Cross-Model Parameter Estimation in Epidemiology

Julia R. Fitzgibbons
University of New Hampshire

Follow this and additional works at: https://scholars.unh.edu/honors

Part of the Analysis Commons, Numerical Analysis and Scientific Computing Commons, and the Ordinary Differential Equations and Applied Dynamics Commons

Recommended Citation
https://scholars.unh.edu/honors/583

This Senior Honors Thesis is brought to you for free and open access by the Student Scholarship at University of New Hampshire Scholars' Repository. It has been accepted for inclusion in Honors Theses and Capstones by an authorized administrator of University of New Hampshire Scholars' Repository. For more information, please contact nicole.hentz@unh.edu.
Cross-Model Parameter Estimation in Epidemiology

By:
Julia Fitzgibbons

Advised By:
Dr. Adam Boucher

A thesis presented for the degree of Bachelor of Science in Mathematics

May 2021

Department of Mathematics and Statistics
University of New Hampshire
Durham New Hampshire, USA
Abstract

With the rise of the COVID-19 pandemic, the media has boomed with news of epidemiology for various purposes: publicity, education, political influence, and more. Many individuals lack knowledge about matters of epidemic modelling or disease transmission, so they will accept unsubstantiated information blindly. Therefore, the terms $R_0$ and exponential growth are used interchangeably, and misconstrued as one another. We investigated if an exponential function can be used to predict the $R_0$ of an SIR Model, and if a functional relationship can be found which connects the parameters of these two different models. By using MATLAB’s ode45 function, we were able to solve Kermack McKendrick models and produce SIR curves that could then be fit exponentially to produce an exponential growth rate value. These values were used in several estimates which attempt to use the exponential growth rate to calculate $R_0$. This was done for a range of $R_0$ values and Kermack McKendrick model parameters.

We found that all estimates grew less accurate as $R_0$ increased. The equations from Heesterbeek produced inaccurate, yet consistent results through the variety of SIR parameters tested: $e^{\gamma}$ grew exponentially as $R_0$ increased, producing an overestimate, while $1 + \frac{\gamma}{\gamma}$ grew at a slower linear rate than direct correspondence, underestimating $R_0$. The estimate of $e^\gamma$ was inconsistent throughout the range of SIR parameters examined, thus it was not a reasonable estimate. This result can provide insight into the independent nature of epidemic modelling techniques. With the definition of $R_0$ for an SIR curve relying on the specific variables used to produce that model, it is unreasonable to assume that an estimation of initial growth can be equivocated to $R_0$, or used to find a reasonable estimate.

All source code and documentation for this project can be found at https://github.com/JuliaFitzgibbons/Senior_Thesis
Contents

1 Introduction 3
   1.1 Epidemiology ......................................................... 3
   1.2 Epidemic Modelling .................................................. 3
   1.3 Trustworthy Modelling .............................................. 5

2 Hypothesis 6
   2.1 Dimensionless Parameters ......................................... 6
   2.2 Exponential Growth Rate vs. $R_0$ .................................. 6

3 Method 8
   3.1 Simulations ............................................................... 8
   3.2 $R_0$ Estimation ....................................................... 8

4 Experimental Results 11
   4.1 Experiential Design .................................................. 11
   4.2 Experiential Results .................................................. 11
   4.3 Discussion of Results ............................................... 11

5 Conclusion and Applications 14
   5.1 Summary ................................................................. 14
   5.2 Applications ........................................................... 14

6 Acknowledgments 15
1 Introduction

1.1 Epidemiology

Epidemiology is the branch of science that explores the spread of diseases, with the intention to understand the factors that contribute to their occurrence [1]. This data-driven topic is reliant upon multiple other scientific fields: biology, economics, mathematics, and psychology. All of these subjects come together to help uncover where diseases originate, how they travel through populations, how we can prevent growth of an infected populous, and countless other questions that arise about how humans are affected by disease.

