Evaluating Public Health Responses: A Comparative Study of Ebola Management in Sierra Leone, Guinea, and Liberia.

Devanshi Shah
Georgia Institute of Technology
Atlanta, Georgia
devanshishah@gatech.edu

Navya Gautam
Georgia Institute of Technology
Atlanta, Georgia
ngautam37@gatech.edu

Moe Kyaw Thu
Georgia Institute of Technology
Atlanta, Georgia
mthu6@gatech.edu

Abstract

The Ebola virus is transmitted through contact with bodily fluids from infected animals to humans. Since its first detection in 1976, it has spread significantly across Africa, particularly the West African countries, namely Sierra Leone, Liberia, and Guinea. To this end, this proposal examines the epidemic's progression with a focus on these three countries between 2014 and 2016. In particular, this project will apply time series analysis, survival analysis, and intervention analysis to understand the patterns of suspected cases and deaths. In particular, the intervention analysis also takes into account public health measures, namely contact tracing, vaccinations, and patient isolation, to understand their impact on the total number of suspected cases and deaths. The outcome of this study will be utilized to provide policy recommendations for other advancing economies that are particularly vulnerable to Ebola epidemics.

Keywords

Ebola, Policy Intervention, Time Series Analysis, Survival Analysis, Epidemics, Sierra Leone, Liberia, Guinea

ACM Reference Format:
Devanshi Shah, Navya Gautam, and Moe Kyaw Thu. 2024. Evaluating Public Health Responses: A Comparative Study of Ebola Management in Sierra Leone, Guinea, and Liberia.. In . ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Response to Milestone Comments

The majority of the comments on our project milestone were positive feedback. For time series analysis, this project made adjustments to the AR, ARIMA, and ARMA models. Furthermore, the model diagnostics were also provided to examine their performance. For the survival analysis, a new column, titled *Deaths Per Week* , was created in each of the countries' datasets; for more information, please see the codes and write-up for survival analysis. This column was used for the survival analysis. For the intervention analysis, the authors collected news sources and then plotted against the numbers from the datasets. Lastly, as noted by the milestone feedback, this project also addressed the strengths and weaknesses of the datasets.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

2 Background

The Ebola virus is caused by *Orthoebolavirus Zairense* and is thought to have infected humans through contact with infected animals. Transmission occurs through contact with bodily fluids such as blood and saliva. It is also thought that contact with bodies that were infected contributed to the transmission of the disease [6]. Symptoms often include fever, aches, fatigue, memory loss, diarrhea, vomiting, etc. Ebola can be treated by oral medicines, blood transfusions, and oral or intravenous fluids. There are also vaccines with antibodies that can be administered to high-risk communities to increase their chance of survival if infected [5].

Ebola was first detected in Sudan and the Democratic Republic of the Congo in 1976. There were around 600 reported cases, the majority of which were fatal. Outbreaks have occurred since then primarily in sub-Saharan Africa [1]. The largest and most widespread Ebola outbreak occurred from 2014-2016 in West Africa, where 28,000 cases were reported and over 11,000 deaths were reported [7]. The outbreak spread to other parts of the world, including Italy, the United Kingdom, the United States, Spain, and other parts of Africa. This outbreak was unexpected in terms of our quickly it spread and how many people it infected.

3 Literature Review

3.1 Post-Ebola Outbreak in West Africa — Unpredictable Risks and Preventable Epidemics

The WHO Ebola Response Team reviewed the 2014-2016 Ebola outbreak in Western Africa. They address the major loss and travesty inflicted on the countries, recount the interventions taken, and introduce tools to prepare for future outbreaks. The paper provides a comprehensive analysis of the epidemic, including a thorough review of factors that contributed to the epidemic and lessons learned. The paper does not go into depth about specific case studies or the effect of specific interventions on the spread of Ebola.

The infection quickly spread from Guinea, Liberia to Sierra Leone and was highly unexpected. There was an exponential increase in the infected populations in each country. Sierra Leone suffered the worst of the epidemic when looking at the total number of reported cases, which can be seen in 1. Following the intervention, the number of reported cases went down much quicker in Liberia compared to Sierra Leone and Guinea. Such interventions included vaccinations, rapid case detection patient isolation, and contact tracing. However, these were not implemented fast enough and could not keep up with the spread of the virus. There was also an issue of cooperation where symptomatic patients were unwilling to seek medical attention, exposed patients would leave quarantine, and

