A SOLUTION TO THE KERMACK AND MCKENDRICK INTEGRO-DIFFERENTIAL EQUATIONS

Ted Duclos, Tom Reichert

Author affiliations:

T. Duclos, Entropy Research Institute, Heidelberg, Germany, ORCHID: 0000-0002-5623-2623, email address: tedduclos56@gmail.com

T. Reichert, Entropy Research Institute, Portland, Oregon. ORCHID: 0000-0003-3006-5385

Author contributions: T. Duclos and T. Reichert: Conceptualization, Writing – Original draft preparation, Writing – Review & Editing. T. Duclos: Methodology, Validation. T. Reichert: Data Investigation.

Funding: This work received no specific funding.

Declarations of interest: none.

Word counts: Abstract, 236 words; text, 7,485 words

Key Words

Kermack and McKendrick integro-differential equations

Herd Immunity

Closed form solution

Epidemic management

SIR model

Abbreviations used in this text

KMES	Kermack McKendrick integro-differential Equation Solution
RCO	Rate of Change Operator
RMM	Residential Mobility Measure
SIR	Susceptible–Infectious–Recovered

Abstract

In this manuscript, we derive a closed form solution to the full Kermack and McKendrick integro-differential equations (Kermack and McKendrick 1927) which we call the KMES. The KMES can be cast in the form of a step function response to the input of new infections; and that response is the time series of the total infections. We demonstrate the veracity of the KMES using independent data from the Covid 19 pandemic and derive many previously unknown and useful analytical expressions for diagnosing and managing an epidemic. These include new expressions for the viral load, the final size, the effective reproduction number, and the time to the peak in infections.

Since the publication of Kermack and McKendrick's seminal paper (1927), thousands of authors have utilized the Susceptible, Infected, and Recovered (SIR) approximations; expressions which are putatively derived from the integro-differential equations, to model epidemic dynamics. Implicit in the use of the SIR approximation are the beliefs that there is no closed form solution to the more complex integro-differential equations, that the approximation adequately reproduces the dynamics of the integro-differential equations, and that herd immunity always exists. However, as we explicate in this manuscript, the KMES demonstrates that the SIR models are not adequate representations of the integro-differential equations, and herd immunity is not guaranteed. Our conclusion is that the KMES obsoletes the need for the SIR approximations; and provides a new level of understanding of epidemic dynamics.

Introduction

Modern epidemiological modeling has its roots in the Kermack and McKendrick epidemic model presented in their 1927 paper (Kermack and McKendrick 1927). Since its publication, well over 10,000 authors have referenced this paper and used it as a foundational starting point. Throughout this vast literature, three basic tenets are held to be true: 1) There are no published closed-form solutions to the full set of Kermack and McKendrick's integro-differential equations; 2) An approximation to the full equations, known as the SIR (Susceptible, Infected, Recovered) model and its variants, are accepted as reasonable representatives of the full equations; and 3) the final number of uninfected individuals, S_∞, is greater than zero (ie, it is not possible for everyone to become infected). This latter property is referred to as "herd immunity".

Despite the durability of these associations, in this manuscript, we explicate that a closed-form solution to the equations can be derived; and based on this solution, we demonstrate that the arguments and mathematics used to justify the use of the SIR model, as well as those advanced to prove that herd immunity exists, rest upon flawed logic. As a preview to our approach, in this introduction, we present an outline for this solution and use this to illustrate the illogical reasoning that supports the existence of herd immunity. In the main body of this manuscript, we derive the entire solution in detail.

We begin by defining $\mathrm{I}(t)$ as the number of infected individuals at a time, t; and then write the following relationship,

$$\frac{dI(t)}{dt} = (K_T(t) - \psi(t))I(t), \qquad (1)$$

where $K_T(\mathrm{t})$ is a function describing the transmissibility of the disease and $\psi(t)$ is a function describing the recovered state of the infected people.

Equation 1 is easily solved to find an expression for $\mathrm{I}(t)$,

$$I(t) = I(0)e^{\int_0^t (K_T(t)-\psi(t))dt}, \tag{2}$$

where the initial number of infected individuals is $\mathrm{I}(0)$,

Since we are interested in determining the epidemic's final size, we first use Equations 1 and 2 to determine the running total of infected people. We define the running total as $\mathrm{N}(t)$ and the initial number of susceptible (or infectable individuals) as $\mathrm{S}(0)$. We then note that ***if*** the condition $N(t) = \mathrm{S}(0)$ can occur, $S_\infty = 0$ is possible; and herd immunity is not guaranteed.

The expression, $K_T(t)I(t)$, on the right side of Equation 1 is the change in new cases, therefore,

$$\frac{dN(t)}{dt} = K_T(t)I(t), \tag{3}$$

and,

$$N(t) = \int_0^t K_T(\mathrm{t})\mathrm{I}(0)\, e^{\int_0^t (K_T(t)-\psi(t))dt}dt \tag{4}$$

We see from Equation 4 that regardless of the size of $\mathrm{S}(0)$, there are finite values of $K_T(\mathrm{t})$ and $\psi(t)$, which allow $N(t) = \mathrm{S}(0)$. Therefore, under that range of conditions, it is possible for $S_\infty = 0$; and herd immunity is not guaranteed. This finding is provocative and suggests that we carefully examine the logic used to reach the conclusion that herd immunity, a sine qua non of conventional epidemiological modeling, always exists.

The conventional approach to deriving an expression for the final size of an epidemic begins with the summarized Kermack and McKendrick equations,

$$\frac{dS(t)}{dt} = -\frac{\varphi(t)S(t)\mathrm{I(t)}}{A_p} \tag{5}$$

$$\frac{dI(t)}{dt} = \frac{\varphi(t)S(t)\mathrm{I(t)}}{A_p} - \psi(t)I(t) \tag{6}$$

where $S(t)$ is the number of people in the population that remain susceptible, $I(t)$ is the number of people infected, $\varphi(t)$ and $\psi(t)$ are the time varying versions of the parameters that Kermack and McKendrick define respectively as the "rate of infectivity" and "the rate of recovery", and A_p is the area that encompasses the population. (We also note here that if $K_T(t)$ is assumed to equal $\frac{\varphi(t)S(t)}{A_p}$ and since $\frac{dN(t)}{dt} = -\frac{dS(t)}{dt}$, Equations 1 and 3 are equivalent to Equations 6 and 5, respectively. Therefore, Equation 4 could also be derived from Equations 5 and 6.)

The conventional derivation proceeds by dividing both sides of Equation 5 by $S(t)$; and within this early step a fundamental problem arises: Equation 5 cannot be divided by $\mathrm{S}(t)$ without first assuming $\mathrm{S}(t)$, and therefore, S_∞ is > 0 for all time. Thus, at its very beginning, the demonstration that $S_\infty > 0$, the essence of herd immunity, begins with the assumption that the conclusion is true.

Adherents of the analysis justify this assumption by either explicitly or implicitly assuming and accepting that $\varphi(t)$ (or its equivalent) is finite. Using the further assumption that $\varphi(t)$ and $\psi(t)$ are constants, and employing Equation 6, they then arrive at the following implicit expression for determining the final size,

$$\log\left(\frac{S_0}{S_\infty}\right) = R_0\left(1 - \frac{S_\infty}{K}\right), \tag{7}$$

where $K = S_0 + I(0)$, I(0) is the initial number of infected people, S_∞ is the final size of the uninfected population, S_0 is the initial size of the uninfected (or susceptible) population, and $R_0 = \frac{\varphi(0)}{\psi(0)}$. When presented in these analyses, Equation 7 is also accompanied by the statement $\varphi(t) < \infty$ and therefore $S_\infty > 0$. There are variations on this theme, but in all cases the conclusion that $S_\infty > 0$ rests on the assumption that $\varphi(t) < \infty$.

The critical assumption, that $\varphi(t) < \infty$, is supported in these analyses by making one or the other of two physical assertions. One assertion is that $\varphi(t)$ is the contact rate between the susceptible and infected populations (Brauer 2005) and therefore can never be infinite. In the second assertion, the quantity I(t)$\varphi(t)$ is referred to as the "force of infection" (Breda, et al, 2021, Diekmann, et al, 2021) which, by its very nature, is assumed to be finite.

A simple dimensional analysis of Equation 5 contravenes these assertions. In Equation 5, the units of $\varphi(t)$ must be: $new\ infections\ x\ area\ x\ (infected\ x\ susceptible\ x\ time)^{-1}$ and I(t)$\varphi(t)$ has the units of $new\ infections\ x\ area\ x\ (susceptible\ x\ time)^{-1}$. Clearly, the units of $\varphi(t)$ are not a simple contact rate and I(t)$\varphi(t)$ is not a force because it simplifies to units of inverse time. Furthermore, based on its units, as S(t) approaches zero, the quantity $\varphi(t)$ should be expected to become increasingly large because each new infection occurs in proportion to an ever-decreasing value of S(t). Therefore, $\varphi(t)$ will surely not be a constant as S(t) approaches zero; but rather will approach infinity by its very definition. Similarly, the quantity, I(t)$\varphi(t)$, with its units of inverse time, will also approach infinity as S(t) approaches zero because I(t)$\varphi(t)$ is the inverse of the time until S(t) reaches zero. Beyond these dimensional problems,

questions such as upper limits and functionality are simply unaddressed and unanswered in these analyses. They are, therefore, unconvincing.

In contrast to the conventional approach, in the first part of this introduction, we re-posed the problem, derived Equation 4, and arrived at the conclusion that herd immunity is not guaranteed. Significantly, we deduced this result without relying on any a priori assumptions about the parameters or the allowable final size. We need now only to define $K_T(\mathrm{t})$ and $\psi(t)$ in Equations 2 and 4 to arrive at a solution describing epidemic dynamics.

It is also instructive to note that when Equation 2 is rewritten as,

$$I(t) = e^{\int_0^t K_T(t)dt} I(0) e^{\int_0^t -\psi(t)dt}, \tag{8}$$

the righthand side of this expression can be interpreted as the decaying step input of the initial infections, $I(0)e^{\int_0^t -\psi(t)dt}$, times the response function, $e^{\int_0^t K_T(t)dt}$, to this input. That the response function is an exponential function, is an expected result because if there were no recovery (i.e., if $\psi(t) = 0$ for all time), then the number of infections would, and should grow exponentially. Therefore, at the outset, we see that our approach yields a logical form of the solution.

In this manuscript, we use the full Kermack and McKendrick equations and basic principles to first define $K_T(\mathrm{t})$ and $\psi(t)$, then find analytical expressions for both $\mathrm{N}(t)$ and $\mathrm{N}(\infty)$. Building on this, we develop a complete solution to the full Kermack and McKendrick equations, which we refer to as the KMES (**K**ermack and **M**cKendrick **E**quation **S**olution). We validate the KMES by correctly projecting data obtained from the Covid-19 pandemic, and we derive many useful, new analytical formulas for diagnosing and managing an epidemic.

Lastly, the availability of a closed form solution enables us to closely examine the assumptions behind the SIR epidemiological compartmental models. In that examination, we find that there are implicit, implausible assumptions, which have not been previously appreciated. We also demonstrate the forced conclusion that the conventional image of people travelling irreversibly from one compartment to the next, even under the assumption of perfect immunity in recovery, has significant flaws.

Section 1: Derivation of the final size

(Note: We use the following equation notation, (X, SY-Z), where X is the equation number in the body, Y is the supplement number and Z is the number of the equation in the supplement)

We begin our analysis by stating two of the integro-differential equations from Kermack and McKendrick in the following form:

$$\frac{dS(t)}{dt} = -\frac{\mathrm{S(t)}}{A_p}\left(\int_0^t A(\theta)V(t-\theta)d\theta + A(t)I(0)\right), \qquad (9, \text{S1-1})$$

$$I(t) = \int_0^t B(\theta)V(t-\theta)d\theta + B(t)I(0)), \qquad (10, \text{S1-2})$$

Where $S(t)$ is the susceptible population, $I(t)$ is the infected population, $V(t-\theta)$ is the new infections, $B(\theta) = e^{-\int_0^\theta \psi(a)da}$, and $A(\theta) = \varphi(\theta)B(\theta)$. Kermack and McKendrick (1927, p. 703) defined $\varphi(\theta)$ as "the rate of infectivity at age θ", and $\psi(\theta)$ as "the rate of removal" (page 703) of the infected population to the recovered population. A_p is the area that contains the population. Equation 9 is the equivalent of Equation 5 from which we have reverted to Kermack and McKendrick's original notation.

Using the prior definition that $N(t)$ is the total number of people that have been infected, and defining $R(t)$ as the portion of $N(t)$ that have recovered and are therefore immune we write the sum:

$$N(t) = I(t) + R(t) \qquad (11, \text{S1-3})$$

and,

$$\frac{dS(t)}{dt} = -\frac{dN(t)}{dt} \qquad (12)$$

If we divide Equation 9 by Equation 10, we can define a new function, $K_T(t)$ as the negative of the change in the susceptible population per infected person:

$$K_T(t) = -\frac{\frac{dS(t)}{dt}}{I(t)} = \frac{\mathrm{S(t)}(\int_0^t A(\theta)V(t-\theta)d\theta + A(t)I(0))}{A_P(\int_0^t B(\theta)V(t-\theta)d\theta) + B(t)I(0)} = S(t)\varphi(t) \qquad (13)$$

Using Equation 12 in Equation 13 and rearranging the terms, we arrive at a fundamental statement of the epidemic:

$$-\frac{dS(t)}{dt} = \frac{dN(t)}{dt} = K_T(t)I(t) \qquad (14, \text{S1-8})$$

Equation 14 describes a direct relationship between the cause of the epidemic, namely, infections, $I(t)$, and the change in the susceptible population. We can also see from Equation 14 that the units of $K_T(t)$ must be: $\frac{New\ Infections}{Infected\ person\ x\ Time}$. Since the righthand side of Equation 14 has no specific reference to susceptibles, it is plausible that $K_T(t)$ could be a parameter of the disease alone.

A solution to Equation 14 can be found by employing a parameter describing the contacts between the people within the affected population. We call this parameter $P_c(t)$ and define it as:

$$P_c(t) = \lim_{\Delta t \to 0} \int_t^{t+\Delta t} P_{cr}(t)dt, \tag{15}$$

where $P_{cr}(t)$ is the contact rate for the subpopulation $N(t)$.

Using $P_c(t)$ and $N(t)$ in the righthand side of Equation 14, we now rewrite it as the following,

$$\frac{dN(t)}{dt} = N(t)P_c(t)\frac{K_T(t)}{P_c(t)}\frac{I(t)}{N(t)} \tag{16}$$

Equation 16 states that the change in the number of people infected is fully determined by the number of interactions of people within the already infected community ($N(t)P_c(t)$), times the effectiveness of the disease transmission per contacted person ($\frac{K_T(t)}{P_c(t)}$), times the fraction of the people still infected within the already infected population ($\frac{I(t)}{N(t)}$). This is a sensible statement in that the interactions, $N(t)P_c(t)$, are discounted by the fraction of $N(t)$ that is still infected and the effectiveness of the transmission per contact, $\frac{K_T(t)}{P_c(t)}$. The latter portion of this statement sharpens the assumption that $K_T(t)$ is a function of the disease transmission alone.

We now need to find an expression for $\frac{I(t)}{N(t)}$ in terms of $K_T(t)$ and $P_c(t)$. In a first step, we define $F_i(t) = \frac{I(t)}{N(t)}$ as the fraction infected and we rewrite Equation 16 as the following difference equation at time $t = 0$,

$$N(\Delta t) = N(0) + N(0)P_c(0)(F_i(0) - F_i(0)(1 - \frac{K_T(0)}{P_c(0)}\Delta t)) \tag{17}$$

In Equation 17, since $N(0)$ is a given and $P_c(0)$ is only a function of people's behavior, the quantity $F_i(0) - F_i(0)(1 - \frac{K_T(0)}{P_c(0)}\Delta t)$ must be the change in $F_i(0)$ during the period Δt. That is, the difference between $F_i(0)$ and $F_i(\Delta t)$ can be written as

$$F_i(0) - F_i(\Delta t) = F_i(0) - F_i(0)(1 - \frac{K_T(0)}{P_c(0)}\Delta t). \quad (18)$$

Rewritten, this is,

$$F_i(\Delta t) = F_i(0)(1 - \frac{K_T(0)}{P_c(0)}\Delta t) \quad (19)$$

and for any time $(n + 1)\Delta t$, Equation 19 can be written as

$$F_i((n + 1)\Delta t) = F_i(n\Delta t)(1 - \frac{K_T(n\Delta t)}{P_c(n\Delta t)}\Delta t). \quad (20)$$

We solve Equation 20 for $F_i(t)$ by forcing $\Delta t \to 0$ to find that,

$$F_i(t) = \frac{I(t)}{N(t)} = F_i(t_0)e^{-\int_0^t \frac{K_T(t)}{P_c(t)}dt} \quad (21, S1\text{-}38)$$

Substituting Equation 21 into Equation 16, we have the following expression for $\frac{dN(t)}{dt}$,

$$\frac{dN(t)}{dt} = N(t)F_i(t_0)K_T(t)e^{-\int_0^t \frac{K_T(t)}{P_c(t)}dt} \quad (22, S1\text{-}9)$$

This equation can then be solved for $N(t)$ in terms of $K_T(t)$ and $P_c(t)$,

$$N(t) = N(t_0)e^{F_i(t_0)\int_0^t K_T(t)e^{-\int_0^t \frac{K_T(t)}{P_c(t)}dt}dt} \quad (23, S1\text{-}10)$$

and, by combining Equations 14, 22, and 23, we obtain an expression for $I(t)$,

$$I(t) = I(t_0)e^{F_i(t_0)\int_0^t K_T(t)e^{-\int_0^t \frac{K_T(t)}{P_c(t)}dt}dt - \int_0^t \frac{K_T(t)}{P_c(t)}dt} \quad (24, S1\text{-}11)$$

If $K_T(t)$ truly is solely a function of the disease, it is likely to be a constant, at least in the initial stages of an epidemic. Using this inferred property and assuming $F_i(t_0) = 1$, we can find an

expression for the potential final size, $N(\infty)$, by also assuming the population does not change its behavior (i.e., P_c is a constant) during the pandemic,

$$N(\infty) = N(t_0)e^{P_c} \tag{25}$$

Using Equation 23, we can see that it is possible for $N(\infty)$ to equal $S(0)$ and, therefore, herd immunity is not guaranteed. If this occurs, the time it would take is given by the following expression,

$$t = -\frac{P_c}{K_T} ln(1 - \frac{\ln\left(\frac{S(t_0)}{N(t_0)}\right)}{P_c}) \tag{26}$$

and the criteria that must be true for the entire population to become infected is,

$$P_c > \ln\left(\frac{S(t_0)}{N(t_0)}\right) \tag{27}$$

Equations 25 through 27 enable the estimation of the level of social interaction that will cause the total population to be infected and they demonstrate very clearly that the existence of herd immunity is not an inherent property of the Kermack and McKendrick model.

Section 2: Solution to the Kermack and McKendrick full equations

From the definition of $N(t)$ and using Equations 23 and 24 we find the expression for $R(t)$,

$$R(t) = (1 - F_i(t_0)e^{-\int_0^t \frac{K_T(t)}{P_c(t)}dt})N(t_0)e^{F_i(t_0)\int_0^t K_T(t)e^{-\int_0^t \frac{K_T(t)}{P_c(t)}dt}dt} \tag{28}$$

If we then assume that both $K_T(t)$ and $P_c(t)$ are constant for a period, we arrive at simplified expressions for $N(t)$, $I(t)$, and $R(t)$,

$$N(t) = N(t_0)e^{-F_i(t_0)P_c(e^{-\frac{K_T}{P_c}t}-1)} \tag{29}$$

$$I(t) = I(t_0)e^{-F_i(t_0)P_c\left(e^{-\frac{K_T}{P_c}t}-1\right)-\frac{K_T}{P_c}t} \tag{30}$$

$$R(t) = (1 - F_i(t_0)e^{-\frac{K_T}{P_c}t})N(t_0)e^{-F_i(t_0)P_c(e^{-\frac{K_T}{P_c}t}-1)} \tag{31}$$

In their 1927 paper, Kermack and McKendrick parameterized their equations with the functions $B(t)$, $B(\theta)$, $\varphi(t)$, $\varphi(\theta)$, $\psi(t)$, and $\psi(\theta)$. In Supplement 1, we use a vector and matrix approach to show that these parameters can be put in terms of $K_T(t)$, $P_c(t)$:

$$B(t) = e^{-\int_0^t \psi(t)dt} = e^{-\int_0^t K_T(t)dt+\int_0^t K_T(t)e^{-\int_0^t \frac{K_T(a)}{P_c(a)}da}dt-\int_0^t \frac{K_T(t)}{P_c(t)}dt} \tag{32, S1-23}$$

$$B(\theta) = e^{-(K_T(t)-K_T(t)e^{-\int_0^t \frac{K_T(a)}{P_c(a)}da}+\frac{K_T(t)}{P_c(t)})\theta} \tag{33, S1-22}$$

$$\psi(t) = \psi(\theta) = \frac{\frac{dR(t)}{dt}}{I(t)} = K_T(t) - K_T(t)e^{-\int_0^t \frac{K_T(a)}{P_c(a)}da} + \frac{K_T(t)}{P_c(t)} \tag{34, S1-21}$$

$$\varphi(t) = \varphi(\theta) = \frac{K_T(t)A_p}{S(t)} \tag{35, S1-28}$$

Equations 32 through 35 along with Equations 23, 24, and 28 form the complete solution of the full Kermack and McKendrick equations with time varying disease transmissibility, $K_T(t)$, and contact behavior, $P_c(t)$. As previously stated, we refer to these as the KMES.

