Int. J. Math. And Appl., 9(3)(2021), 105–110
ISSN: 2347-1557
Available Online: http://ijmaa.in/

International Journal of *Mathematics And its Applications*

The Effect of Model Type on the Accuracy of Predicted Values

Pranav Bandla[1,*]

1 Mills E Godwin High School, Henrico, Virginia, United States.

Abstract: The purpose of this experiment was to determine which model would be the most effective in predicting disease progression given two months of data. If which model is more accurate is known, governments can develop and use the models to respond to disease more effectively. The SIR and SEIR are commonly used compartmental models for diseases. The SIR model incorporates another factor. It was hypothesized the values predicted by the SEIR model would be more accurate than the values predicted by the SIR model. The experiment was conducted by first calculating an SEIR model and an SIR model based on data from the first 2 months of 3 different Ebola outbreaks in 2014. The actual values of data were compared against the predicted values of data by the SIR and SEIR value by use of percent error. The SEIR model had a lower mean percent error than the SIR model, which supports the hypothesis. A t-test performed on the data revealed it was not significant. The null hypothesis of no difference in the accuracy of the data predicted by the SIR model and the data predicted by the SEIR model failed to be rejected. It is believed there is no difference in the accuracy of predictions of the SIR model vs the SEIR model. This may be because there are outside factors such as peoples responses and non-disease related death and births. Research can be done into more adaptable models that are able to accommodate birth and death rates and effects of the responses of people.

Keywords: Disease Modeling, SIR model, SEIR model, Infectious Diseases.

1. Introduction

Due to modernization, diseases are not as common and dangerous as they have been in the past. Diseases like smallpox and polio have been eradicated by vaccines. This has led to people underestimating the impact of such diseases and undermining preventive measures like masks and vaccines. Recently, the Covid-19 pandemic crippled nations and disrupted the lives of everyone, but this could have been prevented. Models can predict the effects of infectious diseases by using inputs like population size, number of people infected, and the contagiousness of a disease. This information can be used to predict what effect an action may have on preventing the spread of a disease. Governments can use this information to make scientifically supported decisions to protect their citizens and prevent an outbreak or epidemic from ever becoming a pandemic.

Compartmental models are a category of disease models. The models of this category consist of groups between which a population is divided into. In this experiment, the SIR and SEIR models were used. The SIR model consists of the categories Susceptible, Infected, and Removed. The Susceptible category includes all the members of the population not included in the Infected or Removed categories. The Infected category includes people currently contagious with the disease while the Removed category includes people who have either recovered or died. If a person becomes infected, they would move from the susceptible category to the infected category, and eventually the removed category [2]. Under the SIR model,

* *E-mail: pbandlas@gmail.com*

it is assumed this is the only pathway, so the model is only applicable for diseases with lasting immunity. The modeled values for the number of people in each category over time for the 2014 Ebola Virus pandemic is shown in Graph 1. This reflects the typical changes in the values for the categories in an SIR model. Normally, the Susceptible category starts high, decreases sharply, and stabilizes. The Infected category starts low and increases sharply at the same time as the decrease in the Susceptible category. It then peaks and decreases. The Recovered category starts low and increase sharply at the same time as the Infected category. It then stabilizes after the Infected category decreases. The change in each category of the SIR model is dictated by 3 differential equations listed below.

$$\frac{dS}{dt} = -\frac{\beta IS}{N}$$
$$\frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I$$
$$\frac{dR}{dt} = \gamma I$$

These equations use time as a variable and have two constants: gamma and beta. The gamma value represents the average rate of removal and is the reciprocal of the weighted average of the average time until death and average time period of contagiousness. The beta value represents the transmission rate [6].

The SEIR model contains the categories of the SIR model, and an additional category: Exposed. After exposure to most diseases, there is a period for which the disease is latent. During this period, people are not contagious. People for which the disease is latent are placed in the exposed category. In the SEIR model, people move from Susceptible category to the Exposed category upon infection. The differential equations for the categories of the SEIR model are listed below [4].

$$\frac{dS}{dt} = -\frac{\beta IS}{N}$$
$$\frac{dE}{dt} = \frac{\beta IS}{N} - \sigma E$$
$$\frac{dI}{dt} = \sigma E - \gamma I$$
$$\frac{dR}{dt} = \gamma I$$

These equations have three constants: gamma, beta, and sigma. The gamma and beta values represent the same values as the gamma and beta values in the SIR model. The sigma value represents the rate of individuals becoming infectious and can be calculated by taking the inverse of the average latency period [3].

The purpose of this experiment is to determine which model is best at predicting the outcome of the disease, given the same data. The hypothesis is if the SEIR model is used, its predicted values will be more accurate than the SIR model, because the SEIR model is able to account for the incubation period present in many disease [8]. The Independent variable is the type of model, and the levels are SIR and SEIR model for outbreaks, which model the 2014 Ebola outbreak in Guinea, Sierra Leone, and Liberia. This will be done by using data from the first 2 months of each outbreak and general information such as latency period, mortality rate, average period of contagiousness, and average time until death. Data will be considered three, four, five, six, and seven months after the outbreak because the data from the first two months will be used to create the model, and after 7 months, there will most likely be efforts to end the epidemic, making the model invalid. The SIR model was used as a control, as it is considered similar as it is simpler [8]. The dependent variable is percent error between the expected value of total cases based on the models and the reported data provided by the World Health Organization. This is found by taking the difference between the predicted value and the actual value and dividing it by the predicted value [7].