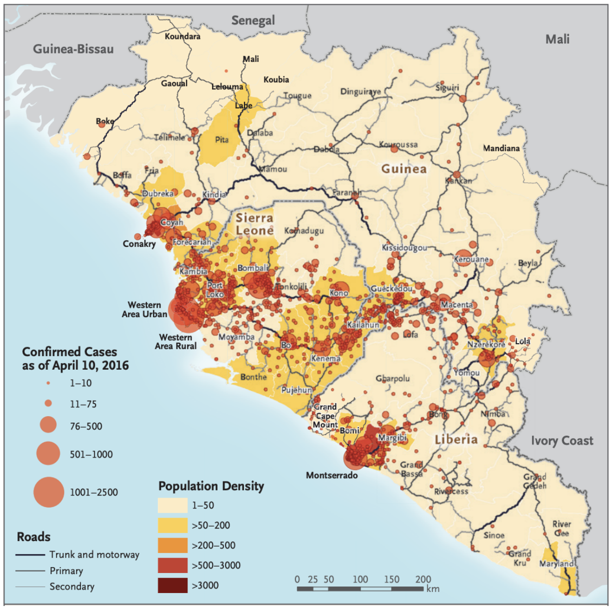


Figure 1: Geographic Distribution of Confirmed Cases of Ebola Virus Disease in Sierra Leone, Guinea, and Liberia as of April 10, 2016.

dead bodies were unsafely buried, all of which led to the continuous spread of Ebola.

The authors suggest preparing West African countries for such an outbreak prior to the event. The Ebola virus has proved to be unpredictable and fatal without proper treatment but can be survived with the right policies set in place. These countries should adopt a "classic, responsive approach to Ebola control: community engagement, early case detection and diagnosis, comprehensive contact tracing, prompt patient isolation, supportive clinical care, and rigorous infection control, including safe burial" [11]. They also need strong financial backing to support such public health needs.

3.2 Methods of interventions and the incidence of Ebola Outbreak in Liberia, Sierra Leone, and Guinea

While the previous section has focused on the spread of Ebola diseases, this section focuses on the methods of interventions in the three countries that are the units of analysis for this project. For example, this paper analyzes the impact of the interventions with respect to the spread of the disease in Liberia from 2014-2016. It investigates interventions such as contact tracing, community engagement, and healthcare facility improvement [10]. Results suggest that contact tracing was found to be vital in reducing transmission and implementing quick isolation of infected individuals. It also discusses how community engagement informed patients about the risks of Ebola and how treatment strategies can help slow the spread. However, while these strategies increased the number of reported cases, this study does not further suggest that if this is due to reduced stigma or an overall increase in cases is unknown.

Studies have shown similar interventions in Guinea as well. For instance, similar to Liberia, this paper also analyzes the impact of contact tracing in Guinea during its height in 2015, focusing on the prefectures of Ratoma and Boké [9]. The paper acknowledges that contact tracing played a vital role in Guinea's Ebola response, particularly by reducing the number of unmonitored cases and limiting transmission within communities. It further highlights how the combination of well-coordinated contact tracing teams, accompanied by supportive measures namely community education and logistical support, helped to improve disease surveillance and management during the outbreak. Despite these contact tracing efforts, logistical difficulties and the need for quality control at all levels.

Similar studies have shown contract tracing to be the main line of defense against preventing Ebola in Sierra Leone as well. This paper, for example, compares the paper-based and mobile-based contact tracing in Port Loko District, Sierra Leone [8]. The comparative approach suggests that the intervention of mobile-based contact tracing enabled faster reporting, more complete data, and a permanent record stored centrally, allowing for better monitoring of the epidemic in real life. However, challenges such as poor network coverage, technical difficulties, and inconsistent use of the apps were identified, which further delayed in rural areas where infrastructure was less developed.

Safe and Dignified Burial (SDB) was an intervention implemented in the 2018-2020 Ebola virus epidemic in the eastern Democratic Republic of Congo (DRC). This epidemic affected the Ituri, North, and South Kivu provinces of the DRC and is said to have caused 3470 confirmed and probable cases. As mentioned earlier, Ebola is spread through the bodily fluids of infected patients. Infected bodies can spread the virus even after death so bodies must be buried safely to avoid further infection from the passed patient. During the epidemic in the DRC, "the International Federation of Red Cross and Red Crescent Societies (IFRC) supported the DRC Red Cross and other local actors to offer SDBs" [13]. When informed of a confirmed or probable case of Ebola death, these organizations dispatched workers to perform an SDB and tested all the involved members to measure the change in the number of cases when SDB is utilized versus when it is not. A change in the number of cases was observed and it was found that SDB was feasible for an epidemic of this scale. This intervention requires the involvement of government organizations and informed families which is not entirely reliable for larger epidemics. SDBs were introduced after the epidemic had already started. Although SDBs can help reduce disease transmission, further work needs to be done to make them more accessible and more quickly implemented once the disease begins to spread.

4 Problem Statement

Based on the existing literature, this project aims to address to predict unexpected disease outbreaks and evaluate public health interventions. In order to do so, there are three specific goals. Firstly, this proposal will closely examine the number of suspected cases and deaths of the Ebola disease from 2014 to 2016. Second, a deeper understanding of the nature of its spread is required for all three countries. To this end, this project will extend the analysis by investigating the interventions used to prevent disease spread in each country.

Lastly, this project will provide policy recommendations on how to address the critical need to predict unexpected outbreaks and create scalable interventions. By incorporating predictive models based on the Ebola outbreak data in these three countries from 2014 to 2016, this proposal will also assess and provide feedback to the early warning systems and data-driven public health responses. The outcomes of the analysis also intend to provide a better understanding of where African states with similar human development, as well as those among advancing economies in Asia and Latin America, should allocate resources to minimize the number of cases and ensure accurate reporting of total cases.

