Ajla Sacic, Hayat Jan and Nataly Ojeda
Infectious Diseases: Preventing and Stopping Epidemics
Professor Stephane Helleringer
Analysis Report

Data analysis for the Ebola outbreaks in Yambuku, 1976 and Sierra Leone, 2014-2015

Data

In this analysis report, we mainly focused on the dataset from the first Ebola outbreak that took place in Yambuku from September 1, 1976 to November 5, 1976 as described in the epidemic profile. The epidemic lasted three months and a total of 318 cases and 280 deaths were recorded during this study period. Since it was first discovered in 1976, there have been many Ebola outbreaks in Africa. The world's largest Ebola outbreak was first reported in March 2014, and it was declared over by the World Health Organization (WHO) on June 10, 2016 ("Ebola outbreak 2014-2016 - West Africa", n.d.). While the epidemic expanded across Africa, Europe, and the United States, countries at the epicentre of the outbreak (Guinea, Sierra Leone, and Liberia) were impacted the most ("2014-2016 Ebola Outbreak Distribution in West Africa", 2021). There were a total of 28,616 suspected and confirmed cases and 11,310 deaths in these three countries over the course of the outbreak ("Ebola outbreak 2014-2016 - West Africa", n.d.). Despite the fact that the 2014-2016 Ebola epidemic affected several West African countries, Sierra Leone was hit the hardest. It had the highest total number of cases (14,124 out of the 28,616 cases) and the second-highest total number of deaths caused by Ebola (3,956 out of the 11,310 deaths) out of Guinea and Liberia, according to the case counts reported by WHO ("Ebola case counts", 2020).

As a result, we will also be using the dataset for Sierra Leone to compare the dynamics of the larger outbreak in Sierra Leone in 2014-2015 to the smaller Yambuku outbreak in 1976. However, there are systematic differences in how the data was collected and documented for the outbreak in Yambuku compared to the outbreak in Sierra Leone. In Yambuku, the data was collected retrospectively, after the disease had already progressed and infected many. This is because it took some time to recognise that this was a new disease and experts had to work backwards to determine what happened and how the disease was spreading. In comparison, when the Ebola outbreak began in Sierra Leone, they already knew what Ebola was and so they started collecting data straightaway. As a result, the Sierra Leone outbreak was monitored prospectively over time in real-time.

For our analysis of the Sierra Leone outbreak, the data was collected from the "Outbreaks" package in R which contains information on confirmed and suspected cases of Ebola in Sierra Leone across one year, from May 18, 2014 to September 13, 2015. Similar to the data for Yambuku, each row represents an individual case in the dataset, which is organised in a linelist format. Age, sex, status (whether confirmed or suspected), date of symptom onset, date of sample collection, district, and chiefdom were recorded for each individual case. A total of 11,903 cases were reported during the study period, with 8,358 of them being confirmed and 3,545 of them being suspected.

In our analysis, we decided to focus on the gender distribution of infected patients in the Yambuku and Sierra Leone outbreaks. This decision was guided by the fact that in Yambuku, disease spread was amplified through hospitals and hospital staff positions are highly gendered, such as nurses. This meant that those who cared for infected Ebola patients and handled their blood or body fluids were mainly women, putting them at a higher risk of contracting the disease.

Methods

The Yambuku Ebola outbreak data was provided in a text (.txt) format so it needed to be converted to a Comma Separated Values (.csv) format through Microsoft Excel (or any other application) and exported. After that, it was loaded into R Studio where it was analysed.

After importing the data, the data needed cleaning as it had cells which contained "NA" labelled data. Furthermore, the date columns, namely disease_onset and disease_ended, were not in proper objects as they were character objects. They were converted to date objects which meant that the data could be properly analysed.

