Modeling the 1995 Ebola Outbreak, Democratic Republic of the Congo

Robert Micatka
University of Michigan

December 19, 2013

Abstract

Using the compartmental epidemiological SIR model as well as an enhanced version to simulate the 1995 Ebola outbreak in the Bandundu region, Democratic Republic of the Congo. Two models are tested, the first is a basic SIR model and the second uses two different infectious classes, a high-risk and low-risk populations to attempt to elucidate any effect, if any, that close contact with infected patients, such as that experienced by health-care workers, had on the outbreak size overall as well as within those classes.

I. Introduction

The Ebola virus is an exceedingly deadly (the 1975 outbreak had a 90% fatality rate [1]) viral hemorrhagic fever and is named after the Ebola River in the Democratic Republic of the Congo where it was discovered during an outbreak in 1975. The namesake symptom of this disease is the coagulopathy the host shows, which can lead to hemorrhaging from the mucous membranes as well as hematemesis (vomiting blood) and hemoptysis (coughing blood). This diesase has had several outbreaks in the years since it was discovered, mostly in the Democratic Republic of the Congo although other nearby regions have also experienced outbreaks. While other species besides humans can be infected by Ebola this is thought to be incidental and not as a source of the virus. There are no confirmed natural reservoirs of the virus, although some research has suggested fruit bats as a probable source [1].

There is no known vaccine for Ebola with the only prevention is to avoid contact with an infected individual's bodily fluids. However, this prevention method requires modern medical equiptment and procedures which are scarce in the region hit most often by this disease.

This disease has struck in only small numbers and although the fatality rate is extraordinarly high the incidence of the disease is not. This means that the model created had a limited data-set and thus inaccuracies are expected. However, the model should be able to shed some light on a potential future outbreak and what could be expected from such an outbreak.

II. Methods

The Susceptible-Infected-Recovered (SIR) Model is a type of compartmental epidemiological model that compartmentalizes a population into three distinct groups. This model was chosen over other compartmental models because no individual during an epidemic has recovered and become reinfected. Population demography was not used due to the short time-span of the outbreaks, the population change during the outbreak time will not affect the outcome noticeably. The coupled differential equations used for the SIR model are as follows:

$$\frac{dS}{dt} = -\beta SI$$
$$\frac{dI}{dt} = \beta SI - \gamma I$$
$$\frac{dR}{dt} = \gamma I$$
$$N = S(t) + I(t) + R(t)$$

A two-class SIR model was explored in due to the data consisting of two seperate group of individuals, health care workers and non-health care workers. The health care workers would come into contact with infected individuals more often than the non-health care workers and so are at a higher risk of infection. However, unlike other high-risk groups of other diseases, health care workers do not have a higher contact rate with uninfected non-health care workers and so do not neccessarily increase the infection rates of the lower-risk group. The coupled differential equations for the two-class SIR model are as follows:

$$\frac{dS_H}{dt} = -(\beta_{HH} S_H I_H + \beta_{HL} S_H I_L)$$
$$\frac{dS_L}{dt} = -\beta_{LL} S_L I_L$$
$$\frac{dI_H}{dt} = \beta_{HH} S_H I_H + \beta_{HL} S_H I_L - \gamma I_H$$
$$\frac{dI_L}{dt} = \beta_{LL} S_L I_L - \gamma I_L$$
$$\frac{dR_H}{dt} = \gamma I_H$$
$$\frac{dR_L}{dt} = \gamma I_L$$
$$S(t) = S_H(t) + S_L(t)$$
$$I(t) = I_H(t) + I_L(t)$$
$$R(t) = R_H(t) + R_L(t)$$
$$N = S(t) + I(t) + R(t)$$

The data used in this model is from surveillance case-reports as analyzed by Khan, et al[3]. This incidence data consisted of the number of new cases during the course of the outreak broken down into health care workers and non-health care workers.

The data was analyzed by calculating the outbreak over size (using a recovery rate calculated by Astacio et. al. from a previous outbreak [Astacio]) and then applying a linear-regression model using MATLAB software in order to create starting points for the estimation of the appropriate parameters for the SIR and the two-class SIR models used.

The linear-regression model consisted of logarithmic-normalizing the data and finding the slope of the early data points. After the starting parameters were found they were optimized using utilizing MATLAB's minimum-search function. The coupled differential equations were solved in order to get a continuous function for the range of the data. This was performed using MATLAB's built in ordinary differential equations solver, ode15s.

After the model was completed the mean-square error was computed between the calculated infected function and the data for the total population for both the SIR and the SIR with two classes as well as for each of the infectious classes in the SIR with two classes model. This allows for a quantitative assement of a model's accuracy.

III. Results

The models were run using parameters from the data itself and estimated parameters as described in above in the methods section. The graph of the infected individuals from the data is overlayed with both models in Figure 1.

Both models behave similarily, however examining the mean-squared error (MSE) the simple SIR model outperforms the two-class SIR model, with calculated MSEs of 60.4386 and 67.4716 respectively.
The infected individual data for the non-health care workers overlayed with the two-class SIR model is shown in Figure 2.

The non-health workers are modelled more accurately than the overall population in the

previous graph. The calculated MSE was 42.9596, much better than the MSEs calculated above.
The infected individual data for health care workers overlayed with the two-class SIR model is shown in Figure 3.

The health workers are modelled more accurately than the overall population, similar to the improvement seen in the previous graph. The calculated MSE was 42.1551, the graph does not look as "good" but the fringes are accurately modelled leading to a low MSE.

IV. Discussion

Overall the SIR and two-class SIR models accurately simulated the data although errors did occur. Suprisingly overall the two-class SIR model performed worse than the SIR model on the overall data when comparing the MSEs of each. However, the individual classes performed better (ie had a lower MSE) than the overall population model.

The results suggest that the health care workers during this outbreak were at a greater risk than the general public because when the contact rates were adjusted results correlated better. This means that in the future, stricter protective measures for health care workers should be a priority. While this would not a have a large impact on the outbreak size, as workers have low contact rates with uninfected individuals, the health care workers themselves would be better protected.

While the models performed well a better fit could have been achieved had more data been available. In the future a different model could be tested such as an SEIR model which takes into account a latent class in order to account for carriers that are not yet symptomatic. Other possible improvements could be better estimation of parameters using more sophisticated methods.

V. Figures

References

[1] "Ebola haemorrhagic fever." *World Health Organization.* Web. 23 Apr 2013. <http://www.who.int/mediacentre/factsheets/fs103/en/>.

[2] Khan, et al "The Reemergence of Ebola Hemorrhagic Fever, Democratic Republic of the Congo, 1995." *Journal of Infectious Diseases.* (1999): 76-86. Web. 23 Apr. 2013.

[3] Astacio, Jaime. "Mathematical Models to Study the Oubreaks of Ebola." Web. 23 Apr. 2013. <http://mtbi.asu.edu/files/Mathematical_Models_to_Study_the_Outbreaks_of_Ebola.pdf>.