After the emergence of the COVID-19 pandemic, the topic of epidemics and epidemic modelling has become popular to both scientists and the public alike. The entire world has been affected by this virus, and the majority of the population has not required a comprehensive knowledge of disease transmission, herd immunity, or other important concepts for public health during such a crisis. Many people are familiar with the basic terms used in epidemiology, but only superficially; words like exposure, risk factor, epidemic, and bias are known to the public by other definitions unrelated to those used to describe disease modelling.

With this high level of interest and dearth of knowledge, there are thousands of sources attempting to inform the public about epidemiology and the most recent findings on COVID-19. With epidemiology being such a diversified study, the quality of information is highly variable. The news covers most variants of the topic, such as the perspectives of public health, statistical analysis, psychological impact, and predictive modelling.

1.2 Epidemic Modelling

Modelling epidemics is an important approach used to comprehend the underlying components of specific diseases and their growth patterns, and can help shape policies put in place to control or eliminate growth. One favored technique is partitioning a population into categories of individuals. By doing so, we can see how these groups grow in relation to time, thus how a disease propagates through a population. The following three classes are the most common partitions used:

- S, the susceptible - those who are healthy and can catch the disease.
- I, the infected - those who have contracted the disease and can transmit it to susceptible individuals.
- R, the recovered/removed - those who have had the disease and are no longer infecting susceptible people. The individuals in this category are now either immune to the disease, or have been removed from spreading the disease in some other fashion, such as quarantine or death.
Many models break the population into other subsections of these groups, such as Exposed and Asymptomatic. This complicates the models by considering factors like disease incubation periods and quarantine precautions. We will investigate the correspondence between the basic SIR and exponential models.

Another important concept to epidemic modelling is the basic reproduction number, also known as $R_0$. This is the number of disease transmissions introduced to a susceptible population by a singular infected individual. When we create an SIR model, the $R_0$ is uniquely determined by a combination of the model’s characteristics. For example, an airborne virus with an infectious period of 3 days will develop differently than a disease transmitted through direct contact that remains active for the duration of an individual’s life. $R_0$ is not simple to estimate. Thus, there is a large portion of biological and mathematical epidemiology that attempts to estimate this value for each disease that is discovered. The model behavior changes at $R_0 = 1$, known as the critical value. This means that it is the tipping point for epidemics to occur. $R_0 < 1$ implies the disease could subside and disappear, while $R_0 > 1$ implies that the disease has the potential to grow into an epidemic.

### 1.2.1 The Kermack-McKendrick Model

One of largest influential papers on the mathematical modelling of epidemics is by Kermack and McKendrick [4]. This paper created a foundation for many modelling techniques used today. They assume the population, $N$, to be constant, and divide this population into the three aforementioned classes, S, I, and R. First, we assume the population is closed and some constant $N$, thus for any point in time, $S + I + R = N$. Secondly, we define $\beta$ to be the rate at which any individual comes into contact with another. This value is correlated with probabilities of infections when viewed statistically. Lastly, we define $\gamma$ to be the rate at which infected individuals recover. Note that both $\beta$ and $\gamma$ are positive. Thus the following system of differential equations is formed:

\[
\begin{align*}
\frac{dS}{dt} &= -\beta SI \\
\frac{dI}{dt} &= \beta SI - \gamma I \\
\frac{dR}{dt} &= \gamma I
\end{align*}
\]

where $I(0) > 0$, $S(0) > 0$, and $R(0) = 0$.

In this case, our basic reproduction number $R_0 = \frac{\beta S(0)}{\gamma}$. We can see that our disease could grow into an epidemic when $\beta S(0) > \gamma$, which is when the contact with the initial populous of susceptibles is greater than the recovery rate.
1.2.2 The Exponential Model

Another popular way of modelling disease is with the use of exponential functions. The initial growth of an infected population is often observed to grow exponentially, with each newly infected individual infecting roughly the same number of susceptible individuals. Hence, it makes sense to use a model of the same order. We let \( N \) be the constant population, and \( C(t) \) is the cumulative number of cases at time \( t \), and therefore the susceptible population is \( S(t) = N - C(t) \). We use an exponential function to estimate the growth of the number of cases:

\[
C(t) = C(0)e^{rt}
\]  

where \( r \) is a positive parameter denoting the growth rate, and \( C(0) \) is the initial number of cases.