Using Equation 32, and as anticipated in the introduction, we can rewrite Equations 23, 24, and 28 as,

$$I(t) = e^{\int_0^t K_T(t)dt}B(t)I(0) \tag{36, S1-34}$$

$$N(t) = e^{\int_0^t (K_T(t) + \frac{K_T(t)}{P_c(t)})dt} B(t)I(0) \quad (37, \text{S1-35})$$

$$R(t) = (e^{\int_0^t (K_T(t) + \frac{K_T(t)}{P_c(t)})dt} - e^{\int_0^t K_T(t)dt})B(t)I(0) \quad (38, \text{S1-36})$$

Since $B(t)$ is the time varying infectiousness input of the original infected group, $I(0)$, the exponential expressions, $e^{\int_0^t K_T(t)dt}$, $e^{\int_0^t (K_T(t) + \frac{K_T(t)}{P_c(t)})dt}$, and $e^{\int_0^t (K_T(t) + \frac{K_T(t)}{P_c(t)})dt} - e^{\int_0^t K_T(t)dt}$, are the step response functions to this input. These step response functions also show that if $B(t)I(0) = 1$, that is, if there were no recovery, then, based on Equation 37, the epidemic would proceed exponentially until the entire population was infected. This result is both intuitive and sensible.

Section 3: Veracity of the solution

We can test the KMES projections using data from the Covid-19 pandemic. To do this, we must first determine the appropriate values of K_T and P_c.

By rearranging the terms of Equation 22 and taking the natural log of both sides we obtain the following expression,

$$RCO = ln\left(\frac{\frac{dN(t)}{dt}}{N(t)}\right) = \ln\left(F_i(t_0)K_T(t)\right) - \int_0^t \frac{K_T(t)}{P_c(t)}dt \quad (39)$$

We labelled this as the "RCO" which stands for Rate of Change Operator because it is a measure of the rate of change of the population, $N(t)$, per person within the group $N(t)$.

If we again assume that both K_T and P_c are at least piece-wise constant for substantial periods of time, Equation 39 becomes the linear equation in time,

Figure 1 (A-F)

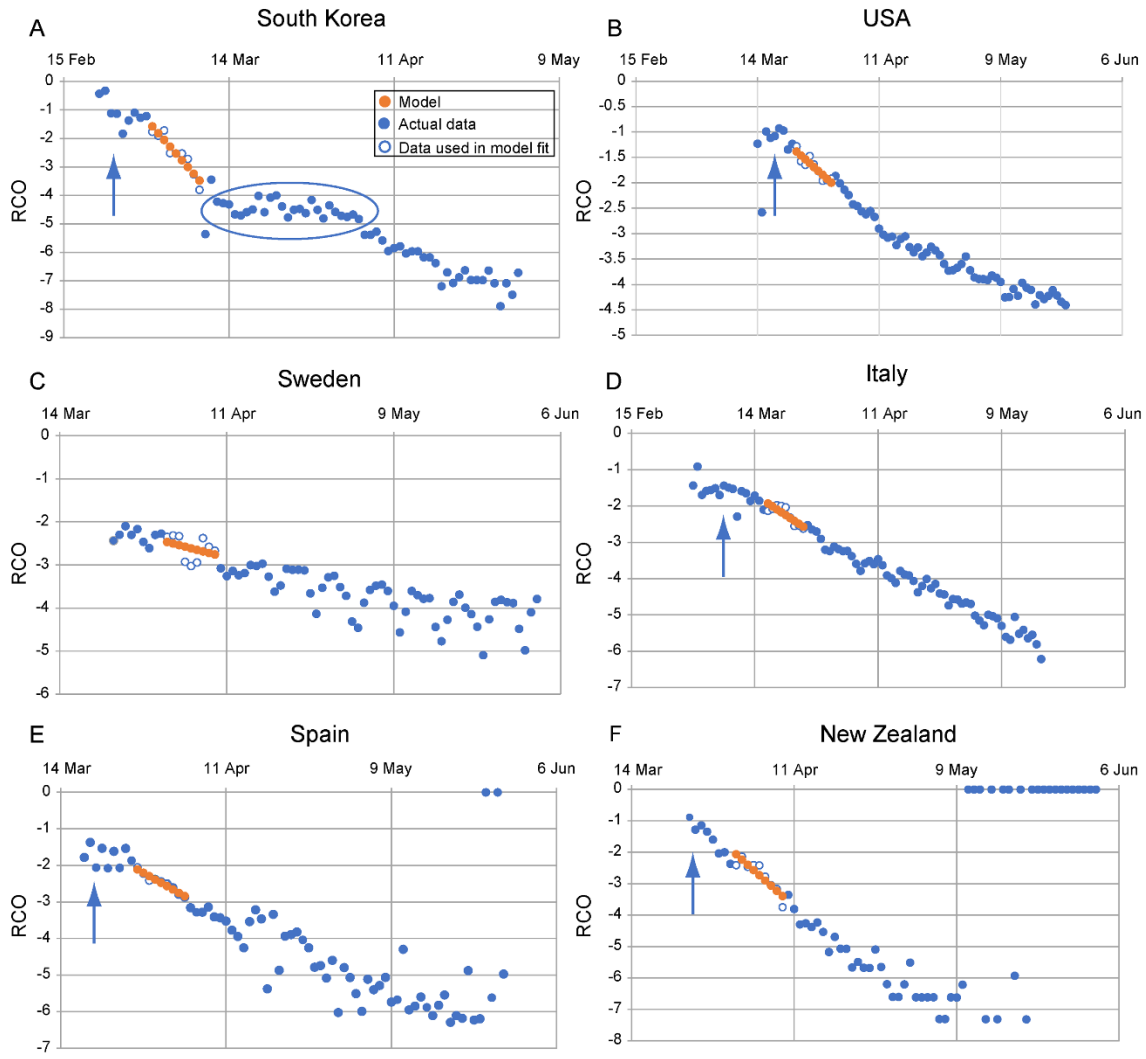

Figure 1. Rate of change operator (RCO) curves for COVID-19 cases in various countries. An epidemic can be described by a piecewise linear model using the RCO (Equation 40). A short segment of orange dots in each graph is a linear fit to the corresponding points (blue/white circles) in the observed data. The slopes and initial points of these dotted-line segments are the values of K_1 and K_2 respectively which are tabulated in Table1. In some countries, RCO curves changed markedly soon after the date containment measures were implemented (arrows): **A)** South Korea, February 21; (the oval highlights a departure of the observed data from the RCO slope, indicating failures in, or relaxations of, social distancing); **B)** USA, March 16; **C)** Sweden did not implement any specific containment measures, so the model calibration was begun on April 1, the date when the slope of the RCO curve first became steady. **D)** Italy, March 8; **E)** Spain, March 14; **F)** New Zealand, March 25. All dates are in 2020.

Table 1. Social containment parameters used to model total cases and new daily cases of infection for different countries (Roser et al 2021).

	$\frac{K_T}{P_c}$	$ln(F_i(t_0)K_T)$	$N_{(to)}$	Date range for RCO fit
South Korea	0.24	−1.58	3,736	March 1–March 9
USA	0.076	−1.39	46,136	March 23–March 31
Sweden	0.036	−2.47	5,320	April 1–April 9
Italy	0.080	−1.93	31,506	March 17–March 25
Spain	0.09	−2.11	65,719	March 27–April 4
New Zealand	0.17	−2.06	708	April 1–April 9

Parameters from linear fit of rate of change operator (RCO) data in Figure 1. $\frac{K_T}{P_c}$, slope; $ln(F_i(t_0)K_T)$, intercept; $N_{(t_0)}$, number of cases at time (t_0), first day of time range used. All dates are in 2020.the data falls in a straight line.

$$RCO = ln\left(\frac{\frac{dN(t)}{dt}}{N(t)}\right) = \ln\left(F_i(t_0)K_T\right) - \frac{K_T}{P_c}t \tag{40}$$

By applying Equation 40 to data (Roser, et al. 2021) from six different countries during the initial stages of the Covid-19 pandemic, we obtained the curves in Figure 1. As can be seen in the figure, before and shortly after the date (indicated by the arrows in the figures) of the imposition of containment actions in the six countries these RCO curves become straight lines. This is a strong indication that it is reasonable to assume K_T and P_c were constants before and after the imposition of the containment actions.

To project the course of the epidemic in the individual countries, we used a small portion of the data immediately after the imposition of the containment measures to determine the values of $ln(F_i(t_0)K_T)$ and $\frac{K_T}{P_c}$ by fitting Equation 40 to short, early portions of the straight segments of the RCO time series. These early portions comprised nine data points each and Table 1 displays the values derived for each country.

Using these values in Equation 29, we then predicted the course of daily total cases (Figure 2) for the six countries. These predictions matched the actual time series of the daily total cases with an $R^2 > 0.97$ in each of the six countries for the 45 days following the date containment measures were introduced. We then used Equation 22 to plot the predicted time series of the daily new cases in Figure 3 for the six countries for the same 45 days. These predictions have an R^2 range of 0.29 to 0.90; and as seen in the figure; the predicted peak of new cases was close to the observed peak for all countries.

It is important to emphasize that the predictions in Figures 2 and 3 are *not* fits to the full-length of the data shown. Rather, the two constants, $ln(F_i(t_0)K_T)$ and $\frac{K_T}{P_c}$ were estimated using only a short, linear, nine-point portion of the epidemic data starting between 7 to 14 days after the imposition of containment measures. These constants were then used to project the data following the nine points.

In an additional demonstration of the veracity of the KMES, we tested the assumption that $K_T(t)$ is a property of the disease, and therefore, should be the same constant for each country. Equation S2-5, derived in Supplement 2, shows that the model parameters, expressed in a purposefully constructed function, $F(N(t))$, should be linearly proportional to time with a constant of proportionality or slope equal to $-K_T(t)$. As illustrated in Figure 4, the fit of Equation S2-5, using the population density data from Table 2, has an $R^2 = 0.956$ and a slope of −0.26 (the slope is equal to $-K_T(t)$). This excellent correlation confirms that $K_T(t)$ can confidently be assumed to be the same for all countries (=0.26), is a constant in the initial stage of the epidemic, and is plausibly, a property of the disease.

Fig.2 (A-F)

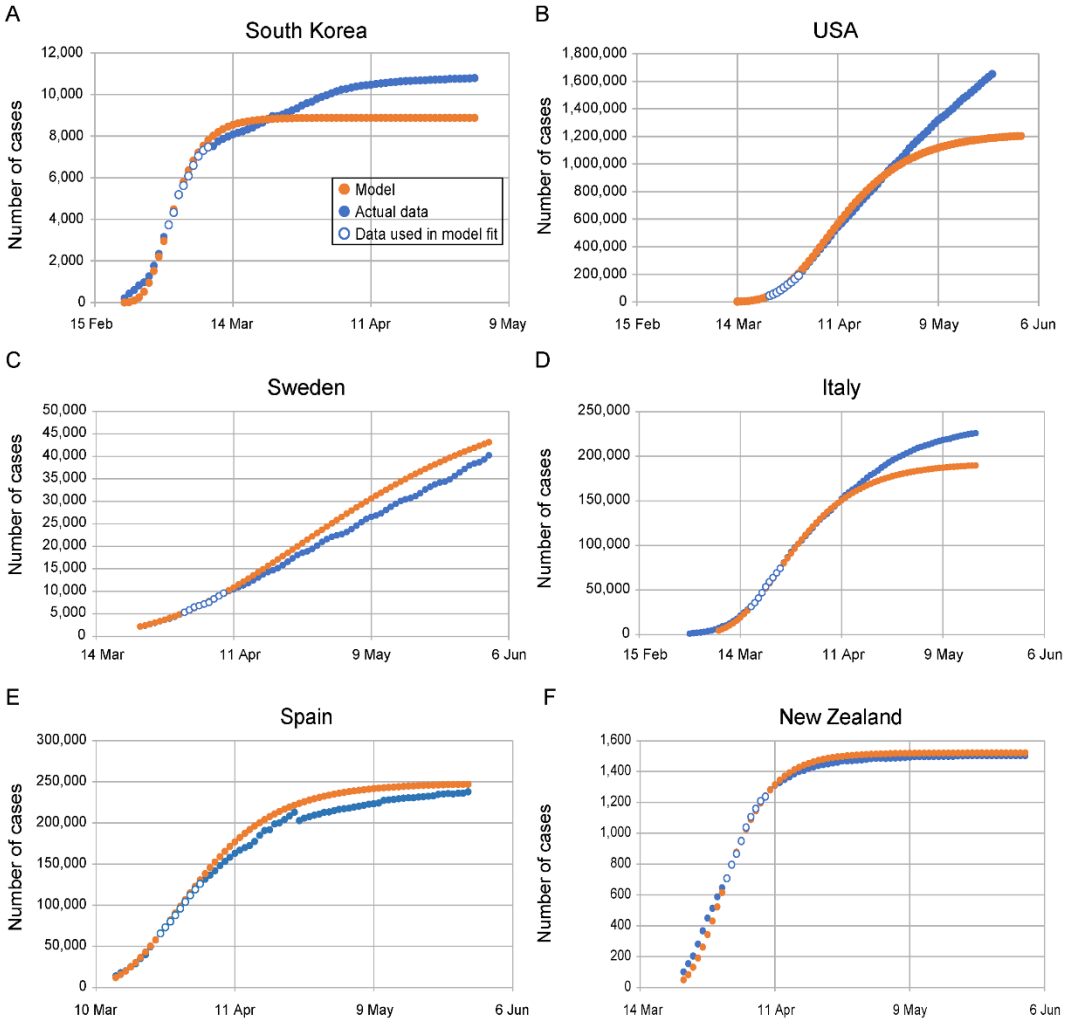

Figure 2. Complete KMES model predictions for daily total case counts. A) South Korea; **B)** USA; **C)** Sweden; **D)** Italy; **E)** Spain; and **F)** New Zealand. Dots are daily data points observed from (white-center and all blue) or calculated (orange) for each country. The KMES model was calibrated using data from the date ranges listed in Table 1 (white-center blue dots). $R^2 > 0.97$ for the model fit for all countries for the 45 days after the containment measures were implemented: South Korea, February 21-April 4; USA, March 16-April 30; Italy, March 8–April 22; Spain, March 14-April 28; New Zealand, March 25-May 9. Sweden did not implement any specific containment measures, so the dates used were March 23- May 7. The deviation of the model from the data in the USA, panel (**B**), after April is elucidated in Supplement 5. All dates are in 2020.

Fig.3 (A-F)

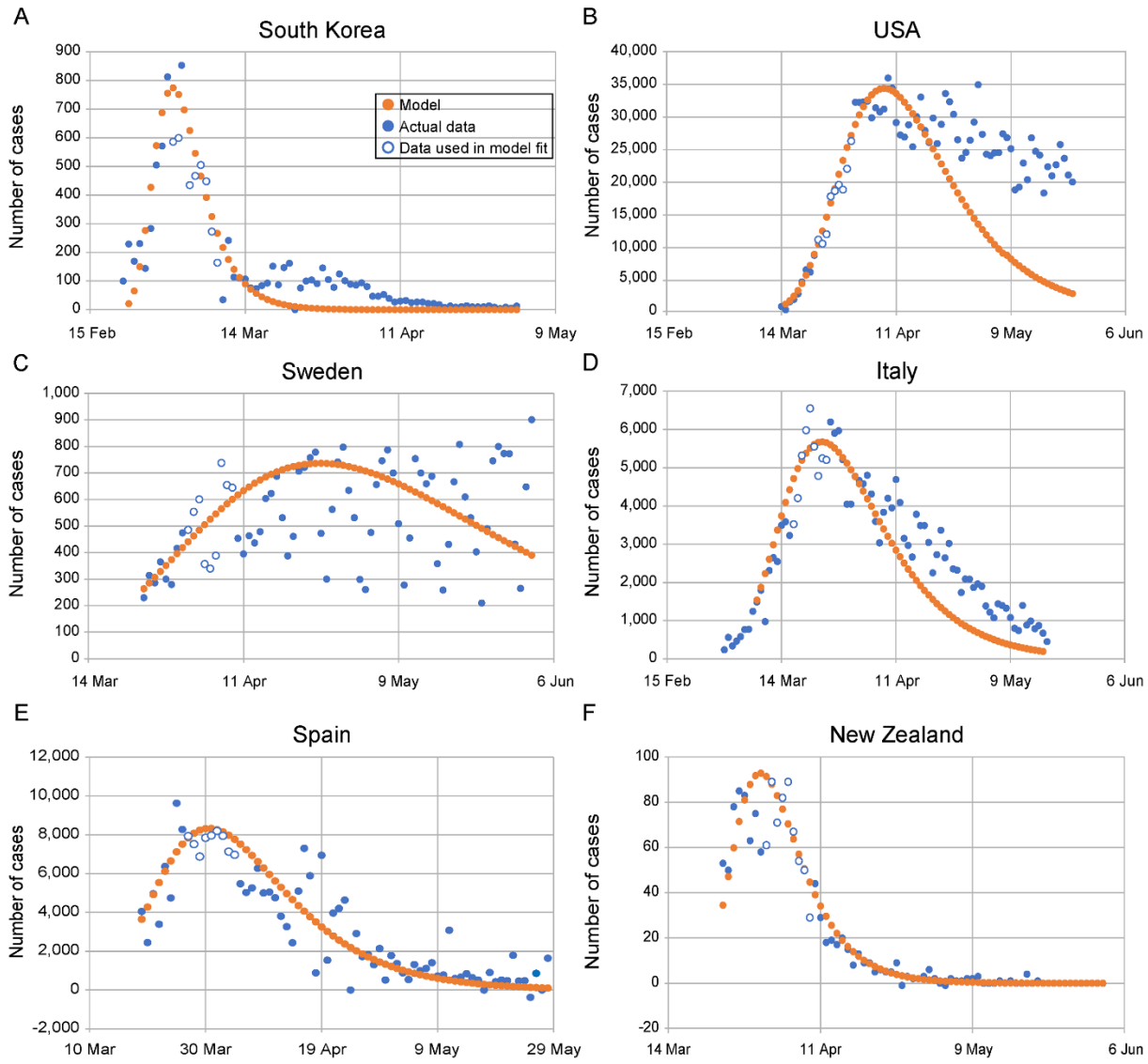

Figure 3. Complete KMES model predictions for number of new daily cases. A) South Korea, $R^2 = 0.86$; **B)** USA, $R^2 = 0.83$; **C)** Sweden, $R^2 = 0.29$; **D)** Italy, $R^2 = 0.69$; **E)** Spain, $R^2 = 0.65$; and **F)** New Zealand, $R^2 = 0.90$. The orange dotted line is the model in all panels. The all-blue and white-center blue dots are data points, daily observations from each country. The white-center blue points are used to determine model parameters. R^2 values are between the model and the data, across countries for the 45 days after containment measures were instituted.

Fig.4

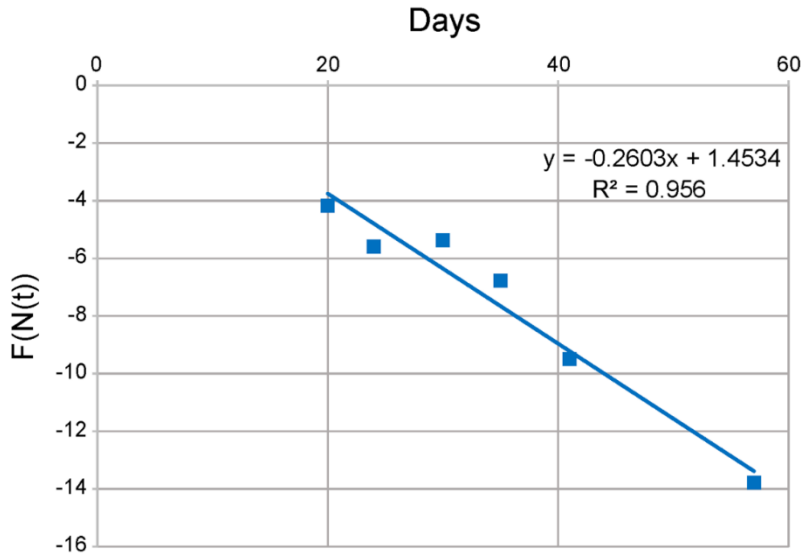

Figure 4. **Verification that K_T is the same for all countries**. The data from Table 2 is plotted using Equation S2-5 and $A_1 = 0.48\ km^2$. Each data point corresponds to a different country. The value of K_T is the negative of the slope of the line, and K_T is closely approximated everywhere by $K_T \approx 0.26$.

