Ebola Virus Disease in West Africa

The First 9 Months of the Epidemic and Forward Projections

Group 9

Lin Jiao (A0098993B)

Stephanie van den Boogaard (A0149081E)

Table of Contents

1 Introduction

The Ebola Virus Disease (EVD) epidemic occurred in the December 2013 is the largest outbreak of Ebola up to this moment. The World Health Organisation declared the epidemic to be a public health emergency of international concern in 2014. The outbreak took more 11 thousands lives in West Africa. The team is concern about the EVD and would to know more details about the EVD in West Africa. Hence, the team chooses to study a paper related to the EVD and to have deeper understand on this epidemic. In this report, the team will discuss the paper, Ebola Virus Disease in West Africa — The First 9 Months of the Epidemic and Forward Projections. Firstly, the team will elaborate on the background and the objective of the paper. After that, the methodologies used in the paper will be discussed, followed by an overview of the results. Lastly, the team will conclude with contributions made by the paper and the limitations relating to the research of the paper.

1.1 Background & Objective of the Paper

The paper studied is called *Ebola Virus Disease in West Africa — The First 9 Months of the Epidemic and Forward Projections*. The paper is written by the World Health Organisation (WHO) Ebola Response Team. It is published on October 16th 2014 in the New England Journal of Medicine, Vol 371, No 16. The dataset used in the paper is based on the first 9 months of the epidemic in Guinea, Liberia, Nigeria and Sierra Leone. Figure 1 presents the timeline from first Ebola cases happened in Guinea to the official end of the outbreaks in February 2016. The paper focusses on the time between the first cases in December 2013 till the beginning of September.

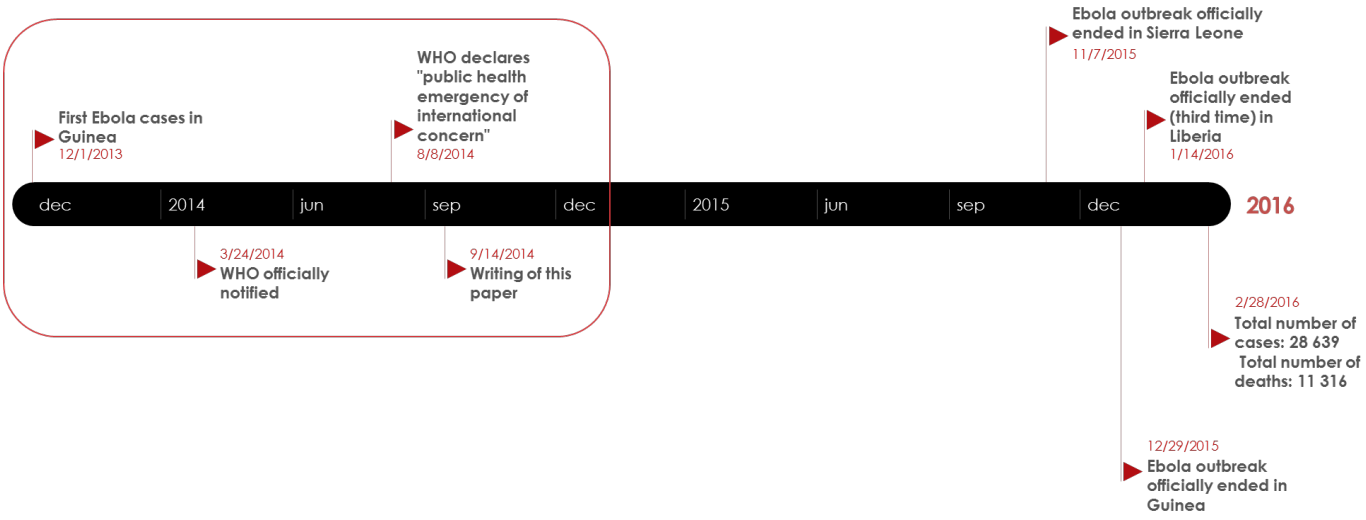

Figure 1: Timeline of EVD Epidemic December 2013 – September 2014

The objectives of the paper are summarized into three main points. First objective is to explore the clinical and epidemiologic characteristics of the epidemic. Second objective is to document trends in the epidemic. Last but not least, the final objective of the paper is to project expected case numbers for the coming weeks if control measures are not enhanced.

1. To explore the clinical and epidemiologic characteristics of the epidemic
   By collecting and analysing all possible data from the countries mentioned in West Africa, the Ebola Response team did the research works to find out a number of characteristics about the epidemic. Examples of those characteristics are the number of cases which includes suspected cases, probable cases and confirmed cases, division of male and female, countries which most affected by EVD, the most common symptoms and so on.

2. To document trends in the epidemic
   By using the methodology to generate reproduction numbers and to identify the incubation period of the epidemic, this data helps the Ebola Response team to well document the trends of the epidemic in the current stage. With the current trends, it will help the research team to better predict the trends in the future and to better control the epidemic in West Africa.

3. To project expected case numbers for the coming weeks if control measures are not enhanced
   By making assumption if the control measures are not enhanced, the Ebola Response team does a comprehensive projection and to project the expected case numbers for the following weeks. This helps to better understand the damages caused by the epidemic if there are no proper solutions and measurements make to control the EVD.

