=1}^{k} m_j'(t)I_j(t) + m_j(t)I_j'(t), \tag{40}$$

and

$$m(t)I(t) = \sum_{j=1}^{k} m_j(t)I_j(t). \tag{41}$$

272 Thereby, using (41) and simplifying the expression of $\eta(t)\sum_{j=0}^{k-1}\frac{(k-1)!}{j!}\frac{1}{\gamma(\gamma t)^{k-j-1}}$, we have that

$$\begin{aligned} &m_j'(t)I_j(t) + m_j(t)I_j'(t) \\ &\quad = -\gamma m_j(t)I_j(t) + \gamma m_{j-1}(t)I_{j-1}(t) + \frac{\beta}{N}\left(\sum_{i=1}^{k} m_i(t)I_i(t)\right)S(t)\left(\sum_{i=0}^{k-1}\frac{(j-1)!}{i!}\frac{1}{(\gamma t)^{j-i-1}}\right) \end{aligned} \tag{42}$$

and

$$m_1'(t)I_1(t) + m_1(t)I_1'(t) = -\gamma m_1(t)I_1(t) + \frac{\beta}{N}\left(\sum_{i=1}^{k} m_i(t)I_i(t)\right)S(t)\left(\sum_{i=0}^{k-1}\frac{(j-1)!}{i!}\frac{1}{(\gamma t)^{j-i-1}}\right)^{-1}. \tag{43}$$

273 Isolating (42)-(43) for $I_j'(t)$ and $I_1'(t)$, respectively, we arrive at

$$I_1'(t) = -\left(\frac{m_1'(t)}{m_1(t)} + \gamma\right)I_1(t) + \frac{1}{m_1(t)}\frac{\beta}{N}\left(\sum_{i=1}^{k} m_i(t)I_i(t)\right)S(t)\left(\sum_{i=0}^{k-1}\frac{(j-1)!}{i!}\frac{1}{(\gamma t)^{j-i-1}}\right)^{-1}, \tag{44}$$

274 and

$$I_j'(t) = -\left(\frac{m_j'(t)}{m_j(t)} + \gamma\right)I_j(t) + \gamma\frac{m_{j-1}(t)}{m_j(t)}I_{j-1}(t) + \frac{1}{m_j(t)}\frac{\beta}{N}\left(\sum_{i=1}^{k} m_i(t)I_i(t)\right)S(t)\left(\sum_{i=0}^{k-1}\frac{(j-1)!}{i!}\frac{1}{(\gamma t)^{j-i-1}}\right) \tag{45}$$

275 Finally, we consider the special case when individual stages feature identically constant waiting-times,

276 which implies that

$$I_1^{'}(t) = -\gamma I_1(t) + \frac{\beta}{N}\left(\sum_{j=1}^{k} I_j(t)S(t)\right)\left(\sum_{i=0}^{k-1}\frac{1}{i!}\frac{1}{(\gamma t)^{-i}}\right)^{-1}, \tag{46}$$

and

$$I_j^{'}(t) = -\gamma I_j(t) + \gamma I_{j-1}(t) + \frac{\beta}{N}\left(\sum_{i=1}^{k} I_i(t)\right)S(t)\left(\sum_{i=0}^{k-1}\frac{(j-1)!}{i!}\frac{1}{(\gamma t)^{j-i-1}}\right)^{-1}. \tag{47}$$

Through (29), (30) and (32), we reduce system (19) to a system of 3 differential equations based on an Erlang distributed duration of infection, or through the 'linear chain trick' a system of $k+2$ differential equations (i.e. (46)-(47), $S^{'}$ and $R^{'}$). An important distinction between (46)-(47) and the traditional system of differential equations (3b) with an Erlang distributed duration of infection is that new infections enter into any given infectious state in (46)-(47), based on a component of the duration of infection distribution, whereas (3b) requires all new infections to enter the first stage of infection. Thereby, (46)-(47) present an approach that is likely to better conserve the variation in infectious period, at least relative to its traditional compartmental model counterparts.