Epidemic Curves

To plot the epidemic curves, the incidence package offered through CRAN repository was used, in particular, the function incidence(). It takes in date object data and a constant time interval and returns an object of class incidence which can be used to plot an epidemic curve. Epidemic curves were plotted for the Yambuku and Sierra Leone Ebola outbreaks after which the two plots were combined to be shown in one window using the plot_grid() function from the cowplot package. The script also contains the plots of epidemic curves which take sex into consideration and plots epidemic curves that display the bars that designate the two sexes, as well as 'NA' in the case of Sierra Leone.

Case-fatality rate

To plot the case fatality rate (CFR), a variable was needed to be produced first which would store the number of deaths for each day of the outbreak in Yambuku. By filtering the status variable in the Yambuku data set which consists of either "died" or "recovered" and dividing it by the total number of cases, a graph which showed the CFR of the first Ebola outbreak was produced. The Sierra Leone data set did not contain indications of death or recovery, but according to Forna et al. (2020), although there was a high number of unreported survival cases, the CFR was approximately 40%.

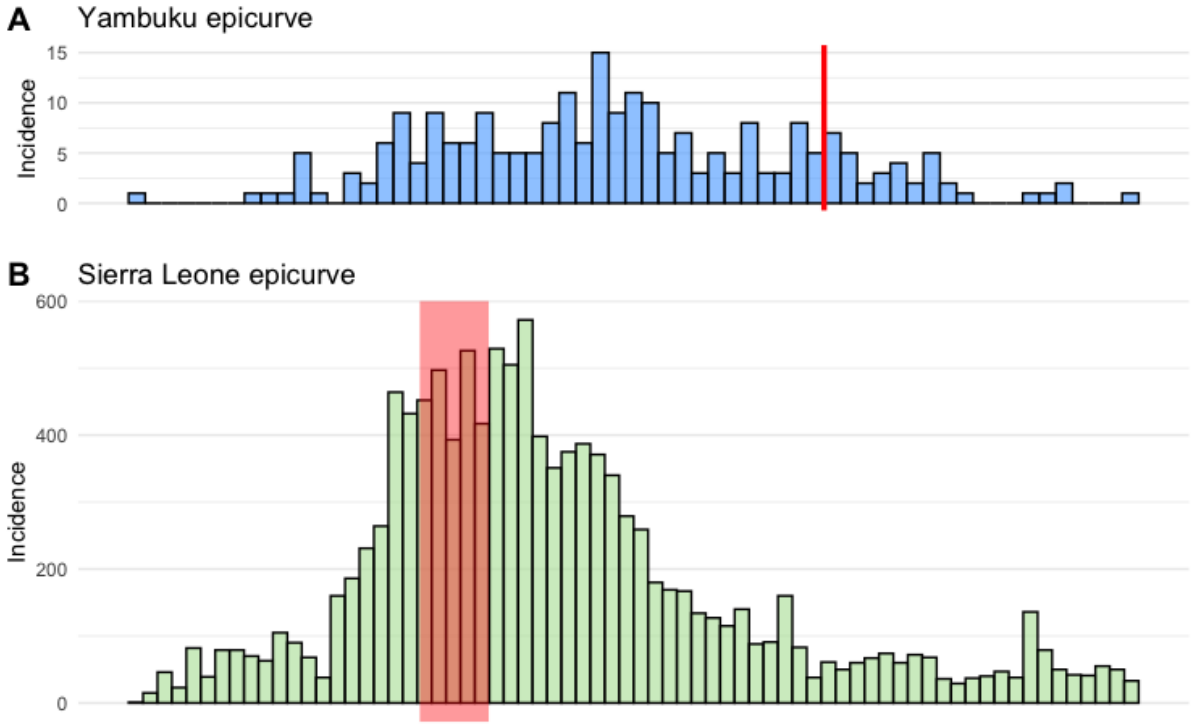
Gender ratio incidence

The incidence of both the Yambuku and Sierra Leone outbreaks was plotted based on the sex/gender ratio by plotting the ratio of females and males (the sum of females divided by the sum of males) against the week of onset of the disease (Yambuku) or month of onset of the disease (Sierra Leone). Within the plots, a dashed blue line was added to designate the even ratio of females and males, i.e., when it is 1. Above the line, the graphs suggest a larger number of females were infected, whereas, underneath the line, it suggests the opposite.

