A Solution for the Principle Integral of the Kermack McKenrick Epidemiological Model

James K. Baird[1], Douglas A. Barlow[2] *, Buddhi Pantha [3]

[1] Departments of Chemistry and Physics, The University of Alabama in Huntsville, Huntsville, Ala. USA

[2] Alderman Barlow Labs, Trenton Fla. USA

[3] Department of Science and Mathematics, Abraham Baldwin Agricultural College, Tifton, Ga. 32111 USA

Abstract

The Kermack–McKendrick integral is accurately computed here without resorting to numerical techniques. This integral appears in the process of solving the so-called SIR epidemiological model. To our knowledge, no closed form solution for this integral has been tabulated. We derive here a closed form solution for the Kermack McKendrick integral valid over the range of the variables useful for epidemiological modeling. The final result contains three closed-form terms one of which contains an infinite series of indexed integrals that can be written to arbitrary order in terms of hypergeometric functions. Using a finite number of terms in the infinite series, this approximation is compared with the numerical results for replacement numbers 1.5 through 5.5, which are typical of the values governing epidemics. Over this range, the analytical forms for $S(t), I(t)$ and $R(t)$ are nearly identical with those obtained by numerical evaluation of the Kermack–McKendrick integral. As the number of terms used from the infinite series is increased, the analytical result appears to converge with the numerical solution. For large values of the replacement number, we show that the Kermack–McKendrick integral can be approximated by only one term.

Keywords: SIR model; Differential equations; Epidemiology; Hypergeometric functions

1 Introduction

In the early twentieth century, Kermack and McKendrick proposed their now famous model to describe the temporal effects of a contagious epidemic [1]. In this model there are a number, $S(t)$, of uninfected, yet susceptible, individuals, which varies with the time, t. The initial number of susceptible individuals is S_o. An initial number, I_o, of infected individuals interact with the susceptible population to spread the disease. The infected population

*doug.barlow@protonmail.com

spreads the disease to the susceptible population at the specific rate, β. The number of people infected at any time, t , is $I(t)$. The infected recover at the specific rate, γ. The number of recovered individuals at time, t , is $R(t)$. The rate of change of the susceptible population is given by,

$$\frac{dS}{dt} = -\beta SI \ . \tag{1}$$

The growth rate of recovered individuals is given by

$$\frac{dR}{dt} = \gamma I \ . \tag{2}$$

The initial number of recovered individuals is $R(0) = 0$. If the total population is conserved then,

$$S(t) + I(t) + R(t) = S_o + I_o \ . \tag{3}$$

Upon differentiation of (3) with respect to time and substitution of (1) and (2) in the result, one obtains,

$$\frac{dI}{dt} = \beta SI - \gamma I \ . \tag{4}$$

(1) through (4) define what is often referred to as the SIR model, and there are many examples of its application to the study of epidemics in the literature. Recenty, reviews of the SIR model, along with applications to current epidemics, have been given by Ketcheson and Duan [2, 3].

In addition to epidemiological applications, the SIR model has recently been used in the field of chemical kinetics to describe a two-step reaction where an autocatalytic step is followed by a unimolecular step [4].

Since the time of Kermack and McKendricks' report, more advanced models based upon the SIR system have been developed including the network, two-group, multi-group and stochastic SIR models [5, 6, 7, 8, 9]. Additionally, SIR models with density dependent birth rates and pulse vaccinations with time delays have been considered [10, 11].

Though the SIR model is one of the simpler epidemic models, a closed form solution for the system given by (1), (2) and (4) has remained elusive. Kermack and McKendrick showed that in the case where β and γ were constants, the system could be reduced to an integral in R and t, which must be computed numerically–the so called *Kermack–McKendrick Integral* [1]. By expanding an exponential term in the argument of the integral, they were able to arrive at an approximate solution for R, and thus I and S, valid for early times. More recently, others have presented solutions for this system, but in each case an integral remains that must be computed numerically [12, 13, 14, 15].

Results of a study by Weinstein and Barlow for the SEIR model were reported [16]. This model assumes that a subset of the population, $E(t)$, are exposed to the infected at any time [1]. They give an analytic solution for $\ln S(t)$ in terms of a power series valid for early times augmented by an asymptotic approximate correction term for larger times. When $E(0) = 0$ the SEIR model reduces to the SIR system and thus this result gives a useful approximate solution for this model as well.

However, our approach here was to find a closed form solution for the Kermack–McKendrick integral, of the SIR model, which yields $t(R)$. In doing so, we show that this result can be expressed, to arbitrary order, in closed form using an infinite series of hypergeometric functions. This result appears to converge to the true solution of the system as the number of terms included from the infinite series is increased.

Recently, a pre-print by Mieghem [17] gives a solution for the Kermack McKendrick integral using a method based upon expansion of the integrand in a Taylor series followed by a term by term integration. By contrast, in our method, we rely upon integration by parts where possible, representing the remaining integrals in terms of the hypergeometric function.

In this work, we consider the situation where β/γ is a constant. We let the initial number of infected be I_o and the initial susceptible population be S_o. Then it must be that $S(t) + I(t) + R(t) = S_o + I_o$ as $R(0) = 0$.

When applying the SIR model a useful definition is the *replacement number* R_o, where $R_o = (\beta S_o)/\gamma$. One should note that $R_o \neq R(0)$. A theorem concerning R_o, given by Hethcote [18], is worth repeating here in abridged form:

Theorem 1.0

Let $(S(t), I(t))$ be solutions for (1) and (4). If $R_o \leq 1$ then $I(t)$ decreases to zero as $t \to \infty$. If $R_o > 1$ then $I(t)$ first increases up to a maximum value $I_{max} = Io + So - \gamma/\beta - [\gamma \ln(R_o)]/\beta$ and then decreases to zero as $t \to \infty$.