Table 2. Initial COVID-19 pandemic data and social interaction parameters for various countries ((Roser et al 2021), case and date data; (Worldometers 2021, population density data)

	Date of first case reported	Date of cases in calculation	Days	Cases on calculation date	Population density (people/km^2)	K_T	ln(K_T)	$\frac{K_T(0)}{P_c(0)}$
South Korea	22 Jan	21 Feb	30	204	527	0.26	−1.3	2.39E-04
USA	22 Jan	19 Mar	57	13,663	36	0.26	−1.3	3.50E-03
Sweden	1 Feb	7 Mar	35	179	25	0.26	−1.3	5.04E-03
Italy	31 Jan	24 Feb	24	229	206	0.26	−1.3	6.11E-04
Spain	1 Feb	13 Mar	41	5,232	94	0.26	−1.3	1.34E-03
New Zealand	28 Feb	19 Mar	20	28	18	0.26	−1.3	6.86E-03

All dates are in 2020.

A third illustration of veracity arises from the ability to correlate independently sourced mobility data to the RCO. As explained in Supplement 2, and based on Equation 40, if the KMES is correct, then the integral of this mobility data should correlate linearly with the measured RCO. Mobility data, available from Google (Google 2020), are a measure of the difference between the amount of time people stayed at home (the Residential Mobility Measure or RMM) during the period modelled and a baseline measured for 5 weeks starting January 3, 2020. Figure 5 shows that, as the KMES predicts, for each country considered, the integral of the RMM and the RCO are linearly correlated to a high degree.

Section 4: Useful expressions derived from the solution

The derivation of the KMES suggests that we should be able to find an expression for the average viral load of the disease. Equation 32 provides a starting point because it describes the evolution of the infectiousness of the initial infected population, $I(t_0)$. If we set $P_c = I(t_0) = 1$ in Equation 32, meaning that there was only one initial infection, assume that this infected person contacted only one person during the epidemic and recall that $K_T(t)$ is a constant, we obtain an expression for the viral load,

$$B(t) = e^{-K_T t + (e^{-K_T} - 1) - K_T t} \tag{41}$$

If we set $K_T = 0.26$ in Equation 41, as derived from figure 4, we obtain the plot in Figure 6.

The interpretation of Equation 41 and its graphical representation in Figure 6 require careful consideration. First, the initial infected person in the epidemic did not become infected at $t = 0$, and therefore, time in Equation 41 must be allowed to be negative. Second, Equation 41 has the appropriate characteristics of a viral load; that is, it grows in an initially exponential fashion,

Fig.5 (A-F)

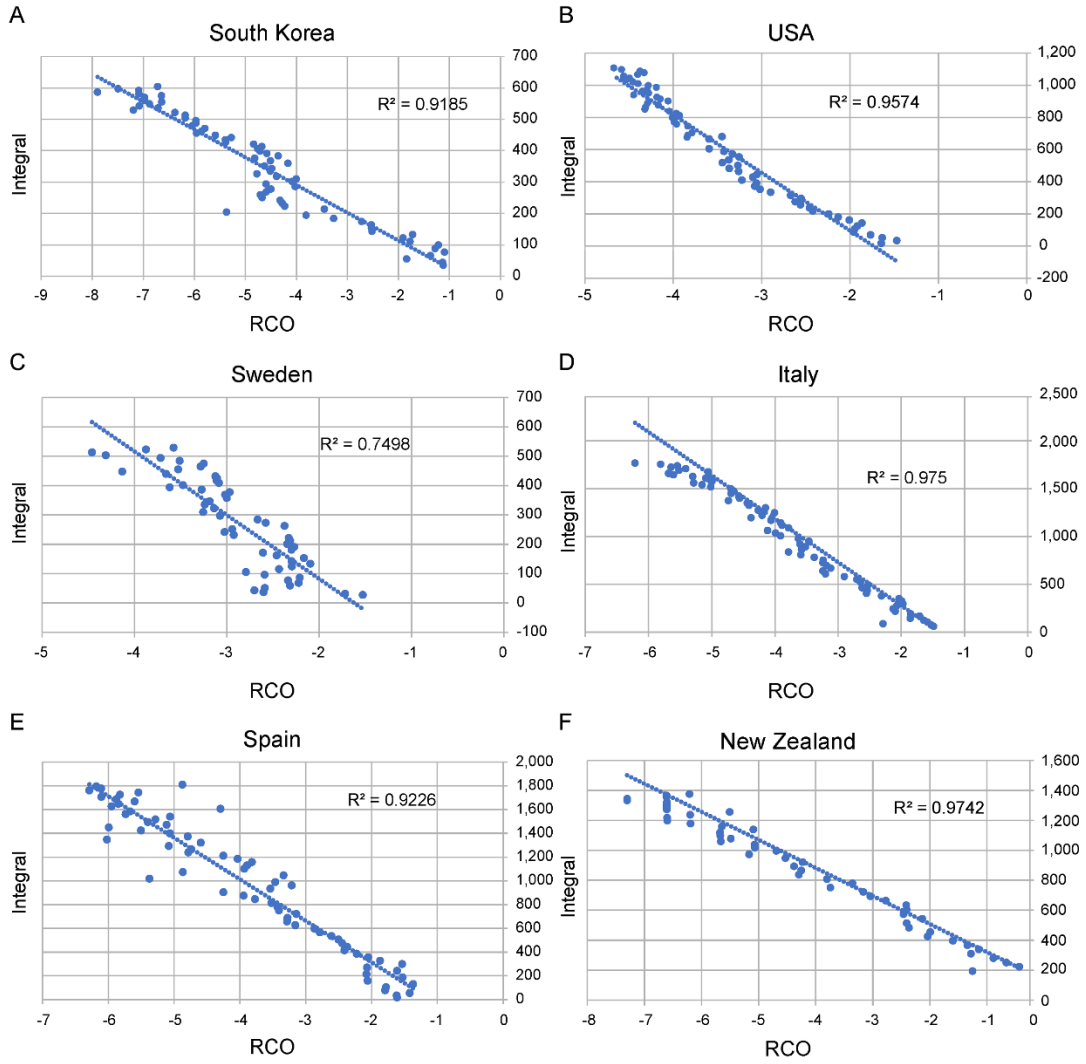

Figure 5. Correlations between the daily value of the rate of change operator (RCO) and the integral of the Google Residential Mobility Measure (RMM) (9). A) South Korea (date range, February 23 to April 23); **B)** USA (March 25 to May 31); **C)** Sweden (March 5 to May 5); **D)** Italy (March 25 to May 31); **E)** Spain (March 25 to May 31); and **F)** New Zealand (March 21 to April 22). All dates are in 2020.

reaches a maximum, and then declines at a slower exponential pace. Third, it has the same overall shape and dynamic change in load from peak to 15 days after the peak as the estimated

average loads obtained by direct measurements of individual Covid-19 patients with Covid-19 (Challenger et al 2021 and Jones et al 2022).

However, while it is tempting to label the expression $B(t)$ in Equation 41 as the viral load, it is more appropriately labeled as the effect that the viral load in one person has on the one person they contact. This is related to the true viral load by some factor (not specified here); and its dynamic shape reflects the dynamics of the viral load.

Fig. 6

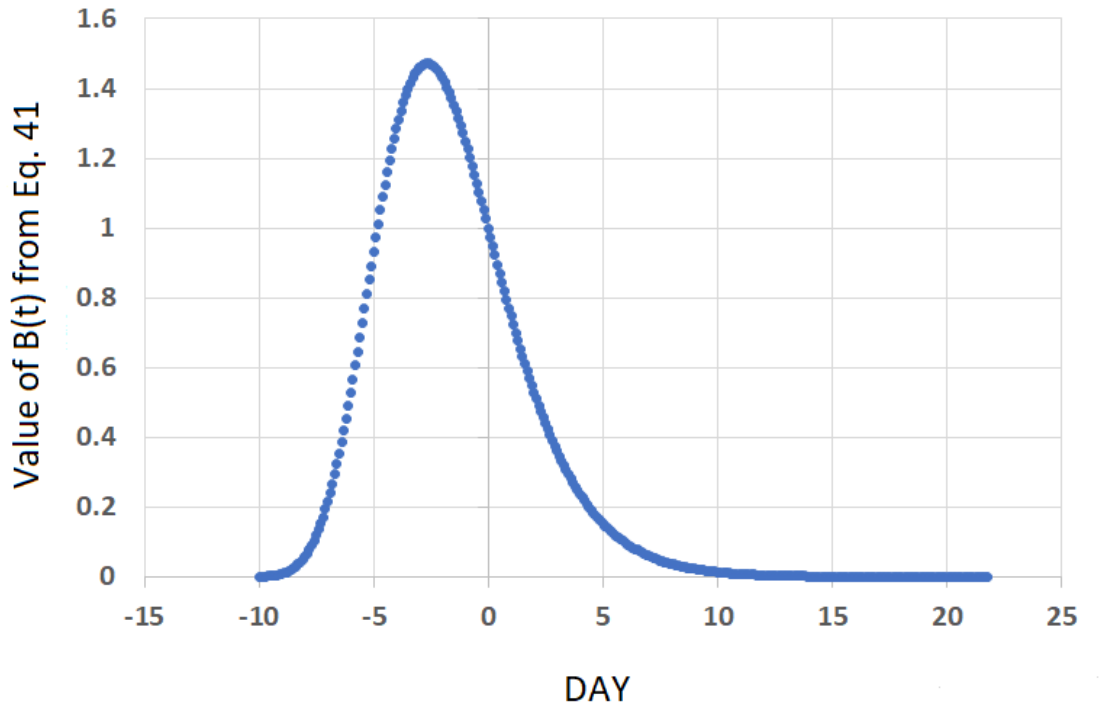

Figure 6. Representation of the average viral load of an infected person in the Covid 19 pandemic. The plot was generated using the values $K_T = 0.26$ and $P_c = 1$ in Equation 41

As we illustrate in the following paragraphs, there are many additional useful expressions that can be derived from the KMES. For simplicity, we assume that $ln(F_i(t_0)K_T)$ and $\frac{K_T}{P_c}$ are

constants in these derivations. We also assume that $F_i(t_0) = I(t_0) = N(t_0) = 1$ and $t_0 = 0$. These simplifying assumptions allow the nature of the expressions to be more easily seen and understood.

Using these assumptions, differentiating Equation 30, and dividing the result by $I(t)$, we obtain an expression for the number of new infections per infected person,

$$\frac{\frac{dI(t)}{dt}}{I(t)} = K_T e^{-\frac{K_T}{P_c}t} - \frac{K_T}{P_c}. \quad (42)$$

From Equation 42, we can see that the number of infections, $I(t)$, will begin to decrease when,

$$e^{-\frac{K_T}{P_c}t} - \frac{1}{P_c} < 0 \quad (43)$$

If we subtract 1 from both sides of Equation 43 and divide by the left side, we find an expression for the Effective Reproduction Number or R_{Eff}, and the criteria for when the epidemic will begin to decline:

$$R_{Eff} = \frac{1}{1 - e^{-\frac{K_T}{P_c}t} + \frac{1}{P_c}}. \quad (44)$$

This expression for R_{Eff} can also be derived by dividing K_T by $\psi(t)$. As seen from Equation 44, R_{Eff} is a function of both the disease and the behavior of the population. When $t = 0$, $R_{Eff} = Basic\ Reproduction\ Number = R_0 = P_c$, and an epidemic begins to decline when $R_{Eff} < 1$.

Using Equation 44, we obtain the following expression for the time when the decline begins:

$$t_{decline} = \frac{P_c \ln(P_c)}{K_T}. \quad (45)$$

Likewise, if we differentiate both sides of Equation 22, we can obtain an identical expression for the time when the change in the new infections is a maximum,

$$t_{max} = \frac{P_c \ln(P_c)}{K_T}, \qquad (46, \text{S3-5})$$

where t_{max} = the time to the peak of new infections. As it should, the time of the peak in new cases coincides with the start of the decline of infections.

Equation 46 demonstrates the relationship between the strength of social intervention measures, P_c, and the time to the peak of new infections. When social interventions are stronger (smaller P_c), the time to the peak will always be shorter.

Another important expression is the rate of acceleration of the epidemic:

$$\frac{d^2N(t)}{d^2t} = \left(K_T e^{-\frac{K_T}{P_c}t} - \frac{K_T}{P_c}\right)\frac{dN(t)}{dt} = K_T\left(\frac{I(t)}{N(t)} - \frac{1}{P_c}\right)\frac{dN(t)}{dt} = \left(e^{RCO(t)} - \frac{K_T}{P_c}\right)\frac{dN(t)}{dt} = K_T\frac{dI(t)}{dt}. \qquad (47)$$

Equation 47, with its four equivalent expressions, demonstrates the power that an authentic model provides. The leftmost expression allows us to compare the acceleration—the potential to change the rate of new infections—at any stage of the epidemic for any two countries, even those with different population densities, using only the daily case rate and the defining constants, K_T and P_c.

Equation 47 is also an immediate determinant of whether the control measures in place, represented by P_c, are effective enough. If the value of the term $\frac{I(t)}{N(t)} - \frac{1}{P_c}$ is positive, then the control measures are not strong enough. Conversely, when this term is negative, the epidemic is being brought under control.

The maximum value of P_c that will begin to bring down the new cases per day occurs when the acceleration is less than zero. If we set the left-hand side of Equation 47 to zero, use the third expression from the left and solve for P_c, we arrive at the defining relationship for this critical parameter of epidemic management:

$$P_c < K_T e^{-RCO(t)}. \tag{48}$$

Since we can easily determine the value of $RCO(t)$ every day during the epidemic and the value of K_T can be determined using the technique illustrated previously, the maximum allowable value of P_c needed to reduce the number of daily cases can always be determined. This value of P_c is the maximum level of infectable social contact allowable if we want the number of new daily cases to continue decreasing. Also, as explained in Supplement 5.1, if the value of $\frac{K_T(t)}{P_c(t)}$ is determined to be less than zero from the graphical analysis, then an outbreak has occurred and immediate reductions in social interactions are needed.

Yet another important relationship can be derived from Equation 22. In that equation, the term $\frac{dN(t)}{dt}$ is the rate of new cases and, in the figures, this is the new cases per day. If we define a desired target for the number of new cases per day at a future time, $t + t_{target}$, then we can derive a new quantity, the desired fraction of the current new cases, D_{tf}, as:

$$D_{tf} = \frac{\frac{dN(t)}{dt}}{\frac{dN\left(t+t_{target}\right)}{dt}}, \tag{49}$$

and using Equations 22 and 29, we arrive at the following expression:

$$D_{tf} = e^{-P_c\left(e^{-\frac{K_T}{P_c}(t_{target}+t)} - e^{-\frac{K_T}{P_c}t}\right)} e^{-\frac{K_T}{P_c}t_{\text{target}}}. \tag{50}$$

If $t \gg t_{target}$, then $e^{-\frac{K_T}{P_c}(t_{target}+\mathrm{t})} - e^{-\frac{K_T}{P_c}t} \approx 0$ and we can obtain the following equation from the remaining terms:

$$t_{target} = -\frac{P_c \ln(D_{tf})}{K_T}. \quad (51)$$

Equation 51 quantitates the number of days, t_{target}, that a level of social containment, P_c, must be imposed to achieve a fraction of daily cases, D_{tf}, compared to the current level.

In Supplement 5, using quantitative examples, we explain the use of Equations 44 through 51 to diagnose and control an epidemic.

Section 5: Comments on the SIR approximations

Due to the previous lack of a closed form solution, approximations to the full Kermack and McKendrick equations appear throughout the literature. Since, by their definition, approximations differ from the exact form, it is instructive to analyze these approximations with an eye to determining whether these behave qualitatively like the solution; and whether conclusions based on them are valid. The approximations have their roots in Kermack and McKendrick's 1927 paper where they proposed the following approximation,

$$\frac{dS(t)}{dt} = -\frac{\beta I(t) S(t)}{N_p}, \quad (52, \text{S3-1})$$

$$\frac{dI(t)}{dt} = \frac{\beta I(t) S(t)}{N_p} - \gamma I(t), \quad (53, \text{S3-2})$$

$$\frac{dR(t)}{dt} = \gamma I(t), \text{ and} \quad (54, \text{S3-3})$$

$$N_p = S(t) + I(t) + R(t), \quad (55, \text{S3-4})$$

where β = rate of contact and transmission, γ = rate of recoveries. It should also be noted that the basic reproduction number is, $R_0 = \frac{\beta}{\gamma}$ by definition.

These equations are the well-known "SIR" equations which can be derived from the full Kermack and McKendrick equations by assuming that the parameters $\varphi(t) = \varphi(\theta) = \frac{A_p \beta}{N_p}$ and $\psi(t) = \psi(\theta) = \gamma$ are constants. The SIR equations and their variants (SEIR, MSEIR, etc.) have been used for decades in attempts to quantitatively and qualitatively model epidemics. These models are known as compartmental models and they all share the common characteristic that the term $\frac{\beta I(t) S(t)}{N_p}$, appears in the equations for $S(t)$ and $I(t)$.

The full Kermack and McKendrick can also be written in the following form,

$$\frac{dS(t)}{dt} = -\mathrm{K}_T(t)\, I(t), \tag{56}$$

$$\frac{dI(t)}{dt} = K_T(t)\, I(t) - \psi(t)\, I(t), \tag{57}$$

$$\frac{dR(t)}{dt} = \psi(t)\, I(t) \text{ and} \tag{58}$$

$$S(t) = N_P - N(t), \tag{59}$$

Where $N(t) = I(t) + R(t)$; and $K_T(t)$ and $\psi(t)$ are defined as:

$$K_T(t) = \frac{\frac{dS(t)}{dt}}{I(t)} = \frac{-\mathrm{S(t)}(\int_0^t A(\theta) v(t-\theta) d\theta + A(t) I(0))}{A_P(\int_0^t B(\theta) v(t-\theta) d\theta + B(t) I(0))} \text{ and} \tag{13}$$

$$\psi(t) = \frac{\frac{dR(t)}{dt}}{I(t)} = \frac{A_P(\int_0^t C(\theta) v(t-\theta) d\theta + C(t) I(0))}{A_P(\int_0^t B(\theta) v(t-\theta) d\theta + B(t) I(0))}, \tag{60}$$

When we compare equations 52 to 54 to equations 56 through 58, we find that if $K_T(t)$ and $\psi(t)$ are approximated as the constants β and γ, the equations are not equivalent because the term $\frac{\beta I(t)S(t)}{N_p}$ does not appear in the approximations of equations 57 or 58. Since the SIR equations are putatively an approximation of the full Kermack and McKendrick integro-differential equations, this demonstrates that the methodology used to derive the SIR equations must be flawed. Therefore, we need to re-examine the logic Kermack and McKendrick used in developing the SIR equations.

A key statement comes early in the derivation of their integro-differential equations where Kermack and McKendrick write (1927, pg. 703), "Now $\mathrm{v}(t)$ denotes the number of persons in unit area who become infected at the interval t, and this must be equal to $\mathrm{x}(t)\sum_1^t \varphi(\theta)v(t,\theta)$ where $\mathrm{x}(t)$ denotes the people per unit area still unaffected, and $\varphi(\theta)$ is the rate of infectivity at age θ." In equation form, this statement reads,

$$\frac{\mathrm{v}(t)}{\Delta t} = \mathrm{x}(t)\sum_1^t \varphi(\theta)v(t,\theta) \qquad (61)$$

Later in their manuscript, Kermack and McKendrick used Equation 61 as the basis for deriving Equation 43 and 44 by assuming that $\varphi(\theta)$ equals the constant $\frac{\beta A_p}{N_p}$.

Although Kermack and McKendrick labeled $\varphi(\theta)$ as a "rate", $\varphi(\theta)$ in Equation 61 cannot merely be a rate with the units of $time^{-1}$ because $\mathrm{v}(t,\theta)$ and $\mathrm{v}(t)$ have the same units: $infected/area$ and $x(t)$ has the units of $susceptible/area$. A proper balancing of the dimensions leads to the conclusion that $\varphi(\theta)$ has the units of $time^{-1}(\frac{susceptible}{area})^{-1}$.