5 Datasets and Methodology

As Ebola in all three West African countries is well known, this project intends to use the open data set on the outbreak of Ebola in Western Africa from 2014 to 2016 from Kaggle. It contains a total of 2483 observations and 3 columns, giving it a large amount of data to analyze. Regarding geographical coverage, the dataset encompasses all West African nations, including the three focal countries, which collectively account for 883 out of 2483 observations. The dataset also offers nearly daily records, starting from August 29, 2014, to March 23, 2016, detailing all confirmed, probable, and suspected cases and deaths.

Since the analysis of this proposal focuses on the Ebola outbreak, several approaches will be considered here. Firstly, since the dataset contains a large number of countries, this proposal will take subsets of the data for Sierra Leone, Guinea, and Liberia. After initial data cleaning, the authors will ensure numerical values are formatted correctly and consistently. Outside of the datasets, using the aforementioned existing literature, the authors will also take into account the dates of contact tracing implementation and other public health measures. In particular, when making interpretations after utilizing different modeling techniques, these measures will be incorporated into the analysis as external variables to examine their impact on the suspected cases and deaths as well.

In terms of methods, this project will use time series analysis to understand the trend of the epidemic over time. Time series analysis is suitable for the Ebola data, where the temporal aspect is critical for predicting future outbreaks, identifying cyclical patterns, and evaluating how the number of cases fluctuates with public health interventions. In particular, within time series analysis, this project will split the dataset into training sets with dates from August 29th, 2014, to December 31st, 2015, and testing with the remaining days to forecast the number of suspected cases and deaths on the dataset and evaluate the model's performance. Secondly, this project will utilize survival analysis to determine the duration of time until certain critical events occur, such as the time from the suspected cases to deaths or recovery. Survival analysis is suitable because it accounts for both the events and the time until the event occurs. This approach will enable us to estimate the survival rates over time, at least for the first 2 years when the Ebola epidemic was the highest, and examine how public health interventions can alleviate the likelihood and timing of death.

Lastly, this project will use intervention analysis to determine the efficacy of public health interventions. Intervention analysis is applicable because the use of interventions has been highly debated. To protect from future Ebola outbreaks, it is vital to know whether or not public health interventions are useful and implementable in regions such as the three West African countries of interest. Overall, this project aims to provide a comparative analysis of the Ebola epidemic in Sierra Leone, Guinea, and Liberia, and by doing so, it hopes to examine insights into the nature of the epidemic and the effectiveness of different public health strategies and interventions deployed during this time.

6 Data Processing

This project utilizes the weekly Ebola outbreak data from West African countries. The dataset provides the number of suspected/probable/confirmed cases, as seen in Figure 2, and the number of suspected/probable/confirmed deaths, as seen in Figure 3, for each week from August 2014 to March 2016. The number of cases and deaths are cumulative, i.e. the values are representative of the total number of cases/deaths since the start of the epidemic. This project pulled the data from Kaggle, which used the published WHO data for this epidemic. The original data had individual columns for numbers of suspected, probable, and confirmed cases and deaths, and included all countries that had Ebola cases during the mentioned time period. This project decided to use the cleaned version of the data for cleaner and more comprehensive interpretations.

When evaluating the dataset, it was observed that the 'Liberia' data had duplicates and had a sudden change in the number of cases and deaths. For the purpose of this project, all duplicates were removed. The data also showed an invariance in the number of cases and deaths for all countries towards the end of 2015 and into 2016. It is thought that numbers were not being reported as consistently towards the end of the epidemic. For the purpose of this project, reports after November 2015 were removed. This project observes Sierra Leone, Liberia, and Guinea, and the data shows Guinea has significantly fewer cases than the other two countries, consistent with literature findings.

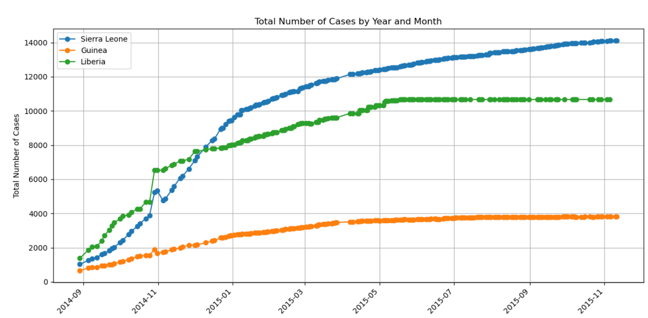


Figure 2: Total Number of Cases by Year and Month

7 Mathematical Intuition and Algorithms

For time series analysis, AR and ARMA are particularly chosen for several reasons. First and foremost, they are excellent at capturing