In other words, when $R_o > 1$ there is an epidemic, and the number of infected increases to a maximum, while when $R_o \leq 1$, the number of infected individuals steadily decreases from the initial number I_o and no epidemic occurs.

We report here on a closed form result for the Kermack McKendrick integral, that is the derivation of $t = t(R)$, and thus a full solution for the SIR system for β/γ a constant. The principle result involves an infinite series. This series contains an indexed integral, which to our knowledge, does not have a solution listed in standard tables. It is shown here that this integral can be resolved in terms of hypergeometric functions and thus the solution for the Kermack McKendrick integral is then known in closed form. It is demonstrated here that the resulting series can be summed finitely to obtain an accurate result, when used within the SIR model, for replacement numbers in the range of 1.5 to 5.5. For large values of the replacement number, say greater than 50, the Kermack McKendrick integral is be shown to be approximated well by only one term.

2 Deriving the Kermack McKendrick Integral

Taking into account the initial condition, $R(0) = 0$, the integral of (2) is

$$R(t) = \gamma \int_0^t I(\tau) d\tau \ . \tag{5}$$

We next turn to finding the form of $S(t)$. Upon separating variables in (1) and exploiting the initial condition, $S(0) = S_o$, integration leads to the result,

$$S(t) = S_o \exp\left[-\beta \int_0^t I(\tau) d\tau \right] \ . \tag{6}$$

Next substitute (5) into (6) to obtain

$$S(t) = S_o e^{-\lambda R(t)} \ , \tag{7}$$

where we have introduced the definition,

$$\lambda = \frac{\beta}{\gamma} \ . \tag{8}$$

To derive an equation for $R(t)$, use (3) along with (2) and (7) to get

$$\frac{1}{\gamma}\frac{dR}{dt} + R + S_o e^{-\lambda R(t)} = S_o + I_o \ . \tag{9}$$

Separating variables in (9), and assuming that γ is a constant, leads to the Kermack–McKendrick integral.

$$\gamma t = \int_0^{R(t)} \frac{dR}{[I_o - R + S_o(1 - e^{-\lambda R})]} \ . \tag{10}$$

The general solution to this integral does not seem to be known in terms of tabulated functions and it is therefore routinely computed using numerical integration techniques.

Here, we will introduce a variable transformation followed by an series expansion of the integrand, which permits evaluation of the leading terms using elementary functions. The remaining terms are represented as an infinite series in inverse powers of the replacement number which acts as a sort of perturbation parameter. Each term in this perturbation series involves an indexed integral, for which a general solution is found. Additionally, we obtain a closed form solution in the limit that the replacement number approaches infinity.

3 Analytic Solution for $(R/S_o) < [(I_o/S_o) + (1 - \exp(-\lambda R))]$

To proceed with the evaluation of (10), we divide the numerator and denominator of the integral on the right by S_o. The result is

$$\gamma t = \int_0^{R(t)} \frac{dR/S_o}{[(I_o/S_o) - (R/S_o) + (1 - e^{-\lambda R})]} \ . \tag{11}$$

Next we introduce the dimensionless variables

$$x = \frac{R}{S_o} \ , \tag{12}$$

$$\sigma = I_o/S_o \tag{13}$$

and

$$R_o = \lambda S_o \ . \tag{14}$$

With these substitutions, (11) transforms to read

$$\gamma t = \int_0^{R/S_o} \frac{dx}{[\sigma - x + (1 - e^{-R_o x})]} \ . \tag{15}$$

Define

$$M(x) = [\sigma + (1 - e^{-R_o x})] \ . \tag{16}$$

Substitute (16) into (15) to obtain

$$\gamma t = \int_0^{R/S_o} \frac{dx}{M(x) - x} \ . \tag{17}$$

Section 3 assumes $(R/S_o) < [(I_o/S_o) + (1 - \exp(-\lambda R))]$, which in dimensionless variables is equivalent to $(x/M) < 1$. Given this restriction, it is feasible to expand the integrand in expression (17) in a geometric series to read

$$\frac{1}{M - x} = \frac{1}{M} \frac{1}{\left(1 - \frac{x}{M}\right)} = \frac{1}{M} \sum_{n=0}^{\infty} \left(\frac{x}{M}\right)^n = \sum_{n=1}^{\infty} \left(\frac{x^{n-1}}{M^n}\right) \ . \tag{18}$$

Upon substitution of (18) into (17), the result is

$$\gamma t = \int_0^{R/S_o} \sum_{n=1}^{\infty} \left(\frac{x^{n-1}}{M^n}\right) dx \ . \tag{19}$$

Introduce the following variable substitution,

$$z = 1 - e^{-R_o x} \ , \tag{20}$$

$$x = -\frac{1}{R_o} \ln(1 - z) \ , \tag{21}$$

$$dx = \frac{dz}{R_o(1 - z)} \ . \tag{22}$$

Since x^{n-1}/M^n is positive, and assumed integrable, the summation in (19) can be performed after integration so that on using (20), (21) and (22) in (19) one obtains

$$\gamma t = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{{R_o}^n} \int_0^{1-e^{-\lambda R}} \frac{[\ln(1 - z)]^{n-1}}{(1 - z)(\sigma + z)^n} dz \ . \tag{23}$$

Perform the integral in (23) by parts by integrating $[\ln(1-z)]^{n-1}/(1-z)$ and differentiating $1/(\sigma + z)^n$. The result of this integration is

$$\gamma t = \sum_{n=1}^{\infty} \frac{1}{n} \left[\frac{(-1)\ln(1-z(x))}{R_o(\sigma+z(x))}\right]^n + \sum_{n=1}^{\infty} \frac{(-1)^n}{{R_o}^n} \int_0^{1-e^{-\lambda R}} \frac{[\ln(1-z)]^n}{(\sigma+z)^{n+1}} dz \ . \tag{24}$$

