A Mathematical Model for Forecasting the Spread of COVID-19 in Pennsylvania

Jack DeGroot, Heather Kwolek, Cody Dosch, Craig Johnson June 26, 2021

The use of mathematical methods to predict the spread of infectious diseases goes back at least to the work of mathematician Daniel Bernoulli in 1760 on smallpox, which at that time was ravaging Italy and Europe. He was the first to create the prototype of the SIR (Susceptible-Infectious-Recovered) model that uses calculus and a 3 x 3 system of first-order nonlinear differential equations. This paper expands that model by adding 2 functions corresponding to Vaccinated persons and asymptomatic Exposed persons. It is known as the **SVEIR** model (Gumel et al. 2006) first used to examine the impact of a vaccine on the spread of the SARS virus in Canada. The important difference in our model is that asymptomatic carriers exist and are considered to be able to transmit the Covid-19 virus, while this was not believed to be true for SARS. Other models (SEIR and SuEIR) have been extensively studied by other universities e.g. MIT, Princeton, University of Illinois, UCLA, University of Geneva, and others to aid in predicting Covid-positive cases and Covid deaths in America and the world. In the remainder of this paper, the following functions are used.

$S(t)$ = number of susceptible people in a fixed population at time t (in days).
$V(t)$ = number of vaccinated people at time t.
$E(t)$ = number of asymptomatic infected people at time t. They can pass the disease to susceptible people and be recovered in a specific period.
$I(t)$ = number of new symptomatic infected people at time t. They can pass the disease to susceptible people and be recovered in a specific period.
$R(t)$ = number of recovered people at time t. We assume they acquire immunity and are no longer susceptible and so no feedback loop returns to S from R.

The rates of transfer between compartments are shown in the following schematic diagram with the parameters defined in Table 1.