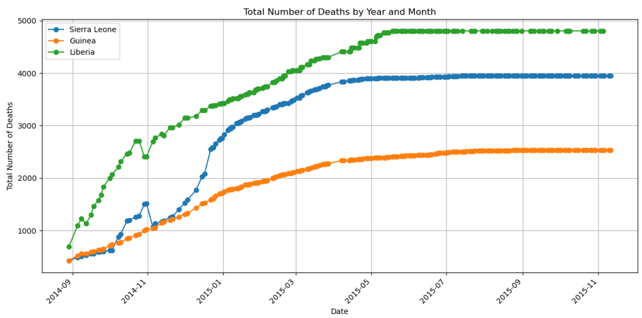


Figure 3: Total Number of Deaths by Year and Month

patterns inside the data that change over time, including trends and seasonality. An auto-regressive (AR) model is a statistical model used to predict a variable based on its own past results [4]. An AR model is self-dependent meaning it is dependent on the past values to predict future values. This is a fairly simple statistical model that can be used to forecast the outcome of an epidemic during the epidemic by utilizing the collected data at that moment in time. Mathematically, to find the value at time X for the AR model:

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + \epsilon$$

where c is a constant, ϕ is a coefficient for lagged observations (based on previous data points), and ϵ is a noise term. The purpose of a simple AR model is to predict the outcome of the Ebola outbreak based on the dataset and compare it to the actual outcome to see if this is a reliable method of prediction.

Auto-Regressive Moving Average (ARMA) is also deployed as it is much more complex than the basic AR model. Overall, ARMA is a more comprehensive approach that extends AR by integrating the differentiation and moving averages [2]. This means that ARMA is particularly valuable when the time series data is non-stationary, i.e., its mean, variance, or autocorrelation structures may change over time. Mathematically, the ARMA model can be described in the following way:

$$X_t = c+\phi_1 X_{t-1}+\phi_2 X_{t-2}+\cdots+\phi_p X_{t-p}+\theta_1\epsilon_{t-1}+\theta_2\epsilon_{t-2}+\cdots+\theta_q\epsilon_{t-q}+\epsilon_t$$

where c is a constant, ϕ is the coefficients for lagged observations based on previous data points, θ represents the coefficients for the moving average, and ϵ_t is the error term at time t. The order parameters p, d, and q in ARMA correspond to the number of AR terms, the number of times the data needs to be differenced to make it stationary, and the number of MA terms, respectively. Lastly, while the project focuses on ARMA, the presence of Auto-Regressive Integrated Moving Average (ARIMA) should also be taken into consideration. Mathematically, while they are the same, ARIMA extends ARMA to handle non-stationary data through differencing (the "I" part of ARIMA), ARMA is a simpler model that can be used when the data is stationary.

To conduct a survivor analysis, this project made use of the Kaplan Meier estimator. The Kaplan Meier estimator is a non-parametric statistic that estimates the survival function of a time-series data [12]. The purpose of this estimator is to construct a survival curve from collected data without assuming an underlying distribution. The survival function can be written as:

$$S(t) = \prod_{i:t_i \leq t} (1 - \frac{d_i}{n_i})$$

where t_i = time, d_i = number of events at time t_i, and n_i = number of individuals known to have survived up to time t_i. This project employs the KaplanMeierFitter from the Python Lifelines library [3]. For each country, the number of deaths per week is defined as the difference between the total number of deaths from this week and the total number of deaths from the previous week. The purpose of using the Kaplan Meier Estimator is to show the number of survivors as the epidemic progresses.

8 Results

8.1 Time Series Analysis

In order to define the appropriate models for AR, ARIMA, and ARMA, an Augmented Dickey-Fuller (ADF) test is conducted to see if they are stationary for both mean cases and deaths. The initial results show that for total cases, all three countries are shown to be stationary, with a p-value less than 0.05. In contrast, for total deaths, Sierra Leone and Guinea are found to be stationary, whereas Liberia does not contain stationary. Outside of the ADF tests, autocorrelation (ACF) and partial autocorrelation (PACF) plots are conducted to analyze the lags within the data. The results show that by using ACF plots for total cases and deaths across the three countries, autocorrelation decreases steadily as the lag increases, and by 15-17 lags, the values drop within the confidence intervals. This suggests that the influence of past values diminishes over time. In contrast, the PACF plots for total cases and deaths show that after two lags, the rest of the lags drop rapidly into the confidence intervals. This demonstrates that for PACF plots, the influence of past values on the current values is primarily limited to the two initial lags.

By using AR modeling with a lag of 2 based on PACF plots, the sharp decline in autocorrelation suggests that it is sufficient to capture significant autocorrelation present at these lag points while avoiding over-fitting with unnecessary higher-order lags. The results show that by fitting the AR model with 2 lags to predict for the next 12 months, the total cases for all three countries show similar trends with constant levels of cases. In a similar manner, by applying the AR model with 2 lags, the total deaths for Guinea and Sierra Leone have similar trends as their mean cases counterparts. As AR relies on the relationship with past data, the trends observed for mean cases and deaths suggest that the model is fairly effective in capturing the patterns without adding additional complexity.