1.2 Issues Addressed in the Paper

By mid-September, 9 months after the first Ebola case which has been occurred in Guinea, the numbers of reported Ebola cases and deaths were still growing from week to week despite multinational and multi sectorial effects to control the spread of infection. The statistic figure in the paper shows by September 14, 2014, a total of 4507 probable and confirmed cases, including 2296 deaths from EVD had been reported from five countries in West Africa. With such high spread of infection, the current outbreak in West Africa is considered as the largest and most complex Ebola outbreak since the Ebola virus was first discovered in 1976. There have been more cases and deaths in this outbreak than all others combined. This has been drawn a huge concern by WHO. Hence, it requests urgent needs to control the cases. On the other hand, it is obvious to tell that without drastic improvements in control measures, the number of cases and deaths from EVD are expected to continue increasing from hundreds to thousands per week in the coming months. In addition, the paper also addresses the concern of analysing the clinical and epidemiologic characteristic of the epidemic to better understand the high spread of the epidemic in West Africa. Moreover, it is also important that to predict expected number of cases for the coming week in order to generate the better control measures to solve the issues mentioned here. All of this will contribute to a better understanding of the Ebola outbreak and will provide insights on the control of the epidemic.

2 Methodology

To be able to reach the objectives of the paper, three different methods are used. First of all surveillance is used to collect sufficient data. The data is analysed by using Logistic Regression Modelling and the Reproduction Number R. Logistic Regression Modelling is used to explore the characteristics of the epidemic. The Reproduction number R will be the base for the projection of future cases.

2.1 Surveillance

Using a standard case investigation form the clinical and demographical data of patients were collected. In addition to this, these data were supplemented by information collected in informal case reports, data from diagnostic laboratories, and from burials. To create the fullest possible picture of the unfolding epidemic. The following aspects are included in the standard case investigation form:

- District of residence
- District of disease report
- Age
- Sex
- Signs and symptoms recorded
- Date of symptom onset
- Name of hospital
- Date of hospitalization
- Date of death or discharge

Some of the cases provided addition information about potentially infectious contacts with other EVD cases, including possible exposure at funerals.

According to the definition is the World Health Organisation[1], which is used in this paper, the cases are categorised in three categories:

- A suspected case
- A probable case
- A confirmed case

A suspected case[1] *is illness in any person, alive or dead, who has (or had) sudden onset of high fever and had contact with a person with a suspected, probable, or confirmed Ebola case or with a dead or sick animal,* or *any person with sudden onset of high fever and at least three of the following symptoms: headache, vomiting, anorexia or loss of appetite, diarrhoea, lethargy, stomach pain, aching muscles or joints, difficulty swallowing, breathing difficulties, or hiccupping,* or *any person who had unexplained bleeding or who died suddenly from an unexplained cause.*

A probable case[1] is *illness in any person suspected to have EVD who was evaluated by a clinician or any person who died from suspected Ebola and had an epidemiologic link to a person with a confirmed case but was not tested and did not have laboratory confirmation of the disease.*

A case[1] is classified as confirmed when *a sample of the patient tested positive for Ebola.*

The analysis is done based on the confirmed and probable cases, so the suspected cases are excluded. This results in an analysis on 3343 confirmed and 667 probable Ebola cases.

2.2 Logistic Regression Modelling

Using the cases with a definitive outcome (either the patient is death or recovered), logistic regression modelling is used to determine potential risk factors for a fatal outcome. The risk factors which are evaluated are:

[1] World Health Organization. Case definition recommendations for Ebola or Marburg virus diseases (http://www.who.int/csr/resources/publications/ebola/ebola-case-definition-contact-en.pdf?ua=1)

- Sex
- Age group (< 15 years; 15 - 44 years; >44 years)
- Symptoms
- Occupation (healthcare worker; not healthcare worker)

For all symptoms the odds ratio was calculated, where only symptoms which had at least five observation were taken into account. Furthermore, the Case Fatality Rate (CFR) is calculated. The CFR is the percentage of fatal EVD cases among reported cases with a known definitive clinical outcome.

2.3 Reproduction Number R

The reproduction number R is the average number of secondary cases which arise, when a primary case is introduced into a population. When R is greater than 1, the disease can spread into the population. If R is smaller than 1, the number of cases will become smaller and smaller and the disease cannot sustain in the population. The calculation of R can be useful to get insights in the magnitude of the epidemic, the efforts which are required to control the virus, and to evaluate the effectiveness of control measures.

To calculate the reproduction number, the *generation time* is determined. The generation time is the time between infection in an index case patient and infection in a patient infected by that index case patient. Besides this, five other key time periods that characterise the spread of the virus are calculated:

- The incubation period, which is the time between infection and the onset of symptoms
- The interval from symptom onset to hospitalization
- The interval from hospital admission to death and the interval from hospital admission to discharge
- The serial interval, which is defined as the interval between disease onset in an index case patient and disease onset in a person infected by that index case patient
- The doubling time, which is the time required for the incidence to double, is estimated on the basis of the reproduction number and the serial interval

The initial value of R, *called the basic reproduction number* or R_0, indicates how fast the virus will spread into the uninfected population. If the virus sustains the early phase, the number of cases will grow exponentially. Once the infection has become established within the population, the number of people still at risk declines, either because people already suffer from the virus, died from the virus, or recovered from it and are resistant now. This results in a decline in the reproduction number R_0 to a smaller number which is called *net reproduction number*, R_t.

Based on the data collected by surveillance, the two methods presented above will lead to the results presented in the next section.

3 Results of the Data Analysis

The results can be split into three parts: the scale and characteristics of the epidemic, the geographic origin and spread of the epidemic, and the projection of the number of future cases. This paragraph will elaborate on this three parts.