Estimating Reproductive Number

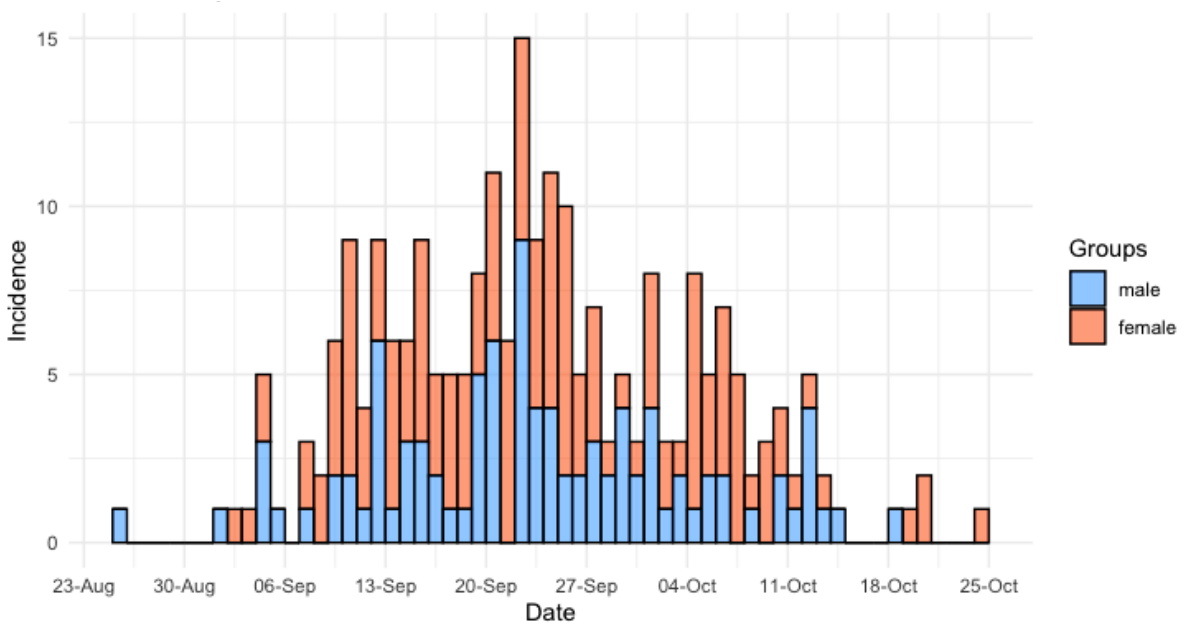
The reproductive (R) number of the Yambuku and Sierra Leone outbreaks was estimated and plotted with the help of the EpiEstim package. The Yambuku data set contains the type of transmission which makes it hard to properly estimate the reproductive number as most of the cases were due to syringes and not person-to-person transmissions. To overcome this, the R number was estimated by extracting the number of cases for each unique disease_onset date and using the estimate_R() function from the aforementioned package.