Finally, in regards to the basic reproductive numbers, we have that $\int_x^{\infty} P(t,x)dx = \frac{k}{\gamma}$, $\lim_{t\to\infty}\frac{m(t)}{m^{'}(t)+1} = \frac{1}{\gamma}$ and $\lim_{t\to 0}\frac{m(t)}{m^{'}(t)+1} = \frac{k}{\gamma}$, so

$$\mathcal{R}_0 = \beta\frac{k}{\gamma}, \text{ and, } \widehat{\mathcal{R}}_0 \in \left[\frac{\beta}{\gamma}, \beta\frac{k}{\gamma}\right]. \tag{48}$$

3.3. The Pearson distribution. If the duration of infection is Pearson distributed, then the associated survival function does not posses a general closed form. Thus, we define the Pearson distribution as

$$P(t,x) = \frac{P(T > t)}{P(T > x)}$$

where (49)

$$\frac{d^2}{dt^2}P(t,x) = h(t;\mu,\sigma,\gamma_1,\gamma_2)\,\frac{d}{dt}P(t,x),$$

289 and

$$h(t;\mu,\sigma,\gamma_1,\gamma_2) = \frac{\left(-12\gamma_1^2 + 10\gamma_2 + 12\right)(t-\mu) + (\gamma_1\gamma_2 + 6\gamma_1)\sigma}{\left(3\gamma_1^2 - 2\gamma_2\right)(t-\mu)^2 - \sigma\gamma_1(\gamma_2 + 6)(t-\mu) + \left(3\gamma_1^2 - 12 - 4\gamma_2\right)\sigma^2}. \quad (50)$$

290 Note, μ,σ,γ_1, and γ_2 are the mean, standard deviation, skewness, and excess kurtosis of the duration of

291 infection [30,34,35].

292 To obtain the mean residual waiting-time in terms of the hazard function, we have that [30,36]:

$$g(t) = \frac{1}{1-2c}\left(\sigma^2 a + \sigma b(t-\mu) + c(t-\mu)^2\right), \quad (51)$$

293 where $a = \frac{13\gamma_1^2 - 4\gamma_2 - 12}{2\ 6\gamma_1^2 - 5\gamma_2 - 6}$, $b = -\frac{1\ \gamma_1(\gamma_2 + 6)}{2 6\gamma_1^2 - 5\gamma_2 - 6}$, and $c = \frac{1\ \ 3\gamma_1^2 - 2\gamma_2}{2 6\gamma_1^2 - 5\gamma_2 - 6}$.

294 It follows from (15) that the mean residual waiting-time is:

$$m(t) = \mu - t + \left(\frac{1}{1-2c}\left(\sigma^2 a + \sigma b(t-\mu) + c(t-\mu)^2\right)\right)\eta(t), \quad (52)$$

295 where $c \neq 1/2$.

296 Through the use of (8), we obtain from (52) a first order differential equation to determine $m(t)$:

$$m'(t) = \left(\frac{1}{1-2c}(\sigma^2 a + \sigma b(t-\mu) + c(t-\mu)^2)\right)^{-1}(m(t)+t-\mu)m(t) - 1, \tag{53}$$

297 provided $\sigma^2 a + \sigma b(t-\mu) + c(t-\mu)^2 \neq 0$, and $c \neq \frac{1}{2}$.

To ensure the mean residual waiting-time is finite we apply l'Hôpital's rule to obtain

$$\lim_{t\to\infty} m(t) = \lim_{t\to\infty} \frac{\mu - \int_0^t P(z,0)dz}{P(t,0)} = \lim_{t\to\infty} -\frac{P(t,0)}{\frac{dP}{dt}} = \lim_{t\to\infty} -\frac{\frac{dP}{dt}}{\frac{d^2P}{dt^2}} = \lim_{t\to\infty} -\frac{1}{h(t;\mu,\sigma,\gamma_1,\gamma_2)}. \tag{54}$$

298 Thus, we require that $3\gamma_1^2 - 2\gamma_2 = 0$ to force the terms involving t^2 to drop out. Under this assumption,

299 we have that

$$\lim_{t\to\infty} -\frac{1}{h\left(t;\mu,\sigma,\gamma_1,\frac{3}{2}\gamma_1^2\right)} = \frac{1}{2}\sigma\gamma_1, \tag{55}$$

300 Given (55), the two formulations of the basic reproductive number are:

$$\mathcal{R}_0 = \beta\mu, \tag{56}$$

301 and

$$\hat{\mathcal{R}}_0 \in \left[\frac{1}{2}\beta\sigma\gamma_1, \beta\mu\right]. \tag{57}$$

302 Note, the lower bound of (57) is consistent with the next-generation method estimate for the basic

303 reproductive numbers (28) and (48), as the Erlang distribution has moments $\sigma = \frac{\sqrt{k}}{\gamma}$ and $\gamma_1 = \frac{2}{\sqrt{k}}$, which

304 implies $\frac{1}{2}\sigma\gamma_1 = \frac{1}{\gamma}$.