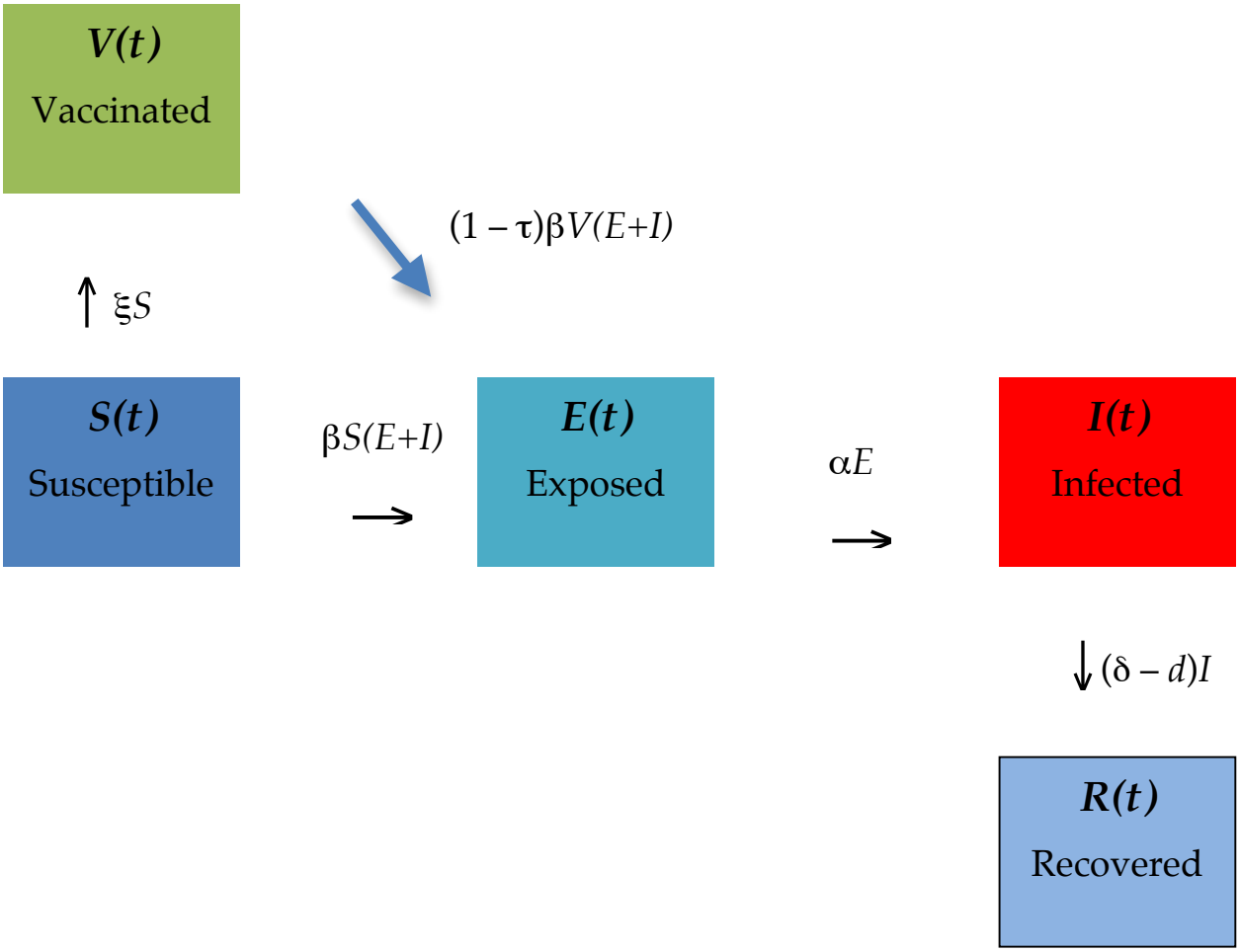

Figure 1. Rates of infections traced from susceptible to recovered.

Parameter	Description	Value
β	Effective contact rate	R_0 δ
ξ	Vaccination coverage rate	0.00642
τ	Vaccine efficacy	0.8 – 0.95
α	Rate of development of symptoms	0.6
δ	Recovery rate	0.10
d	Disease-induced mortality rate	0.015

Table 1. Parameter Descriptions

Figure 1 corresponds to the following 5 x 5 nonlinear system of first-order differential equations.

$$\frac{dS}{dt} = -\beta S(E+I) - \xi S$$
$$\frac{dV}{dt} = \xi S - (1-\tau)\beta V(E+I)$$
$$\frac{dE}{dt} = \beta S(E+I) + (1-\tau)\beta V(E+I) - \alpha E$$
$$\frac{dI}{dt} = \alpha E - (\delta - d)I$$
$$\frac{dR}{dt} = (\delta - d)I$$

We used the 4th order Runga-Kutta numerical method to solve this system for various values in the ranges of the parameters. (The program code is included at the end.) Here the total population N is assumed to be constant and at any time

$$N = S(t) + V(t) + E(t) + I(t) + R(t)$$

Determining values for the constant parameters is the key to enable this model to give reasonable forecasts. The value of t he recovery rate δ is computable because in (Li, 2018) it is shown that infectious period is exponentially distributed. Therefore, the density function giving the distribution of I is $f(t) = \delta e^{-\delta t}$ for $t \geq 0$ (0 elsewhere) and so the expected value $E[X]$, or mean, of the associated random variable X is given by the improper integral:

$$E[X] = \int_{-\infty}^{\infty} t f(t)\,dt = \int_{-\infty}^{\infty} t\,\delta e^{-\delta t}\,dt$$
$$= \lim_{a\to\infty} \int_{0}^{a} t\,\delta e^{-\delta t}\,dt$$
$$= \lim_{a\to\infty} [-ae^{-\delta a} - \frac{1}{\delta}(e^{-\delta a} - 1)] = \frac{1}{\delta}.$$

Therefore the mean infectious period $D = E[X] = \frac{1}{\delta}$. The most recent value for the mean infectious period according to CDC is $D = 10$ days. [**3**] So we have used $\delta = 1/10 = 0.10$.