Hence, from the description given and our analysis in Supplement 1, it is clear that $\varphi(\theta)$ can be written as,

$$\varphi(\theta) = \frac{\varphi_r(\theta)}{x(t)}, \tag{62}$$

where $\varphi_r(\theta)$ is a rate of new infections, $\varphi(\theta)$ is the rate of new infections *per susceptible density*, and Equation 61 can be rewritten as,

$$\frac{\mathrm{v}(t)}{\Delta t} = \sum_1^t \varphi_r(\theta) v(t,\theta). \tag{63}$$

In Supplement 1 we demonstrate that $\varphi(\theta) = \varphi(t)$. Therefore, Equations 61 and 63 are merely restatements of Equation 14 and the units of $\varphi(t)$ are $time^{-1}(\frac{susceptible}{area})^{-1}$.

Since the parameter $\varphi(\theta)$ is a function of a rate of new infections, $\varphi_r(\theta)$, divided by the susceptible population density, $x(t)$, and $x(t)$ is an ever-decreasing quantity as the epidemic progresses, $\varphi(\theta)$ cannot itself be considered a constant as the epidemic progresses, especially as the affected population grows very large. Rather, the quantity, $\varphi_r(\theta)$ (which is equivalent to $K_T(t)$) is a better choice to be considered a constant when seeking to simplify the equations without losing their essence.

Other authors too, have misinterpreted and misused $\varphi(\theta)$. For example, in Hethcote (2000, pg 602), the term, $\frac{\beta I(t)S(t)}{N_p}$ is derived using the argument that "If β is the average number of adequate contacts (i.e., contacts sufficient for transmission) of a person per unit time, then $\frac{\beta \mathrm{S(t)I(t)}}{N_p} = \beta N_p s(t) i(t)$ is the number of new cases per unit time due to $S(t) = N_{\mathrm{P}} s(t)$ susceptibles." Notable in the statement is that β is defined as, "…the average number of

adequate contacts (i.e., contacts sufficient for transmission) of a person per unit time …". Because it defines β as the average number of adequate contacts per unit time, ***and per person***, this statement also implies that β is not merely a contact rate with the units of $contacts\ x\ time^{-1}$. Since Heathcote goes on to imply that the "person" is a susceptible, β must have the units $contacts\ x\ time^{-1} susceptible^{-1}$ and the conclusions that $\frac{\beta I(t) S(t)}{N_p} = K_T(\mathrm{t})\mathrm{I}(t)$ and $\beta = \frac{K_T(\mathrm{t}) N_p}{S(t)}$ immediately follow.

The case when $S(t) \approx N_p$, which is early in an epidemic, further illuminates the fundamental flaws in the SIR formulation. When $S(t) \approx N_p$, Equations 52 through 55 become,

$$\frac{dS(t)}{dt} = -\beta I(t), \tag{64}$$

$$\frac{dI(t)}{dt} = \beta I(t) - \gamma I(t), \tag{65}$$

$$\frac{dR(t)}{dt} = \gamma I(t), \text{ and} \tag{66}$$

$$N_p = S(t) + I(t) + R(t), \tag{67}$$

Inspection of Equation 65 makes it immediately clear that this set of equations will not adequately model epidemics. Equation 65 predicts a continual exponential increase in $I(t)$ at a constant rate of $\beta - \gamma$; and therefore, these equations predict a peak will never occur. However, since peaks do occur in epidemics while $S(t) \approx N_p$ and since β can be approximated as the constant K_T under these conditions, we conclude that γ cannot be adequately modelled as a constant while $S(t) \approx N_p$. This conclusion is supported by Equation 34 which shows that $\psi(t)$,

and therefore γ, cannot be a constant because the term, $e^{-\int_0^t \frac{K_T(a)}{P_C(a)} da}$, decays exponentially with time.

From the preceding observations, we conclude that all SIR constructs are inappropriate approximations. In Supplement 3, we support this conclusion by detailing the mathematically implausible assumptions that underlie the SIR approximations and make the case that these assumptions are not even qualitatively correct. We also analyze the inherent flaws within the SIR formulation that led to the erroneous "flatten-the-curve" projection by explicitly demonstrating that the flatten-the-curve projection illustrated throughout the literature (see, for example, Di Lauro et al 2021) is caused by hidden, inherent and implausible assumptions about both the populace and the disease.

Discussion

We began this manuscript with a simple observation: A fundamental tenet of modern epidemiology, the presumed existence of herd immunity, which states that the final size of an epidemic is always smaller than the total population, is premised on a circular argument. Prior authors (Breda, Brauer, Diekmann) contend that $S(t)$ cannot ever reach zero in a finite time because to do so would require $\varphi(t)$ to become infinite. Since the assertion that $\varphi(t)$ must remain finite depends on $S(t)$ remaining greater than zero, the resultant conclusion, that $S(t)$ must be greater than zero, is based on a circular argument because it essentially rests on the initial assertion that the conclusion is true. As we show with a simple dimensional analysis, the flaw in the analysis is the assertion that $\varphi(t)$ must remain finite, because this is a contradiction of the definition of $\varphi(t)$.

In a departure from the conventional analyses, we started with the same equations, but took an alternative path, used basic principles, and successfully derived a solution to the full Kermack and McKendrick equations, the KMES. The KMES accurately projects phenomenon which arose in the Covid epidemic, even under the simplifying assumption that $K_T(t)$ and $P_c(t)$ are constants for periods of time. These successful projections contrast sharply with SIR model projections which must violate the assumption that β is constant to produce any relevant results. Also, the high degree of correlation between the proxy of the population behavior in the KMES and the independently measured actual population mobility in the Google data contrasts with the weak correlations to the SIR construct of mobile phone mobility data found by prior authors (Wesolowski 2015). These accurate projections and the contrasts with the SIR approximations, strongly suggest that the KMES should replace the SIR models.

Our expression for final epidemic size, an important finding, disproves one of the accepted tenets of epidemiology. Equations 25 through 27, in their mathematical simplicity, state what is an intuitive conclusion: if the people contact each other infectiously at a high enough rate for a long enough time, herd immunity is not guaranteed, and an epidemic can spread to an entire population. This conclusion must become a new tenet of epidemiology

The KMES also has an intuitive form, exemplified by Equation 36:

$$I(t) = e^{\int_0^t K_T(t)dt} B(t) I(0) \qquad (36, \text{S1-34})$$

Equation 36 states that the input of infections, $B(t)I(0)$, is transformed into the time varying output of infections, $I(t)$, through an exponential step response function, $e^{\int_0^t K_T(t)dt}$. Thus, our analysis, leads us to an irrefutable mathematical statement of the epidemic dynamics.

As further support to the veracity of our solution, we note that Equations 23 and 29, are both Gompertz equations. This form is supported by Onishi et al (2021) who demonstrate that the Covid-19 epidemic time course in many countries was well fit by a Gompertz model. These authors do not offer a basic principles argument as to why this is so, but they demonstrate a strong correlation to our model structure.

With the availability of an analytical solution, we derived previously unknown, pragmatically useful expressions of important epidemiological relationships: Time course of the epidemic size, Final epidemic size, Time to peak infections, Effective Reproduction number, Viral load, and targets for reducing the epidemic along a planned path. In addition, in the supplements, we demonstrate how to detect and model outbreaks.

The analytical expressions are intuitive and sensible. For instance, the expression for time to maximum new cases, in Equations 45 or 46, passes smoothly through the epidemic peak. This contrasts with the expression for the time to the peak derived for the SIR models (Koger and Schlickeiser, 2020). In this reference, the expressions for the SIR models are only valid when $R_{Eff} > 1$, and they have the peculiar property that the time to the peak in new infections becomes increasingly larger as R_{Eff} approaches one before suddenly plunging to negative infinity just as R_{Eff} reaches one. This mathematically describes the claimed phenomena behind the concept of "flattening the curve", but it is unsettlingly nonintuitive. How can the peak in new infections move away from attainment as people interact less infectiously?

Equations 45 and 46 have none of this peculiar behavior and their behavior is supported by data from different countries which imposed very different containment strategies. As we explain in Supplement 3, when social containment is increased, the peak number of infections is much

lower, and it occurs earlier. The stronger the containment actions, the shorter the epidemic, as one would intuit.

The reason the conventional SIR models project epidemic phenomenon incorrectly is that beginning with Kermack and McKendrick, numerous authors have misunderstood the units of φ on the way to deriving the SIR equations. Even authors (Hethcote 2000), who derived the SIR equations without the use of Kermack and McKendrick's equations, replicate this mistake. Furthermore, $\psi(t)$ cannot be assumed constant early in the epidemic and $\varphi(t)$ cannot be assumed constant as the epidemic becomes large relative to the total population without imposing subtle, implausible assumptions on the model. Analytic expressions for these quantities, Equations S3-8 and S3-9, derived in Supplement 3, make this clear.

It should not be surprising that a solution to the epidemic equations produces an expression that can be interpreted as the viral load. We say that the righthand side of Equation 41, is a proxy to the viral load, because it is more accurately defined as a measure of the infectiousness of a person who is in infectious contact with only one other person. This expression surely is a measure of the viral load modified by some intermediate and unmeasured transmission impedance between the two people. With these caveats in mind, the curve in Figure 6, wholly derived from the Covid-19 data from several countries, certainly has the characteristics many authors have expected a viral load to have (Challenger et al 2022, Jones et al 2021). While these authors reached their conclusions through direct measurement of the viral load of thousands of patients, we have derived the same form using only the country case data. In retrospect, this is remarkable.

Since Equation 32 clearly shows that $B(t)$ is dependent on $K_T(t)$, a property of the disease and $P_C(t)$, a function of the population behavior, this, in turn, means that the time variations of $I(t)$

and $R(t)$ also depend on these two parameters. It is natural to assume that both $I(t)$ and $R(t)$ will depend on properties of the disease, but it may be somewhat surprising to see that their values also depend on the behavior of the population.

In supplement 1.1 we explain this dependency by showing that $I(t)$ is best interpreted as the total infectiousness within the infected population $N(t)$. As a complementary interpretation, $R(t)$ is best thought of as the degree of recovery from infectiousness within $N(t)$. Therefore, a previously infected individual is simultaneously a part of both the infected and recovered populations with the degree of membership determined by the parameter $\psi(t)$.

As time goes on, the degree of membership inevitably moves the infected individuals towards membership in the recovered community, but during this time, the infectiousness of all individuals vary as their viral load and number of contacts vary. An increase in social contact causes an increase in infectiousness, which, in turn, decreases the degree to which the person remains in the recovered population and vice versa. Therefore, as an individual's viral load changes, based on time and the disease dynamics, so too, does this individual's ability to infect others change based on their level of social interaction.

With this concept of variable membership in mind, then, we see that the idea of a compartmental model where people irreversibly move from being infected to recovered is an inadequate model construct. Rather, assuming immunity exists, the proper compartment construct is that there are only two compartments: 1) not yet infected, $S(t)$; and 2) previously infected, $N(t)$; and only from the latter of these is there no escape.

Concluding remarks

We recognize that the mathematics and resulting conclusions described in this manuscript disprove long-accepted tenets of epidemiology, the valid representation of epidemic dynamics by SIR models, and the guarantee of herd immunity. While the abandonment of these concepts is a difficult proposition, it is nevertheless, a necessary conclusion derived from our analysis. Fortunately, the replacement for these concepts, the KMES is consistent with data gathered from the Covid-19 pandemic and with basic epidemiological notions such as the Effective Reproduction number and viral load. The expressions are well behaved under all epidemic conditions and the KMES accurately predicts correlations among a variety of independent data sources.

In sum, this is a hopeful message to the epidemiological community. Logical, analytical tools are available to easily diagnose the state of an epidemic and provide guidance to public health officials. These tools clearly show that with stronger initial measures, an epidemic can be stopped more quickly with much less economic damage than predicted by conventional models. Although the disease will have its own dynamics, the overall epidemic dynamics can ultimately be controlled by the behavior of the population.

References

Brauer, F, The Kermack-McKendrick epidemic model revisited. Mathematical Biosciences, 198, August 2005, Pages 119-131

Breda D, Diekmann O, de Graaf W F, Pugliese A, Vermiglio R On the formulation of epidemic models (an appraisal of Kermack and McKendrick). Journal of Biological Dynamics, 2021, 6:sup2, 103-117, DOI: 10.1080/17513758.2012.716454

Campbell C. South Korea's health minister on how his country is beating coronavirus without a lockdown. TIME. https://time.com/5830594/south-korea-covid19-coronavirus/ (2020). Published April 30, 2020. Accessed February 9, 2021.

Carvalho, MA, Goncalves, S. An Analytical Solution for the Kermack-McKendrick Model, Physica A. 566 (2021) 125659

Challenger JD, Foo CY, Wu Y, Yan AWC, Marjaneh MM, Liew F, Thwaites RS, Okell LC, Cunnington AJ. Modelling upper respiratory viral load dynamics of SARS-CoV-2, BMC Medicine (2022) 20:25 https://doi.org/10.1186/s12916-021-02220-0

Di Lauro F, Berthouze L, Dorey MD, Miller JC, Kiss IZ. The Impact of Contact Structure and Mixing on Control Measures and Disease-Induced Herd Immunity in Epidemic Models: A Mean-Field Model Perspective, Bulletin of Mathematical Biology (2021) 83:117 https://doi.org/10.1007/s11538-021-00947-8

Diekmann O, Othmer H G, Planque R, Bootsma, M C. The discrete-time Kermack–McKendrick model: A versatile and computationally attractive framework for modeling epidemics. PNAS 2021 Vol. 118 No. 39 e2106332118. https://doi.org/10.1073/pnas.2106332118

Elassar A. This is where each state is during its phased reopening. CNN. https://edition.cnn.com/interactive/2020/us/states-reopen-coronavirus-trnd. Published May 27, 2020. Accessed February 9, 2021.

Field A. New Zealand isn't just flattening the curve. It's squashing it. The Washington Post. https://www.washingtonpost.com/world/asia_pacific/new-zealand-isnt-just-flattening-the-curve-its-squashing-it/2020/04/07/6cab3a4a-7822-11ea-a311-adb1344719a9_story.html. Published April 7, 2020. Accessed February 9, 2021.

Google community mobility reports. https://www.google.com/covid19/mobility/. Published December 24, 2020. Accessed December 24, 2020.

Hethcote HW. The Mathematics of Infectious Diseases. SIAM Review December 2000, Vol. 42 No. 4, 599-653

Jones TC, Biele G, Mühlemann B, Veith T, Schneider J, Beheim-Schwarzbach J, Bleicker T, Tesch J, Schmidt ML, Sander LE, Kurth F, Menzel P, Schwarzer R, Zuchowski M, Hofmann J, Krumbholz A, Stein A, Edelmann A, Corman VM, Drosten C. Estimating infectiousness throughout SARS-CoV-2 infection course, Science 9 July 2021 373, eabi5273, https://doi.org/10.1126/science.abi5273

Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics, Proc R Soc Lond A. 1927, 115(772):700–721. https://doi.org/10.1098/rspa.1927.0118

Kröger M, Schlickeiser R. Analytical solution of the SIR-model for the temporal evolution of epidemics. Part A: Time-independent reproduction factor. J Phys A Math Theor. 2020;53:505601

Ohnishi A, Namekawa Y, Fukui T. Universality in COVID-19 spread in view of the Gompertz function. Prog. Theor. Exp. Phys. 2020, 12:123J01. https://doi.org/10.1093/ptep/ptaa148

Orlowski EKW, Goldsmith DJA. Four months into the COVID-19 pandemic, Sweden's prized herd immunity is nowhere in sight. J R Soc Med. 2020, 113(8): 292–298. https://doi.org/10.1177/0141076820945282

Ritchie H, Roser M. Land use. Our world in data. https://ourworldindata.org/land-use Published September 2019. Accessed February 9, 2021.

Roser M, Ritchie H, Ortiz-Ospina E, et al. Coronavirus pandemic (COVID-19). Our World in Data. https://ourworldindata.org/coronavirus. Published February 2, 2021. Accessed February 2, 2021.

Sanfelici M. The Italian response to the COVID-19 crisis: Lessons learned and future direction in social development. Int J Commun Soc Dev. 2020, 2(2):191–210. https://doi.org/10.1177/2516602620936037

Schlickeiser R, Kröger M. Dark numbers and herd immunity of the first COVID-19 wave and future social interventions. Epidemiology Int J. 2020;4(4):000152.

Toda AA. Early draconian social distancing may be suboptimal for fighting the COVID-19 epidemic. VoxEU. https://voxeu.org/article/early-draconian-social-distancing-may-be-suboptimal-fighting-covid-19-epidemic. Published April 21, 2020. Accessed February 9, 2021.

Wesolowski A, Metcalf CJE, Eagle N, Kombich J, Grenfell BT, Bjørnstad ON, Lessler J, Tatemc AJ, Buckee CO Quantifying seasonal population fluxes driving rubella transmission dynamics using mobile phone data. PNAS 2015 Vol. 112 No. 35 11114-11119. www.pnas.org/cgi/doi/10.1073/pnas.1423542112

Worldometers. https://.worldometers.info. Published August 21, 2021. Accessed August 21, 2021

Supplement 1. The Solution to the Full Kermack and McKendrick Equations

In this supplement, we demonstrate that Equations 23, 24, 28, and 32 to 35 are a complete solution to the full Kermack and McKendrick equations. As a starting point, we write out the full Kermack and McKendrick equations as,

$$\frac{dS(t)}{dt} = -\frac{\mathrm{S(t)}}{A_p}\left(\int_0^t A(\theta)V(t-\theta)d\theta + A(t)I(0)\right), \qquad \text{(S1-1, 9)}$$

$$I(t) = \int_0^t B(\theta)V(t-\theta)d\theta + B(t)I(0)), \qquad \text{(S1-2, 10)}$$

$$\frac{dR(t)}{dt} = \int_0^t C(\theta)V(t-\theta)d\theta + C(t)I(0)), \qquad \text{(S1-3)}$$

$$N(t) = I(t) + R(t) \qquad \text{(S1-4, 11)}$$

Where $S(t)$ is the susceptible population, $I(t)$ is the infected population, $V(t-\theta)$ is the new infections at time = $t-\theta$, $B(\theta) = e^{-\int_0^\theta \psi(a)da}$, $A(\theta) = \varphi(\theta)B(\theta)$, and $C(\theta) = \psi(\theta)B(\theta)$. Kermack and McKendrick (1927, p. 703) defined $\varphi(\theta)$ as "the rate of infectivity at age θ" (page 703), and $\psi(\theta)$ as "the rate of removal" (page 703) of the infected population. A_p is the area that contains the population.

In sections 1 and 2 of the manuscript, we derive expressions for $N(t)$, $I(t)$, and $R(t)$ directly from Equations S1-1 and S1-2 employing the parameters $K_T(t)$ and $P_c(t)$. These expressions project the progression of an epidemic, but they are not a complete solution to the system of Equations S1-1 to S1-4 in terms of the parameters originally defined by Kermack and McKendrick. Therefore, in this supplement, we derive expressions for $B(t)$, $B(\theta)$, $\varphi(t)$, $\varphi(\theta)$, $\psi(t)$, and $\psi(\theta)$ in terms of $K_T(t)$, $P_c(t)$ and time. We accomplish this by first re-deriving the

expressions for $N(t), I(t)$, and $R(t)$, using the discrete forms of the equations. We do this using a vector notation to expand the expressions into the θ dimension, which then enables us to determine the proper forms of $B(t)$, $B(\theta)$, $\varphi(t)$, $\varphi(\theta)$, $\psi(t)$, and $\psi(\theta)$.

We begin by first defining our notation. We note that any of the variables $I(t)$, $N(t)$, $R(t)$ and their derivatives are the integrals or sums of these variables over θ. These can be expressed as summations of vectors over all θ. For example, $\frac{dN}{dt}(t)$ can be written as,

$$\frac{dN}{dt}(t) = \sum_{\theta=0}^{t} \frac{dN}{dt}(t,\ \theta) \qquad \text{(S1-5)}$$

In other words, when the variables are expressed as a function of time only, it is understood that this representation implicitly contains the sum of the vector elements in the θ-vector of the variable. The θ-vector of the variable is expressed as $I(t, \theta)$ and is the vector of that variable over θ at a specific time, t. The format for the vector notation is: $N(t, \theta) = \{N(t, 0); N(t, \Delta t); N(t, 2\Delta t) \ldots N(t, \theta - \Delta t); N(t, \theta)\}$. We also note that $\Delta\theta = \Delta t$ based on Kermack and McKendrick's definition of θ (Kermack and MKendrick 1927).