3.1 The Scale and Characteristics of the Epidemic

The scale of the epidemic is made visual in different plots, where the weekly number of cases is plotted. One plot provides us with an overview of all the cases in the most affected countries, Sierra Leone, Nigeria, Liberia and Guinea (figure A). Figure B, C and D show the number of cases divided into the three categories suspected, probable and confirmed for Guinea, Liberia and Sierra Leone. As can be seen from the figures, the first cases started at the end of 2013 in Guinea. From July 2014 onwards there was a sharp increase in the number of cases in all three countries. Especially Liberia suffers from many cases. A visualisation of these results can be found in figure 2.

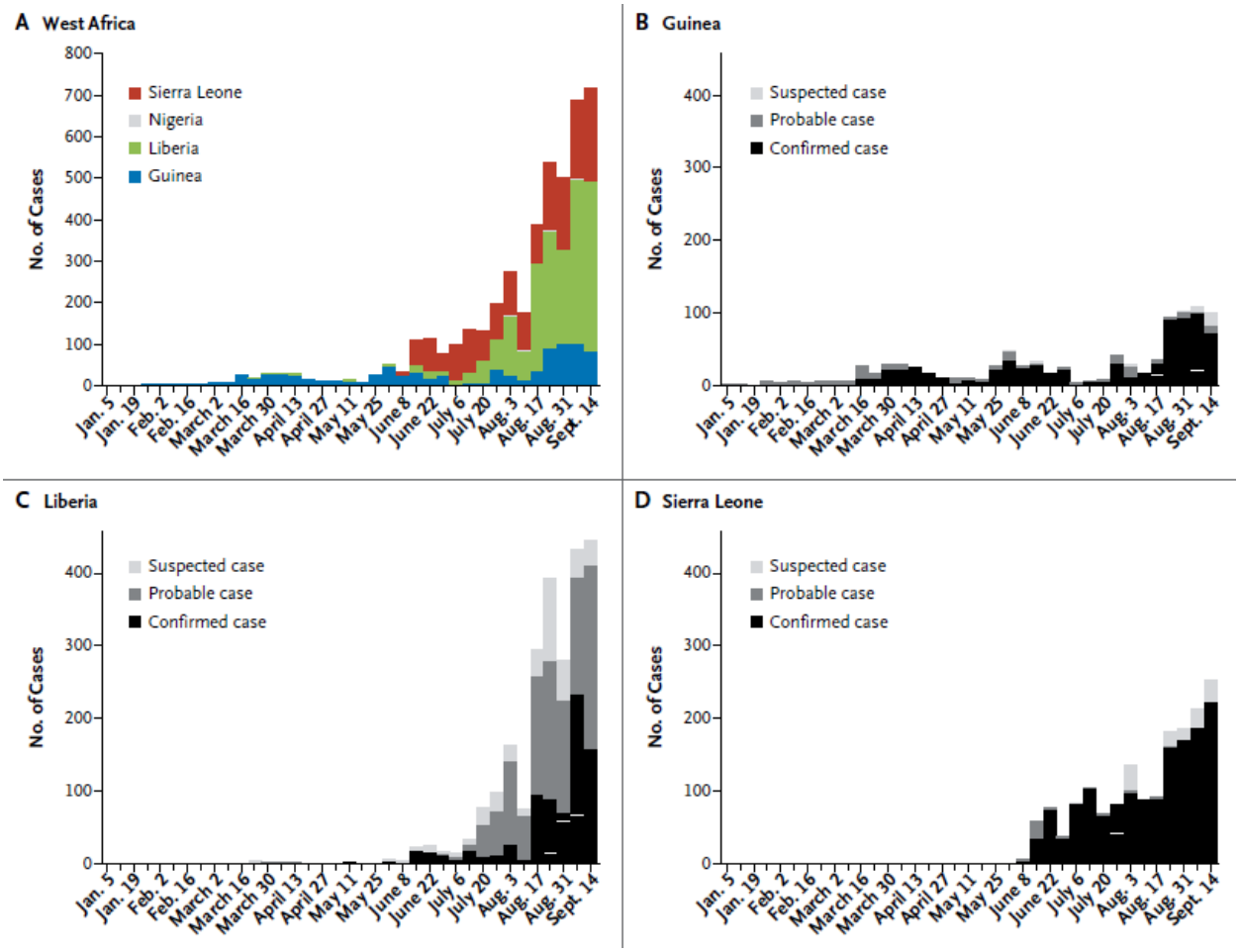

Figure 2: Weekly Incidence of Confirmed, Probable, and Suspected Ebola Virus Disease Cases

The demographic characteristics and most common symptoms of the epidemic are summarised in a table. There is no separation made into the different countries. The results can be found in table 1.

Table 1: Demographic Characteristics and Signs and Symptoms in Confirmed and Probable Ebola Case Patients with a Definitive Clinical Outcome