If we let $\xi = [(-1)\ln(1 - z)]/[R_o(\sigma + z)]$, then the first sum on the right of (24) can be recognized as [19]

$$\sum_{n=1}^{\infty} \frac{\xi^n}{n} = \ln\left(\frac{1}{1 - \xi}\right) \ , \tag{25}$$

for $|\xi| < 1$, which is true in this case when $I_o < S_o$. Using (25), (24) can be rewritten as

$$\gamma t = \ln\left[\frac{R_o(\sigma+z(x))}{R_o(\sigma+z(x))+\ln(1-z(x))}\right] + \sum_{n=1}^{\infty} \frac{(-1)^n}{{R_o}^n} \int_0^{1-e^{-\lambda R}} \frac{[\ln(1-z)]^n}{(\sigma+z)^{n+1}} dz \ . \tag{26}$$

(26) can be rewritten as

$$\gamma t = \ln\left[\frac{R_o(\sigma + z(x))}{R_o(\sigma + z(x)) + \ln(1 - z(x))}\right] + I(x) \ , \tag{27}$$

where we have defined,

$$I(x) = \sum_{n=1}^{\infty} \frac{(-1)^n}{{R_o}^n} \int_0^{1-e^{-\lambda R}} \frac{[\ln(1-z)]^n}{(\sigma + z)^{n+1}} dz \ . \tag{28}$$

Writing the sum in (28), with the first term shown explicitly, one obtains

$$I(x) = \tfrac{(-1)}{R_o} \int_0^{1-e^{-\lambda R}} \tfrac{\ln(1-z)}{(\sigma+z)^2} dz + \sum_{n=2}^{\infty} \tfrac{(-1)^n}{{R_o}^n} \int_0^{1-e^{-\lambda R}} \tfrac{[\ln(1-z)]^n}{(\sigma+z)^{n+1}} dz \ . \tag{29}$$

By use of the variable substitution,

$$y = 1 - z \ , \tag{30}$$

$$dz = -dy \ , \tag{31}$$

the first integral on the right side of (29) can then be rewritten as

$$\frac{(-1)}{R_o} \int_0^{1-e^{-\lambda R}} \frac{\ln(1-z)}{(\sigma + z)^2} dz = \frac{1}{R_o} \int_1^{1-z(x)} \frac{\ln y}{(\sigma + 1 - y)^2} dy \tag{32}$$

This integral can be resolved as [20]

$$\int_1^{1-z(x)} \tfrac{\ln y}{(\sigma+1-y)^2} dy = \left\{ \tfrac{\ln(1-z(x))}{(\sigma+z(x))} - \tfrac{1}{(\sigma+1)} \ln\left[\tfrac{\sigma(1-z(x))}{\sigma+z(x)}\right] \right\} \ . \tag{33}$$

Upon substitution of (33), (29) reads

$$I(x) = \tfrac{1}{R_o} \left\{ \tfrac{\ln(1-z(x))}{(\sigma+z(x))} - \tfrac{1}{(\sigma+1)} \ln\left[\tfrac{\sigma(1-z(x))}{\sigma+z(x)}\right] \right\} + \sum_{n=2}^{\infty} \tfrac{(-1)^n}{{R_o}^n} \int_0^{1-e^{-\lambda R}} \tfrac{[\ln(1-z)]^n}{(\sigma+z)^{n+1}} dz \ . \tag{34}$$

Finally, we substitute (34) into (27) with the result,

$$\gamma t = \ln\left[\frac{R_o(\sigma+z(x))}{R_o(\sigma+z(x))+\ln(1-z(x))}\right] + \frac{1}{R_o}\left\{\frac{\ln(1-z(x))}{(\sigma+z(x))} - \frac{1}{(\sigma+1)}\ln\left[\frac{\sigma(1-z(x))}{\sigma+z(x)}\right]\right\}$$

$$+ \sum_{n=2}^{\infty}\frac{(-1)^n}{{R_o}^n}\int_0^{1-e^{-\lambda R}}\frac{[\ln(1-z)]^n}{(\sigma+z)^{n+1}}dz \; . \tag{35}$$

We can return to the original dimensioned variables by consulting (12) through (14) so that (35) is written as

$$\gamma t = \ln\left[\frac{(I_o/S_o)+1-e^{-\lambda R}}{(I_o/S_o)+1-e^{-\lambda R}-(R/S_o)}\right] + \frac{1}{R_o}\left\{\frac{-\lambda R}{[I_o/S_o+(1-e^{-\lambda R})]} - \frac{1}{(I_o/S_o+1)}\ln\left[\frac{(I_o/S_o)e^{-\lambda R}}{I_o/S_o+(1-e^{-\lambda R})}\right]\right\}$$

$$+ \sum_{n=2}^{\infty}(-1)^n\int_0^{1-e^{-\lambda R}}\frac{[\ln(1-z)]^n}{{R_o}^n(I_o/S_o+z)^{n+1}}dz \; . \tag{36}$$

A general formula valid for any positive integer n, for the remaining integral in (36), is not known by the authors to be tabulated. However, in Appendix A we derive a closed form solution for this integral in terms of hypergeometric functions. With this result known, (36) can be used to find $t(R)$ to arbitrary order in n. This can then be numerically inverted to arrive at values for $R(t)$.

Now, $I(t)$ can be found by combining (2) and (9) to get

$$I(t) = (S_o + I_o) - R(t) - S_o e^{-\lambda R(t)} \; , \tag{37}$$

while $S(t)$ can be determined from (7).

4 Analysis

More can be learned about the conditions for the applicability of the results given in the previous section by exploring the nature of the functions $M(R)$ and $x(R)$. We consider plots of both of these functions for four different replacement numbers over the range of interest. From reports, we take a reasonable range that would be of interest to epidemiologists to be 1.5, to 5.5 [2, 18], but we examine M and x here for R_o as small as 0.5 and as large as 10.0. Plots of M and x versus R for the replacement numbers 0.5, 2, 5 and 10 are depicted below in Figure 1.