Results

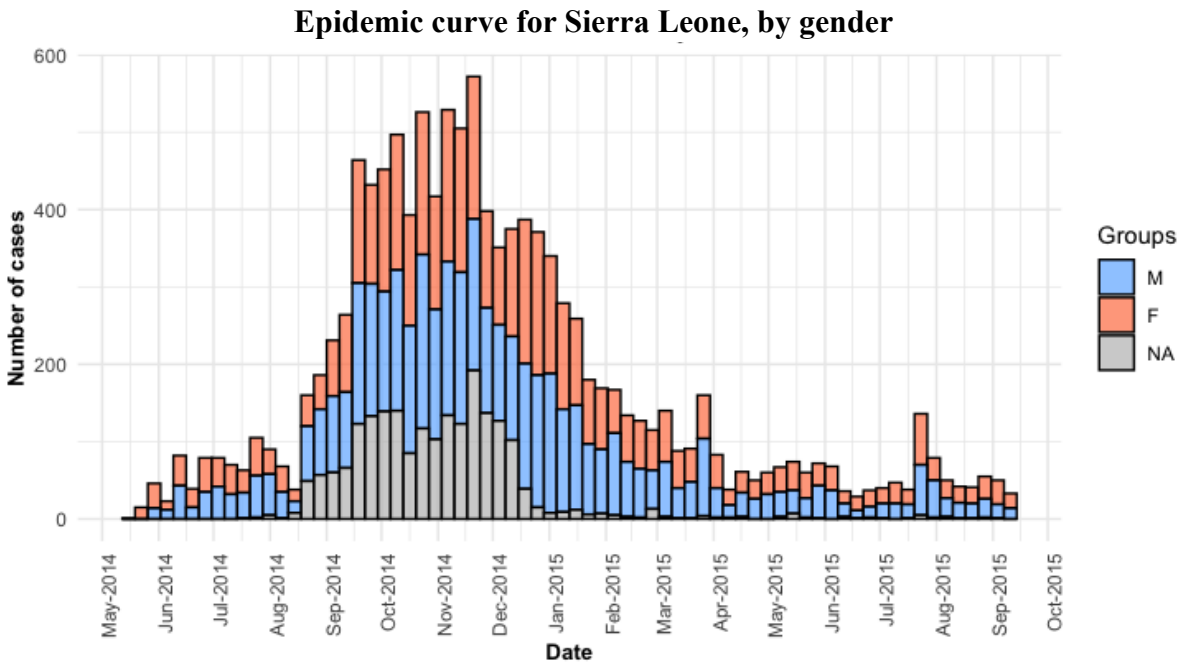


The epidemic curves show how the Ebola outbreak unfolded over time in Yambuku (A) and in Sierra Leone (B). The red line and area designates the implementation of preventive measures in Yambuku (October 6th) and Sierra Leone (October to November 2014), respectively. By removing the date labels from the x-axis, we can analyse the general progression of each outbreak. The Yambuku outbreak reaches the maximum number of incidences before the intervention took place whereas in Sierra Leone, the intervention measures occured before the maximum incidence was recorded. The CDC intervention in Sierra Leone, which included expanded airport screening and improved hospital readiness, helped to minimise illness spread over time, as seen in the graph. In Yambuku, the closure of the hospital, which was considered the epicentre of the outbreak, helped to fully reduce the spread of the disease. Considering the scale of the outbreaks, it is noticeable that incidence in Sierra Leone was much higher, reaching numbers as high as 600 per week, while the peak incidence rate in Yambuku was 16. Furthermore, the progression of the Yambuku outbreak seems to vary much less than in Sierra Leone.