Variable	All Patients	Patients Who Died	Patients Who Recovered	Odds Ratio (95% CI)†
		no./total no. (%)		
Demographic characteristics				
Male sex	685/1415 (48.4)	515/1056 (48.8)	170/359 (47.4)	0.93 (0.73–1.19)
Age group				
<15 yr	190/1378 (13.8)	145/1021 (14.2)	45/357 (12.6)	1.18 (0.83–1.71)
15–44 yr	838/1378 (60.8)	577/1021 (56.5)	261/357 (73.1)	0.48 (0.36–0.62)
≥45 yr	350/1378 (25.4)	299/1021 (29.3)	51/357 (14.3)	2.47 (1.79–3.46)
Health care worker	158/1429 (11.1)	112/1067 (10.5)	46/362 (12.7)	0.86 (0.60–1.27)
Signs and symptoms				
General symptoms				
Fever‡	1002/1151 (87.1)	746/846 (88.2)	256/305 (83.9)	1.34 (0.92–1.95)
Fatigue	866/1133 (76.4)	633/829 (76.4)	233/304 (76.6)	0.94 (0.68–1.28)
Loss of appetite	681/1055 (64.5)	498/778 (64.0)	183/277 (66.1)	0.92 (0.69–1.23)
Vomiting	753/1114 (67.6)	566/816 (69.4)	187/298 (62.8)	1.19 (0.89–1.59)
Diarrhea	721/1099 (65.6)	555/813 (68.3)	166/286 (58.0)	1.42 (1.06–1.89)
Headache	553/1035 (53.4)	407/757 (53.8)	146/278 (52.5)	1.03 (0.78–1.36)
Abdominal pain	439/992 (44.3)	311/715 (43.5)	128/277 (46.2)	0.85 (0.64–1.13)
Muscle pain	385/990 (38.9)	293/728 (40.2)	92/262 (35.1)	1.24 (0.92–1.67)
Joint pain	374/950 (39.4)	283/695 (40.7)	91/255 (35.7)	1.32 (0.98–1.80)
Chest pain	254/686 (37.0)	196/488 (40.2)	58/198 (29.3)	1.53 (1.07–2.20)
Cough	194/655 (29.6)	150/462 (32.5)	44/193 (22.8)	1.74 (1.18–2.61)
Difficulty breathing	155/665 (23.3)	123/472 (26.1)	32/193 (16.6)	1.68 (1.10–2.63)
Difficulty swallowing	169/514 (32.9)	138/375 (36.8)	31/139 (22.3)	2.22 (1.41–3.59)
Conjunctivitis	137/658 (20.8)	109/465 (23.4)	28/193 (14.5)	2.03 (1.29–3.29)
Sore throat	102/467 (21.8)	82/339 (24.2)	20/128 (15.6)	1.94 (1.13–3.46)
Confusion	84/631 (13.3)	68/446 (15.2)	16/185 (8.6)	2.00 (1.14–3.71)
Hiccups	108/947 (11.4)	91/699 (13.0)	17/248 (6.9)	2.15 (1.27–3.82)
Jaundice	65/627 (10.4)	52/443 (11.7)	13/184 (7.1)	1.83 (0.99–3.63)
Eye pain	48/622 (7.7)	39/438 (8.9)	9/184 (4.9)	1.95 (0.95–4.40)
Rash	37/642 (5.8)	30/453 (6.6)	7/189 (3.7)	1.90 (0.86–4.83)
Coma or unconsciousness	37/627 (5.9)	34/445 (7.6)	3/182 (1.6)	4.59 (1.61–19.34)
Unexplained bleeding	168/932 (18.0)	140/693 (20.2)	28/239 (11.7)	1.83 (1.20–2.90)
Hematemesis	26/670 (3.9)	20/503 (4.0)	6/167 (3.6)	1.07 (0.44–3.01)
Blood in stool	48/843 (5.7)	35/614 (5.7)	13/229 (5.7)	0.98 (0.52–1.96)
Bleeding gums	19/837 (2.3)	18/608 (3.0)	1/229 (0.4)	6.69 (1.35–121.32)
Bloody nose	16/836 (1.9)	15/610 (2.5)	1/226 (0.4)	8.02 (1.54–148.62)
Bloody cough	20/831 (2.4)	16/605 (2.6)	4/226 (1.8)	1.63 (0.58–5.82)
Other bleeding	8/657 (1.2)	5/493 (1.0)	3/164 (1.8)	0.45 (0.11–2.23)
Bleeding at injection site	20/833 (2.4)	19/605 (3.1)	1/228 (0.4)	6.51 (1.32–118.04)
Blood from vagina§	14/431 (3.2)	13/290 (4.5)	1/126 (0.8)	6.0 (1.11–112.4)
Blood in urine	10/827 (1.2)	9/601 (1.5)	1/226 (0.4)	5.14 (0.90–98.73)
Bleeding under skin	5/827 (0.6)	5/604 (0.8)	0/223	NA

The paper highlights that the case fatality rate is 70.8% (95%CI = 68.8 - 72.8). The median age of the patients is 32 years old. There were no significant differences in the median age between the different countries. Most patients are between 15 and 44 years old, 60.8%. Patients older than 44 have a higher chance to die from the Ebola virus, which can be concluded based on the odds ratio of 2.47 for this group of patients, which is a lot higher than for the other age categories. There is no significant difference in gender for the patients, so male and female have the same chance to get infected by the Ebola virus. Also for gender there is no significant difference between the different countries. The most common symptoms are Fever (87.1%), Fatigue (76.4%), Vomiting (67.6%), Diarrhoea (65.6%), Loss of Appetite (64.5%) and Headache (53.3%). For less common symptoms such as bloody nose and bleeding gums, the odds ratio is very high. More details can be found in table 1.

3.2 The Geographic Origin and Spread of the Epidemic

The first cases occurred in Guéckédou and Macenta districts in Guinea. During March 2014 the first cases were reported in Lofa and other districts in Liberia, and were followed by the discovery of cases in the capital of Guinea, Conakry. During May the epidemic in Guinea expanded to the neighbouring districts of Kenema and Kailahun in Sierra Leone. From July onwards there was a sharp increase in the number of cases in all three countries. Out of the 67 districts in the three countries, 43 reported Ebola cases. More than 90% of the cases has been reported in 14 of the districts. Figure 2 shows the geographic spread of the Ebola virus. Old cases are marked grey, where cases reported the last 21 days are marked red.

