

Evaluating the Application of the SIR Model to a Localized Population

Alexandra Close

University of Maryland, College Park, Maryland, USA

Objective: Using data collected from the Johns Hopkins COVID-19 Repository, I investigated the reliability of the SIR (Susceptible-Infected-Removed) model.

Summary of Background Data: Modern pandemic responses have evolved into effective defenses when implemented correctly but present issues in epidemiological modeling efforts. Intervention strategies such as quarantining and masking limit population size (N), which affects the accuracy of modeling population-based rate systems especially in highly transmissible diseases.

Methods: I begin by reviewing the SIR model retroactively to the initial SARS-CoV-2 Wuhan strain. I compared the parameters available in the published literature (N = 2,717,000, β = 2, $\gamma = 1/14$) to the best-fitting SIR-yielded values by minimizing the root mean squared error function. Subsequently, I evaluated its predictive capabilities on the Delta variant using early surge data, which was later compared against a retroactive analysis.

Results: Using a least-squares error best-fit analysis allowed me to retroactively define remarkably accurate model parameters for the Wuhan waves. Parameters including N = 730, $\beta = 0.46$, $\gamma = 0.043$ in the first wave and N = 11,200, $\beta = 0.198$, $\gamma = 0.07$ in the second reflected effective intervention strategies. I show it is an effective predictive tool regarding the Delta variant, yielding parameters N = 50,900, $\beta = 0.87$, $\gamma = 1/3.7$ that proved accurate when compared with parameters from a full retroactive analysis (N = 60,000, $\beta = 0.94$, $\gamma = 1/3.6$).

Conclusion: The similarity of the yielded parameters in my results supports the SIR model's utility in epidemiological monitoring of high-transmissibility, low-mortality outbreaks vis-à-vis various containment measures.

Key words: SIR model – Epidemiology – COVID-19 – SARS-CoV-19 – Pandemic

SARS-CoV-2 is a coronavirus that emerged in formally recognized as an epidemic before the Wuhan, China, in December 2019. It was first World Health Organization classified it as a

Corresponding author: Alexandra Close, University of Maryland, 8700 Baltimore Avenue, 505-A, College Park, MD 20740, USA. Tel.: 301 980 8809; E-mail: altara2c@gmail.com

pandemic in March 2020 because of its rapid and widespread transmission.

The Susceptible-Infected-Removed (SIR) model is one of the most basic computational epidemiological models. The concept was first proposed by Sir Ronald Ross in 1902 to simulate malarial transmission and demonstrate how reducing the mosquito population below a given level was sufficient to control the epidemic.¹ The mentioned "given level" is otherwise denoted as the basic reproduction value, R0. The mathematics supporting this concept was expanded in the late 1920s when the Kermack-McKendrick² equation system used in this study was developed. These models have grown in complexity over time with the addition of more populations, a defined R0 value, and more refined mathematics.³ More recently, computational models have been used to model the Ebola, Zika, and severe acute respiratory syndrome (SARS) outbreaks.⁴

In this study, I analyzed the efficacy of this simplest of epidemiological models, the SIR model, as a reliable forecast tool for containment measures when various interventions such as masks, quarantine, and later vaccines were implemented. Would it work in an environment of containment efforts?

Methods

This SIR model performs a predictive analysis of an infectious disease's spread by dividing the population of a confined area into 3 different groups after assuming all members have homogeneous characteristics.

S = the susceptible population—those who have yet to be infected

I = the infected population—those who have caught the disease and are contagious

R = the removed population—those who are recovered and cannot be infectious again

Several limiting assumptions are applied to simplify the mathematics:

- 1. The epidemic begins with 1 person.
- 2. The population impacted is large and located in the same region, without significant land barriers so consistent mixing can be assumed.
- 3. All who are infected recover and become part of the "removed" group and cannot become reinfected. In other words, the recovered do not return to the susceptible population.

- 4. People are exposed to the disease at equal rates. This means the frequency and varied methods of social interaction all have the same impact.
- 5. Natural immunity is not present. Nobody is more or less vulnerable to the disease than another. For example, a child is just as likely to catch the disease as an adult.
- 6. The population remains constant for the epidemic's duration. In other words: zero births and zero deaths.
- 7. There is no fixed length of time that a person remains infected. All are assumed to have a fixed probability of recovering each day they remain infected.

When it is assumed no births or deaths, this can be written mathematically as follows:

$$S(t)+I(t)+R(t)\,{=}\,N,$$

where N = the total population.