However, due to changes such as government intervention, medical innovation, and societal action, the growth curve flattens and the rate of infections could gradually reduce \([4]\). This causes the deviation of the data points from an exponential curve, and therefore an exponential model is only applicable in the first growing stages of an epidemic \([3]\).

1.3 Trustworthy Modelling

Epidemiology depends on a systematic and unbiased approach to the collection, analysis, and interpretation of data. However, with the rapid growth of the COVID-19 pandemic, this rigorous standard may not always be met. Issues arise when there is conflicting and unreliable information being rushed to satisfy an uninformed, knowledge hungry audience. The pedagogical intent of the media is more harmful than good, when they do not have the resources to back their claims. Therefore, the public must hold a healthy amount of doubt on the reliability of these sources. The data and modelling approaches utilized may be justified, but the assumptions taken to achieve the reported output have to be verified.
2 Hypothesis

2.1 Dimensionless Parameters

The title "basic reproduction number" is already slightly misleading in its name. It is not a ratio, or a metric one should use directly when mapping out disease growth. It is a dimensionless parameter used to observe whether a disease has the infectious capacity for unrestrained growth. Another parameter similar to this can be found in fluid dynamics: The Reynolds Number, known as $R_e$. It is dimensionless, and is used as a threshold in classifying whether a fluid is in laminar or turbulent flow [6].

We can see parallels in the use of these parameters. Both are calculated to determine what kind of behavior is expected within a model. $R_e$ classifies the type of flow in pipelines, while $R_0$ distinguishes the growth of a disease. Both parameters are generated from the system being used, and have different definitions based off of the model in place. The expressions which correlate to $R_e$ rely on the model applied, as $R_0$ relies on the parameters in the system in which it is defined.

Thus, the parameter $R_0$, like $R_e$, can be descriptive to how the model to which it is applied will act. Note that it is calculated by using the system it is defined. Therefore, trying to generate $R_0$ using a model other than the one it was generated for seems problematic. This concern comes to light when there is ambiguity in the use of the exponential growth rate and basic reproductive number. Although they both describe disease growth, they are established and used in different contexts. Let us refer to the exponential growth rate as $r$.

2.2 Exponential Growth Rate vs. $R_0$

$R_0$ does not dictate how quickly things are changing, because $R_0$ is not a rate; there is no timescale involved. An example of this concept can be seen in the spread of COVID-19 in Africa: the exponential growth rate was estimated to be .22 people per day, while $R_0 = 2.37$ [5]. As previously mentioned, $R_0$ is the threshold for unrestrained growth, but it does not describe how quickly this growth will happen. For diseases such as HIV or Tuberculosis, where there can be months or years between one person infecting the next person, an $R_0 > 1$ will still have slow growth over time. On the other hand, for influenza or measles, $R_0 > 1$ means very rapid growth, due to the methods of transmission.

By this reasoning, there is little justification to assume any parallel in the usage of either value when modelling an epidemic. Unfortunately, whether it be due to the similar names or the definitions both pertaining to epidemiological growth, there is a misconception that the exponential growth rate $r$ can be used to describe or even produce $R_0$.

Our research aims to investigate this misinterpretation of parameters. We hypothesize that there may be functional relationships between the initial exponential growth rate of
an SIR model and its $R_0$, however this relationship is not an accurate way to estimate an SIR model’s $R_0$. Furthermore, we specifically conjecture that a direct relationship used to estimate $R_0$ with the exponential growth rate will be unreliable.
3 Method

3.1 Simulations

To generate simulations of the Kermack McKendrick model, we used MATLAB’s ode45() function, which employs a standard RK45 algorithm. The initial conditions were kept constant for all simulations: $S(0) = 100$, $I(0) = 1$, and $R(0) = 0$.

From there, we found the first inflection point of the data generated for the infected populous in order to identify the initial growth period of the model. This portion of the data was then fit with a variety of possible exponential curves in order to obtain the exponential growth rate, $r$, to use in the estimations of $R_0$ for that model.