We now start with $\frac{dN(t)}{dt}$ in discrete vector form and derive the equations and solutions. We first note that at any time, t, $N(t, \theta) = I(t, \theta) + R(t, \theta)$ and the magnitude, N, of each θ group never changes size after it is formed. Therefore, $\Delta N(t, \theta) = 0$ when $\theta > 0$; and, using Kermack and McKendrick's concept of $\varphi(\theta)$, we can write the vector, $\frac{dN}{dt}(t, \theta)$ in discrete form as,

$$\Delta N(t, \theta) = \left\{ \frac{S(t-\Delta t)}{A_p} \sum_{\theta=0}^{t-\Delta t} \varphi(t - \Delta t, \theta) I(t - \Delta t, \theta) \Delta t; 0;\ 0; \ldots 0; 0 \right\} \qquad \text{(S1-6)}$$

With no loss of generality, we can then define a function, $K_T(t)$, where:

$$K_T(t-\Delta t)\sum_{\theta=0}^{t-\Delta t} I(t-\Delta t,\theta)\Delta t = \frac{S(t-\Delta t)}{A_p}\sum_{\theta=0}^{t-\Delta t}\varphi(t-\Delta t,\theta)I(t-\Delta t,\theta)\Delta t =$$

$$\Delta N(t,0) = \Delta N(t) = K_T(t-\Delta t)I(t-\Delta t)\Delta t \qquad \text{(S1-7)}$$

Equation S1-7 is the discrete form of Equation 14,

$$\frac{dN(t)}{dt} = K_T(t)I(t) \qquad \text{(S1-8, 14)}$$

Using the derivation in the body of the manuscript we arrive at the following expressions,

$$\frac{dN(t)}{dt} = K_T(t)N(t)F_i(0)e^{-\int_0^t \frac{K_T(t)}{P_c(t)}dt}. \qquad \text{(S1-9, 22)}$$

and,

$$N(t) = N(t_0)e^{F_i(t_0)\int_0^t K_T(t)e^{-\int_0^t \frac{K_T(a)}{P_c(a)}da}dt}. \qquad \text{(S1-10, 23)}$$

Solving Equation S1-8 for I(t) and substituting Equations S1-9 and S1-10 as appropriate, we find that:

$$I(t) = I(t_0)e^{F_i(t_0)\int_0^t K_T(t)e^{-\int_0^t \frac{K_T(a)}{P_c(a)}da}dt-\int_0^t \frac{K_T(t)}{P_c(t)}dt}, \qquad \text{(S1-11, 24)}$$

If we assume $N(0) = I(0)$, then $F_i(0) = 1$, and:

$$\frac{dI(t)}{dt} = I(0)(K_T(t)e^{-\int_0^t \frac{K_T(t)}{P_c(t)}dt} - \frac{K_T(t)}{P_c(t)})e^{\int_0^t K_T(t)e^{-\int_0^t \frac{K_T(a)}{P_c(a)}da}dt-\int_0^t \frac{K_T(t)}{P_c(t)}dt} \qquad \text{(S1-12)}$$

By combining Equations S1-9 and S1-12, we arrive at this expression for $\frac{dR(t)}{dt}$,

$$\frac{dR(t)}{dt} = I(0)(K_T(t) - K_T(t)e^{-\int_0^t \frac{K_T(t)}{P_c(t)}dt} + \frac{K_T(t)}{P_c(t)})e^{\int_0^t K_T(t)e^{-\int_0^t \frac{K_T(a)}{P_c(a)}da}dt-\int_0^t \frac{K_T(t)}{P_c(t)}dt} \qquad \text{(S1-13)}$$

Using these equations, we can now determine $\psi(\theta), \psi(t), B(\theta)$ and $B(t)$.

First, we start with a discrete formulation of $\frac{dI(t)}{dt}$ using Equation S1-12:

$$\Delta I(t+\Delta t) = \Delta t(K_T(t)e^{-\int_0^t \frac{K_T(t)}{P_c(t)}dt} - \frac{K_T(t)}{P_c(t)})I(t) \quad \text{(S1-14)}$$

From the prior defined notation, we know that:

$$\Delta I(t+\Delta t) = \sum_{\theta=0}^{t+\Delta t} \Delta I(t+\Delta t,\ \theta) \quad \text{(S1-15)}$$

We also know from Equation S1-7 that the new infections, $\Delta N(t+\Delta t, 0) = K_T(t)I(t)\Delta t = \Delta I(t+\Delta t, 0)$. Therefore, from Equation S1-15:

$$\Delta I(t+\Delta t) = K_T(t)I(t)\Delta t + \sum_{\theta=1}^{t+\Delta t} \Delta I(t+\Delta t,\ \theta) \quad \text{(S1-16)}$$

Note that θ begins at 1 in the summation in Equation S1-16 because $K_T(t)I(t)\Delta t$ is the term for when $\theta = 0$

We now write Equation S1-16 using the parameter $\psi(\theta)$,

$$\Delta I(t+\Delta t) = K_T(t)I(t)\Delta t - \sum_{\theta=0}^{t} \psi(t,\theta)I(t,\ \theta)\Delta t \quad \text{(S1-17)}$$

We also note that the vector, $\Delta I(t+\Delta t, \theta)$, can be written as:

$$\Delta I(t+\Delta t, \theta) = \{K_T(t)I(t)\Delta t;\ -\psi(t,0)I(t,0)\Delta t;\ -\psi(t,1)I(t,1)\Delta t;\ \ldots;\ -\psi(t,t)I(t,t)\Delta t\} \quad \text{(S1-18)}$$

If we now write Equation S1-14 in summation notation, it looks like this:

$$\Delta I(t+\Delta t) = \Delta t(K_T(t)e^{-\int_0^t \frac{K_T(t)}{P_c(t)}dt} - \frac{K_T(t)}{P_c(t)}) \sum_{\theta=0}^{t} I(t,\ \theta) \quad \text{(S1-19)}$$

Equating Equation S1-17 to Equation S1-19, applying full summation notation, and some algebra, we get:

$$(K_T(t) - K_T(t)e^{-\int_0^t \frac{K_T(t)}{P_c(t)}dt} + \frac{K_T(t)}{P_c(t)}) \sum_{\theta=0}^{t} I(t,\ \theta)\Delta t = \sum_{\theta=0}^{t} \psi(t,\theta) I(t,\ \theta)\Delta t \qquad \text{(S1-20)}$$

It is clear from Equation S1-20 that $\psi(t,\theta) = (K_T(t) - K_T(t)e^{-\int_0^t \frac{K_T(t)}{P_c(t)}dt} - \frac{K_T(t)}{P_c(t)})$ and it is only a function of time. In equation form:

$$\psi(\theta) = \psi(t) = K_T(t) - K_T(t)e^{-\int_0^t \frac{K_T(t)}{P_c(t)}dt} + \frac{K_T(t)}{P_c(t)} \qquad \text{(S1-21, 34)}$$

The forms of $B(\theta)$ and $B(t)$ follow from Equation S1-21,

$$B(\theta) = e^{-\int_0^\theta \psi(b)db} = e^{-(K_T(t) - K_T(t)e^{-\int_0^t \frac{K_T(a)}{P_c(a)}da} + \frac{K_T(t)}{P_c(t)})\theta} \qquad \text{(S1-22, 33)}$$

$$B(t) = e^{-\int_0^t \psi(b)db} = e^{-\int_0^t K_T(t)dt + \int_0^t K_T(t)e^{-\int_0^t \frac{K_T(a)}{P_c(a)}da}dt - \int_0^t \frac{K_T(t)}{P_c(t)}dt} \qquad \text{(S1-23, 32)}$$

It should also be noted that based on the preceding, this is the form of the $\Delta R(t,\theta)$ vector:

$$\Delta R(t,\theta) = \{0;\ \psi(t-\Delta t)I(t-\Delta t,0)\Delta t;\ \psi(t-\Delta t)I(t-\Delta t,1)\Delta t;\ \ldots;\ \psi(t-\Delta t)I(t-\Delta t, t-\Delta t)\Delta t\} \qquad \text{(S1-24)}$$

We will now find the expressions for $\varphi(\theta)$ and $\varphi(t)$ using the first two portions of Equation S1-7,

$$K_T(t-\Delta t) \sum_{\theta=0}^{t-\Delta t} I(t-\Delta t,\theta)\Delta t = \frac{S(t-\Delta t)}{A_p} \sum_{\theta=0}^{t-\Delta t} \varphi(t-\Delta t,\theta) I(t-\Delta t,\theta)\Delta t \qquad \text{(S1-25)}$$

Expanding the expression in the righthand side summation, we obtain the following vector,

$$\sum_{\theta=0}^{t-\Delta t} \varphi(t-\Delta t,\theta)I(t-\Delta t,\theta)\Delta t = \{\varphi(t-\Delta t,0)I(t-\Delta t,0)\Delta t;\ \varphi(t-\Delta t,\Delta t)I(t-\Delta t,\Delta t)\Delta t;\ \ldots;\ \varphi(t-\Delta t,\theta)I(t-\Delta t,\theta)\Delta t\} \quad \text{(S1-26)}$$

In Equation S1-26, each $\varphi(t-\Delta t,\theta)$ is the number of new infections caused by the number of people infected within each θ group during the time interval from $t-\Delta t$ to t. If we denote the inherent disease transmissibility per contact in each θ group as $\frac{K_T(t-\Delta t)}{P_c(t-\Delta t,\theta)}$, then we can write,

$$\varphi(t-\Delta t,\theta) = \frac{K_T(t-\Delta t)}{P_c(t-\Delta t,\theta)} \frac{P_c(t-\Delta t,\theta)A_p}{S(t-\Delta t,\theta)} \quad \text{(S1-27)}$$

where the term $\frac{P_c(t-\Delta t,\theta)A_p}{S(t-\Delta t,\theta)}$ has been inserted into the expression to denote the number of infectious contacts that occur per member of the susceptible population density. We assume that the disease transmissibility per infectious contact is the same for a given time, and the contact parameter $P_c(t)$ is uniform throughout the population. Since time is held constant along the θ vectors, all the terms on the righthand side of Equation S1-27 are constant over all θ at a given time. Therefore, the function $\varphi(t-\Delta t,\theta)$ is also constant for all θ at a given time and its value is $\frac{K_T(t-\Delta t)A_p}{S(t-\Delta t,\theta)}$. Extending this argument for all time, we arrive at the expression,

$$\varphi(\theta) = \varphi(t) = \frac{K_T(t)A_p}{\mathrm{S}(t)} \quad \text{(S1-28, 35)}$$

The Kermack and McKendrick model in discrete form can also be shown using a matrix notation where the matrices of the variables $N(t,\theta), I(t,\theta), R(t,\theta)$ and their derivatives are defined over t and θ according to this map,

$$\begin{vmatrix} (t,0) & (t,\Delta t) & (t,2\Delta t) & \cdots & (t,\theta-\Delta t) & (t,\theta) \\ (t-\Delta t,0) & (t-\Delta t,\Delta t) & (t-\Delta t,2\Delta t) & \cdots & (t-\Delta t,\theta-\Delta t) & (t-\Delta t,\theta) \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ (3\Delta t,0) & (3\Delta t,\Delta t) & (3\Delta t,2\Delta t) & \cdots & (3\Delta t,\theta-\Delta t) & (3\Delta t,\theta) \\ (2\Delta t,0) & (2\Delta t,\Delta t) & (2\Delta t,2\Delta t) & \cdots & (2\Delta t,\theta-\Delta t) & (2\Delta t,\theta) \\ (\Delta t,0) & (\Delta t,\Delta t) & (\Delta t,2\Delta t) & \cdots & (\Delta t,\theta-\Delta t) & (\Delta t,\theta) \\ (0,0) & (0,\Delta t) & (0,2\Delta t) & \cdots & (0,\theta-\Delta t) & (0,\theta) \end{vmatrix}$$

This notation uses the vector notation previously defined for vectors in θ where each row of the matrix is a vector in θ at a different time.

Keeping this convention, we can use Equation S1-7 and the knowledge that $\Delta N(t,\theta) = 0$ when $\theta > 0$ to write the matrix for $\Delta N(t,\theta)$,

$$\Delta N(t,\theta) = \begin{vmatrix} K_T(t-\Delta t)\sum_{\theta=0}^{t-\Delta t} I(t-\Delta t,\theta)\Delta t & 0 & 0 & \cdots & 0 & 0 \\ K_T(t-2\Delta t)\sum_{\theta=0}^{t-\Delta t} I(t-2\Delta t,\theta)\Delta t & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ K_T(2\Delta t)\sum_{\theta=0}^{t-\Delta t} I(2\Delta t,\theta)\Delta t & 0 & 0 & \cdots & 0 & 0 \\ K_T(\Delta t)\sum_{\theta=0}^{t-\Delta t} I(\Delta t,\theta)\Delta t & 0 & 0 & \cdots & 0 & 0 \\ K_T(0)\sum_{\theta=0}^{t-\Delta t} I(0,\theta)\Delta t & 0 & 0 & \cdots & 0 & 0 \\ N(0) & 0 & 0 & \cdots & 0 & 0 \end{vmatrix}$$

We can also write the following matrices for $\Delta R(t,\theta)$, $\Delta I(t,\theta)$, and $I(t,\theta)$

$$\Delta R(t,\theta)=\begin{vmatrix} 0 & \psi(t-\Delta t)I(t-\Delta t,0)\Delta t & \psi(t-\Delta t)I(t-\Delta t,\Delta t)\Delta t & \cdots & \psi(t-\Delta t)I(t-\Delta t,\theta-2\Delta t)\Delta t & \psi(t-\Delta t)I(t-\Delta t,\theta-\Delta t)\Delta t \\ 0 & \psi(t-2\Delta t)I(t-2\Delta t,0)\Delta t & \psi(t-2\Delta t)I(t-2\Delta t,\Delta t)\Delta t & \cdots & \psi(t-2\Delta t)I(t-2\Delta t,\theta-2\Delta t)\Delta t & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & \psi(2\Delta t)I(2\Delta t,0)\Delta t & \psi(2\Delta t)I(2\Delta t,\Delta t)\Delta t & \cdots & 0 & 0 \\ 0 & \psi(\Delta t)I(\Delta t,0)\Delta t & \psi(\Delta t)I(\Delta t,\Delta t)\Delta t & \cdots & 0 & 0 \\ 0 & \psi(0)I(0,0)\Delta t & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{vmatrix}$$

$$\Delta I(t,\theta)=\begin{vmatrix} K_T(t-\Delta t)\sum_{\theta=0}^{t-\Delta t} I(t-\Delta t,\theta)\Delta t & -\psi(t-\Delta t)I(t-\Delta t,0)\Delta t & -\psi(t-\Delta t)I(t-\Delta t,\Delta t)\Delta t & \cdots & -\psi(t-\Delta t)I(t-\Delta t,\theta-2\Delta t)\Delta t & -\psi(t-\Delta t)I(t-\Delta t,\theta-\Delta t)\Delta t \\ K_T(t-2\Delta t)\sum_{\theta=0}^{t-\Delta t} I(t-2\Delta t,\theta)\Delta t & -\psi(t-2\Delta t)I(t-2\Delta t,0)\Delta t & -\psi(t-2\Delta t)I(t-2\Delta t,\Delta t)\Delta t & \cdots & -\psi(t-2\Delta t)I(t-2\Delta t,\theta-2\Delta t)\Delta t & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ K_T(2\Delta t)\sum_{\theta=0}^{t-\Delta t} I(2\Delta t,\theta)\Delta t & -\psi(2\Delta t)I(2\Delta t,0)\Delta t & -\psi(2\Delta t)I(2\Delta t,\Delta t)\Delta t & \cdots & 0 & 0 \\ K_T(\Delta t)\sum_{\theta=0}^{t-\Delta t} I(\Delta t,\theta)\Delta t & -\psi(\Delta t)I(\Delta t,0)\Delta t & -\psi(\Delta t)I(\Delta t,\Delta t)\Delta t & \cdots & 0 & 0 \\ K_T(0)\sum_{\theta=0}^{t-\Delta t} I(0,\theta)\Delta t & -\psi(0)I(0,0)\Delta t & 0 & \cdots & 0 & 0 \\ N(0) & 0 & 0 & \cdots & 0 & 0 \end{vmatrix}$$

$$
I(t,\theta)=\begin{vmatrix}
K_T(t-\Delta t)\sum_{\theta=0}^{t-\Delta t} I(t-\Delta t,\theta)\Delta t & I(t\text{-}\Delta t,0)(1\text{-}\psi(t\text{-}\Delta t)\Delta t) & I(t\text{-}\Delta t,\Delta t)(1\text{-}\psi(t\text{-}\Delta t)\Delta t) & \cdots & I(t\text{-}\Delta t,\theta\text{-}2\Delta t)(1\text{-}\psi(t\text{-}\Delta t)\Delta t) & I(t\text{-}\Delta t,\theta\text{-}\Delta t)(1\text{-}\psi(t\text{-}\Delta t)\Delta t) \\
K_T(t-2\Delta t)\sum_{\theta=0}^{t-\Delta t} I(t-2\Delta t,\theta)\Delta t & I(t\text{-}2\Delta t,0)(1\text{-}\psi(t\text{-}2\Delta t)\Delta t) & I(t\text{-}2\Delta t,\Delta t)(1\text{-}\psi(t\text{-}2\Delta t)\Delta t) & \cdots & I(t\text{-}2\Delta t,\theta\text{-}2\Delta t)(1\text{-}\psi(t\text{-}2\Delta t)\Delta t) & 0 \\
\vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\
K_T(2\Delta t)\sum_{\theta=0}^{t-\Delta t} I(2\Delta t,\theta)\Delta t & I(2\Delta t,0)(1\text{-}\psi(2\Delta t)\Delta t) & I(2\Delta t,\Delta t)(1\text{-}\psi(2\Delta t)\Delta t) & \cdots & 0 & 0 \\
K_T(\Delta t)\sum_{\theta=0}^{t-\Delta t} I(\Delta t,\theta)\Delta t & I(\Delta t,0)(1\text{-}\psi(\Delta t)\Delta t) & I(\Delta t,\Delta t)(1\text{-}\psi(\Delta t)\Delta t) & \cdots & 0 & 0 \\
K_T(0)\sum_{\theta=0}^{t-\Delta t} I(0,\theta)\Delta t & I(0,0)(1\text{-}\psi(0)\Delta t) & 0 & \cdots & 0 & 0 \\
N(0) & 0 & 0 & \cdots & 0 & 0
\end{vmatrix}
$$

It is clear from this notation that $\Delta I(t,\theta) = \Delta N(t,\theta) - \Delta R(t,\theta)$. It is also interesting to note that $\frac{\Delta N(t,\theta)}{I(t)}$ and $\frac{\Delta R(t,\theta)}{I(t)}$ are, respectively, impulse and step functions in θ.

We now use the preceding to demonstrate that Equations S1-10 and S1-11, along with expressions S1-21, S1-22, S1-23 and S1-28 form a complete solution to the Kermack and McKendirck Equations S1-1 to S1-4. We first note that in Kermack and McKendrick's formulation, they separated the initial infections from the consequent infections using the time varying terms, $A(t)I(0))$, $B(t)I(0)$, and $C(t)I(0)$ and we see from the preceding matrices that these terms are merely the terms along the diagonal of the matrices. Therefore, with no loss of generality, we recast Equation S1-2 in the following form,

$$I(t) = \int_0^t B(\theta)I(t-\theta)d\theta, \quad \text{(S1-29)}$$

noting that the term $I(t-\theta)$ is the value in the $(t-\theta, 0)$ place in the $I(t,\theta)$ matrix and $I(0,0) = N(0) = I(0)$ as shown in the matrix. Our goal is to now find the solution to S1-29.

Referring to the $I(t,\theta)$ matrix and keeping in mind that $\Delta t = \Delta\theta$, we know that

$$I(t) = \sum_{\theta=0}^{t} I(t,\ \theta) = \sum_{\theta=0}^{t} B(\theta)I(t-\theta, 0) \quad \text{(S1-30)}$$

Equation S1-30 is merely the summation form of S1-29. We use the reference to the $I(t,\theta)$ matrix to show where it comes from in the (t,θ) matrices.

It is also clear from the $I(t,\theta)$ matrix that,

$$I(t) = K_T(t-\Delta t)\Delta t \sum_{\theta=0}^{t-\Delta t} I(t-\Delta t,\ \theta) + (1-\psi(t-\Delta t)\Delta t) \sum_{\theta=0}^{t-\Delta t} I(t-\Delta t,\ \theta) \quad \text{(S1-31)}$$

This operation can be repeated all the way back through the matrix to finally obtain the expression,

$$I(t) = (K_T(t-\Delta t)\Delta t + 1 - \psi(t-\Delta t)\Delta t)(K_T(t-2\Delta t)\Delta t + 1 - \psi(t-2\Delta t)\Delta t) \ldots (K_T(0)\Delta t + 1 - \psi(0)\Delta t)I(0) \quad \text{(S1-32)}$$

Equating S1-32 with S1-30 and taking the limit as $\Delta t = \Delta\theta$ go to zero, we arrive at the expression,

$$I(t) = \int_0^t B(\theta)I(t-\theta)d\theta = e^{\int_0^t (K_T(t)-\psi(t))dt} I(0) = e^{\int_0^t K_T(t)dt} B(t)I(0) \quad \text{(S1-33)}$$

Since the right-hand terms of Equation S1-33 are all known from expressions S1-21 and S1-28, Equation S1-33 is a solution to Equation S1-2, and Equations S1-1 and S1-3 can be easily solved using Equation 33 to obtain the solutions depicted in equations S1-10 and S1-13. Because Equation S1-33 was derived directly from the Kermack and McKendrick equations, this

demonstrates that Equations S1-10, S1-11, and S1-13 along with the expressions in Equations S1-21, S1-22, S1-23, and S1-28 are a solution to the Kermack and McKendrick equations.