To complete the SIR model, the following parameters are needed:

$$t = time in days$$

 $\beta = rate of transmission$

 $\gamma = rate of recovery$

S, I, and R are variables dependent on time *t*, and β and γ are constants defined by the parameters of the disease of interest. These variables and parameters form the 3 differential equations that outline the movement among the 3 populations.

$$\begin{split} S'(t) &= (-\beta S(t))/NI(t)\\ \mathrm{I}'(t) &= (\beta S(t)/N-\gamma)\mathrm{I}(t)\\ \mathrm{R}'(t) &= \gamma \mathrm{I}(t) \end{split}$$

To apply the SIR model to SARS-CoV-2 spread, data were obtained from the Johns Hopkins University COVID-19 Repository,⁵ which stored free and accessible transmission information for all reporting areas in the United States. Miami-Dade County, Florida, was chosen as the population of interest because of accessible and valid data; it is a localized point that controlled for geographic and policy exposure differences, and is a high-transmission area for all outbreaks, implying that there were sufficient data to apply the model. No permissions were necessary for the attainment of these data.





Fig. 1 Miami-Dade SARS-CoV-2 Wuhan initial peak compared against limited population SIR projection (±1 person).

N = 2,717000 $\beta = 2 (2 \text{ persons infected/person/day})$ $\gamma = 1/14 = 0.071 (1 \text{ person every 14 days recovers})$

It must be noted that in the population of interest, reporting switched from daily to weekly between the Wuhan and Delta outbreaks because of policy changes. This led to the Wuhan data having much higher raw data variability because of the daily nature of the reporting versus that seen in the Delta variant.

In this study, the SIR model was used in 2 ways: to retroactively describe SARS-CoV-2 Wuhan outbreaks and then to forecast the Delta variant. In both cases, estimated values of the initial parameters were used as a starting point before deriving the best-fit values. These estimated values were obtained via a meta-analysis of the literature to determine what a zero-intervention outbreak would look like. Given that interventions did occur, the best-fit parameters β , γ , and N were determined by minimizing the root mean squared error function between the actual and modeled values.

In all outbreaks, I compared my best-fit values with the baseline to evaluate the efficacy of containment policies in place.

In addition, for the Delta variant, I used data from the beginning stages of the outbreak to create a forecast of how I expected the outbreak to continue. This was compared with the parameters of an additional retroactively fitted SIR model to determine the efficacy of this modeling approach as a potential forecasting tool for health services and professionals to use in an environment in which various interventions could be applied.

Results

Setting a baseline in which no containment policies (masks, quarantine, and later vaccines were readily available) were in place for the initial Wuhan outbreak, and using SIR outbreak parameters derived from the literature meta-analysis ($\beta = 2,^6 \gamma = 1/14,^7 N = 2.7$ million) displayed an unsustainable outbreak that would overwhelm health services (Fig. 1).

Lowering the total susceptible population N (at t = 0) as a consequence of quarantining represented a more sustainable outbreak (Fig. 2), but it was still not a good approximation of the reality of the outbreak.

Although the population constant (N = 684) equalized the magnitudes of the peaks, the characteristics of the SIR I(t) curve itself still varied from the spread of the actual outbreak. The real I(t) curve was both slower to change and broader. I hypothesized that the reasons were as follows:

- **Slower rise:** The transmission rate on average was slower—lower β—due to intervention presence
- Slower fall: The recovery rate on average was slower—lower γ —due to the reported cases being more severe (on average severe COVID-19 infections took 8 days longer to fully recover⁷)

Now varying β , γ , and N in order to minimize the root mean squared error function to find the best-fitting parameters, I obtained the result in Fig. 3.







Both the β and N values were significantly lower than the standard baseline (2 and 2.7 million, respectively), signifying that the intervention methods were being accurately represented by the retroactive modeling. The γ value found, 0.043, was smaller than the literature 0.071. As γ equals the inverse of days to recover, the average recovery time was longer than expected for the Wuhan outbreak.

During the second Wuhan outbreak, a different set of interventions was in place, resulting in a significant increase in the initial population. As can be seen in Fig. 4, the SIR model was still effective at



Fig. 3 Best fit: Miami-Dade SARS-CoV-2 Wuhan initial peak compared against SIR projection (±1 person).

N = 730 $\beta = 0.46 \ (0.46 \text{ persons infected/person/day})$ $\gamma = 1/23 = 0.043 \ (1 \text{ person every } 23 \text{ days recovers})$



Fig. 4 Best fit: Miami-Dade SARS-CoV-2 Wuhan second peak compared against SIR projection (±1 person).

 $N \,{=}\, 11200$ $\beta \,{=}\, 0.198 \,\, (0.2 \,\, persons \,\, infected/person/day)$ $\gamma \,{=}\, 1/14 \,{=}\, 0.07 \,\, (1 \,\, person \,\, every \,\, 14 \,\, days \,\, recovers)$

finding a set of parameters that when applied, would create an accurate fit to the data.