From these figures it can be seen that the curves cross at the point of maximum recovered. Therefore, in all cases, $x < M$ over the domain of $t(R)$: $[0, R_{max})$ where R_{max} is the positive root of $M - x = 0$. Therefore, expression (36) should give an accurate solution for the SIR system, with constant β/γ, for all replacement numbers in the range 0.5 to 10.0. However, the difference between M and x increases as the replacement number becomes larger so it is expected that fewer terms involving the indexed integral of (36) will be required, to obtain an accurate approximate solution, as the replacement number increases.

To demonstrate the utility of the solution outlined in the previous section, we will now consider the numerical solution for the system of (1), (2) and (4) and compare this to that given by (36) along with (37) and (7) using a finite number of terms in the infinite series.

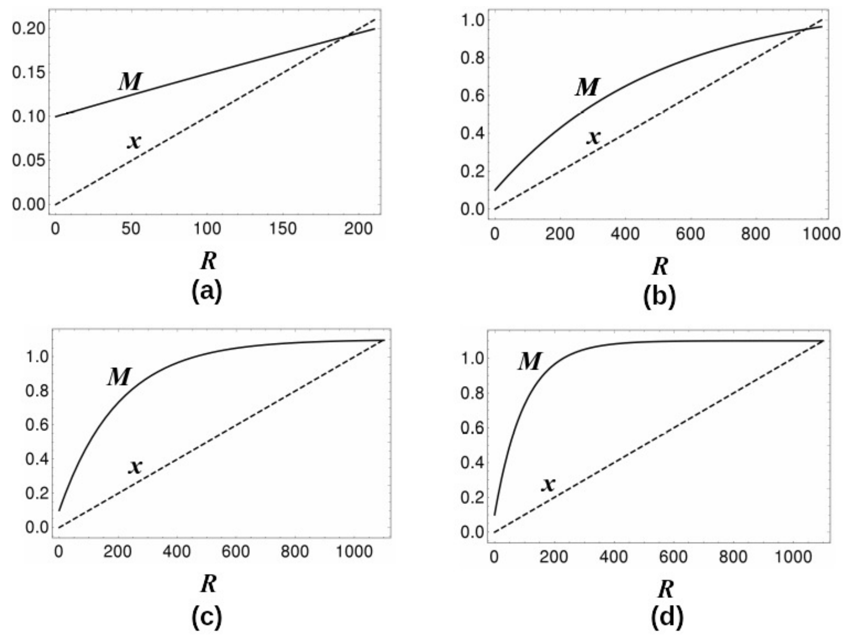

Figure 1: From upper left corner clockwise: **a)** M and x versus R for replacement number 0.5. **b)** Replacement number 2.0. **c)** replacement number 5.0. **d)** replacement number 10.0. In each case $S_o = 1000$, $\beta = 4.95 \times 10^{-4}$ people/time and $I_o = 100$. M is the solid curve while x is given by a dashed curve.

Replacement numbers of 1.5 to 6.0 are considered. The accuracy of the approximation given in the previous section will be examined visually and quantitatively.

To compare the two solutions, we focus on the similarity between the curves generated for the infected, $I(t)$. Consider a set of N $I(t)$ values taken at the times $t_1, t_2, t_3 \cdots t_N$ generated by using (36) and (37) for a given replacement number. For comparison, an interpolated function for $I(t)$ is created from the numerical solution to (1), (2) and (4), and used to generate N values for $I(t)$ at the same times $t_1, t_2, t_3 \cdots t_N$. Then, we have the two sets $\{I_1\}$ and $\{I_2\}$ both with N elements, I_1 from the method shown in this report and I_2 values come from the numerical solution. As a test for equivalence between the two sets, we define what we term the *mean relative error*. Let the relative error between an I value in each set, both taken at the same time t_k, be $\langle I_k \rangle$ then

$$\langle I_k \rangle = \frac{|I_{2_k} - I_{1_k}|}{I_{2_k}} \; . \tag{38}$$

From this we define the test statistic ζ, that is, the mean relative error, as

$$\zeta = \frac{1}{N} \sum_{k=1}^{N} \frac{|I_{2_k} - I_{1_k}|}{I_{2_k}} \; , \tag{39}$$

where, we let $N = 10R_{max}$. Here R_{max} is determined, as defined above, and then rounded to the nearest integer.

By trial and error, we find that a good match between the approximate and true numerical solution is obtained when $\zeta < 0.05$. As expected, ζ decreases as the number of terms included for the series in expression (36) is increased. Let the total number of terms used in this series be n_{max}. We found n_{max} required to first achieve $\zeta < 0.05$ for a variety of replacement numbers between 1.5 and 5.5. Curves generated for S, I and R for three of these cases are depicted in Figure 2 below. Also included in this figure is a plot showing n_{max} versus R_o. n_{max} is 13 for $R_o = 1.5$. However, at approximately $R_o = 4.5$, only third order terms are required in the indexed integral to generate excellent agreement which we suspect remains true for all $R_o \geq 4.5$.