4. Discussion

In this work, we presented a class of novel differential equation compartmental models by modifying the classical assumptions that simplify nonlinear Volterra integral equations into systems of differential equations. To do this, we generalize the notion of the total infectivity of a disease, and the representation of its duration of infection. We illustrate the consistency of our class of novel models to the traditional models for exponential and Erlang distributed durations of infections, present a new class of differential equation compartmental models based a Pearson distributed duration of infection, and provide equilibria and basic reproductive numbers for our approach.

The requirement that the duration of infection follows an exponential distribution is often a source of weakness with regards to the biological validity of differential equation compartmental models. While the extension of such models through the linear chain trick [9,32,33] to a duration of infection that follows the Erlang distribution alleviates this weakness to some degree, it does so at the cost of inflating the size of the compartmental model, and thereby increasing the computational complexity of the system. Our new class of models avoids this inflation, while retaining the benefits of having a distribution of infection that follows an Erlang distribution. Thereby our new class of models offer an approach to reduce model complexity in an era when the complexity of compartmental models is ever increasing. Furthermore, if one generalizes the concept of a function to include the survival function of the Gamma distribution, our new class of models advantageously accounts for any Gamma distribution parameter values, including non-integer cases.

A main advantage of the new class of models is that they are ODE based. Therefore, like the traditional compartmental models, they do not require specialist knowledge to use, and possess well-developed numerical methods for their simulation. Furthermore, the theoretical extensions of the traditional

328 models to include alternative formulations of the force of infection, state-dependent recovery rates, in
329 addition to the inclusion of additional disease compartments are easily implemented in the new class of
330 models. In addition, many of the applications of the traditional models, such as the study of multi-strain
331 dynamics, health benefit analysis, and cost effectiveness analysis, should naturally carry over without
332 the need to reinvent the procedures of each analysis.
333
334 A potentially fruitful avenue for future applications of our new class of models is in the study of
335 virulence and disease evolution. To elaborate, because our new class of model includes both the
336 quantity of infected individuals and their duration of infection, it may serve as a better paradigm for
337 investigating selective pressures that pathogens face, at least relative to traditional models. Similarly,
338 our idea to track both the quantity of infected individuals along with their duration of infection could be
339 adapted to study species competition, as a similar modification should provide stronger intuition on
340 species fitness.
341
342 A surprising outcome of our work is the discovery that the lower bound provided by the next-generation
343 method estimate of the basic reproductive number for the Pearson distributed example depends on
344 standard deviation and skewness, instead of the mean. As standard deviation and skewness indicate the
345 spread and lean of a distribution, it seems reasonable that their combination makes for a decent proxy
346 for the location of the middle of a distribution. While this could be a consequence of assuming that the
347 duration of infection follows the Pearson distribution, it also highlights a potentially new approach to
348 bound the basic reproductive number for a disease directly from data.
349
350 The use of a duration of infection that is Pearson distributed may also open up an interesting avenue
351 into bifurcation analysis. In particular, with modifications to incorporate demographic turnover or loss of

352 immunity, and the use of the Pearson distribution, one could investigate if a relationship exists between
353 the occurrence of periodic cycles and the first four moments of the duration of infection. Thereby, one
354 may be able to gain intuition as to whether bifurcations are likely to occur simply by examining
355 statistical moments and the diseases transmission rate. In addition, through such modified models, it
356 may be possible to determine the existence (or non-existence) of hopf bifurcations by examining
357 whether $m(t)$ is periodic, as this seems like it would a requirement for such behavior in reality.
358
359
360 As our new class of models are based on the integral equation version of the Kermack and McKendrik
361 model, it shares this model's limitations. Namely, the assumptions of a sufficiently large and well-mixed
362 population, the compartmentalization of diseases into distinct stages, and the transmission assumption
363 of the law of mass-action. While these limitations may seem numerous, they do not impeded research
364 on traditional models, and thereby should inhibit the theoretical extension and application of the
365 broader class of models proposed here.
366
367 The traditional assumptions that reduce the integral equation version of the Kermack and McKendrik
368 model to a system of differential equations provides disease modellers with a rich source for
369 mathematical and scientific discovery. Here, we proposed a generalization of these traditional
370 assumptions to a biologically more accurate description of the total infectivity of a disease. By imposing
371 these new assumptions, we provide a more descript picture of how a disease propagates throughout a
372 population, while retaining the convenience and simplicity of differential equation compartmental
373 models.
374
375