A comparison between the 2 waves of Wuhan showed that the second outbreak was significantly more severe, reflecting the relaxation of multiple containment measures around this time frame. According to my SIR fitting in Fig. 4, the descriptive parameters for the real infected population reflected a significant increase in population despite a reduction in the β value. I hypothesized that the standardization of γ was due to less biased reporting or better therapy options.⁸

The retroactive analysis of Wuhan provided evidence that the SIR model is able to accurately describe



Fig. 5 Miami-Dade SARS-CoV-2 lifespan with Delta variant SIR prediction (±1 person).

N = 50900 $\beta = 0.87 (0.87 \text{ persons infected/person/day})$ $\gamma = 1/3.7 = 0.27 (1 \text{ person every } 3.7 \text{ days recovers})$



Fig. 6 Miami-Dade SARS-CoV-2 Delta variant magnified peak model (±1 person).

N = 60000 $\beta = 0.94 (0.94 \text{ persons infected/person/day})$ $\gamma = 1/3.6 = 0.28 (1 \text{ person every } 3.6 \text{ days recovers})$

an outbreak in spite of varying interventions in effect. To determine if it can accurately predict an outbreak with intervention methods, I modeled the Delta variant in the early stages of its outbreak (Fig. 5). Because of its severity, the magnitude of the infected population I graphed was much larger than either of the previous outbreaks mentioned, despite a large percentage of the Miami-Dade County population being at least partially vaccinated.

Reported parameters with no containment were used initially:

- $\geq \beta = 5$ (5 persons infected/person/day)⁶
- > Much higher viral loads
- Earlier start to, and longer duration, of shedding
- $\succ \gamma$ = No reported value likely due to vaccine prevalence

These parameter values were then adjusted using the same least-squares error minimization method as before but limited to a 2-month-long trend of data points to fit a prediction of the expected course of the outbreak (Fig. 5). This yielded the parameters $\beta =$ 0.87, $\gamma = 0.27$, and N = 50,900. These values still imply that containment methods were effective because β is far smaller than the baseline value of 5 and N is far smaller than the potential susceptible population. An important distinction is the low γ value, which most likely reflected the effect of vaccinations.

To analyze the accuracy of my prediction, I once again retroactively fitted the real infected population to compare parameters once the outbreak had passed (Fig. 6).

Int Surg 2024;108

Optimizing the fit via the parameters for the first month created a very good forecast. The values found from modeling I(t) movement were extremely close to those yielded from my prediction.

Discussion

Computational epidemiological models have proven to be very useful in understanding the nature of infectious disease spread. They are helpful tools when discussing time constraints and predicting the consequences and course of an epidemic. The SIR model is one of the simplest mathematical models and, as such, is considered a less accurate forecasting tool than other more sophisticated variants.

- 1. The constant population limits its efficacy in predicting these spikes, as the population of Miami-Dade fluctuated throughout the SARS-CoV-2 outbreak.
- 2. Susceptible-Infected-Removed does not account for Immunity, Exposure, and Death.
- 3. Reductions in populations may be due to immunities or reduced susceptibility in children rather than containment measures.
- 4. We cannot know the exact efficacy of containment measures through the SIR model as human behavior is also a variable the SIR model cannot fully estimate.
- 5. SIR models require at least some data for parameter estimation, and so are not good methods for

pure forecasting unless standard baseline parameters are already applicable.

Limiting assumptions are frequently used in modeling to remove factors that are quantitatively less significant and to keep the problem mathematically tractable. The difficulty comes in recognizing which can be removed without loss of predictive quality. Regarding the SIR model, the 5 preceding assumptions not only led to a simplified set of equations to solve but also, in many cases, do not negatively impact its predictive power.

In this article, I have applied the SIR model to the SARS-CoV-2 Wuhan and Delta outbreaks in Miami-Dade County, Florida, to analyze its efficacy in an environment of intervention methods. In reviewing the literature, various studies have examined the SIRbased modeling approach to the COVID-19 crisis. Some concentrate on the characteristics and consequences of SARS-CoV-2 itself and use the SIR model to provide quantitative data, as shown in Talukder's⁹ analysis of the COVID-19 crisis in Bangladesh. Others concentrate on improving the SIR model forecast, with Ajbar et al¹⁰ investigating a nonlinear removal rate and Cooper et al¹¹ adjusting the basic assumptions mentioned previously in this study. However, this study used the SIR model in its most basic form with no changes as an investigation into its forecast effectiveness in an environment in which various intervention methods are applied while also evaluating the efficacy of these containment measures by using a root mean squared error function. This is an important result since there is debate in the literature as to its suitability for modeling diseases such as SARS-CoV-2 with high transmission rates. What this report has shown is that, with certain constraints, it remains very useful if the appropriate data are available.