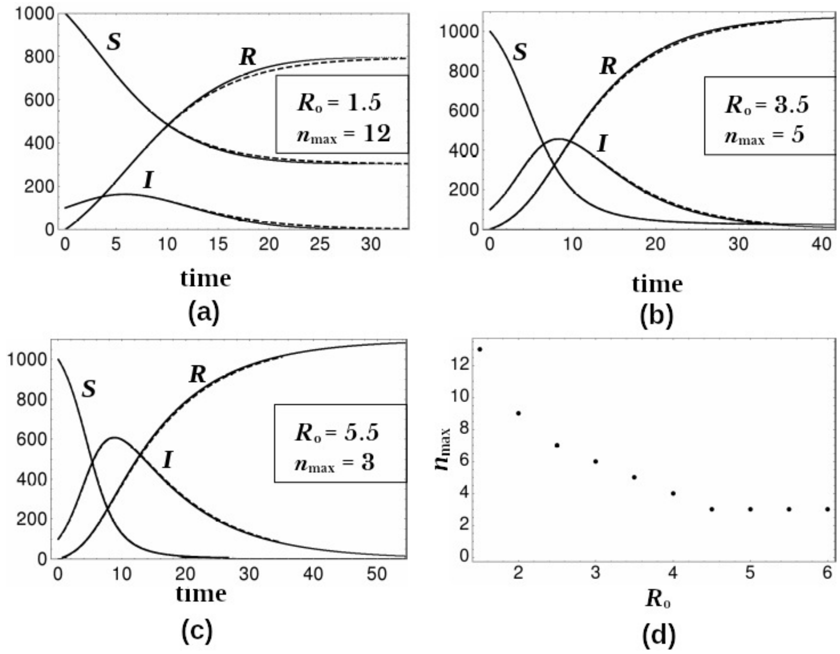

Figure 2: In each plot $S_o = 1000$, $\beta = 4.95 \times 10^{-4}$ people/time and $I_o = 100$. In Figures (a), (b) and (c) plots of R, S and I estimated by (36), (7) and (37) are given by solid curves, numerical solutions to (1), (2) and (4) are given by the dashed curves. **a)** Sum in (36) taken to twelfth order for $R_o = 1.5$ and $\zeta = 0.0512$. **b)** Same as in a) up to order 5 for replacement number 3.5 and $\zeta = 0.0242$. **c)** Same as in a) up to order 3 for replacement number 5.5 and $\zeta = 0.0149$. **d)** Plot of smallest n_{max} required to achieve $\zeta < 0.05$, versus replacement number.

5 A Solution for Large R_o

Let the sum on the right hand side of (36) be identified by

$$\Theta = \sum_{n=2}^{\infty} (-1)^n \int_0^{1-e^{-\lambda R}} \frac{[\ln(1-z)]^n}{{R_o}^n (I_o/S_o + z)^{n+1}} dz \ , \tag{40}$$

where Θ plays the role of the remainder. We propose to show that Θ approaches zero for sufficiently large values of R_o. Reintroduce the variable, x, defined by equations (20) through (22) so that (40) reads,

$$\Theta = \sum_{n=2}^{\infty} \int_0^{x(t)} \frac{x^n R_o e^{-R_o x}}{(I_o/S_o + 1 - e^{-R_o x})^{n+1}} dx \ . \tag{41}$$

In the limit $t \to \infty$, the population reaches an equilibrium state where $x(t)$ takes on its equilibrium value $x(t \to \infty) = x(\infty)$, satisfying,

$$x(\infty) - M(x(\infty)) = 0 \ . \tag{42}$$

By virtue of (16), expression (42) implies that

$$x(\infty) = I_o/S_o + 1 - e^{-R_o x(\infty)} \ . \tag{43}$$

According to (43), as $R_o \to \infty$, the equilibrium value, $x(\infty)$, satisfies $x(\infty) = M(x(\infty)) = (I_o/S_o)+1$. In this extreme case, $\beta \to \infty$, and every susceptible individual gets infected. The ultimate number of recovered individuals equals the initial number of infected individuals, I_o, plus the initial number of susceptible individuals, S_o.

For finite values of R_o, Figure 1 indicates that $x < M(x) < 1 + (I_o/S_o)$ over the domain $0 \le x < x(\infty)$. In the case of the larger values of R_o, $M(x)$ rapidly rises and approaches unity for values $x << x(\infty)$. This implies that we can safely replace $x(t)$ in the upper limit of the integral in (41) by x_o, where x_o is the solution to

$$M(x_o) = (I_o/S_o) + 1 - e^{-R_o x_o} = 1 \ . \tag{44}$$

Solving equation (44), one obtains

$$x_o = \frac{\ln(S_o/I_o)}{R_o} \ . \tag{45}$$

After substitution of x_o for $x(t)$, (41) reads

$$\Theta = \sum_{n=2}^{\infty} \int_0^{x_o} \frac{x^n R_o e^{-R_o x}}{(I_o/S_o + 1 - e^{-R_o x})^{n+1}} dx \ . \tag{46}$$

Since we seek to establish an upper bound satisfied by Θ, we are safe in replacing x^n in the integrand in (46) by ${x_o}^n$. Moreover, it does no harm to extend the sum to include the $n = 1$ term. With these changes, (46) reads,

$$\Theta = \sum_{n=1}^{\infty} {x_o}^n \int_0^{x_o} \frac{R_o e^{-R_o x}}{(I_o/S_o + 1 - e^{-R_o x})^{n+1}} dx \ . \tag{47}$$

The remaining integral in (47) can be evaluated with the result,

$$\Theta = \sum_{n=1}^{\infty} \frac{{x_o}^n}{(-n)} \left[(I_o/S_o + 1 - e^{-R_o x})^{-n} \right] \Big|_0^{x_o} \ . \tag{48}$$

Since x_o corresponds to $(I_o/S_o) + 1 - e^{-R_o x_o} = 1$, (48) simplifies to read,

$$\Theta = \sum_{n=1}^{\infty} \frac{{x_o}^n}{(n)} \left[\frac{1}{(I_o/S_o)^n} - 1 \right] . \tag{49}$$

If we assume that $x_o < (I_o/S_o) < 1$, then by use of expression (25), the sum in (49) can be evaluated to read [19],

$$\Theta = \ln \left(\frac{1}{1 - x_o/(I_o/S_o)} \right) - \ln \left(\frac{1}{1 - x_o} \right) . \tag{50}$$

Upon substitution of (45) into (50), we obtain the final result,

$$\Theta = \ln \left[\frac{1 - (1/R_o) \ln(S_o/I_o)}{1 - (1/R_o)(S_o/I_o) \ln(S_o/I_o)} \right] . \tag{51}$$

Because $(S_o/I_o) > 1$, then according to (51), Θ approaches zero when $\ln(S_o/I_o) < (S_o/I_o) \ln(S_o/I_o) << R_o$. This explains why the number of terms in (36) required for an accurate representation of $R(t)$ decreases as the value of R_o increases.