By extracting residuals, model diagnostics for AR modeling show that for total cases across all three countries, the AR models effectively capture temporal dependencies, with no significant autocorrelation, though the residuals are not normally distributed. However, while heteroscedasticity is present in Sierra Leone and Liberia, it is absent in Guinea. For total deaths, Sierra Leone and Guinea show no significant autocorrelation and some heteroscedasticity in Sierra Leone, while Liberia has significant autocorrelation and heteroscedasticity, indicating the model fails to account for some temporal structure. Residual plots for both cases and deaths show spikes in the early months for Sierra Leone, Guinea, and Liberia,

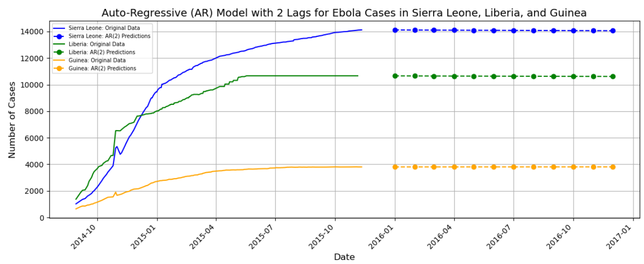


Figure 4: AR Model for Total Cases

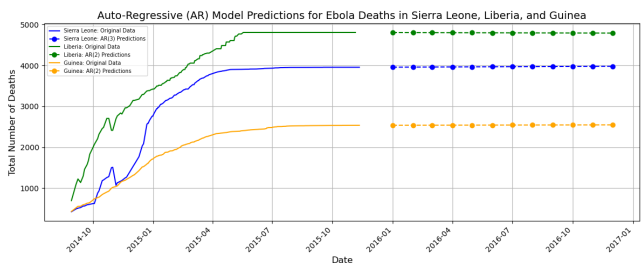


Figure 5: AR Model for Total Deaths

suggesting varying model fit (for residual plots' full visualization, due to the limit of the report, please refer to the time series Python file).

Based on the ACF plots, ARIMA and ARMA models with a large number of autoregressive and moving average terms might be appropriate. However, it is crucial to note that in order to choose between ARIMA and ARMA, it depends on the outcomes of the ADF results on whether the results are stationary or non-stationary; in this case, only Liberia's total number of deaths is non-stationary, thus making it possible to use ARIMA here. Additionally, for the ARMA and ARIMA models, the chosen order parameters, p, d, q, depend on the observations from the differencing needed to achieve stationary and autocorrelation patterns.

For the total number of cases, the ARMA model parameters are chosen with an order of p as 2, d as 0, and q, where the value of q varies between 15, 16, and 18 across the three countries. The prediction for the next 12 months using ARMA modeling shows that only Sierra Leone has fallen down in terms of total cases, as seen in Figure 6. For total deaths' ARMA modeling, Sierra Leone and Guinea have an order of p as 2, d as 0, and q varies between 17 and 18, respectively. For Liberia, the ARIMA model has an order of p as 2, d as 1, and q as 15. The results show that between the three models' predictions for the next 12 months, Liberia has seen a rise in total deaths, Guinea remains constant, and only Sierra Leone could expect a decrease in death rates.

Model diagnostics were also conducted for ARMA modeling by extracting the residuals. The residuals for the ARMA models on total cases and deaths in Sierra Leone, Liberia, and Guinea reveal generally favorable results. For total cases, all three show no significant autocorrelation at the tested lags, with residuals normally distributed, and with no evidence of heteroscedasticity in Liberia and Guinea, while Sierra Leone does exhibit some of it. Total deaths show similar results. Sierra Leone, Guinea, and Liberia show no significant autocorrelation at the tested lags, and the residuals are normally distributed. However, Liberia's model displays heteroscedasticity, suggesting that the variance in residuals may vary over time. For ARMA residual plots for total cases, there is an initial spike, but it remains stable for all three countries. However. the residual plots for the ARMA model of total deaths for Sierra Leone and Guinea capture the overall trend, but the ARIMA model for Liberia exhibits instability (for residual plots' full visualization, due to the limit of the report, please refer to the time series Python file).