My analysis has shown that, for the Wuhan outbreak with successful intervention mechanisms, the SIR model, using a root mean squared error function to adjust the parameters from baseline, is an effective tool for analysis. It also showed for the Delta variant scenario an accurate forecasting ability when using initial data from an outbreak to predict the future movement of an infected population. The accuracy of the analytics and prediction was surprising and may have been influenced by the characteristics of SARS-CoV-2 itself. The virus has an incredibly high transmission rate (minimizing exposure variances) compared with a relatively low mortality rate (minimizing deaths) once effective treatments are in place. The SIR model does not account for exposure or death, and perhaps would not be as accurate in predicting diseases that have larger exposure differences or a higher mortality rate. In these cases, more complicated variations of the model such as the SEIR, SIRD, or SEIRD models (E = Exposure, D = Death) may be more applicable.

Despite this, the SIR model remains a promising and easily deployable tool even when interventions are in effect, given that baseline parameters are defined especially in terms of clinical and administrative policies. For example, in the surgical sector, elective surgical procedures had a median reduction of 54% compared with the pre-COVID-19 period.¹² Emergency and oncological surgeries were also reduced, although less significantly.¹² Having effective mathematical models for infectious diseases enables improved preparation and implementation of intervention policies, which in turn reduces the likelihood that surgery and care are delayed. Knowing which model is effective in which scenarios is a key first step to this. For example, more rigorous intervention methods could be used for diseases with larger β values before widespread infection, thereby protecting public health.

Although this marks the end of the scope of this work, it may be interesting to investigate how early the SIR model can be applied and remain accurate for forecasting. In this article, when the model was applied there was a significant amount of data on the parameters of each strain due to research on transmissibility. There was also already a trend in place for how well intervention mechanisms were being executed in limiting the population. Whether or not the SIR model would remain an accurate forecasting tool without needing additional information remains a question that needs further exploration.

© 2024 Close; licensee The International College of Surgeons. This is an Open Access article distributed under the terms of the Creative Commons Attribution Noncommercial License which permits use, distribution, and reproduction in any medium, provided the original work is properly cited, the use is noncommercial and is otherwise in compliance with the license. See: http:// creativecommons.org/licenses/by-nc/3.0

References

- Bacaër N, Ross and malaria (1911). In: A Short History of Mathematical Population Dynamics. *Springer*, London. 2011, https://doi.org/10.1007/978-0-85729-115-8_12
- Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 1927;115(772):700–721. Available at: www.jstor. org/stable/94815. Accessed April 10, 2020
- Brauer F. Mathematical epidemiology: past, present, and future. Infect Dis Model 2017;2(2):113–127

Downloaded from https://prime-pdf-watermark.prime-prod.pubfactory.com/ at 2025-07-07 via Open Access. This is an Open Access article distributed under the terms of the Creative Commons Attribution Noncommercial License which permits use, distribution, and reproduction ... https://creativecommons.org/licenses/by-nc/3.0/

- Saxena R, Mahipal J, Vikrant B. Propagation analysis of COVID-19: an SIR model-based investigation of the pandemic. Arabian Journal for Science and Engineering 2021;1–13
- Johns Hopkins Coronavirus Resource Center. COVID-19 map. Available at: coronavirus.jhu.edu/map.html. Accessed April 11, 2020
- Liu Y, Rocklöv J. The reproductive number of the Delta variant of SARS-CoV-2 is far higher compared to the ancestral SARS-CoV-2 virus. J Travel Med 2021;28(7):taab124
- 7. Baristaite B, Gurwitz D. Effects of age and sex on recovery from Covid-19: analysis of 5769 Israeli patients. *J Infect* 2020
- Dos Santos WG. Natural history of COVID-19 and current knowledge on treatment therapeutic options. *Biomed Pharmac*other 2020;**129**(9):110493

- 9. Talukder A. Susceptible-Infectious-Recovered (SIR) modelbased forecasting of COVID-19 outbreak in Bangladesh. *Int J Clin Pract* 2020;**74**(11):e13648
- Ajbar A, Alqahtani RT, Boumaza M. Dynamics of an SIR-Based COVID-19 Model With Linear Incidence Rate, Nonlinear Removal Rate, and Public Awareness. *Front. Phys.* 2021;9:634251. doi: 10.3389/fphy.2021.634251
- Mondal A, Antonopoulos C. A SIR model assumption for the spread of COVID-19 in different communities. *Chaos Solitons Fractals* 2020;**139**:110057
- Mylonakis A, Kalfoutzou A, Panagakis A, Despotidis M, Yfantopoulos J. The impact of the COVID-19 pandemic on surgical activities: a single-center experience and literature review. *Cureus* 2022;14(10):e30785