In the limit of $R_o > (S_o/I_o) \ln(S_o/I_o)$, (51) justifies our ignoring the sum in (36). Likewise for sufficiently large values of R_o, the closed form term proportional to $1/R_o$ can be ignored. In this limit, (36) assumes the simple form,

$$\gamma t = \ln \left[\frac{(I_o/S_o) + 1 - e^{-\lambda R}}{(I_o/S_o) + 1 - e^{-\lambda R} - (R/S_o)} \right] . \tag{52}$$

As a demonstration of the validity of the above claim we use (52) and (37) to develop $I(t)$ for $R_o = 50$. Then, this result is compared with the numerical result for $I(t)$ found using (1), (2) and (4). These two curves are plotted in Figure 3.

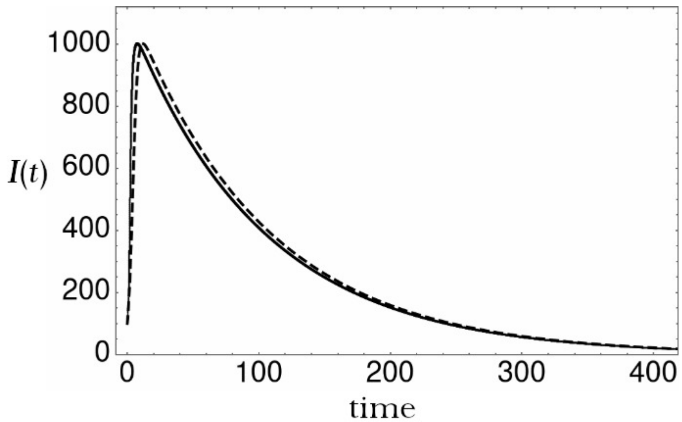

Figure 3: Here $S_o = 1000$, $\beta = 4.95 \times 10^{-4}$ people/time and $I_o = 100$, $R_o = 50$. Dashed curve is the numerical solution from (1), (2) and (4). Solid curve was made using (52) and (37)

6 Conclusions

In this report a description is given of an accurate approximation to the Kermack–McKendrick integral which in turn can be used to determine values for $R(t), I(t)$ and $S(t)$ in the SIR epidemic model. The result is shown to be effective for situations where $1.5 \leq R_o \leq 10$ with no need to numerically compute an integral. The expression for $t(R)$ contains an infinite series, though only a finite number of terms are required to obtain an accurate approximate solution.

The definite integral in expression (36) can be resolved in terms of hypergeometric functions for any integer $n > 2$, thus the solution for $t(R)$, is written in closed form. Further, it is shown that $x < M$, over the relevant range of R, for all replacement numbers greater than zero, so that it seems likely that $t(R)$, as given by (36), approaches the exact solution for the Kermack–McKendrick integral as $n \to \infty$. This possibility can be explored somewhat by comparing values for $I(t)$, at some fixed value for $R_o > 1$, coming from (36) and (41) with those from the numerical solution to (1), (2) and (4). A plot of the mean relative error between the two results, versus the number of terms included in the sum of (36), is shown in Figure 4.

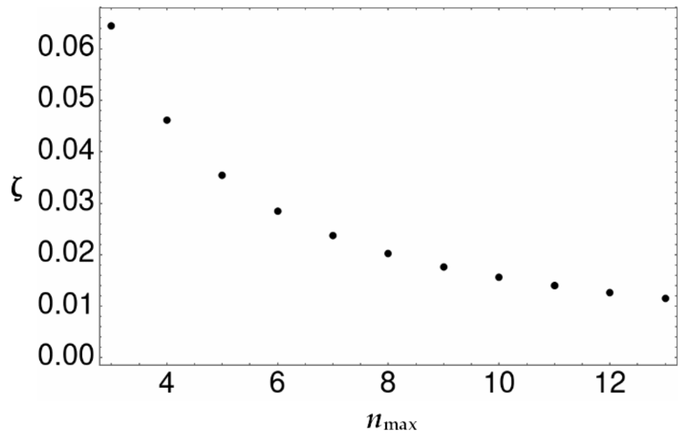

Figure 4: Mean relative error between the numerical result for $I(t)$ and the result coming via (36) versus number of terms used in the sum of (36). $\beta = 4.95 \times 10^{-4}$, $S_o = 1000$, $I_o = 100$ and $R_o = 3$.

Further work is required to determine if the trend observed in Figure 4 will continue as the number of terms in the sum of (36) is increased beyond 13.

7 Appendix A

Consider the integral from (36) in indefinite form where $n \geq 2$ while letting $\sigma = I_o/S_o$,

$$I_A = \int \frac{[\ln(1-z)]^n}{{R_o}^n(\sigma+z)^{n+1}}\ dz\ . \tag{53}$$

Integrate by parts making the choice that $dv = \int [\ln(1-z)]^n/{R_o}^n\ dz$ and $u = \frac{1}{(\sigma+z)^{n+1}}$. Upon integrating dv for $n = 2, 3, 4 \cdots$ one finds that

$$v = (1-z)\sum_{k=0}^{n} \frac{n!}{k!{R_o}^n}[\ln(1-z)]^k(-1)^{n+k}\ . \tag{54}$$

Using this along with du/dz the above integral can be written as

$$I_A = \tfrac{1-z}{(\sigma+z)^{n+1}} \textstyle\sum_{k=0}^{n} \tfrac{n!}{k!{R_o}^n}[\ln(1-z)]^k(-1)^{n+k} + \int \tfrac{(n+1)(1-z)}{(z+\sigma)^{n+2}} \textstyle\sum_{k=0}^{n} \tfrac{n!}{k!{R_o}^n}[\ln(1-z)]^k(-1)^{n+k}\ dz\ . \tag{55}$$

Assuming that the order of integration and summation in the second term of (55) can be interchanged, we rearrange and label this term as

$$I_{A_{2_k}} = (n+1)\sum_{k=0}^{n} \frac{n!}{k!}(-1)^{n+k}\int \frac{(1-z)[\ln(1-z)]^k}{{R_o}^n(z+\sigma)^{n+2}}\ dz \tag{56}$$

As with the integral in (36), the solution for this integral is not known by the authors to have been tabulated. We can however, arrive at an overall result by considering the integral for successive values of the integer k.