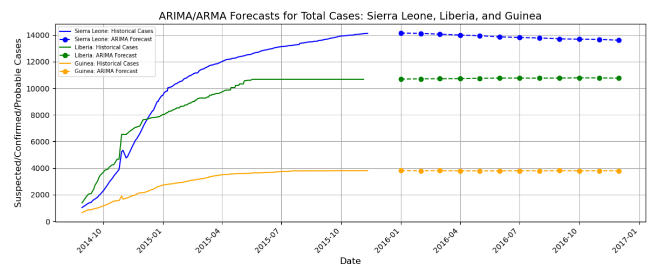


Figure 6: ARMA/ARIMA Model for Total Cases

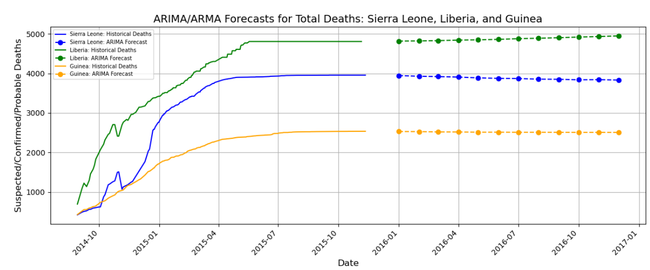


Figure 7: ARMA/ARIMA Model for Total Deaths

8.2 Survival Analysis

A Kaplan Meier Fitter was defined to fit survivor data for each country. The number of deaths per week was used to find the number of survivors overall. This is done by subtracting the cumulative number of deaths for the current week from the cumulative number of deaths for the previous week. The difference will be the number of deaths that occurred over that week. The number of deaths per week can be used as a survival metric because as the number of deaths decreases, the number of survivors increases and vice versa. Using the Kaggle dataset described, the project added a column that defined the number of deaths for each week. The Liberia dataset has a couple of outliers due to the updated reports. The updated data reported the number of suspected/probable/confirmed cases and suspected/probable/confirmed deaths per week, rather than the cumulative values like the rest of the dataset. The number of deaths for these outliers was defined as the reported number of suspected/probable/confirmed deaths. The updated Liberia data was used to complete the following survival analysis.

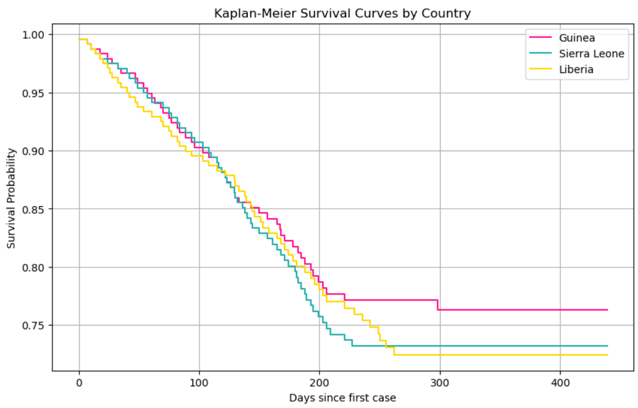


Figure 8: Survival Modeling for Guinea, Sierra Leone, and Liberia

This new data was used as the number of events and the time was defined as the number of days since the first reported case in the data. This process was then repeated separately for each country. A threshold of 10 deaths per week was defined as the number of deaths per week that constitute an event. An event is defined as 1 if the number of deaths for that week is above the previously defined threshold and 0 if there were less than 10 deaths for that week. As shown in Figure 8, all three countries show an initial decrease in survival probability that plateaus after around 200 days since the beginning of the epidemic.

The mean number of cases and the mean number of deaths for each month of the epidemic were plotted for each year, as shown in Figure 9. All three countries displayed an exponential increase in the number of cases. The number of deaths also began to increase exponentially and plateaued towards the end of 2015. In other words, the number of cases is still increasing at the end of 2015, but the number of deaths is not changing much, indicating that the number of survivors is increasing.

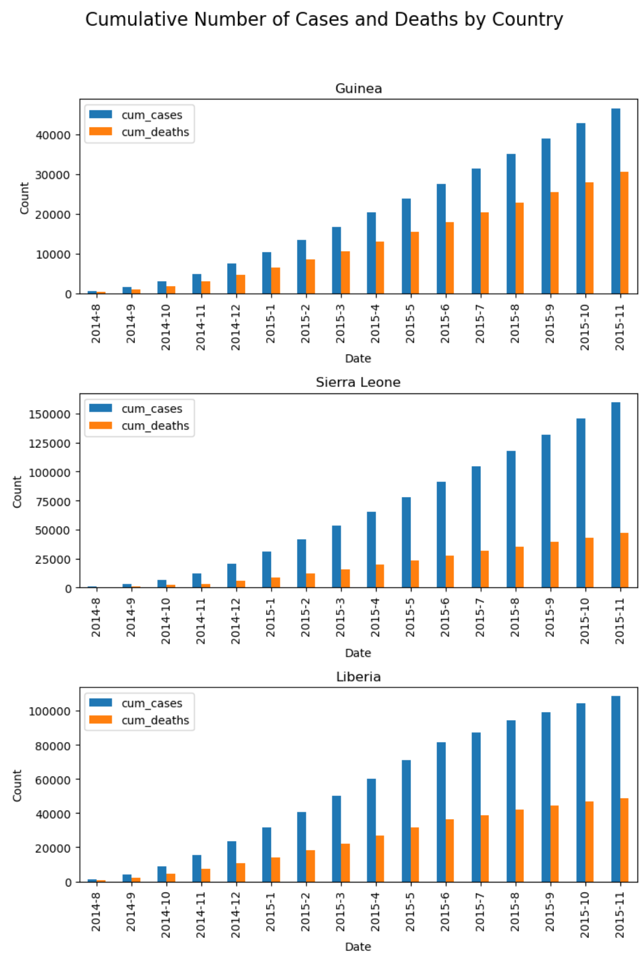
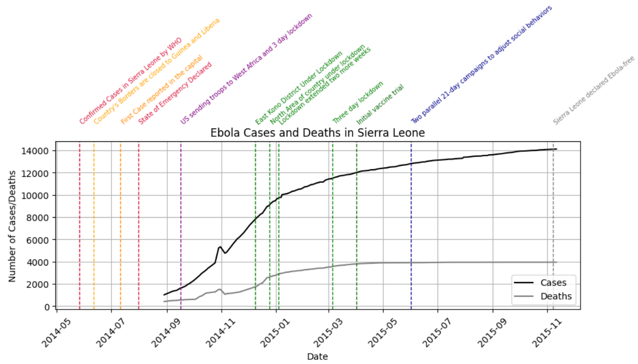