![SIR Model](image)

**Figure 1: Influenza Model**

In Figure 1, the result of this process can be seen. The blue curve is the function $I(t)$, representing an infected populous over time, modelled with the variables $\beta = \frac{1}{400}$ and $\gamma = \frac{1}{14}$. The red curve is an exponential fit of the initial growth period of $I(t)$.

3.2 $R_0$ Estimation

As previously stated, $R_0$, the basic reproduction number, is a difficult value to estimate due to its many contributing factors. In order to see the impact that the SIR parameters $\beta$ and $\gamma$ have on the estimations of $R_0$, we calculated $R_0$ using four different methods: The Kermack McKendrick SIR modelling method, two methods used by Heesterbeek, and then a rough exponential approximation.
3.2.1 Kermack McKendrick $R_0$

The Kermack McKendrick epidemic modelling technique, as described previously, has two variables which relate to the growth of different population sectors: $\beta$ and $\gamma$. The system gives $R_0 = \frac{\beta S(0)}{\gamma}$. Since both are of the dimension "people per time ", this satisfies the dimensionless characteristic of $R_0$, and gives a reasonable estimate of how many people one infected individual can contaminate during their infectious period.

3.2.2 Heesterbeek $R_0$

We examine a technique from a paper by Roberts and Heesterbeek that evaluates the value based upon two other more accessible values: $r$, the exponential growth rate of our disease, and $T_G$, the observed mean generation interval of the epidemic [7]. They estimate $R_0 = 1 + rT_G$ when $rT_G$ is small, but otherwise $R_0 = e^{rT_G}$. We obtained the exponential growth rate, $r$, as previously mentioned. The $T_G$, however, may be determined from the formula provided by Heesterbeek:

$$T_G = \int_0^\infty t f(t) dt$$

(5)

where we are given $f(t) = \gamma e^{-\gamma t}$ for an SIR model. Now, we shall plug this function into our integral and solve for $T_G$:

$$T_G = \int_0^\infty t\gamma e^{-\gamma t} dt$$

(6)

$$= \lim_{M \to \infty} \gamma \int_0^M te^{-\gamma t} dt$$

Using Integration by Parts, we choose $u = t$, $dv = e^{-\gamma t} dt$, $du = 1 dt$, and $v = -\frac{e^{-\gamma t}}{\gamma}$.

$$= \lim_{M \to \infty} \left( -te^{-\gamma t} \bigg|_0^M + \int_0^M e^{-\gamma t} dt \right)$$

$$= \lim_{M \to \infty} \left( -te^{-\gamma t} \bigg|_0^M - \frac{e^{-\gamma t}}{\gamma} \bigg|_0^M \right)$$

$$= \lim_{M \to \infty} \left( \frac{-e^{-\gamma t} (\gamma t + 1)}{\gamma} \bigg|_0^M \right)$$

$$= \lim_{M \to \infty} \left( \frac{-\gamma t - 1}{\gamma e^{\gamma t}} \bigg|_0^M \right)$$

Now we apply L’Hôpital’s Rule due to the indeterminate form the bounds create.

$$= \lim_{M \to \infty} \left( \frac{-\gamma}{\gamma^2 e^{\gamma t}} \bigg|_0^M \right)$$
\[
\lim_{M \to \infty} \left( \frac{-1}{\gamma} \right) \\
= \lim_{M \to \infty} \left( \frac{-1}{\gamma e^{\gamma t}} + \frac{1}{\gamma e^{\gamma t_0}} \right) \\
= \frac{-1}{\gamma e^{\infty}} + \frac{1}{\gamma} \\
= 0 + \frac{1}{\gamma} \\
= \frac{1}{\gamma}
\]

Therefore, we find that for any SIR model, \( T_G = \frac{1}{\gamma} \). Thus Heesterbeek’s estimations become \( R_0 = e^{\frac{r}{\gamma}} \), or \( R_0 = 1 + \frac{r}{\gamma} \) for small values of \( \frac{r}{\gamma} \). Again, we must note that since \( r \) and \( \gamma \) both have units of “people per time”, both of Heesterbeek’s estimations are dimensionless. We shall consider both of these equations, and assess if they are reasonable ways to estimate the basic reproductive number of an SIR model.