References

1. Kermack WO, McKendrick AG. Contributions to the mathematical theory of epidemics--III. Further studies of the problem of endemicity. 1933. Bull Math Biol. 1991;53: 89–118. Available: http://www.ncbi.nlm.nih.gov/pubmed/2059743

2. Kermack WO, McKendrick AG. Contributions to the mathematical theory of epidemics--II. The problem of endemicity.1932. Bull Math Biol. 1991;53: 57–87. Available: http://www.ncbi.nlm.nih.gov/pubmed/2059742

3. Kermack WO, McKendrick AG. Contributions to the mathematical theory of epidemics--I. 1927. Bull Math Biol. 1991;53: 33–55. Available: http://www.ncbi.nlm.nih.gov/pubmed/2059741

4. Diekmann O, Metz H, Heesterbeek H. The legacy of Kermack and McKendrick. In: Mollison D, editor. Epidemic Models: Their Structure and Relation to Data. Cambridge, UK: Cambridge University Press; 1995. pp. 95–115.

5. Brauer F. General compartmental epidemic models. Chinese Ann Math Ser B. 2010;31: 289–304. doi:10.1007/s11401-009-0454-1

6. Blackwood JC, Childs LM. An introduction to compartmental modeling for the budding infectious disease modeler. Lett Biomath. 2018;5: 195–221. doi:10.1080/23737867.2018.1509026

7. Roberts M, Andreasen V, Lloyd A, Pellis L. Nine challenges for deterministic epidemic models. Epidemics. 2015;10: 49–53. doi:10.1016/j.epidem.2014.09.006

8. Hethcote HW, Tudor DW. Integral equation models for endemic infectious diseases. J Math Biol. 1980;9: 37–47. doi:10.1007/BF00276034

9. Lloyd AL. Realistic Distributions of Infectious Periods in Epidemic Models: Changing Patterns of Persistence and Dynamics. Theor Popul Biol. 2001;60: 59–71. doi:10.1006/tpbi.2001.1525

398 10. Feng Z, Xu D, Zhao H. Epidemiological models with non-exponentially distributed disease stages
399 and applications to disease control. Bull Math Biol. 2007;69: 1511–36. doi:10.1007/s11538-006-
400 9174-9

401 11. Kelly-Hope LA, McKenzie FE. The multiplicity of malaria transmission: a review of entomological
402 inoculation rate measurements and methods across sub-Saharan Africa. Malar J. 2009;8: 19.
403 doi:10.1186/1475-2875-8-19

404 12. Bailey NTJ, others. The mathematical theory of infectious diseases and its applications. Charles
405 Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE.; 1975.

406 13. Bailey NTJ. On Estimating the Latent and Infectious Periods of Measles: I. Families with Two
407 Susceptibles Only. Biometrika. 1956;43: 15. doi:10.2307/2333574

408 14. Gough KJ. The estimation of latent and infectious periods. Biometrika. Biometrika Trust; 1977;64:
409 559–565.

410 15. Earn DJ. A Simple Model for Complex Dynamical Transitions in Epidemics. Science (80-).
411 2000;287: 667–670. doi:10.1126/science.287.5453.667

412 16. Hethcote HW, van den Driessche P. Some epidemiological models with nonlinear incidence. J
413 Math Biol. 1991;29: 271–87. Available: http://www.ncbi.nlm.nih.gov/pubmed/2061695

414 17. Ponciano JM, Capistrán MA. First Principles Modeling of Nonlinear Incidence Rates in Seasonal
415 Epidemics. Pascual M, editor. PLoS Comput Biol. 2011;7: e1001079.
416 doi:10.1371/journal.pcbi.1001079