Figure 9: Number of Total Cases and Deaths in Guinea, Sierra Leone, and Liberia for Survival Trends

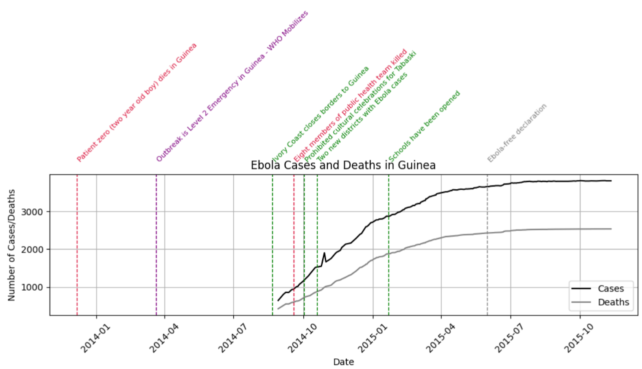
8.3 Actions, Events, and Interventions Analysis

For the intervention analysis, the authors searched online to examine information associated with the Ebola outbreak as well as that of the public health interventions taken by the respective governments in Sierra Leone, Liberia, and Guinea, particularly those in response to events of high magnitude. After organizing them into three tables respective to the dates of the intervention and public health actions, these new pieces of information were plotted against the overall trends and numbers of the epidemic from the original datasets, as seen in Figure 10. The analysis of the plots demonstrates that in Sierra Leone, the government relied heavily on strict lockdown in an attempt to curb the spread of the outbreak. Despite their efforts, it is important to note that they are not effective in controlling the spread; in fact, the cases and death tolls continued to rise in a rapid number, as demonstrated in Figure 10 (a). This may be caused by the avoidance and resistance against the lockdown measures. While patient zero was originally from Guinea, by the time Sierra Leone closed its border to the former and Liberia, cases were already increasing en masse. Despite the vaccine trials occurring as early as February and March 2015, it is possible that the authorities finally recognized the role of social behaviors, thus giving rise to two coordinated campaigns. This included promoting social behaviors, namely social distancing, vaccination, hand-washing, and burial practices.

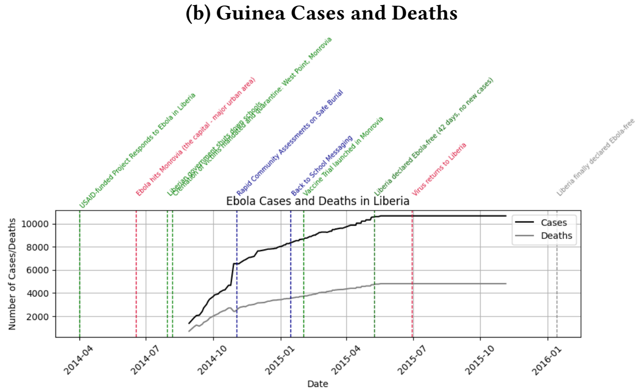
In contrast, in Guinea, the total number of cases and deaths rose in a similar manner. Based on the intervention analysis, it was evident that there was a community mistrust towards public health officials and organizations. This community mistrust eventually gave rise to riots, which ended with the murder of eight members of a public health team. While there were efforts to combat misinformation in Guinea and beyond, in Liberia, there were multiple coordinated efforts when cases peaked. For example, Liberian public health officials led community sessions on safer burial practices by taking account into the cultural differences and traditions between different ethnic groups. These types of efforts continued even as cases began slowing down, which included back-to-school messaging.



(a) Sierra Leone Cases and Deaths



(b) Guinea Cases and Deaths



(c) Liberia Cases and Deaths

Figure 10: Number of Total Cases and Deaths in Guinea, Liberia, and Sierra Leone Combined with Action/Intervention/Event Findings

For all three cases, in general, after the first 42 days with no new cases, they were declared Ebola-free. However, the analysis showed that the Ebola-free declaration occurred twice, which could be due to misreporting or under-reporting of cases. This highlights the importance of maintaining vigilance and caution during periods of uncertainty, ensuring that public health officials remain alert.

9 Discussion and Conclusion

This study explored the progression of the Ebola epidemic in Sierra Leone, Liberia, and Guinea from 2014 to 2016, utilizing time series analysis, survival analysis, and intervention analysis to examine patterns in suspected cases and deaths. For time series, there are several key findings. In general, all the models are stationary with the exception of total deaths for Liberia, which was examined using the ARIMA model; in particular, compared to ARMA, ARIMA is more suitable for non-stationary data series since it takes account into the differencing as well. Although the models performed reasonably well overall, as indicated by the diagnostic evaluations, the predictions for the next 12 months, up to March 2017, indicated a constant number of cases rather than showing a clear increase or decrease.