For $k = 0$, (56) yields the term,

$$I_{A_{2_0}} = \frac{(n+1)n!(-1)^n}{{R_o}^n}\left[\frac{nz-n+z+\sigma}{(z+\sigma)^{n+1}n(n+1)}\right]\ . \tag{57}$$

Setting $k = 1$ and integrating the integral in (56) by parts leads to,

$$I_{A_{2_1}} = \tfrac{(n+1)n!(-1)^{n+1}}{{R_o}^n}\left\{\tfrac{\ln(1-z)}{(\sigma+z)^n}\left[\tfrac{1}{n} - \tfrac{1+\sigma}{(n+1)(\sigma+z)}\right] + \tfrac{1}{n}\int \tfrac{dz}{(1-z)(\sigma+z)^n} - \tfrac{1+\sigma}{n+1}\int \tfrac{dz}{(1-z)(\sigma+z)^{n+1}}\right\}\ . \tag{58}$$

Two integral formulas from Gradshteyn and Ryzhik combine to give [21]

$$\int \frac{dX}{X{z_1}^m} = \sum_{l=1}^{m-1} \frac{1}{a^l {z_1}^{m-l}(m-l)} - \frac{1}{a}\ln\frac{z_1}{X}\ . \tag{59}$$

Here m is an integer greater than one and $z_1 = a + bX$ with a and b constants. Letting $X = 1 - z$, then using (59) to evaluate the integrals in (58), and combining the logarithmic terms yields

$$I_{A_{2_1}} = \frac{(n+1)n!(-1)^{n+1}}{{R_o}^n} \left\{ \frac{\ln(1-z)}{(\sigma+z)^n} \left[\frac{1}{n} - \frac{1+\sigma}{(n+1)(\sigma+z)} \right] + \frac{1+\sigma}{n+1} \sum_{l=1}^{n} \frac{1}{(1+\sigma)^l(\sigma+z)^{n+1-l}(n+1-l)} \right.$$

$$\left. - \frac{1}{n} \sum_{l=1}^{n-1} \frac{1}{(1+\sigma)^l(\sigma+z)^{n-l}(n-l)} + \frac{1}{n(n+1)(1+\sigma)^n} \ln \frac{\sigma+z}{1-z} \right\} . \tag{60}$$

Using commercial computing software to examine the integral in (56) for cases where $k \geq 2$, reveals the trend,

$$I_{A_{2_k}} = \sum_{k=2}^{n} \left\{ \frac{(-1)^{n+k} n! \left[nz - n + z + \sigma - (z+\sigma)\left(\frac{z+\sigma}{1+\sigma}\right)^n \right] [\ln(1-z)]^k}{{R_o}^n k! n (z+\sigma)^{n+1}} \right.$$

$$+ \frac{(-1)^{n+2k}(n+1)n!(1-z)}{{R_o}^n (k-1)!(1+\sigma)^{n+1}} \sum_{m=0}^{k-1} \frac{k!}{m!k} [\ln(1-z)]^m \times$$

$$\Big((-1)^m \ {}_pF_q \left[\{1_1, \ldots 1_{k+1-m}, 1+n\}; \{2_1, \ldots 2_{k+1-m}\}; \tfrac{1-z}{1+\sigma} \right]$$

$$\left. + (-1)^{m+1} \ {}_pF_q \left[\{1_1, \ldots 1_{k+1-m}, 2+n\}; \{2_1, \ldots 2_{k+1-m}\}; \tfrac{1-z}{1+\sigma} \right] \Big) \right\} , \tag{61}$$

where ${}_pF_q$ denotes the generalized hypergeometric function and $p = k + 2 - m$ and $q = k + 1 - m$. Since $p = q + 1$, the series defining the hypergeometric functions in (61) are convergent since $0 < 1 - z \leq 1$, [22]. The result given in (61) for $I_{A_{2_k}}$, can be verified for any integer $n \geq 2$ using known differentiation formulas for ${}_pF_q$, [23].

Therefore, the solution for the integral in (53) can be given by the sum of (57), (58), (60), (61) and the first term on the right of (55) so that

$$\int \frac{[\ln(1-z)]^n}{{R_o}^n(\sigma+z)^{n+1}} \, dz = \frac{1-z}{(\sigma+z)^{n+1}} \sum_{k=0}^{n} \frac{n!}{k!{R_o}^n} [\ln(1-z)]^k (-1)^{n+k} + \frac{(n+1)n!(-1)^n}{{R_o}^n} \left(\frac{nz-n+z+\sigma}{(z+\sigma)^{n+1} n(n+1)} \right)$$

$$+ \frac{(n+1)n!(-1)^{n+1}}{{R_o}^n} \left\{ \frac{\ln(1-z)}{(\sigma+z)^n} \left[\frac{1}{n} - \frac{1+\sigma}{(n+1)(\sigma+z)} \right] + \frac{1+\sigma}{n+1} \sum_{l=1}^{n} \frac{1}{(1+\sigma)^l(\sigma+z)^{n+1-l}(n+1-l)} \right.$$