The vaccine coverage rate ξ is the distribution of vaccines over the area of Pennsylvania. We arrived at this number by taking the current number of partially and fully vaccinated individuals (as of March 31, 2021) and divided it by the number of days the vaccine was available to individuals in Pennsylvania up to that same date. With this total we then divided by the total population of Pennsylvania minus the individuals who have already been partially and fully vaccinated. At the time of our model, we use

$$\xi = [((5{,}154{,}718 / 105)) / (12{,}802{,}000 - 5{,}154{,}718)] = 0.00642$$

Vaccine efficacy is the effectiveness of a vaccine. This number indicates the percentage of reduction a vaccinated individual is from becoming infected. Since three predominant vaccinations for Covid-19 are currently being used, there is not just one efficacy. The Pfizer vaccine was found to have an efficacy of 95%, the Moderna 94.1%, and the Johnson & Johnson single shot 85%. We cannot take a simple average as the shots are not being given at the same rate. As of February 2021, the Moderna has supplied the United States with the majority of vaccinations with 41 million. Moderna is followed by Pfizer with 29 million, and Johnson & Johnson has supplied 20 million. Assuming each vaccination continues to be produced and distributed at the same rate, we can use these numbers as weights. So,

$$\tau = [41(.941) + 29(.95) + 20(.85)]/90 \approx .92$$

Finding β is trickier but one common technique is to make an estimate by fitting the model with actual confirmed cases (infected) over a particular time period. The major difficulty with this though is that the value is a moving target as various influencing factors keep changing on almost a daily basis. A second method is to find the value of the very important **basic reproduction number** R_0 which can be estimated by β/γ . So given R_0, we can use $\beta = R_0 \bullet \gamma$. R_0 is the average number of people infected from one other infected person. For example, a typical flu strain has an $R_0 = 1.2$. That means that, on average, 5 infected people will pass it on to 6 new people. For Ebola it was 2.0, for SARS it was 3.5, for smallpox in the past it was 7.0. Studies on early cases in China indicated the Covid R_0 to be between 2 and 2.5, but current

estimates for each country can be seen on the CDC website. (The goal, of course, is to reduce it to below 1.0, when the number of new cases then begins to decline.

The https://rt.live website (R_t means the same as R_0 in the literature) computed and posted the R_0 value for each state on a daily basis until the end of February 2021. We used these values for preliminary forecasts when the model was being fine-tuned. We must be aware that computing the value of R_0 is a very complex process with the parameters in constant flux. The CDC gave this caution at the end of a report on determining its value.

> R_0 is an estimate of contagiousness that is a function of human behavior and biological characteristics of pathogens. R_0 is not a measure of the severity of an infectious disease or the rapidity of a pathogen's spread through a population. R_0 values are nearly always estimated from mathematical models, and the estimated values are dependent on numerous decisions made in the modeling process. The contagiousness of different historic, emerging, and reemerging infectious agents cannot be fairly compared without recalculating R_0 with the same modeling assumptions. Because R_0 can be misrepresented, …, in a variety of ways that distort the metric's true meaning, … it must be applied and discussed with caution in research and practice. (https://www.cdc.gov/coronavirus/2019-ncov/index.html)

Several key factors affect the values of the parameters.

1. Asymptomatic carriers. Studies reported in the New York Times have indicated that a large majority of young people show no signs of the disease when they contract it. In Iceland where most of the population was tested, 50% of those that tested positive experienced no symptoms. The question is: What percent of infections are caused by asymptomatic carriers?

2. Airborne transmission from atomization. The vast majority of studies in the past year identified airborne transmission from human atomization (coughing and breathing) as the dominant mode of transmission of COVID-19 (Zhang, 2020).

3. Transmission by vaccinated people. Since the efficacy of vaccines treating the disease is not 100%, there will be individuals who are vaccinated that become infected. Further, since not all infections become symptomatic, it is possible for a vaccinated individual to become an asymptomatic carrier of the disease. At this writing, is not precisely known how vaccines affect the potential to spread the Covid-19 virus.

SVEIR model output (December 16, 2020 - March 31, 2021)

Parameter Values:

N = 12,794,971 (population of Pennsylvania on Dec. 16, 2020)

R_o = 0.981468 (average of available values from 12/16/20 to 01/26/21)

$\frac{\beta}{N} = \frac{R_o\gamma}{N} = 7.671 \times 10^{-9}$ $\xi = 0.00642$ $\tau = 0.92$ $\alpha = 0.6$

$\delta = 0.10$ $d = 0.015$

Date	**12/16/21 (t = 0)**	**03/31/2021 (t = 105)**
$S(t)$	12,260,574	5,892,878
$V(t)$	0	5,768,082
$E(t)$	61,859	435
$I(t)$	92,789	4379
$R(t)$	426,144	1,129,197