Secondly, the survival analysis demonstrates the relative effectiveness of measures and the spread of the virus over the 2014-2015 time period. Although the survival curve plateaus after around 200 days, indicating survival probability is consistent, the sharp decrease in survival probability in the first 200 days is worrisome and presents the need for effective and prompt actions for both disease control and prevention.

Lastly, the intervention analysis demonstrates the relative effectiveness of various measures, with vaccines, distances, and safe burial practices emerging as a critical factor in controlling the epidemic. These findings offer valuable insights for policymakers in developing economies that may face similar outbreaks, emphasizing the need for social behaviors and societal cooperation.

9.1 Challenges

However, the project is not without challenges, mainly with respect to the dataset. For one, the dataset is relatively small and focuses on the total number of suspected cases and deaths across three countries. However, it does not provide details at the state, prefecture, provincial, or local levels. This lack of detail limits the ability to model regional disparities. The small size also means that all three approaches rely on small observations, thus introducing a potential for overfitting and limiting the external validity of the models. Initially, this project proposes neural networks. However, due to the data size, it was determined that neural network models would be inappropriate, as the limited data could also result in overfitting and poor generalization. The dataset also starts a good amount later than the start of the epidemic in these countries, which partially obscures some potential analysis efforts.

9.2 Future Work

Building on these findings, further research could explore the long-term effects of public health interventions on post-epidemic recovery and resilience. Comparative studies with other epidemic-prone regions may also provide broader generalization. Such regions may include other African countries prone to Ebola outbreaks. Also, increasing the size of the dataset and including individual case data for each country could lead to more robust results. A comprehensive study on the effects of the socioeconomic status of countries in relation to the public health measures taken can provide important insight into how effective and plausible such studied interventions are in third-world countries. Finally, modeling studies incorporating additional variables such as population mobility and healthcare infrastructure could refine the understanding of outbreak patterns and inform targeted interventions.

References

[1] [n. d.]. Ebola Disease. https://www.afro.who.int/health-topics/ebola-disease

[2] [n. d.]. Introduction to ARIMA: nonseasonal models. https://people.duke.edu/~rnau/411arim.htm

[3] [n. d.]. KaplanMeierFitter. https://lifelines.readthedocs.io/en/latest/fitters/univariate/KaplanMeierFitter.html

[4] [n. d.]. What are AutoRegressive Models. https://aws.amazon.com/what-is/autoregressive-models

[5] 2023. Ebola Virus Disease. https://www.who.int/news-room/fact-sheets/detail/ebola-virus-disease

[6] 2024. Ebola Disease Basics. https://www.cdc.gov/ebola/about/index.html#:~:text=Ebola%20disease%20is%20caused%20by%20a%20group%20of%20viruses%2C%20known,primarily%20in%20sub%2DSaharan%20Africa.

[7] 2024. Ebola Outbreak History. https://www.cdc.gov/ebola/outbreaks/index.html

[8] L. O. Danquah, N. Hasham, M. MacFarlane, F. E. Conteh, F. Momoh, A. A. Tedesco, A. Jambai, D. A. Ross, and H. A. Weiss. 2019. Use of a mobile application for Ebola contact tracing and monitoring in northern Sierra Leone: a proof-of-concept study. *BMC Infectious Diseases* 19, 1 (2019), 810. https://doi.org/10.1186/s12879-019-4354-z

[9] L. Reques, I. Bolibar, E. Chazelle, L. Gomes, V. Prikazsky, F. Banza, M. Keita, and B. Diallo. 2017. Evaluation of contact tracing activities during the Ebola virus disease outbreak in Guinea, 2015. *International Health* 9, 2 (2017), 131–133. https://doi.org/10.1093/inthealth/ihx004

[10] Kirsch TD, Moseson H, Massaquoi M, Nyenswah TG, Goodermote R, Rodriguez-Barraquer I, Lessler J, Cumings DA, and Peters DH. 2017. Impact of interventions and the incidence of ebola virus disease in Liberia—implications for future epidemics. *National Library of Medicine* 32, 2 (March 2017), 205–214. https://www.ncbi.nlm.nih.gov/pmc/issues/291272/

[11] WHO Ebola Response Team. 2016. After Ebola in West Africa — Unpredictable Risks, Preventable Epidemics. *New England Journal of Medicine* 375, 6 (Aug. 2016), 587–596. https://www.nejm.org/doi/pdf/10.1056/NEJMsr1513109

[12] Shae Wang. [n. d.]. Fitting a KaplanMeier estimator: SURVIVAL ANALYSIS IN PYTHON. https://projector-video-pdf-converter.datacamp.com/25923/chapter2.pdf

[13] A. Warsame, G. Eamer, A Kai, and et al. 2023. Performance of a safe and dignified burial intervention during an Ebola epidemic in the eastern Democratic Republic of the Congo, 2018–2019. *BMC Medicine* 484, 21 (2023). https://doi.org/10.1186/s12916-023-03194-x