$$\left. - \frac{1}{n} \sum_{l=1}^{n-1} \frac{1}{(1+\sigma)^l(\sigma+z)^{n-l}(n-l)} + \frac{1}{n(n+1)(1+\sigma)^n} \ln \frac{\sigma+z}{1-z} \right\}$$

$$+ \sum_{k=2}^{n} \left\{ \frac{(-1)^{n+k} n! \left[nz - n + z + \sigma - (z+\sigma)\left(\frac{z+\sigma}{1+\sigma}\right)^n \right] [\ln(1-z)]^k}{{R_o}^n k! n (z+\sigma)^{n+1}} \right.$$

$$+ \frac{(-1)^{n+2k}(n+1)n!(1-z)}{{R_o}^n (k-1)!(1+\sigma)^{n+1}} \sum_{m=0}^{k-1} \frac{k!}{m!k} [\ln(1-z)]^m \times$$

$$\Big((-1)^m \ {}_pF_q \left[\{1_1, \ldots 1_{k+1-m}, 1+n\}; \{2_1, \ldots 2_{k+1-m}\}; \tfrac{1-z}{1+\sigma} \right]$$

$$\left. + (-1)^{m+1} \ {}_pF_q \left[\{1_1, \ldots 1_{k+1-m}, 2+n\}; \{2_1, \ldots 2_{k+1-m}\}; \tfrac{1-z}{1+\sigma} \right] \Big) \right\} . \tag{62}$$

References

[1] W. Kermack and A. McKendrick, *A Contribution to the Mathematical Theory of Epidemics*, Proc. Roy. Soc. A, 115 (1927), pp. 700-721.

[2] D. I. Ketcheson, *Modeling the Spread of COVID-19*, SIAM News, May (2020), pp. 6-7.

[3] Y. Duan, *Simulations of the COVID-19 Epidemic in Nigeria Using SIR Model*, J. Phys.: Conf. Ser. 1893, (2021), 012016.

[4] J. K. Baird and D. A. Barlow, *Phase Plane Analysis and Theory of the Product Yield in a Simple Autocatalytic Reaction Mechanism*, unpublished work, 2022.

[5] L. M. Stolerman, D. Coombs and S. Boatto, *SIR-Network Model and its Application to Dengue Fever*, SIAM J. Appl. Math., 75(6), (2015), pp. 2581-2609.

[6] P. Magal, O. Seydi and G. Webb, *Final Size of an Epidemic for a Two-Group SIR Model*, SIAM J. Appl. Math., 76(5), (2016), pp. 2042-2059.

[7] T. Kuniya, *Global Behavior of a Multi-Group SIR Epidemic Model with Age Structure and an Application to the Chlamydia Epidemic in Japan*, SIAM J. Appl. Math., 79(1), (2019), pp. 321-340.

[8] H. C. Tuckwell and R. J. Williams, *Some Properties of a Simple Stochastic Epidemic Model of SIR Type*, Math. Bio. 208, (2007), pp. 76-97.

[9] Q. Zhang and K. Zhou, *Extinction and Persistence of a Stochastic SIRS Model with Nonlinear Incidence Rate and Transfer from Infectious to Susceptible*, J. Phys.: Conf. Ser. 1324, (2019), 012016.

[10] M. Song, W. Ma and Y. Takeuchi, *Permanence of a Delayed SIR Epidemic Model with Density Dependent Birth Rate*, J. Comput. Appl. Math., 201, (2007), pp. 389-394.

[11] M. Sekiguchi and E. Ishiwata, *Dynamics of a Discretized SIR Epidemic Model with Pulse Vaccination and Time Delay*, J. Comput. Appl. Math., 236, (2011), pp. 997-1008.

[12] T. Harko, F. S. N. Lobo and M. K. Mak, *Exact Analytical Solutions of the Susceptible-Infected-Recovered (SIR) Epidemic Model and of the SIR Model with Equal Death and Birth Rates*, Appl. Math. Comput., 236 (2014), pp. 184-194.

[13] M. Kröger and R. Schlickeiser, *Analytical Solution of the SIR-model for the Temporal Evolution of Epidemics. Part A: Time-Independent Reproduction Factor*, J. Phys. A: Math. Theor., 53 (2020), p. 505601.

[14] R. Schlickeiser and M. Kröger, *Analytical Solution of the SIR-model for the Temporal Evolution of Epidemics. Part B: Semi-Time Case*, J. Phys. A: Math. Theor., 54 (2021), p. 175601.

[15] A. M. Carvalho and S. Gonçalves, *An Analytic Solution for the Kermack-McKendrick Model*, Physica A, 566 (2016), p. 125659.

[16] S. J. Weinstein, M. S. Holland, K. E. Rogers and N. S. Barlow, *Analytic Solution of the SEIR Epidemic Model via Asymptotic Approximant*, Physica D, 411 (2020), p. 132633.

[17] P. Van Mieghem, *Exact Solution of the Kermack and McKendrick SIR Differential Equations*, arXiv:2010.03253v1, (2020).

[18] H. W. Hethcote, *The Mathematics of Infectious Diseases*, SIAM Rev., 42(4), (2000), pp. 599-653.

[19] L. B. W. Jolley, *Summation of Series* 2^{nd} ed., Dover Publications Inc., New York, 1961, p. 22.

[20] G. Petit Bois, *Tables of Indefinite Integrals*, Dover Publications Inc., New York, 1961, p. 142.

[21] I. S. Gradshteyn and I. M. Ryzhik in *Tables of Integrals Series and Products*, A. Jeffrey, D. Zwillinger eds., Amsterdam, Academic Press, 2007, p. 70.

[22] F. W. J. Olver, *Asymptotics and Special Functions*, Academic Press Inc., Boston, 1974, p. 168.

[23] R. A. Askey and A. B. Olde Daalhuis: Generalized Hypergeometric Functions and Meijer G-Function, in *NIST Digital Library of Mathematical Functions*, National Institute of Standards, Washington D.C., 2010-2022.