2. Methods and Materials

Data from 3 outbreaks were compiled into a spreadsheet. The information was collected from World Health Organization reports and various peer reviewed sources. Next, for each disease, the average transmission rate, latency period, and average duration of infection were found from peer reviewed literature. Finally, initial case numbers were taken from the compiled data. By inputting this information into the equations that dictate the change in each category of the SIR model, the SIR model was created for each outbreak. Next, the incubation rate was also researched and noted for each disease. By inputting these values into the differential equations of the categories of the SEIR model, the SEIR model was created for each outbreak by inputting the values for the required constants into the. The list of constants for each disease was listed in table 3. The predicted values from each model for three, four, five, six, and seven months after the outbreak were compared against the actual values using percent error as the statistic. These values were listed in table 1. Percent error was calculated by finding the difference between the predicted value and the actual value and dividing by the predicted value. Statistical analysis was done on this and listed in table 2. A t-test was used to calculate the significance of the data.

3. Results

The accuracy of the SIR and SEIR models was tested and the results of the statistical analysis are shown in table 2. A research hypothesis was formulated that if the SEIR model was used, it would be more accurate. The mean percent error was calculated for each model. The mean percent error for the SIR model (39.5%) was slightly higher than the mean percent error for the SEIR model (35.7%). This implies the type of model has an effect on the accuracy of the prediction. As the mean percent error for the SIR model is higher than the mean percent error of the SIR model, this implies the SEIR model is more accurate than the SIR model, as predicted by the research hypothesis. The variance and standard deviations were also calculated for each level of the Independent Variable. The standard deviation was high indicating the data was most likely spread out. None of the data points were outside the 2 standard deviation range for either level of independent variable, indicating there are likely no outliers.

A t-test was performed on the data at a significance level of 0.05 with 28 degrees of freedom. The null hypothesis was there is no difference in the percent accuracy of the predictions by the SIR model and the SEIR model. The calculated t value (0.346) was lower than the table t value of 2.048. This implies the null hypothesis cannot be rejected and there was no significant difference between the accuracy between the values calculated by the SIR model and the SEIR model. The probability of the results being due to chance is larger than 0.05 and implies the difference in percent accuracy determined in the experiment was most likely not due to the difference in model.

4. Discussion and Conclusions

The purpose of this experiment was to determine whether there was a difference in the accuracy of values predicted by the SEIR and SIR models. Using data from the first two months of the 2014 Ebola epidemic in Guinea, Sierra Leone, and Liberia, an SIR model and an SEIR model were formed. A research hypothesis was formulated that the future values predicted by the SEIR model would be more accurate than the values predicted by the SIR model. As the SEIR model had a lower mean percent error, it was implied the SEIR model was more accurate in predicting the outcome of an epidemic than the SIR model. A t-test was performed on the data. It revealed the data for percent accuracy of the SIR and SEIR models were not statistically significant. This means the difference in accuracy was probably due to chance. This is implies the addition of the exposed category does not significantly affect the accuracy of the model.

The results may be because other factors may largely influence the model. Firstly, not all people may report they have the virus, making the actual case count higher. In addition, birthrates and non-disease related death rates are not being considered [8]. There are also many social factors, such as staying at home when sick or meeting with others, that affect the spread of the disease. The SIR and SEIR models are too simple to accommodate all these factors. There were some sources of error in this experiment. Values such as transmission rate, period of latency and period of contagiousness were taken from peer reviewed sources. However, different sources gave different values for the coefficients. In addition, measures to reduce the effect of the virus may have been used. This would result in the calculated SIR and SEIR models to be inaccurate. For continued study, such models should be used with controllable population with little immigration and emigration. Alternative models that adapt to birth and death rates and that can easily adjust to preventative measures should be developed and used.

Acknowledgements

I would like to thank Dr. Guillermo Goldsztein for introducing me to the topic and his guidance on the paper.

References

[1] ———, *Case Counts Error processing SSI file*, (2020, February 19), Retrieved January 08, 2021, from https://www.cdc.gov/vhf/ebola/history/2014-2016-outbreak/case-counts.html

[2] T. Johnson and B. McQuarrie, *Mathematical modeling of diseases: Susceptible-infected-recovered (sir) model*, University of Minnesota, Morris, Math 4901 Senior Seminar, (2009).