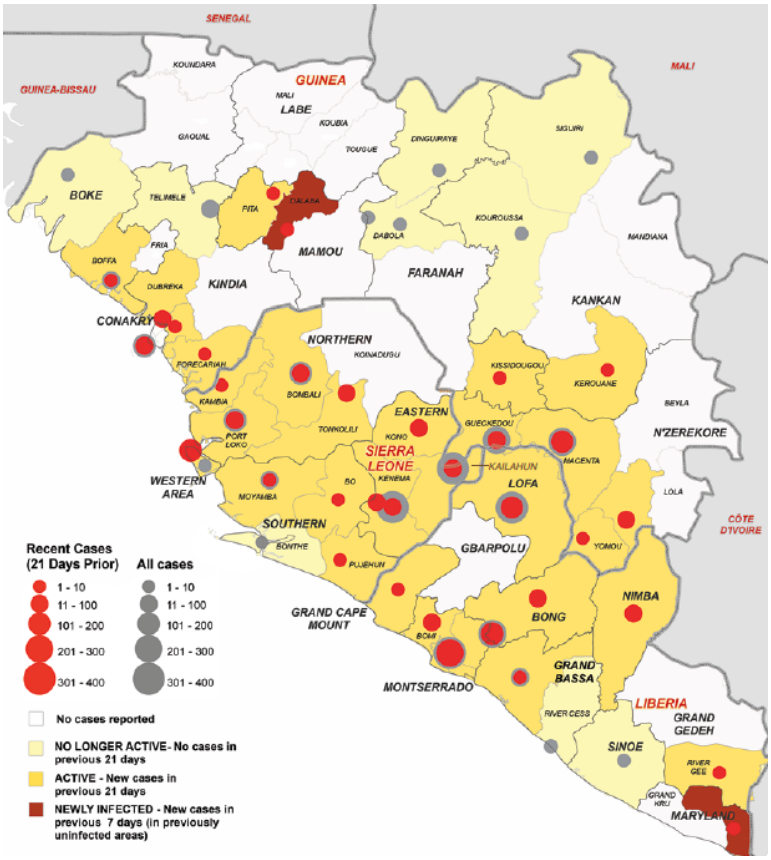

Figure 3: Districts Affected by Ebola Virus Disease

3.3 Projection of Future Cases

The main incubation time of the virus is determined to be 11.4 days, and this did not differ in the different countries. The mean serial interval is calculated to be 15.3 days. The mean time from the onset of symptoms to hospitalization was estimated to be 5.0 days. The mean time to death after admission to the hospital was 4.2 days, and the mean time to discharge was 11.8 days.

The doubling time of the epidemic was 15.7 days in Guinea, 23.6 days in Liberia, and 30.2 days in Sierra Leone in the beginning of September 2014. Estimates of the basic reproduction number, R_0, were 1.71 for Guinea, 1.83 for Liberia, 1.20 for Nigeria, and 2.02 for Sierra Leone. Since the number of cases is still growing in Guinea, Liberia and Sierra Leone this suggests that R_t remains larger than 1. The estimations for R_t are 1.81 for Guinea, 1.51 for Liberia, and 1.38 for Sierra Leone. More detailed numbers with accompanying confident intervals presenting the key time periods and the value of R can be found in table 2.

Based on this number the spread of the epidemic is estimated. Assuming no changes in control efforts, the cumulative number of confirmed and probable cases by the beginning of November 2014 will be 5740 in Guinea, 9890 in Liberia, and 5000 in Sierra Leone, exceeding 20,000 cases in total. The visualisation of this projection can be found in figure 4.