Table 2. Output of SVEIR Mathematical Model

Data from New York Times as of March 31, 2021:

New Infections: 4457

Approximate Vaccinations: 6,060,000

Approximate Recovered: 987,356

Total Infected to date: 1,024,149 to be compared to $E(105) + I(105) + R(105) = 1,134,011$

Percent Error:

New Infected $I(t)$: 2%

Vaccinated $V(t)$: 4%

Recovered $R(t)$: 14%

Combined (E + I + R): 11%

This is an ongoing research project with several purposes. We wish to maintain predictions of the total number of Covid positive test cases in Pennsylvania within a relative error of 5%. Then when further data has been published, we wish to calibrate the value of the reproduction number R_0 for use in successfully forecasting this important value for any resurgence of the Covid-19 virus in the future. Finally, we would like to explore whether dates of peak infections can be predicted.

Program Source Code:

```
// Array-based implementation of Runge Kutta
1 reference
public double[] RunRungeKutta(double t, double[] currentValues)
{
    // Calculate slopes
    var kn1 = derivatives(t, currentValues);
    var kn2 = derivatives(t + (h / 2), currentValues.Zip((kn1.Select(e => e * (h / 2.0)).ToArray()), (x, y) => x + y).ToArray());
    var kn3 = derivatives(t + (h / 2), currentValues.Zip((kn2.Select(e => e * (h / 2.0)).ToArray()), (x, y) => x + y).ToArray());
    var kn4 = derivatives(t + h, currentValues.Zip((kn3.Select(e => e * h).ToArray()), (x, y) => x + y).ToArray());

    // Calculate new array of values to return
    var newValues = currentValues.Zip(((kn1.Zip((kn2.Select(e => e * 2.0).ToArray()), (x, y) => x + y).ToArray().Zip((kn3.Select(e => e * 2.0)
        .ToArray()), (x, y) => x + y).ToArray().Zip(kn4, (x, y) => x + y).ToArray())).Select(e => e * (h / 6.0)).ToArray(), (x, y) => x + y).ToArray();

    // Return the new values
    return newValues;
}
```

Screenshot 1. Runge-Kutta function

```
4 references
private double[] derivatives(double t, double[] currentValues)
{
    // Get the value for S(t), V(t), E(t), I(t), and R(t)
    // NOTE: Indexes - 0 = S, 1 = V, 2 = E, 3 = I, 4 = R
    var tempS = (-1 * BETA * currentValues[0] * (currentValues[2] + currentValues[3])) - (EPSILON * currentValues[0]);
    var tempV = (EPSILON * currentValues[0]) - ((1 - TAU) * BETA * currentValues[1] * (currentValues[2] + currentValues[3]));
    var tempE = (BETA * currentValues[0] * (currentValues[2] + currentValues[3])) + ((1 - TAU) * BETA * currentValues[1] * (currentValues[2] + currentValues[3])) - (ALPHA * currentValues[2]);
    var tempI = (ALPHA * currentValues[2]) - (DELTA * currentValues[3]) - (D * currentValues[3]);
    var tempR = (DELTA * currentValues[3]) + (D * currentValues[3]);

    // Store updated values in a new array
    double[] updatedValues = { tempS, tempV, tempE, tempI, tempR };

    // Return updated values
    return updatedValues;
}
```

Screenshot 2. Derivative function

WORKS CITED

Li, Michael Y. 2018. *An Introduction to Mathematical Modeling of Infectious Diseases*. Springer International Publishing.

Harko, T., F. Lobo, and M. K. Mak. 2014. *Exact Analytical Solutions of the SIR Epidemic Model*. arXiv:1403.2160v1, March 10.

Gumel, A. B., C. Connell McClusky, and James Watmough. 2006. A SVEIR Model for Assessing Potential Impact of an Imperfect Anti-SARS Vaccine. *Mathematical Biosciences and Engineering*, Vol. 3, Num. 3.

Zhang, Renyi, Yixin Li, Annie L. Zhang, Yuan Wang, and Mario J. Molina. 2020. Identifying airborne transmission as the dominant route for the spread of COVID-19. *Proceedings of the National Academy of Sciences of the United States of America*, June 11. https://doi.org/10.1073/pnas.2009637117