417 18. Alexander ME, Moghadas SM. Bifurcation Analysis of an SIRS Epidemic Model with Generalized
418 Incidence. SIAM J Appl Math. 2005;65: 1794–1816. doi:10.1137/040604947

419 19. Liu WM, Hethcote HW, Levin SA. Dynamical behavior of epidemiological models with nonlinear

incidence rates. J Math Biol. 1987;25: 359–80. Available: http://www.ncbi.nlm.nih.gov/pubmed/3668394

20. Liu WM, Levin SA, Iwasa Y. Influence of nonlinear incidence rates upon the behavior of SIRS epidemiological models. J Math Biol. 1986;23: 187–204. Available: http://www.ncbi.nlm.nih.gov/pubmed/3958634

21. Ruan S, Wang W. Dynamical behavior of an epidemic model with a nonlinear incidence rate. J Differ Equ. 2003;188: 135–163. doi:10.1016/S0022-0396(02)00089-X

22. Greenhalgh S, Day T. Time-varying and state-dependent recovery rates in epidemiological models. Infect Dis Model. 2017; doi:10.1016/j.idm.2017.09.002

23. Greenhalgh S, Galvani AP, Medlock J. Disease elimination and re-emergence in differential-equation models. J Theor Biol. 2015;387: 174–80. doi:10.1016/j.jtbi.2015.09.037

24. Greenhalgh S, Hobbs C V., Parikh S. Brief Report: Antimalarial Benefit of HIV Antiretroviral Therapy in Areas of Low to Moderate Malaria Transmission Intensity. J Acquir Immune Defic Syndr. 2018;79: 249–254. doi:10.1097/QAI.0000000000001783

25. Greenhalgh S, Ndeffo M, Galvani AP, Parikh S. The epidemiological impact of HIV antiretroviral therapy on malaria in children. AIDS. 2015;29: 473–82. doi:10.1097/QAD.0000000000000550

26. Day T, Proulx SR. A general theory for the evolutionary dynamics of virulence. Am Nat. 2004;163: E40-63. doi:10.1086/382548

27. Casagrandi R, Bolzoni L, Levin SA, Andreasen V. The SIRC model and influenza A. Math Biosci. 2006;200: 152–169. doi:10.1016/j.mbs.2005.12.029

28. Tchuenche J, Garira W, Gumel A, Mukandavire Z. Mathematical analysis of a model for HIV-malaria co-infection. Math Biosci Eng. 2009;6: 333–362. doi:10.3934/mbe.2009.6.333

442 29. Heffernan JM, Smith RJ, Wahl LM. Perspectives on the basic reproductive ratio. J R Soc Interface.
443 2005;2: 281–93. doi:10.1098/rsif.2005.0042

444 30. Gupta RC, Bradley DM. Representing the mean residual life in terms of the failure rate. Math
445 Comput Model. 2003;37: 1271–1280. doi:10.1016/S0895-7177(03)90038-0

446 31. Finkelstein MS. On the reversed hazard rate. Reliab Eng Syst Saf. 2002;78: 71–75.
447 doi:10.1016/S0951-8320(02)00113-8

448 32. Krylova O, Earn DJD. Effects of the infectious period distribution on predicted transitions in
449 childhood disease dynamics. J R Soc Interface. 2013;10: 20130098–20130098.
450 doi:10.1098/rsif.2013.0098

451 33. Hurtado PJ, Kirosingh AS. Generalizations of the 'Linear Chain Trick': incorporating more flexible
452 dwell time distributions into mean field ODE models. J Math Biol. 2019; doi:10.1007/s00285-019-
453 01412-w

454 34. Pearson K. Contributions to the Mathematical Theory of Evolution. Proc R Soc London. 1893;54:
455 329–333. doi:10.1098/rspl.1893.0079

456 35. Pearson K. Contributions to the Mathematical Theory of Evolution. II. Skew Variation in
457 Homogeneous Material. Philos Trans R Soc A Math Phys Eng Sci. 1895;186: 343–414.
458 doi:10.1098/rsta.1895.0010

459 36. Papathanasiou V. A characterization of the Pearson system of distributions and the associated
460 orthogonal polynomials. Ann Inst Stat Math. 1995;47: 171–176. doi:10.1007/BF00773421

461