### 3.2.3 Exponential \( R_0 \)

Lastly, we shall consider one more way of calculating \( R_0 \). The estimation is \( R_0 = e^r \), where \( r \) is the exponential growth estimation term. It is obvious that this estimation is inaccurate by the fact that \( e^r \) is not dimensionless. We consider this estimation to investigate the relationship between \( R_0 \) and \( r \) implied by the media.
4 Experimental Results

4.1 Experiential Design

To analyze our results, we created several plots illustrating the correspondence between $R_0$ and the estimates of $R_0$. This was done by generating SIR models where the value of beta was fixed, and $R_0$ was changed in increments of 0.5. The first convex interval was exponentially fit, and the resulting exponential growth rate was used to compute the estimated $R_0$. These estimates were plotted against the true $R_0$ used in their generation.

Figure 2 demonstrates this experimental design by showcasing nine sample plots. These plots have been generated with a variety of beta values, listed below each image. The blue line in each plot represents perfect linear correspondence. This is the ideal relationship; when the estimates would be equal to their generating $R_0$ value. The yellow and purple lines display the Heesterbeek estimates $R_0 = 1 + \frac{\gamma}{\gamma}$ and $R_0 = e^{\gamma}$ respectively. The red line illustrates $R_0 = e^{\gamma}$, which is our hypothetical basic functional estimate.

4.2 Experiential Results

First, let us examine the yellow Heesterbeek estimate, $R_0 = 1 + \frac{\gamma}{\gamma}$. This was the most accurate estimate out of the three tested, however, it was a constant underestimate for all beta values. It trended linearly, with a shallower slope than the direct linear correlation.

Next, we expound the second Heesterbeek estimate plotted in purple: $R_0 = e^{\gamma}$. This estimate increased exponentially as $R_0$ increased. It provided a consistent overestimate for all $\beta$ values examined.

Finally, we interpret the estimate analyzing the basic functional relationship plotted in red: $R_0 = e^{\gamma}$. This estimate grew almost linearly at an extremely slow pace for small values of $\beta$. Diversely, at larger values of $\beta$, it grew unpredictably, and diverged more aggressively as $R_0$ increased.

4.3 Discussion of Results

A few interesting observations are illustrated by the results from this experiment. The first is that the Heesterbeek estimates were not affected by changes made to the parameters beta and gamma with a constant $R_0$. The exponential growth rate changes proportionally with the chosen beta value, and thus inversely proportional to the gamma value due to how they are defined. This relationship can be seen in the consistency of the yellow and purple curves in all graphs of Figure 2, and sample of parameter relationships can be seen Figure 3.
(a) $\beta = .00025$

(b) $\beta = .0005$

(c) $\beta = .00075$

(d) $\beta = .0025$

(e) $\beta = .005$

(f) $\beta = .0075$

(g) $\beta = .025$

(h) $\beta = .05$

(i) $\beta = .075$

Figure 2: Estimates plotted along their $R_0$ values

<table>
<thead>
<tr>
<th>$R_0$</th>
<th>$\beta$</th>
<th>$\gamma = \frac{\beta S(0)}{R_0}$</th>
<th>$r$</th>
<th>$R_0 = 1 + \frac{r}{\gamma}$</th>
<th>$R_0 = e^{r\gamma}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.5</td>
<td>.0005</td>
<td>0.033</td>
<td>0.0145</td>
<td>1.4446</td>
<td>1.5598</td>
</tr>
<tr>
<td>1.5</td>
<td>.005</td>
<td>0.33</td>
<td>0.1482</td>
<td>1.4446</td>
<td>1.5598</td>
</tr>
<tr>
<td>1.5</td>
<td>.05</td>
<td>3.33</td>
<td>1.4819</td>
<td>1.4446</td>
<td>1.5598</td>
</tr>
<tr>
<td>3.5</td>
<td>.0005</td>
<td>0.0142857</td>
<td>0.0299</td>
<td>3.0938</td>
<td>8.1155</td>
</tr>
<tr>
<td>3.5</td>
<td>.005</td>
<td>0.142857</td>
<td>0.2991</td>
<td>3.0938</td>
<td>8.1155</td>
</tr>
<tr>
<td>3.5</td>
<td>.05</td>
<td>1.428571</td>
<td>2.9911</td>
<td>3.0938</td>
<td>8.1155</td>
</tr>
</tbody>
</table>