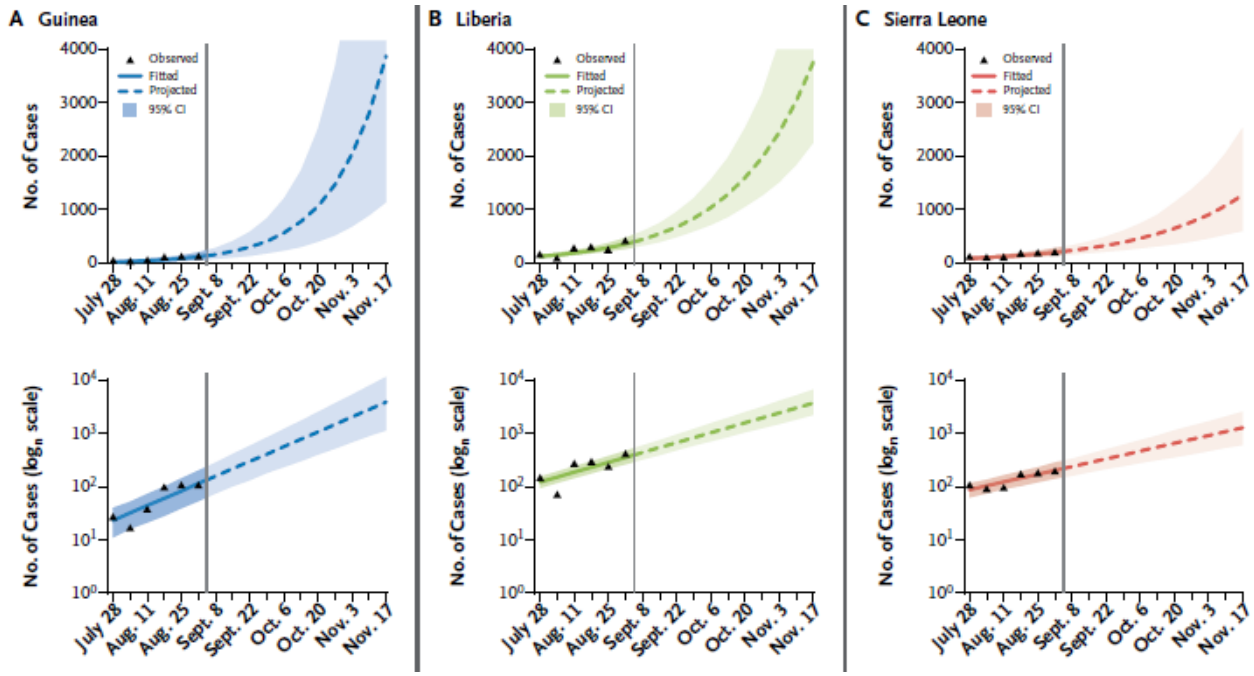

Figure 4: Projection of Future Ebola Cases

Table 2: Estimates of Epidemiologic Variables

Variable	All Countries		Guinea		Liberia		Nigeria		Sierra Leone	
	no. of days	no. of patients with data	no. of days	no. of patients with data	no. of days	no. of patients with data	no. of days	no. of patients with data	no. of days	no. of patients with data
Incubation period										
Single-day exposures										
Observed†	9.4±7.4	500	10.7±8.7	35	9.5±6.6	259	NC	<10	9.0±8.1	201
Fitted‡	9.1±7.3	500	9.9±9.8	35	9.4±6.7	259	NC	<10	8.5±7.6	201
Multi-day exposures										
Observed†	11.4±NA	155	10.9±NA	20	11.7±NA	79	NC	<10	10.8±NA	48
Fitted‡	9.7±5.5	155	8.3±4.5	20	9.9±5.7	79	NC	<10	9.9±5.6	48
Serial interval§										
Observed	15.3±9.1	92	19.0±11.0	40	13.1±6.6	26	NC	<10	11.6±5.6	25
Fitted¶	15.3±9.3	92	19.0±11.2	40	13.1±7.8	26	NC	<10	11.6±6.3	25
*R*0‖										
Mean (95% CI)	—		*1.71 (1.44–2.01)*		*1.83 (1.72–1.94)*		*1.2 (0.67–1.96)*		*2.02 (1.79–2.26)*	
Doubling time — days (95% CI)	—		*17.53 (13.18–26.64)*		*15.78 (14.4–17.37)*		*59.75 (13.27–∞)*		*12.84 (10.92–15.66)*	
*R***										
Mean (95% CI)	—		*1.81 (1.60–2.03)*		*1.51 (1.41–1.60)*				*1.38 (1.27–1.51)*	
Doubling time — days (95% CI)	—		*15.7 (12.9–20.3)*		*23.6 (20.2–28.2)*		*NC*		*30.2 (23.6–42.3)*	
Interval from symptom onset										
To hospitalization	5.0±4.7	1135	5.3±4.3	484	4.9±5.1	245	4.1±1.4	11	4.6±5.1	395
To hospital discharge	16.4±6.5	267	16.3±6.1	152	15.4±8.2	41	NC	<10	17.2±6.2	70
To death	7.5±6.8	594	6.4±5.3	248	7.9±8.0	212	NC	<10	8.6±6.9	128
To WHO notification	6.1±8.5	2185	7.5±10.4	743	6.0±8.7	797	3.9±2.3	11	4.5±5.0	634
Interval from WHO notification										
To hospital discharge	11.8±7.2	312	11.1±5.8	164	11±8.0	41	NC	<10	12.7±8.4	102
To death	−3.0±13.8	584	−4.4±14.4	300	−1.8±13.6	221	NC	<10	−1.6±9.2	58
Interval from hospitalization										
To hospital discharge	11.8±6.1	290	11±5.4	159	12.8±8.1	40	NC	<10	12.4±5.8	86
To death	4.2±6.4	121	2.5±3.4	36	4.5±6.0	63	NC	<10	4.4±6.0	17
Duration of hospital stay — days††	6.42		4.99		6.72		NC		6.88	
	rate (95% CI)	no. of patients with data	rate (95% CI)	no. of patients with data	rate (95% CI)	no. of patients with data	rate (95% CI)	no. of patients with data	rate (95% CI)	no. of patients with data
Case fatality rate										
All cases, based on current status	37.7 (36.1–39.2)	3747	57.5 (53.7–61.1)	677	34.7 (32.4–37.1)	1616	40.0 (19.8–64.3)	15	31.6 (29.3–34.1)	1439
All cases, based on definitive outcome	70.8 (68.6–72.8)	1737	70.7 (66.7–74.3)	542	72.3 (68.9–75.4)	739	45.5 (21.3–72.0)	11	69.0 (64.5–73.1)	445
Before August 18	71.3 (68.7–73.7)	1244	68.7 (64.3–72.8)	454	79.8 (75.7–83.4)	416	50.0 (23.7–76.3)	10	65.4 (60.4–70.1)	364
August 18–September 14	59.9 (54.7–64.9)	354	80.7 (71.2–87.6)	88	41.1 (34.3–48.2)	190	NC	<10	84.0 (74.1–90.6)	75
All hospitalized cases, based on definitive outcome	64.3 (61.5–67.0)	1153	64.7 (60.1–68.9)	450	67.0 (62.0–71.7)	361	40.0 (16.8–68.7)	10	61.4 (56.1–66.5)	332
According to sex										
Male	72.2 (69.1–75.1)	874	68.5 (62.6–73.9)	254	74.9 (70.4–79.0)	395	NC	<10	71.9 (65.7–77.5)	221
Female	69.9 (66.7–73.0)	818	72.7 (67.3–77.6)	286	71.6 (66.4–76.3)	317	NC	<10	64.4 (57.7–70.6)	208
According to age group										
<15 yr	73.4 (67.2–78.8)	218	78.1 (67.3–86.0)	73	70.7 (60.1–79.5)	82	NC	<10	71.4 (59.3–81.1)	63
15–44 yr	66.1 (63.1–69.0)	1012	64.9 (59.5–69.9)	319	70.6 (66.1–74.8)	422	NC	<10	61.4 (55.4–67.0)	264
≥45 yr	80.4 (76.2–84.0)	398	78.6 (71.1–84.6)	140	81.1 (74.4–86.4)	164	NC	<10	82.2 (73.1–88.8)	90
According to occupation										
Health care worker	69.4 (62.1–75.8)	170	56.1 (41.0–70.1)	41	80.0 (68.7–87.9)	65	NC	<10	68.4 (55.5–79.0)	57
Non–health care worker	70.9 (68.6–73.1)	1567	71.9 (67.8–75.6)	501	71.5 (68.0–74.8)	674	NC	<10	69.1 (64.3–73.5)	388