[3] ———, *SEIR and SEIRS models (n.d.)*, Retrieved January 08, 2021, from https://docs.idmod.org/projects/emod-hiv/en/latest/model-seir.html

[4] C. L. Althaus, *Estimating the Reproduction Number of Ebola Virus (EBOV) During the 2014 Outbreak in West Africa*, PLoS Currents, doi:10.1371/currents.outbreaks.91afb5e0f279e7f29e7056095255b288

[5] A. S. Bagbe, *Statistical Analysis of Ebola Virus Disease outbreak in Some West Africa Countries using S-I-R Model*, Annals of Biostatistics & Biometric Applications, 2(3)(2019), doi:10.33552/abba.2019.02.000540

[6] G. Chowell and H. Nishiura, *Transmission dynamics and control of Ebola virus disease (EVD): A review*, BMC Medicine, 12(1)(2014), doi:10.1186/s12916-014-0196-0

[7] D. Mayer and D. Butler, *Statistical validation*, Ecological Modelling, 68(1-2)(1993), 21-32.

[8] J. Tolles and T. Luong, *Modeling Epidemics With Compartmental Models*, JAMA, 323(24)(2020), 2515-2516.

[9] G. E. Velásquez, *Time From Infection to Disease and Infectiousness for Ebola Virus Disease, a Systematic Review*, Clinical Infectious Diseases, 61(7)(2015), 1135-1140.

Appendix

EDD

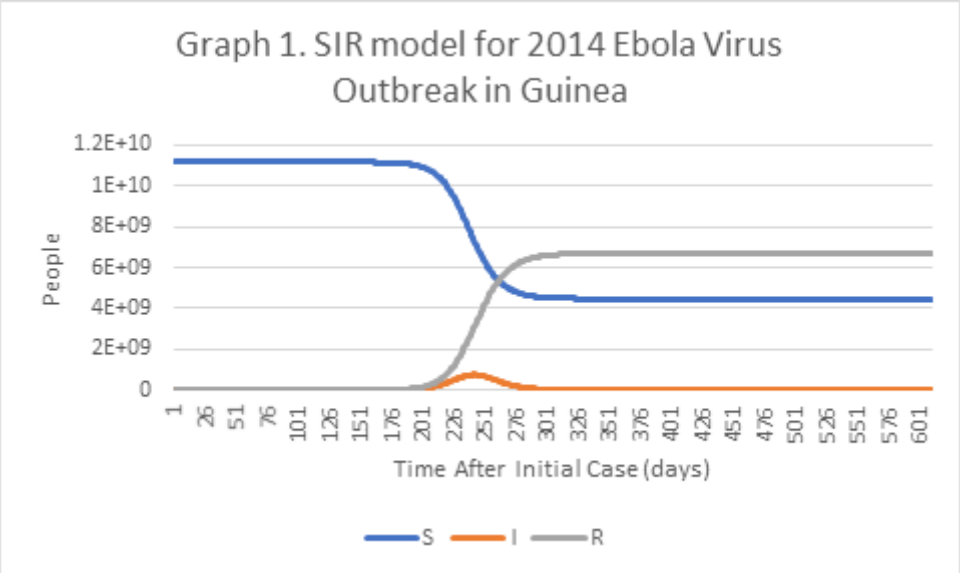
Title: The Effect of Model Type on the Accuracy of Predicted Values.

Hypothesis: If the SEIR model is used to predict the future values of a disease based on data from the first 2 months of an outbreak, then it will be more accurate than the values predicted by the SIR model.

Independent Variable: Type of model	
SIR model (control)	SEIR model
15 trials	15 trials

Dependent Variable: Percent error between the values predicted by the SIR and SEIR models and the actual values. This will be calculated by subtracting the actual value by the predicted value and dividing by the predicted value.

Constants: Beta value, Gamma value, N value, timeframe of data used to create model, time of data collection after outbreak, disease modeled, and



Trial	Actual value	Predicted value (SIR)	Predicted value (SEIR)
1.	390	203	294
2.	460	312	462
3.	648	588	688
4.	1157	1250	1065
5.	1667	2413	1560
6.	1026	764	523
7.	2304	1934	1613
8.	5338	3362	4545
9.	7109	6591	8023
10.	9446	15991	18023
11.	107	268	78
12.	329	632	174
13.	1378	1093	727
14.	3696	1425	2539
15.	6535	7970	6823

Table 1. Table 1. List of Actual Value and Values Predicted by the SIR and SEIR Models for Total Cases of Ebola in Guinea, Sierra Leone, and Liberia

Descriptive Information	Type of Model	
	SIR	SEIR
Mean	39.5%	35.7%
Range	84.7%	95.8%
Maximum	92.1%	96.2%
Minimum	7.44%	0.433%
Variance	750%	1110%
Standard Deviation	274%	333%
1 SD	12.1%- 66.9%	2.4% - 69.0%
2 SD	0% - 94.3%	0% - 102%
3 SD	0% - 122%	0% - 135.6%
Number	15	15

Table 2. **Statistical Table for the Effect of Different Types of Models on Percent Error of Estimated Values**

Results of t-test: $t = 0.346$; $df = 28$; $\alpha = 0.05$; t of $0.346 < 2.048$; $p > 0.05$.

Country & epidemic	β	γ	σ	N
Guinea – Ebola Trials (1-5)	0.200	0.178	0.167	11,150,000,000
Sierra Leone – Ebola Trials (6-10)	0.217	0.178	0.167	7,017,000,000
Liberia – Ebola Trials (11-15)	0.206	0.178	0.167	4,360,000,000

Table 3. **List of constants used to create each model**