Equation S1-33 also provides a convenient shorthand for writing the entire solution as,

$$I(t) = e^{\int_0^t K_T(t)dt} B(t) I(0) \qquad \text{(S1-34, 36)}$$

$$N(t) = e^{\int_0^t (K_T(t) + \frac{K_T(t)}{P_C(t)})dt} B(t) I(0) \qquad \text{(S1-35, 37)}$$

$$R(t) = (e^{\int_0^t (K_T(t) + \frac{K_T(t)}{P_C(t)})dt} - e^{\int_0^t K_T(t)dt}) B(t) I(0) \qquad \text{(S1-36, 38)}$$

These expressions also give us intuitive insight into the epidemic dynamics. Since $B(t)$ is the time varying infectiousness of the original infected group, $I(0)$, the exponential expressions, $e^{\int_0^t K_T(t)dt}$, $e^{\int_0^t (K_T(t) + \frac{K_T(t)}{P_C(t)})dt}$, and $e^{\int_0^t (K_T(t) + \frac{K_T(t)}{P_C(t)})dt} - e^{\int_0^t K_T(t)dt}$, are the step response functions to the input $B(t)I(0)$. If the function $B(t)I(0)$ equaled 1 for all time, that is, if there was no recovery, these expressions demonstrate that the epidemic would proceed exponentially until the entire population was infected, an intuitive result. Also, since $B(t)$ can only equal 1 if $P_C = \infty$, this case implies that $I(t) = N(t)$, which is an additional intuitive result; and provides a check on the formulation.

Lastly, since $B(t)I(0) = I(0)$ is the upper limit of the function and it implies that the entire population would become infected, it stands to reason that there is a value of $B(t)I(0)$ below which entire population will not become infected. This limit is defined by the expression,

$$B(t)\mathrm{I}(0) = e^{-\int_0^t (K_T(t) + \frac{K_T(t)}{P_C(t)})dt} S(0) \qquad \text{(S1-37)}$$

As Equation S1-37 clearly shows, preventing the entire population from becoming infected depends upon both the transmission characteristics of the disease, K_T, and the behavior of the population, P_C. The final size is not an inherent property of the disease alone.

Supplement 1.1 Insights developed during the solution derivation

As we derived the KMES, we did not stop to discuss insights provided by some of the important expressions. In this section, we will provide those insights.

The first expression to understand is the relationship described by the derivative of Equation 21:

$$\frac{dF_i(t)}{dt} = \frac{d\left(\frac{I(t)}{N(t)}\right)}{dt} = -F_i(t)\frac{K_{\mathrm{T(t)}}}{P_c(t)}. \tag{S1-38}$$

This seemingly simple expression is a fundamental statement of an infectious epidemic.

If we evaluate the derivative $\frac{d\left(\frac{I(t)}{N(t)}\right)}{dt}$ in Equation S1-38, we obtain:

$$\frac{d\left(\frac{I(t)}{N(t)}\right)}{dt} = \frac{d(I(t)}{dt}\frac{1}{N(t)} - \frac{dN(t)}{dt}\frac{I(t)}{N(t)^2} = -\frac{I(t)}{N(t)}\frac{K_{\mathrm{T}}(t)}{P_c(t)} \tag{S1-39}$$

With some rearrangement of the terms, we find the following expression:

$$\frac{d\left(\frac{I(t)}{N(t)}\right)}{dt} = \frac{\frac{dI(t)}{dt} - \frac{dN(t)}{dt}\frac{I(t)}{N(t)}}{N(t)} = \frac{-I(t)\frac{K_{\mathrm{T(t)}}}{P_c(t)}}{N(t)} \tag{S1-40}$$

In Equation S1-40, the expression $\frac{d(I(t)}{dt} - \frac{dN(t)}{dt}\frac{I(t)}{N(t)}$ describes how the ratio $\frac{I(t)}{N(t)}$ changes with time as a function of the changes in $I(t)$ and $N(t)$. This expression also makes intuitive sense because if $\frac{d(I(t)}{dt} = \frac{dN(t)}{dt}$ then $\frac{d\left(\frac{I(t)}{N(t)}\right)}{dt} = \frac{\frac{dI(t)}{dt}\left(1-\frac{I(t)}{N(t)}\right)}{N(t)} = \frac{\frac{d(I(t)R(t)}{dt}}{N(t)}$. This means that the size of the

infected populations, $I(t)$, in the ratio $\frac{I(t)}{N(t)}$ only changes to the extent that the change in infections affects the ratio $\frac{R(t)}{N(t)}$.

The more interesting case is when $I(t) = N(t)$ and $\frac{d(I(t)}{dt} < \frac{dN(t)}{dt}$. This case, of course, occurs when $t = 0$. Setting $I(t) = N(t)$ and $t = 0$ in Equation S1-40 and recognizing that $\frac{dN(t)}{dt} - \frac{dI(t)}{dt} = \frac{dR(t)}{dt}$ we obtain the following:

$$\frac{dR(0)}{dt} = I(0)\frac{K_{\mathrm{T}}(0)}{P_c(0)} \tag{S1-41}$$

Equation S1-41 forces us to reach a startling conclusion: The recovered population begins to grow the instant the epidemic starts!

Further reflection on Equation S1-41 offers deeper insight. What Equation S1-41 is telling us is that the infectiousness of the individuals in the population $I(0)$ (and by extension $I(t)$) is not a constant while they are infected. This is immediately obvious in retrospect because, as the viral load changes, the infectiousness of a person changes. What Equation S1-40 is telling us is that the infectiousness of population $I(0)$ is changing at the rate $\frac{K_{\mathrm{T}}(0)}{P_c(0)}$. We have already seen that K_{T} is a measure of the disease infectiousness, but a reasonable question to ask is: Why should the infectiousness change be a function of the population behavior, P_{c}?

The answer to this question is quite straightforward. The infectiousness of a person is not just dependent on the viral load, it is also dependent on the contacts a person has with other, not-yet-infected people. After all, if an infected person never contacts another noninfected person, they are never truly infectious in the sense that they cannot advance the disease. Thus, Equation S1-

40 casts an important light on the meaning of $I(t)$: $I(t)$ is the infectiousness of the infected subpopulation within the larger population, $N(t)$, subject to the disease transmissibility, K_T, and the population behavior, P_c. We can describe $R(t)$ in a similar manner: $R(t)$ is the reduction in the total viral load that has occurred in the population, $N(t)$, and this is also subject to the disease transmissibility and the population behavior.

We gain further insight into the meaning of the KMES by also looking at Equation S1-12 in more detail. Equation S1-11 can be used to rewrite Equation S1-12 as

$$\frac{1}{I(t)}\frac{dI(t)}{dt} = K_T(t)\frac{I(t)}{N(t)} - \frac{K_T(t)}{P_c(t)} = K_T(t) - \frac{R(t)}{N(t)}K_T(t) - \frac{K_T(t)}{P_c(t)}. \quad \text{(S1-42)}$$

The left-hand side of Equation S1-42 is the rate of change in the number of new infections per person currently infected. The first two terms on the furthest right-hand side of Equation S1-42, $K_T(t) - \frac{R(t)}{N(t)}K_T(t)$, describe the net rate of successful infection. Since $K_T(t)$ is the rate at which an infected person causes infections per infectable contact, the terms $-\frac{R(t)}{N(t)}K_T(t) - \frac{K_T(t)}{P_c(t)}$ must represent the rate of recovery per infected person.

We gain an additional intuitive insight about the solution from the following relationship, derived from Equations 22 and 29:

$$\frac{dN}{dt} = N_\infty N_\infty{}^{-e^{-\frac{K_T}{P_c}t}} K_T e^{-\frac{K_T}{P_c}t}. \quad \text{(S1-43)}$$

In words, the form of Equation S1-43 is

$$Rate\ of\ change\ in\ cases = Population\ that\ will\ ultimately\ be\ infected \times Probability\ of\ infection \times Rate\ of\ transmission \times\ Fraction\ of\ cases\ still\ infected \quad \text{(S1-44)}$$

or

$$Rate\ of\ change\ in\ cases = Number\ of\ cases \times Rate\ of\ transmission \times Fraction\ of\ cases\ still\ infected \quad \text{(S1-45)}$$

Equations S1-44 and S1-45 illustrate the logic of the solution in terms of probabilities.

Finally, we can use Equation 21 to write this simple expression for the solution for total cases if $K_T(t)$ and $P_c(t)$ remain constant:

$$N(t) = N_\infty^{(1-F_i(t))}. \quad \text{(S1-46)}$$

Supplement 2. Verification of the solution

To demonstrate that $K_T(t)$ is indeed a constant, we need to first further refine the concept of $P_c(t)$. As previously defined, $P_c(t)$ is the number of specific infectious contacts a member of subpopulation *N*(t) has across the entire population. This is a function of the population's behavior. Initially, we assume this is a function of population density and further, that people's mobility extends over a constant average effective area per unit of time. We define this area as the effective area rate, $A_{1r}(t)$. Using these definitions, we can write an expression for $P_{cr}(t)$:

$$P_{cr}(\mathrm{t}) = \frac{A_{1r}(t)N_p}{A_p} = contact\ rate, \quad \text{(S2-1)}$$

where N_p = the entire population of the region with the infection, = the area of the region, and $\frac{N_p}{A_p}$ = the population density. From Equation S2-1, we can see that $P_{cr}(t)$ is proportional to both the population's behavior, $A_{1r}(t)$, and the population density, $\frac{N_p}{A_p}$.

Similar to the way we defined $P_c(t)$ using $P_{cr}(t)$, we now define a quantity, $A_1(t)$, in terms of $A_{1r}(t)$:

$$A_1(t) = \lim_{\Delta t \to 0} \int_t^{\Delta t} A_{1r}(t)dt, \qquad \text{(S2-2)}$$

where $A_1(t)$ is the effective specific area traversed by an individual. In this case, "specific" has the same meaning as it has for $P(t)_c$; that is, each person traverses the same area for the duration of the time under consideration. We also call $A_1(t)$ the "effective area" because the population is typically only dispersed within ~1% of the land within a given region of a country (Ritchie and Roser 2019). If we take this into account, then $A_1(t) = \frac{Area\ actually\ traversed\ by\ a\ person}{0.01}$.

From the preceding discussion, we can now write an expression for $\frac{K_T(t)}{P_c(t)}$:

$$\frac{K_T(t)}{P_c(t)} = \frac{K_T(t)A_p}{A_1(t)N_p}. \qquad \text{(S2-3)}$$

As defined, neither $P_c(t)$ nor $A_1(t)$ are rates, but they can both vary in time and their values depend on the population's behavior. They are, respectively, the number of specific infectious people who have been contacted and the specific effective area traversed by any index person within a given time interval. $P_c(t)$ and $A_1(t)$ are constant when the number of specific infectious people or the traversed area remain constant. However, if different people are contacted within a given time interval, the rates they depend on change, and therefore, $P_c(t)$ and $A_1(t)$ may change even if the total number of people contacted or area covered did not change during that time interval.

We can now check the assumption that $K_T(t)$ is a constant by substituting Equation S2-3 into Equation 29 and solving for $K_T(t)$t. Doing this, we find the following expression:

$$\frac{A_1 N_p}{A}\ln\left(1+\frac{\ln(N(t))}{-\frac{A_1 N_p}{A}}\right)=-K_T(t)\mathrm{t}. \quad \text{(S2-4)}$$

If we define $F\big(N(t)\big)=\frac{A_1 N_p}{A}\ln\left(1+\frac{\ln(N(t))}{-\frac{A_1 N_p}{A}}\right)$, then we can also write this expression as

$$F\big(N(t)\big)=-K_T(t)\mathrm{t}. \quad \text{(S2-5)}$$

If $K_T(t)$ is a constant, then Equation S2-5 predicts that $F\big(N(t)\big)$ is a linear function of time. Excepting A_1, all the quantities on the left-hand side of Equation S2-4 can be found for each country in the time before containment measures were enacted; and these are listed for the sample of countries addressed in this paper, in Table 2. An implicit assumption in this process is that the behavior of the population, $P_c(t)$, is constant and therefore A_1 is constant at least during the initial phase of the epidemic before containment measures were put in place. Because of this linkage, it is necessary to frame the problem as co-determining a value of A_1 which produces a straight line for the country data; and separately determining whether the slope of that line is a rational value for $K_T(t)$. Through a process of iteration, a value was found for $A_1(=0.48\ \mathrm{km}^2)$ which created a straight line with a correlation coefficient of 0.96 (Figure 4). Using Equation S2-4 (or S2-5) we then determined the slope of the line in Figure 4 indicating the value of $K_T(t)$ as 0.26. This value is completely consistent with the country data; and we take this analysis as strong support for the plausibility that $K_T(t)$ is a constant and represents the transmissibility of the disease in the early stage of the epidemic.

Independent evidence that $P_c(t)$ is a measure of the population behavior was developed by first using Equation 39 to show that, if $K_T(t)$ is a constant, the RCO measure will be proportional to $K_T\int_{t_0}^{t}\frac{1}{P_c(t)}dt$. We reasoned that if an independent measure of people's mobility during the

epidemic could be found and was linearly related to the RCO, we could have additional confidence in the veracity of the KMES.

Google has compiled different measures, derived from mobile phone data, of people's mobility (Google 2020). One of these measures is termed the Residential Mobility Measure (RMM). The RMM is a measure of the percentage change in the degree to which people stayed in their residence during the pandemic relative to a baseline measured over 5 weeks starting on January 3, 2020. Since $\frac{1}{P_c(t)}$ and the RMM are both inversely proportional to the population's mobility, we hypothesized that the RMM would be a good proxy for the value of $\frac{1}{P_c(t)}$. To test this, we plotted the integral over time of the daily RMM for the six countries whose data we analyzed, against the daily RCO. These plots appear in Figure 5, which clearly validates the hypothesized linear relationship.

Supplement 3. An analysis of the SIR model

The SIR model, with non-time-varying parameters β and γ, is described by the following equations:

$$\frac{dS(t)}{dt} = -\frac{\beta I(t) S(t)}{N_p}, \quad \text{(S3-1, 52)}$$

$$\frac{dI(t)}{dt} = \frac{\beta I(t) S(t)}{N_p} - \gamma I(t), \quad \text{(S3-2, 53)}$$

$$\frac{dR(t)}{dt} = \gamma I(t), \text{ and} \quad \text{(S3-3, 54)}$$

$$N_p = S(t) + I(t) + R(t), \quad \text{(S3-4, 55)}$$

where N_p = total number of people in the population, β = rate of contact and transmission, and γ = rate of recoveries. These equations can be derived from Equations S1-1 to S1-4 by assuming that the parameters $\varphi(t) = \varphi(\theta) = \frac{A_p \beta}{N_p}$ and $\psi(t) = \psi(\theta) = \gamma$ are constants.

3.1 "Flattening the Curve"

Since $\beta = \frac{\varphi N_p}{A_p}$, and φ was defined by Kermack and McKendrick (1927) as "the rate of infectivity at age θ" (page 703), β has generally been interpreted as an inverse measure of social containment in the at-risk population, i.e., modelers have assumed that a lower β indicates higher social containment. Likewise, since $\gamma = \psi$ and Kermack and McKendrick defined ψ as "the rate of removal" (page 703) of infected persons to a recovered state or death, γ is generally interpreted as a measure of persistence of infectiousness, a constant associated with the agent of the disease; a lower γ has been assumed to represent longer-lasting disease.

A simulation, depicted in Figures 7A and B, shows that the SIR model projects that an increase in social containment (decreasing β) causes a later end to the epidemic and a lower and progressively later peak in cases per day. This is the so-called "Flatten the Curve" phenomenon predicted by Equations S3-1 through S3-4 which is oft referenced in the literature (see Di Lauro, et al, 2021, as a recent example). In contrast, a plot of the KMES in Figures 7C and D exhibits the opposite phenomenology: an increase in social containment (higher $\frac{K_T}{P_c}$) causes an *earlier* end to the epidemic and a lower and progressively *earlier* peak in cases per day. As social containment measures increase, the positions of the peak in new cases per day move in *opposite* directions for the two models.

We can also mathematically compare the trends projected by the SIR model with trends predicted by the full Kermack and McKendrick equations using the following expression derived from the KMES:

$$t_{max} = \frac{P_c \ln (P_c)}{K_T} \qquad \text{(S3-5, 46)}$$

As can be deduced from Equation S3-5, and in contrast to published analytical solutions of the SIR equations (Kroeger and Schlickeiser 2020), the KMES mathematically predicts that the time of the peak in daily cases will occur earlier with increased social containment (i.e., higher $\frac{K_T}{P_c}$). Therefore, SIR projections differ qualitatively from those of the KMES both graphically and mathematically.

The simplest test of the utility of a model is whether it projects the same trends present in actual data. If the projected trends are similar to those found in reality, then free parameters within the model are plausibly of value in achieving a higher degree of fit and utility. It is fortuitous, then, that the progression of the COVID-19 pandemic has been well documented in multiple countries which took different paths while attempting to contain the spread of the virus. This dataset affords the opportunity to test the veracity of the trends predicted by both the SIR and KMES models against actual data.

In plots E to H in Figure 7, we can compare the SIR and KMES projected trends to COVID-19 pandemic case data (Roser et al 2021) for total cases and for daily new cases in Sweden and New Zealand (Figure 7E and F), and in South Korea and Italy (Figure 7G and H). The paired countries have comparable population densities but implemented mitigation measures with different intensities (Campbell 2020, Field 2020, Orlowski and Goldsmith 2020, and Sanfelici.

2020). New Zealand and South Korea introduced stronger containment measures much earlier than Italy and Sweden.

In support of the KMES and in contradistinction to the SIR model, the country data in Figures 7E–H show that stronger containment measures are associated with an *earlier* levelling off at a *lower* total number of cases and an *earlier* and *lower* peak in new infections. Other authors (Schlickeiser and Kroeger 2020), too, have noted that the peak of cases in countries with stronger containment measures occurred earlier than in countries with weaker measures.

Trends in both peak position and height demonstrate that SIR models are not merely inaccurate, a tolerable trait in an approximation, but project epidemic data to trend in the *opposite* direction to the reported data; a behavior that no amount of free parameter fitting can correct. Therefore, the SIR model both contradicts the KMES and fails the simplest test of model veracity: the projection of qualitative trends

3.2 Understanding the Implications of the "Flatten the Curve"

As seen in the preceding section, though both β and P_c are posited to represent social interaction in their respective models, the trend in the movement of the daily cases peak with decreasing social interaction (decreasing β) in the SIR model is opposite to that with decreasing social interaction (decreasing P_c) in the KMES. Since the KMES reproduces the observed trends and the SIR model does not, it seems likely that the nature and implications of the SIR assumptions may not be sufficiently understood.