4 Conclusion

Based on the results of the analysis presented above, the authors of the paper conclude that the epidemic will keep expanding in the coming weeks if the control measures – improvements in contact tracing, adequate case isolation, increased capacity for clinical management, safe burials, greater community engagement, and support from international partners – do not change. The cumulative number of confirmed and probable cases will be 5740 in Guinea, 9890 in Liberia, and 5000 in Sierra Leone, exceeding 20,000 cases in total by the beginning of November.

The scale of this epidemic is exceptional large, even though the characteristics such as incubation period, the reproduction number and the most common symptoms are similar to those in previous Ebola outbreaks. Therefore the main cause of this large scale is not the virus itself, but rather the characteristics of the population and the (lack of) control measures. Due to the high interconnectedness between the population of Guinea, Liberia and Sierra Leone may have led to the rapid spread of the epidemic. However, Nigeria proved that a large outbreak can be avoided. The critical success factor seems to be the speed of the implementation of strict control measures.

4.1 Main Contributions

The results from this paper provide us with several valuable insights regarding this deadly and large scale Ebola outbreak. First of all, it provides us with insights into the characteristics of the epidemic. By for example determining risk factors, groups who have a high risk of dying can be identified. Furthermore, insights into the most common symptoms and their odds ratio can make the treatment of patients more efficient, and can also indicate the need for medical supply. Based on the future projections, the need for hospital beds can be estimated. Since the current number of beds will probably be not sufficient, action is needed to increase the number of hospital beds. Besides this, the paper also provides insights into the spread of the epidemic, and in this way make it possible for control measures to be implemented more efficiently. Two control measures the paper suggests are to track Ebola cases for minimum 21 days and to lower the period from symptom onset to hospitalization.

Next to this valuable insights, the paper creates a general sense of urgency by showing how serious this Ebola outbreak is in number of patients and the fatality ratio. In this way a fast introduction of control measures, additional research to vaccines and help from international partners is stimulated.

4.2 Limitations

The conclusions drawn by this paper are based upon the collected data. This causes some limitations and interpretation of the data requires knowledge and care. It is certain that not all cases are reported or have been detected. The collected case records are often incomplete. This can lead to biased data, and can for example result in an underestimation of the future cases.

There are two main points of interest for future research, since this can be influenced by the incomplete case records. First, the serial interval of 15.3 days is longer than past estimations. This can be caused by the biased data, but if this serial interval turns out to be reliable the case isolation used so far has been less effective than expected. Second, the case fatality rate appears to be the same for all countries, based on the data with definitive outcomes. If all cases are taken into consideration,

differences between the countries become visible. Further research is necessary to create more certainty about this point.

5 Replication of Plot

Since we were curious how the Ebola epidemic developed after the writing of this paper, we decided to replicate the plot which shows the number of Ebola cases in Guinea, Liberia and Sierra Leone in West Africa. This is the upper left picture of figure 2. Due to limitations of data, we executed the plot with data from November 2014 till March 2015. The data on which the original plot was based, was unfortunately not available anymore, as well as the data from September 2014 till November 2014. The data was obtained from the WHO data and statistics website[2].

5.1 Replication Result

Using the *ggplot package*, a bar plot was created to replicate the original plot. For the readability of the plot, we adjusted the position of the legend to the upper right corner instead of the upper left corner which is used in the original. The dataset that used to plot the bar plot below can be found in appendix, and the source code used to create the plot can also be found in appendix for details.