Figure 3: Table of Heesterbeek estimates along varied generating parameters
This relationship illustrates that each HeesterBeek estimate will produce equivalent estimations for all Kermack McKendrick models which have the same generating $R_0$. However, these estimates become less accurate as $R_0$ grows beyond the critical point $R_0 = 1$. Thus, these estimates would not be reasonable to use to obtain $R_0$ for any disease modelled using the Kermack McKendrick method that has $R_0 ≥ 1.5$.

The final estimate we proposed, $R_0 = e^r$, produced significant results. For SIR models produced with smaller $\beta$ values, the estimates were very close to 1, increasing at an almost negligible rate as the generating $R_0$ increased. However, when the beta value used to generate the SIR models was larger, we saw unpredictable growth in the estimates produced. This growth did seem to become faster for changing $R_0$ as beta values were increased. The described behavior can be observed in Figure 2: (g), (h), and (i).

The interpretation of these results led us to several conclusions. First, the Heesterbeek estimations are only reasonable to use for estimating the $R_0$ of SIR models with $R_0 > 1$ but very close to 1. Therefore, these estimates are only applicable for a small window of SIR models. Additionally, the unpredictable behavior displayed by the estimate $R_0 = e^r$ exemplifies that the parameters $R_0$ and $r$, exponential growth rate, do not have a basic functional correspondence which could be used to predict the $R_0$ of an SIR model accurately.

These findings show that while there are functional relationships between the exponential rate of the initial growth in an SIR model and its $R_0$, they do not correspond directly. This inference declares that our research supports our thesis. Thus, it is unsubstantiated to claim that one can estimate the $R_0$ of any SIR model by using its exponential growth rate.
5 Conclusion and Applications

5.1 Summary

The COVID-19 pandemic resulted in the public generating a high demand for information regarding epidemiology. With so many sources of news, issues with communication and validity arise. The epidemiological terms $R_0$ and exponential growth rate are often used interchangeably as a result of these complications. Our research aimed to investigate if an exponential function can be used to predict the $R_0$ of an SIR model. Through our simulations, none of the estimates obtained an accurate $R_0$ value through the wide variety of SIR parameters tested. Some estimates produced consistently inaccurate results, and furthermore, the estimate which modeled the direct functional relationship between $R_0$ and the exponential growth rate was unpredictable, and imprecise. Thus, we found it is unreasonable to assume that these estimates could be used to approximate $R_0$ for an SIR model. This result can provide insight into the independent nature of epidemic modelling techniques.

5.2 Applications

The results of this thesis can provide insight into the independent nature of epidemic modelling techniques. Since $R_0$ was defined by Kermack and McKendrick for the unique purpose of analyzing the threshold behavior in an SIR model, it is expected that we are unable to produce it using parameters from other modelling techniques. The results of our experiments can hold as support of this separation between modelling approaches. The specification of this disassociation between parameters has not been publicized as strongly as many other facets of epidemiology during the COVID-19 pandemic. An informed populous is essential to halt disease spread; differentiation between exponential growth and $R_0$ emphasizes the importance of social distancing and quarantine during a pandemic such as COVID-19. With a better distinction between the definitions of these model parameters, there might be an improved response to social restrictions put in place to combat disease growth.
6 Acknowledgments

The completion of this thesis was possible due to the contributions from the following people and organizations:

- Dr. Adam Boucher for advising this project.

- University of New Hampshire College of Engineering and Physical Sciences and The Department of Mathematics and Statistics.

- Landon Buell and Claire Ling for consultation on LaTeX and MATLAB, respectively.
References