Fig.7 (A-H)

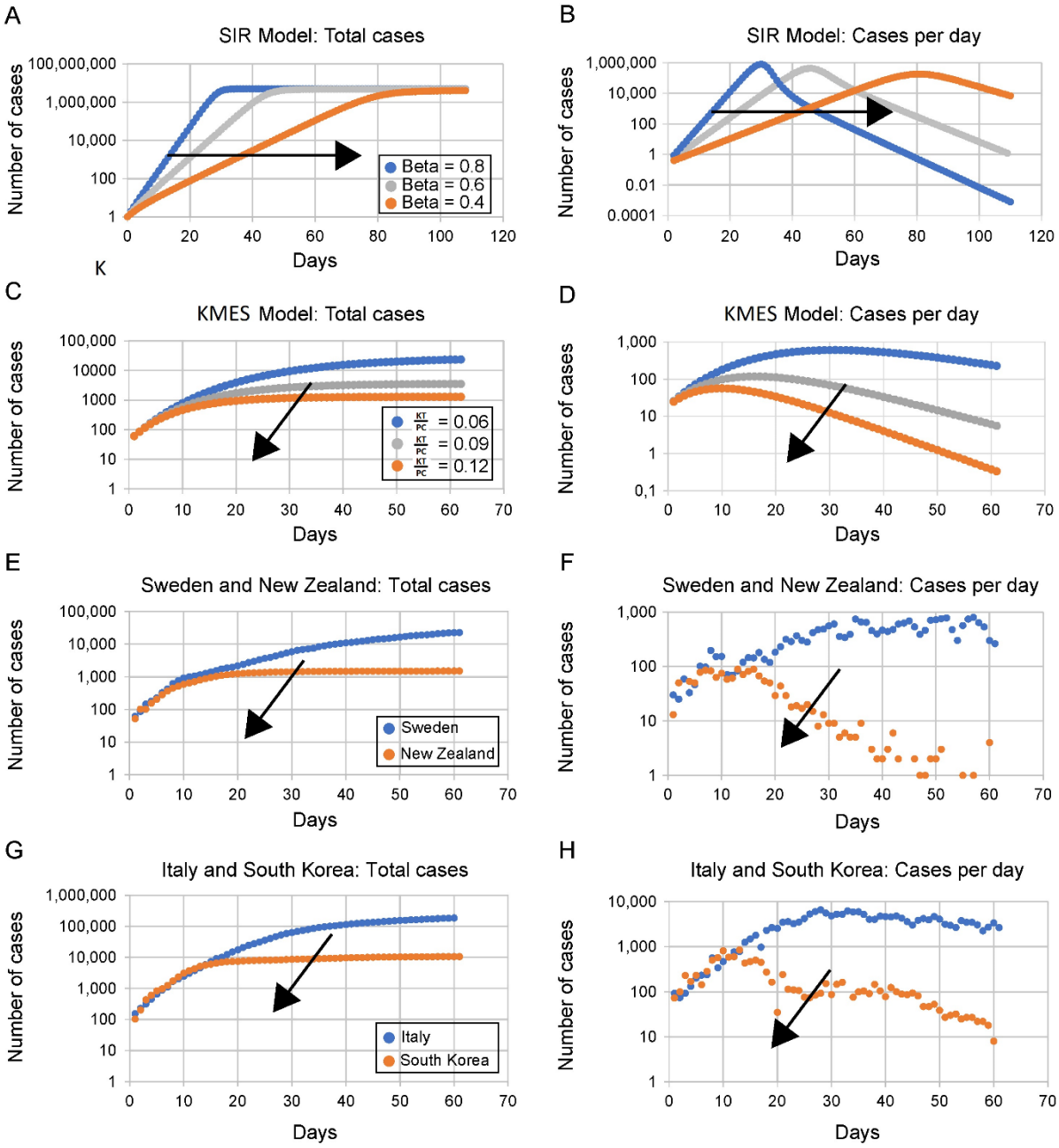

Figure 7. Comparisons of predictions of the approximate SIR (ASIR) and complete SIR (CSIR) models with observed data from four countries

Note: Containment measures increase in all panels from blue to grey to orange dot curves. The arrow on each graph indicates the direction of more social distancing.

SIR model trend predictions: (A) Total cases; **(B)** Daily new cases. Rate of contact and transmission (β) decreases with increasing social distancing (from blue to grey to orange curves). Rate of recoveries (γ) = 0.2 for both sets of plots. As β decreases, the daily total of cases increases more slowly and plateaus later (**A**). Daily new infections project to later, but only slightly lower peaks (**B**).

KMES model trend projections: (C) Total cases; **(D)** Daily new cases.

As containment measures increase (higher $\frac{K_T}{P_c}$, blue to grey to orange curves), Equation 23 projects that total cases will rise to lower levels; and reach these levels earlier **(C).** Similarly, Equation 22 projects that new daily cases will peak earlier to lower values with increasing containment **(D).** $K_T = 0.2$ for plots **(C)** and **(D).**

Data reported from different countries during the COVID-19 pandemic.

The remaining graphs contain data from pairs of countries with differing containment measures referenced to a day when each member of the pair had nearly equal numbers of new cases.

(E) Total cases in Sweden (no containment measures, blue) and New Zealand (strict containment, orange).**(F)** Daily new cases in Sweden (blue) and New Zealand (orange).**(G)** Total cases in Italy (loose containment measures, blue) and South Korea (strict containment, orange).

(H) Daily new cases in Italy (blue) and South Korea (orange).

The trends in the observed data, panels (**E – H)**, are the opposite of those exhibited by the ASIR model for increasing containment (decreasing β) in panels (**A, B**).

The ASIR model trends in (**A**) and (**B**) have completely different shapes; and vary with increasing containment in an opposite sense to those in the country data.

The CSIR model trends in **(C)** and **(D)** are highly similar to those in the country data **(E – H)**.

To understand these implications, we start by examining the conventional assumption that $\varphi(t)$ and $\psi(t)$ can be constants. Using Equations S1-21 and S1-28, and the prior definition that $F_i(t) = \frac{I(t)}{N(t)}$, we can find expressions for the time varying $K_T(t)$ and $P_c(t)$ when $\varphi(t)$ and $\psi(t)$ are assumed to be the constants φ and ψ:

$$K_T(t) = \frac{\varphi S(t)}{A_p} = \frac{\beta S(t)}{N_p}, \text{ and} \tag{S3-6}$$

$$P_c(t) = \frac{1}{\frac{A_p \psi}{\varphi S(t)} - (1 - F_i(t))} = \frac{1}{\frac{\gamma N_p}{\beta S(t)} - (1 - F_i(t))}. \tag{S3-7}$$

From Equation S3-6 we see that $\varphi(t)$ can only remain constant if $K_T(t)$ decreases in direct proportion to the decreasing size of the susceptible population, $S(t)$. This is implausible on its face because, as discussed in the derivation of the solution, $K_T(t)$ is solely a function of the disease agent and thus, is likely a constant for a substantial time at the beginning of the epidemic; at least until the disease agent itself is modified by mutation or selection.

Kermack and McKendrick (1927) themselves, in their introduction on pp. 702**,** note that it is implausible to assume that disease transmissibility will *decrease* as the disease spreads. Furthermore, even if transmission were to decrease over time within the infected population, it is improbable that this decrease would, as required by Equation 21, occur in a fixed linear proportion to the remaining number of susceptible people. Thus, the assumption within the SIR model that φ can be modelled as a constant requires an implausible additional assumption.

It is not possible to state how $P_c(t)$ must vary to maintain ψ as a constant by merely inspecting Equation S3-7. We can, however, elucidate the behavior of $P_c(t)$ required by Equation S3-7 by plotting the time series of Equation S3-7. The time series of both $N(t)$ and $\frac{dN(t)}{dt}$ were simulated using an Euler approximation of Equations 22 and 23 with a time step of 0.1 day. Using the same values of β and γ employed in Figure 6A and B, Equations S3-6 and S3-7 were then used to determine the values of $K_{\mathrm{T}}(t)$ and $P_c(t)$ employed in the simulation.

The purpose of this simulation was first to demonstrate that imposing the conditions of Equations S3-6 and S3-7 on $K_{\mathrm{T}}(t)$ and $P_c(t)$ will cause the KMES to produce the same results as the SIR approximation. The second purpose was to determine and demonstrate the actual temporal behavior the SIR approximation imposes on both $K_T(t)$ and $P_c(t)$.

The time series plots of the simulation of both $N(t)$ (cases) and $\frac{dN(t)}{dt}$ (cases per day) appear in Figure 8. The close approximation of the solution and the SIR curves in Figure 8 demonstrates that the SIR model is, indeed, a subset of the KMES when the constraints of Equations S3-6 and S3-7 are applied to $K_{\mathrm{T}}(t)$ and $P_c(t)$.

To elucidate the behavior of $K_T(t)$and $P_c(t)$ required to create the KMES curves in Figure 8, we plotted the values of $K_T(t)$and $P_c(t)$ in Figure 9. The figure shows that the constraints on

$K_T(t)$and $P_c(t)$, imposed by the SIR model, compel the acceptance of unlikely phenomena; namely, that $K_T(t)$ decreases with time and $P_c(t)$ must behave in an unrealistic manner.

As Figure 9B illustrates, the consequent, implicit assumption of applying the SIR approximation is that, early in the epidemic, the population increases its contacts, and then suddenly and symmetrically (in time), reverses course and reduces the number of contacts. At each value of β, this up and down spike in contacts (Figure 9B) precedes a plunge in the value of $K_T(t)$ (Figure 9A), and the steep decline is immediately followed by the peak in daily cases (seen in Figure 8B), tailing to the eventual end of the epidemic.

These implied consequences of a constant β in the SIR model make the clear points that a constant β does *not* represent constant social interaction; and a higher β does *not* represent a consistently higher level of social interaction. Also, a symmetric spike in social interaction ($P_c(t)$), higher and earlier, proportional to the value of β, followed by an immediate collapse in transmissibility ($K_T(t)$), is simply unfathomable.

The consequences of the approximations in the SIR model become even more clear when we make manifest the time varying nature of $\beta(t)$ and $\gamma(t)$ (and therefore of $\varphi(t)$ and $\psi(t)$) required when the quantities K_T and P_c are held constant. Like the preceding analysis, we explored these consequences using simulations of the KMES and the SIR model. Since the KMES predicts the country data well, we also compared the simulation results to two of the country results (Italy and New Zealand).

Fig.8 (A-B)

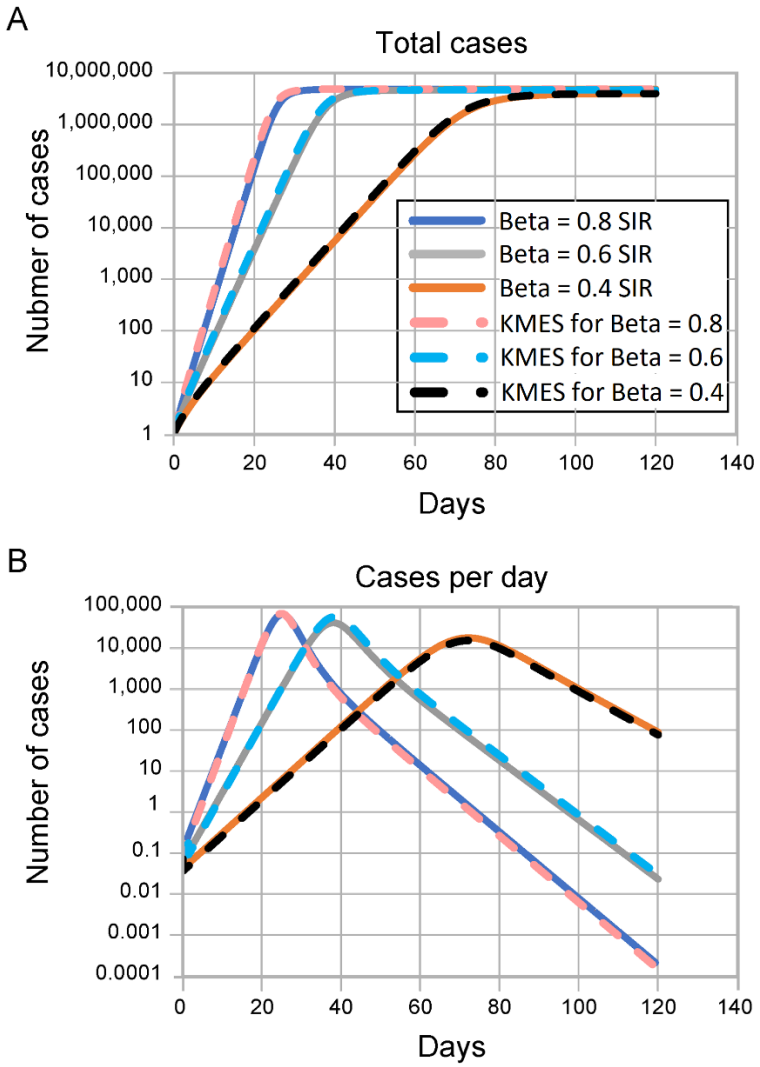

Figure 8. Demonstration that KMES can be modified to produce the SIR model results. Both plots contain 6 lines. For the same β and γ, both the SIR and KMES simulations overlay each other. The KMES simulations were produced by imposing the criteria in Equations S3-6 and S3-7, connecting the SIR and KMES constants. This plot demonstrates that the SIR approximation provides the same result as the KMES provided the constraints of these equations are imposed. γ is 0.2 for all plots. These plots are the same as the SIR plots in figure7A and 7B on a log scale.

Fig.9 (A-B)

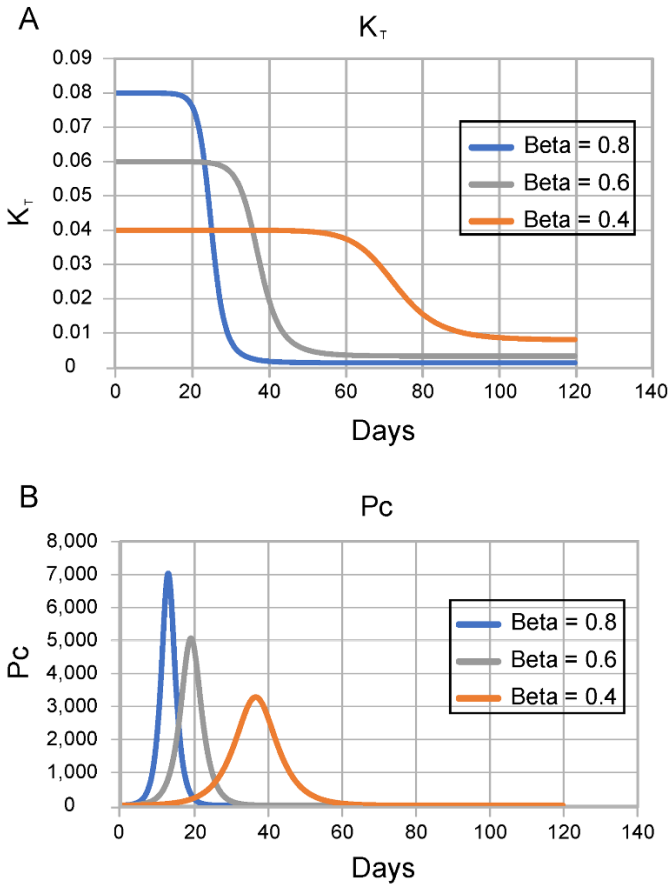

Figure 9. **Time series for creating solution curves.** These graphs show how Pc and KT are forced to vary within the KMES simulation shown in Figure 8 when the constraints of Equations S3-6 and S3-7 (constant φ and ψ) are imposed. γ is 0.2 for all plots. X-axis is days.

As a first step, we used the values of $ln(F_i(t_0)K_T)$ and $\frac{K_T}{P_c}$ in Table 1 to derive values of $\frac{K_T}{P_c}$ and K_T for Italy and New Zealand and used the solution to project the results. We then simulated the SIR model with the assumption that the values of β and γ (and therefore φ and ψ) were constant and equal to the values of $\frac{K_T}{P_c}$ and K_T used in the solution. The results of both the KMES and SIR simulations are plotted in Figure 10, along with the country data. As can be easily seen, the KMES model accurately models the country data, and the SIR model does not.

In a second step, we recast Equations S1-21 and S1-28 in terms of $\beta(t)$ and $\gamma(t)$:

$$\beta(t) = \frac{\varphi(t)N_p}{A_p} = \frac{K_T N_p}{S(t)} \text{ and} \quad \text{(S3-8)}$$

$$\gamma(t) = \psi(t) = K_T \frac{R(t)}{N(t)} + \frac{K_T}{P_c}. \quad \text{(S3-9)}$$

Using Equations S3-8 and S3-9, we then calculated the time series of $\beta(t)$ and $\gamma(t)$ necessary to generate the KMES curves in Figure 10. Those time series, plotted in Figure 11, show that under the conditions present in the countries, $\beta(t)$ is nearly a constant, while $\gamma(t)$ clearly is not. This is because, early in the epidemic when $S(t) \approx N_p$, $\beta(t) \approx K_T$, and $\beta(t)$ can be approximated as a constant.

As a last step in the analysis, we used the values of $\beta(t)$ and $\gamma(t)$ plotted in Figure 11 in the SIR model to generate the curves in Figure 12. This figure shows that when β and γ are forced to vary according to Equations S3-8 and S3-9, the SIR model fits the country data quite well.

Equations S3-8 and S3-9 themselves elucidate why the SIR model, with constant β and γ, is not an accurate or appropriate approximation. Equation S3-8 makes it clear that when $S(t)$ decreases by a significant percentage, neither β nor φ can be appropriately modelled as constants. Also, as can be seen in Equation S3-9, the assumption that γ, and therefore ψ, is a constant ignores the effect of the growing recovered (and therefore resistant) fraction, $\frac{R(t)}{N(t)}$, of the subpopulation, $N(t)$, on the epidemic dynamics.

Figures 8 and 12 provide another validation of the veracity of the solution. In Figure 8, we show that the solution can be configured to replicate SIR simulations by embedding the SIR approximations within the solution framework. In that setting, the two piecewise and logically

Fig.10(A-D)

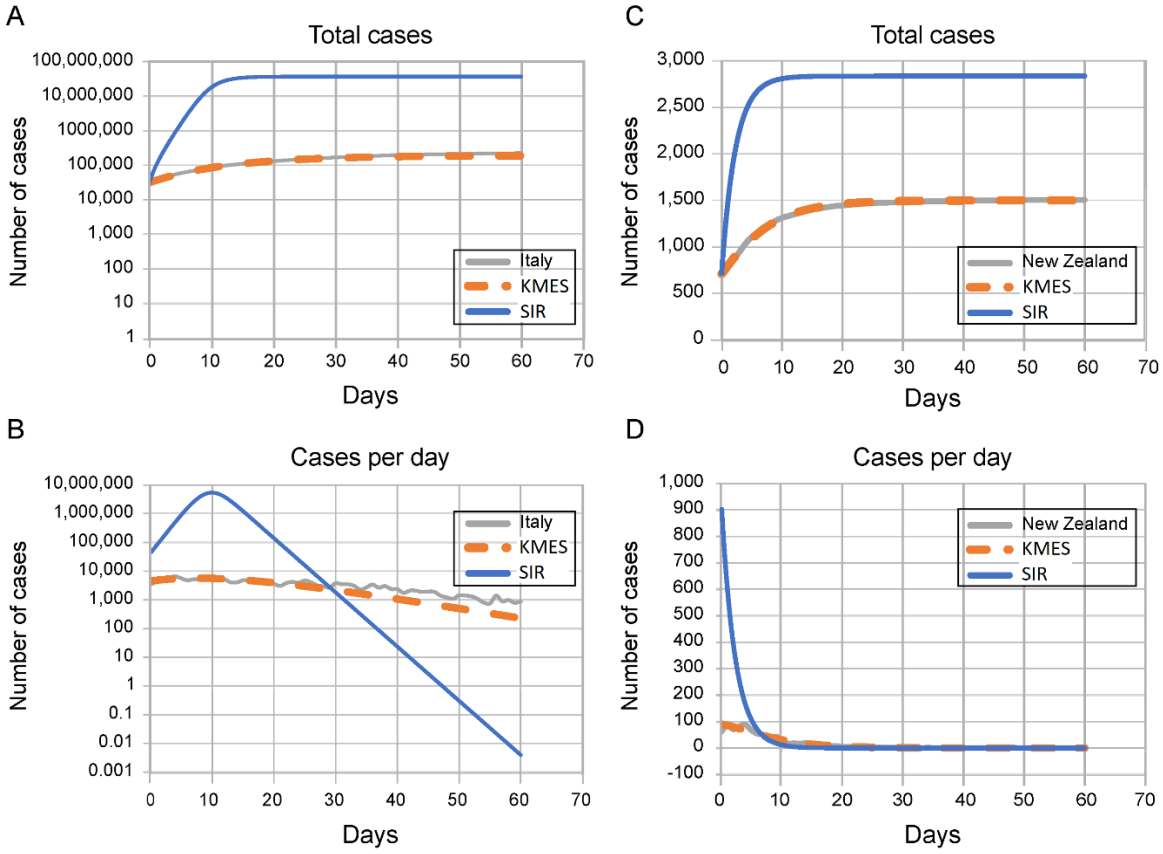

Figure 10. SIR and solution simulations of the Italy (10A & B) and New Zealand (10C & D) data. $\beta = e^{ln(F_i(t_0)K_T)}$ where $ln(F_i(t_0)K_T)$ is from Table 1 and γ is equal to the $\frac{K_T}{P_c}$ constant from Table 1.

invariant solution parameters, K_T and P_c, are forced to take implausible and unrealistic time courses. Figure 12 demonstrates, in counterpoint, that an SIR model can produce results identical to the KMES if β and γ, the analogs of $\varphi(t)$ and $\psi(t)$, are permitted to vary in time according to Equations S3-8 and S3-9. The SIR model fits reality only when "coached" to a time variation for its two parameters using insight derived from the KMES.