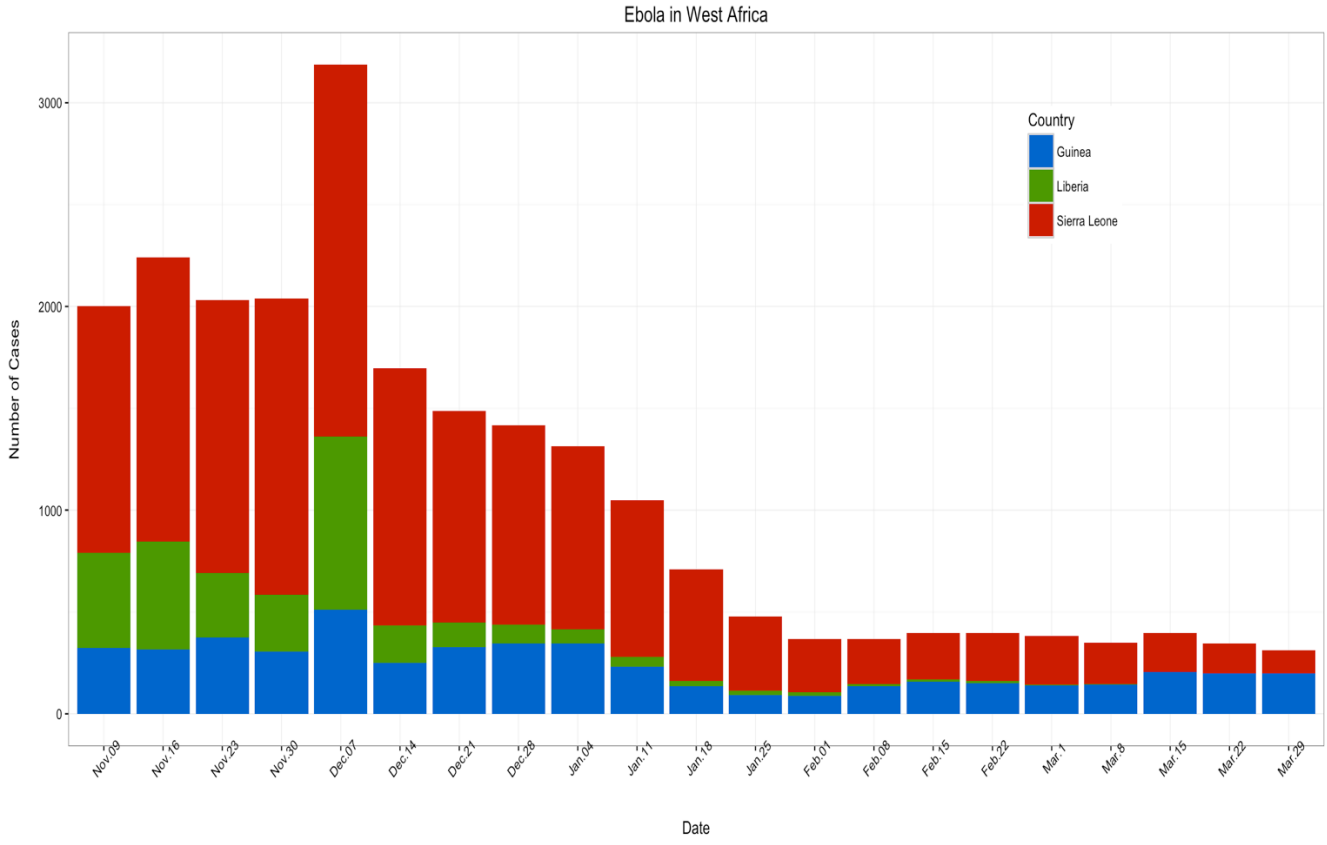

Figure 5: Replication of Plot

[2] WHO website for data and statistics of Ebola, see http://apps.who.int/gho/data/view.ebola-sitrep.ebola-summary-20150105?lang=en

5.2 Reflection on Projection

Out of interest in the development of the Ebola epidemic, we compared the projections of the paper with the actual development of the epidemic. The paper projected 5740 in Guinea, 9890 in Liberia, and 5000 in Sierra Leone, exceeding 20,000 cases in total by the beginning of November. Based on the data that we found from the 9th of November to 29th of March, the number of cumulative cases was 1878 in Guinea, 6822 in Liberia and 5368 in Sierra Leone. This shows that the projections of the paper were not accurate for Guinea and Liberia, but quite accurate for Sierra Leone. The difference in the number of cases in Guinea and Liberia can be causes by the implementation of enhanced control measures, where the paper assumed that the control measures were not enhanced. The paper could have created the urgency and given the knowledge to implement those measures. In this way, the paper still proves it value, even though the projections were not completely accurate.

Appendix

1. Dataset of Replication

	Nov.09	Nov.16	Nov.23	Nov.30	Dec.07	Dec.14	Dec.21	Dec.28	Jan.04	Jan.11
1	325	315	374	306	511	249	328	346	344	230
2	466	532	319	278	852	185	121	92	70	48
3	1211	1394	1339	1455	1824	1261	1039	979	900	769

Jan.18	Jan.25	Feb.01	Feb.08	Feb.15	Feb.22	Mar.1	Mar.8	Mar.15	Mar.22	Mar.29
136	92	89	134	156	152	138	144	204	198	197
25	20	17	12	11	10	6	4	0	1	1
549	366	262	221	230	235	240	202	194	146	113

2. Source code R

```
# Input dataset

Data.Ebola_histogram <- read.csv("C:/Users/Stephanie/Dropbox/3. Studie - Master SEPAM/Singapore/IS4242 Healthcare Analytics/Papers/Ebola database.csv", sep=";")


library(ggplot2)

library(reshape2)


data <- Data.Ebola_histogram

data$row <- seq_len(nrow(data))

data2 <- melt(data, id.vars = "row")

data2$row <- as.factor(data2$row)
```

```
# Store the graph in variable gg1

gg1 <- ggplot(data2, aes(x=variable, y=value, fill=row)) + geom_bar(stat="identity") + ggtitle("Ebola in West Africa") + xlab("\nDate") + ylab("Number of Cases\n") + theme_bw() + theme(axis.text.x=element_text(angle=45), legend.position =  c(0.8, 0.8))


# Add a Legend

legend <- scale_fill_manual(guide = guide_legend(title = "Country"), labels = c("Guinea", "Liberia", "Sierra Leone"), values=c("#0066CC", "#4C9900", "#CC0000") )


# Render the graph plus guide together at once

gg2 <- gg1 + legend

gg2
```