Fig.11 (A-B)

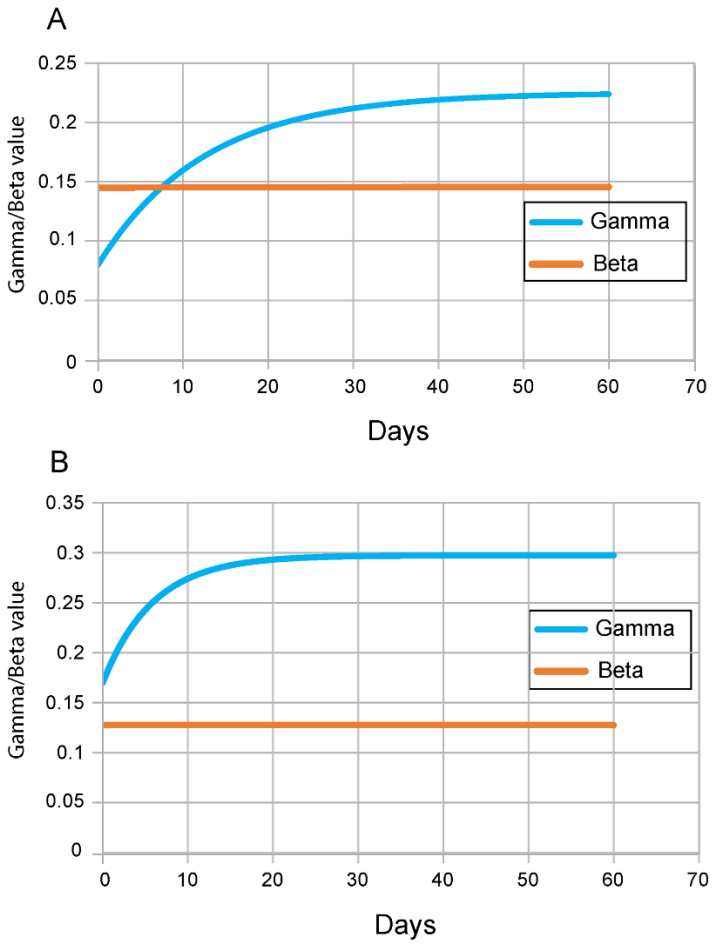

Figure 11. Time series for γ and β. **A)** Italy. **B)** New Zealand. These are the values of γ and β necessary for the SIR approximation to accurately model the country data.

The foregoing discussion has utilized one of the simplest models of the SIR types; nevertheless, the conclusions apply to all the variations of SIR models.

Fig.12 (A-D)

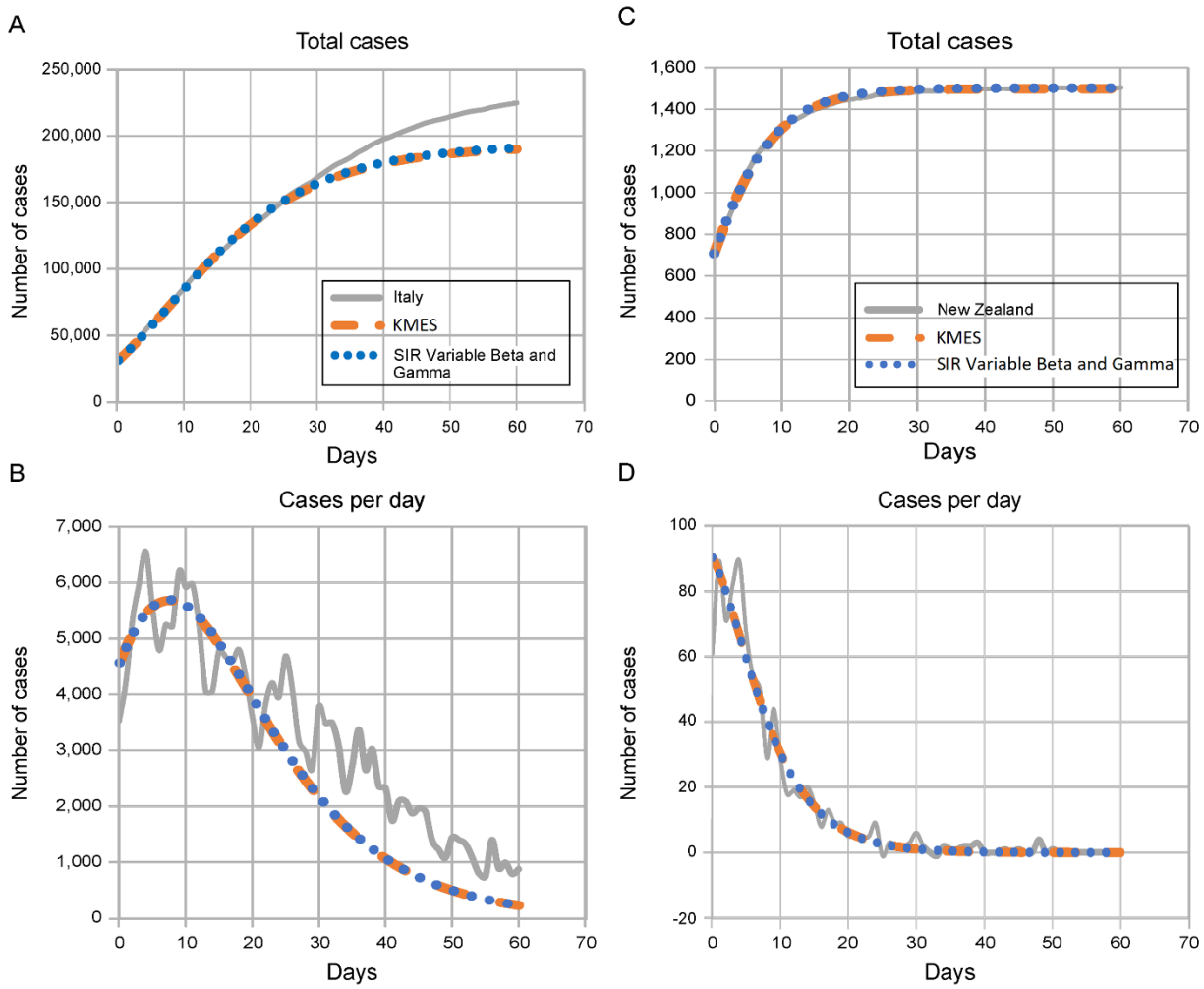

Figure 12. Total cases and new cases per day for Italy and New Zealand. The country data and KMES plots are the same as in Figure 10. The SIR Variable β and γ plot uses the β and γ values from Figure 10 in the SIR equations.

Supplement 4. Controlling epidemics early

The quantitative mathematical relationships derived from the KMES in Supplements 2 and 4 characterize the dynamics of an epidemic and illustrate that strong and early intervention is critical. Equation 25 quantifies that the ultimate number of individuals infected in an epidemic, N_∞, will be exponentially dependent on the number of people with whom each person interacts.

The real-world country data provide vivid examples. Both South Korea and New Zealand enacted strong and early interventions compared to other countries (5,6), as reflected by their $\frac{K_T}{P_c}$ values (Table 1). These strong interventions led to earlier peaks in new cases and to far fewer total cases than in other countries (Figures 2 and 3): the peak number of new cases in both South Korea and New Zealand was 90–99% lower than in other countries, a compelling validation of the explicit statement in the KMES that strong intervention leads to *exponentially* more favorable outcomes.

In the USA, interventions initiated on March 16 began to have an effect around March 23, 2020 (Figure 3B); the number of active cases on March 23, 2020 (Roser et al 2021) was 46,136 (Table 1). Using the values of $\ln(K_T(\mathrm{t}))$ and $\frac{K_T}{P_c}$ from Table 1, Equation 25 predicts that the ultimate number of cases would have been approximately 1.22 million. If the same intervention had been implemented and sustained starting on March 10, when there were 59 times fewer (782) cases (Roser et al 2021), the model predicts that the ultimate number of cases would also have been 59 times lower, or 20,725. Thus, earlier action could have reduced the ultimate number of projected cases by more than 98%. Of course, the projected estimate of approximately 1.22 million total USA cases would only have occurred if the effectiveness of the interventions that were launched on March 16 had been sustained. Unfortunately, a marked reduction in effective interventions occurred in many parts of the USA in mid-April, well before the official reopening of the economy (Elasser 2020). This caused a second surge in new cases in late April and is why the observed data and the model prediction diverge in Figure 3B.

As shown in the main body of the paper, Section 4, the KMES provides an estimate of the time to the peak of new cases, $t_{\max}$. Using Equation 46 and the values of $\ln(K_T(\mathrm{t}))$ and $\frac{K_T}{P_c}$ from Table

1, the predicted peak in new cases in the USA would have occurred near March 24 if the intervention had begun on March 10. Instead, a 6-day delay in effective intervention shifted the initial peak to April 11, 16 days later, as projected, and that peak was much higher (Figure 3B).

As shown, too, in Section 4, epidemic acceleration, the instantaneous potential to change the pace of the epidemic, can be determined at any point in the epidemic and depends on the social containment actions in effect at that time (Equation 47). What is perhaps less apparent, but predicted by the KMES, is that two countries with identical numbers of cases on a given day can, in fact, have different accelerations on the same day, and will, therefore, exhibit different dynamics immediately after that day.

South Korea and New Zealand (Figure 2A and F) had nearly identical case counts when each imposed strong containment measures (204 cases in South Korea on February 21, and 205 in New Zealand on March 25). Their models suggest that their interventions were about equally effective ($\frac{K_T}{P_c}$ = 0.24 in South Korea and 0.17 in New Zealand; see Table 1). However, since South Korea has a much higher population density than New Zealand ((Worldometers 2021), data in Table 2), it had a much higher number of interactions when the interventions were imposed and, therefore, a higher rate of acceleration, as evidenced by its higher RCO at the time of intervention. Indeed, the rate of change of new cases *was* higher in South Korea than in New Zealand, and the later number of cases in South Korea *was* higher than in New Zealand (Figure 2A and 2F).

Equation 44 clearly illustrates these lessons. As social distancing is strengthened (lower P_c), the Effective Replication Number decreases, and the epidemic slows. Early and strong interventions, especially in countries with indigenously high levels of social interaction, are necessary to stop

an epidemic in the initial stages. Reopening, enacted too early, can reignite the epidemic, dramatically increasing the number of cases. The astonishing magnitude of the effects, driven by only a few days of delay, derives from the doubly exponential nature of the underlying relationships.

Supplement 5. Ending an ongoing epidemic

We can use the KMES to design measures to end an epidemic in an advanced stage. The management plan is built by first using Equation 51 to predict the number of days a given level of intervention, $\frac{K_T}{P_c}$, is needed to reduce the new daily cases by a target fraction.

For example, using Equation 51, we see that a country targeting a 90% reduction of new cases per day (e.g., from 50,000 to 5,000 cases per day, $D_{tf} = 0.1$), can attain its target in about 12 days by imposing a containment level of $\frac{K_T}{P_c} = 0.2$. The South Korea and New Zealand data demonstrate that Equation 51 is valid and that $\frac{K_T}{P_c} = 0.2$ is achievable for this duration. Both countries achieved a value of $\frac{K_T}{P_c}$ close to 0.2 for the time necessary to produce a 90% reduction. It took 13 days in South Korea (March 3–16) and 15 days in New Zealand (April 2–15), New Zealand (6).

The criteria, $\frac{K_T}{P_c} = 0.2$, characterizes a lockdown in which people in a country can each have only one plausibly infectious contact with a little over one specific person for the containment duration. This does not mean they cannot contact anyone other than the one person; but they must use care, masks, and proper distancing, to ensure there is no plausibly infectious contact with anyone other than the one person.

Returning to the planning example, after achieving the initial 90% reduction, a reasonable next step might be to relax social containment to a level that allows the economy to remain viable, while preventing the epidemic from erupting again. We can again find the level of $\frac{K_T}{P_c}$ necessary to achieve a chosen target, using Equation 51. If an additional 90% reduction in new cases per day is desired, and a period of 90 days is tolerable for that reduction, then a new level of approximately $\frac{K_T}{P_c} = 0.025$ is needed. This equates to a 90-day period during which each person can be in contact with seven specific people, in an infectable way. Note that this is three times *less* stringent than the original USA shutdown level in April 2020 as shown by the level of $\frac{K_T}{P_c}$ calculated for the United States in that period (Table1). Thus, with a well-planned approach, a country can reduce its new daily cases by 99% in approximately 100 days, enabling the country to control, and essentially end the epidemic, while simultaneously maintaining economic viability.

If even 0.025 is too restrictive, we can choose a still lower $\frac{K_T}{P_c}$, but it must be large enough to avoid a new outbreak. A lower bound for the new value of $\frac{K_T}{P_c}$, high enough to prevent an outbreak, can be found using Equation 48.

We can easily monitor the progress of interventions using the RCO, as the curve for South Korea illustrates (Figure 1A). Had this country maintained the implemented level of distancing measures, the data would have followed the initial slope. However, the actual data departed from the slope, heralding failures in (or relaxation of) social distancing, which were later documented to have occurred during the indicated time frame (Campbell 2020) (circled data, Figure 1A).

Because it summarizes epidemic dynamics, we can use the RCO to continuously determine the effectiveness of implemented measures and whether they need adjustment.

Supplement 5.1 Outbreaks

We can see from Equations 39 and 40 that if the social interventions are strengthened (lower P_c) the slope of the RCO curve will steepen and if the interventions are relaxed, the slope will become shallower. Therefore, if the value of $K_T(t)$ does not change due to a change in the disease transmissibility, the RCO is a metric for monitoring the population interactions. It is also clear that the slope can never become positive, because $\frac{K_T}{P_c}$ must always be greater than zero. However, this only remains true if these three conditions remain true: 1) immunity persists, 2) no new infections are introduced from outside the area, and 3) the epidemic remains contiguous (see Supplement 6).

If new infections are introduced into a portion of the population that has thus far been disconnected from the previously infected area, then the condition of contiguity is violated. This is a common situation when infected people travel from an infected area to a previously uninfected area and cause an outbreak.

Equation S1-29 must be modified to predict the number of cases in an epidemic affected by an outbreak. Assuming that $t_0 = 0$, $N(t_0) = 1$, and introducing the notation P_{cx} where x denotes the number of the outbreak, Equation 29 can be written as:

$$N(t) = e^{-P_{c1}(e^{-\frac{K_T}{P_{c1}}\mathrm{t}}-1)}. \qquad \text{(S5-1)}$$

If a new outbreak occurs in a previously unaffected area of a country, then Equation S5-1 can be modified as follows:

$$N(t) = e^{-P_{c1}(e^{-\frac{K_T}{P_{c1}}\mathrm{t}}-1)} + N_2 e^{-P_{c2}(e^{-\frac{K_T}{P_{c2}}(\mathrm{t}-t_2)}-1)}, \quad \text{(S5-2)}$$

where N_2 is the number of infectious people who initiated the new outbreak, P_{c2} is the social interaction parameter in the new outbreak area, and t_2 is the time the new outbreak occurs. We have assumed that the disease transmissibility remains the same throughout this illustration. If the transmissibility changes in a subset of the population, then a similar formulation, using the notation, K_{Tx}, can be utilized to track the populations with the new transmissibility.

Equation S5-2 can be written in a general form as

$$N(t) = e^{-P_{c1}(e^{-\frac{K_T}{P_{c1}}\mathrm{t}}-1)} + N_2 e^{-P_{c2}(e^{-\frac{K_T}{P_{c2}}(\mathrm{t}-t_2)}-1)} \ldots + N_x e^{-P_{cx}(e^{-\frac{K_T}{P_{cx}}(\mathrm{t}-t_x)}-1)}, \quad \text{(S5-3)}$$

where x denotes the outbreak number and $\mathrm{t} > t_2 > t_3 > \cdots > t_x$. For each outbreak t_x, P_{cx}, and N_x need to be determined independently.

While an epidemic is underway, we can detect an outbreak by monitoring the slope of the RCO curve. A positive slope detected in an RCO curve indicates that an outbreak has occurred. This is an indication that immediate action, within days, is required from policy makers to strengthen intervention measures and prevent the outbreak from overwhelming prior progress in controlling the epidemic.

By monitoring the RCO curve, we can also detect if the disease changes its transmissibility through mutation. In this situation, a proper fit of the parameters in Equation 40 is not possible and a modification of K_T is required to accommodate the change.

Supplement 6: Understanding contiguousness

To enhance the understanding of the KMES and explain what it means for an epidemic to be contiguous, we applied a new perspective to Equations S1-1 to S1-5. Rather than look at the epidemic as affecting the total population, N_p, from the outset, this perspective focuses on only that portion of the population that will eventually become infected in the epidemic, the subpopulation N_∞. We also introduce a subpopulation of N_∞ called $N_S(t)$, which we define as the sub-population at time t that is in contact with the epidemic and under threat of infection. This change in perspective views the epidemic mathematically from inside the bounds of the ever expanding, already infected population, rather than describing what is happening within a fixed total population.

The reinterpretation begins by first recognizing that in Equations S1-1 to S1-3, the initial number of infections introduced to the population—which we designate I_i—is merely the starting value of the epidemic and can be any value at all. The second step is to imagine that at every increment in time, Δt, the epidemic starts again, and the number of initial infections introduced into the population, $I_i(t)$, is equal to the then-current number of infected. The third step is to recognize that the values of the integrals in Equations S1-1 to S1-3 are equal to zero and $B(t) = 1$ whenever the epidemic starts. This can be seen by also recognizing that the diagonals in the matrices in Supplement 1 are the time history of the initial infections, therefore, when these values are subtracted from the matrices, the first rows of every matrix are zero and therefore, their sums are equal to zero.

With these perspectives in mind, Equations S1-1 to S1-3 can be rewritten as,

$$\frac{dSu(t)}{dt} = -\mathrm{Su(t)}\frac{\sigma(t)}{N_S(t)}I_i(t), \qquad \text{(S6-1)}$$

$$\frac{dI(t)}{dt} = \mathrm{Su(t)}\frac{\sigma(t)}{N_S(t)}I_i(t) - \omega(t)I_i(t), \text{ and} \quad \text{(S6-2)}$$

$$\frac{dR(t)}{dt} = \omega(t)I_i(t). \quad \text{(S6-3)}$$

where $\sigma(t) = \frac{\varphi(t)N_S(t)}{A_p}$ and $\omega(t) = \psi(t)$,

$Su(t)$ is the remaining portion of the population that will become infected, and since there are only susceptible people in $N_S(t)$ when new infections are introduced,

$$N_S(\mathrm{t}) = Su(t), \quad \text{(S6-4)}$$

and of course, $\frac{dSu(t)}{dt} = \frac{dS(t)}{dt}$.

If we also now define a quantity, $R_i(t)$, as the number of recovered persons introduced to the population at the same time as $I_i(t)$, we can write an equation for N_∞ and define a new quantity, $N(t)$, as the number of people either currently infected or recovered:

$$N_\infty = Su(t) + I_i + R_i, \quad \text{(S6-5)}$$

$$N(t) = I_i(\mathrm{t}) + R_i(t), \text{ and} \quad \text{(S6-6)}$$

$$N_\infty - N(t) = N_S(t) = Su(t). \quad \text{(S6-7)}$$

We have not assumed that any of the quantities are constant; therefore, we can now explicate the time-varying nature of these quantities. We begin by considering what happens to Equations S6-1 to S6-3 during the time interval Δt from a time t to $t + \Delta t$, and by rewriting these equations as difference equations:

$$\mathrm{Su(t} + \Delta t) = \mathrm{Su(t)} - \sigma(t + \Delta t)I_i(t + \Delta t)\Delta t, \quad \text{(S6-8)}$$

$$I_i(t + \Delta t) = I_i(t) + (\sigma(t + \Delta t)I_i(t + \Delta t) - \omega(t + \Delta t)I_i(t + \Delta t))\Delta t), \text{ and} \quad \text{(S6-9)}$$

$$R_i(t+\Delta t) = R_i(t) + \omega(t+\Delta t) I_i(t+\Delta t)\Delta t. \quad \text{(S6-10)}$$

Taking the limit as $\Delta t \to 0$, we obtain the following differential equations:

$$\frac{dSu(t)}{dt} = -\sigma(t) I_i(t), \quad \text{(S6-11)}$$

$$\frac{dI_i(t)}{dt} = \sigma(t) I_i(t) - \omega(t) I_i(t), \quad \text{(S6-12)}$$

$$\frac{dR_i(t)}{dt} = \omega(t) I_i(t) \text{ and} \quad \text{(S6-13)}$$

$$Su(t) = N_\infty - N(t). \quad \text{(S6-14)}$$

From the preceding, it can be easily seen that $I_i(t) = I(t)$ and $R_i(t) = R(t)$. Note that $Su(\infty)$, unlike $S(\infty)$, is always equal to zero.

The perspective in the immediately preceding part of the analysis is that the epidemic can be considered to start over again at each instant in time. In this perspective, the susceptible population $Su(t)$ is not fixed by initial conditions, but rather is the population that will eventually become infected. Embedded in this concept is the assumption that during each Δt, the susceptible population is always in contact with those people who have been previously infected or who will become infected. That is, the epidemic remains contiguous.

We have recast the equations by defining the susceptible portion of the population during the epidemic as those who will eventually become infected under the conditions in place at each instance in time. This shift in perspective retains the mathematical equivalence to the equations derived by Kermack and McKendrick (1927) because the portion of N_p that is not a part of N_∞ never becomes infected. Therefore, the solutions to equations S6-11 to S6-14 are identical to the KMES and the KMES assumes the epidemic is contiguous.