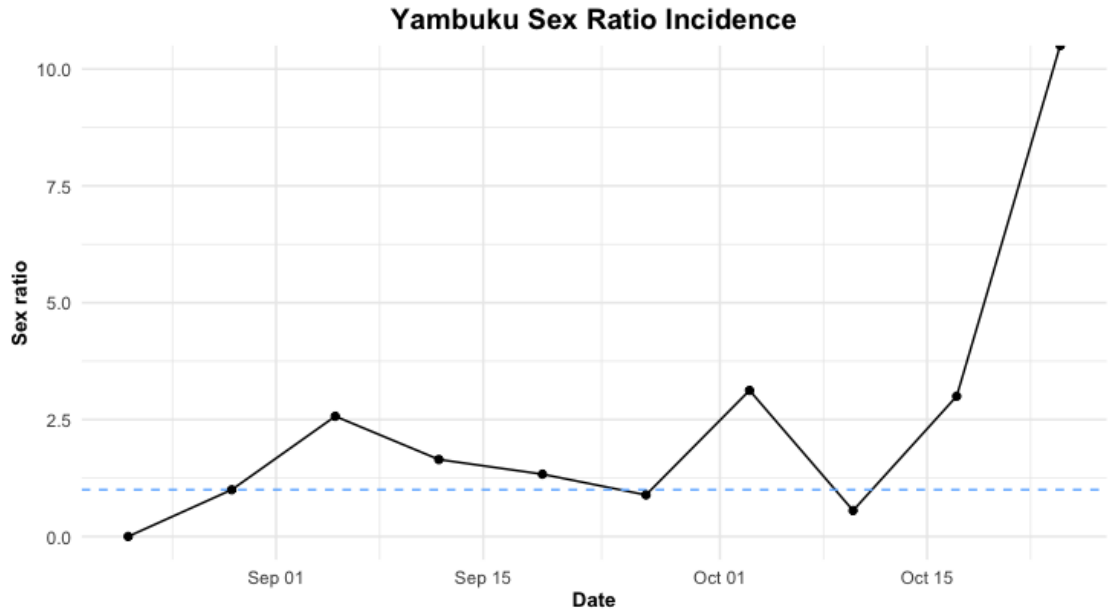
Epidemic curve for Yambuku, by gender



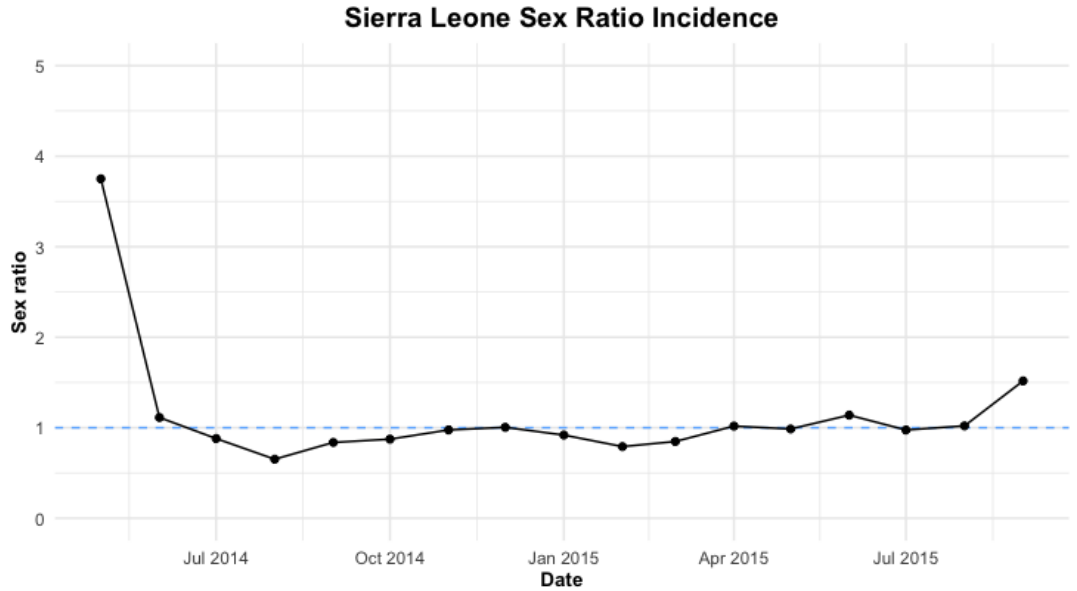
The epidemic curve of the Yambuku outbreak, when divided by gender, shows that there were significantly more Ebola cases among women than men. For example, on September 22, 1976, there was an incidence rate of ~6 where only women contracted Ebola. Similarly, as the cases start to decline towards the end of October, we can see that only women were becoming infected. This makes sense as, in addition to syringe use, coming into contact with infected blood or bodily fluids was contributing to the spread of Ebola among the Yambuku population. Those infected would visit clinics and hospitals where nurses, the majority of whom were female, cared for sick patients and were in close proximity to those infected. This made them particularly vulnerable to contracting the disease, and without proper protective gear, allowed for high rates of transmission which we can see in the gender divide of Ebola cases.



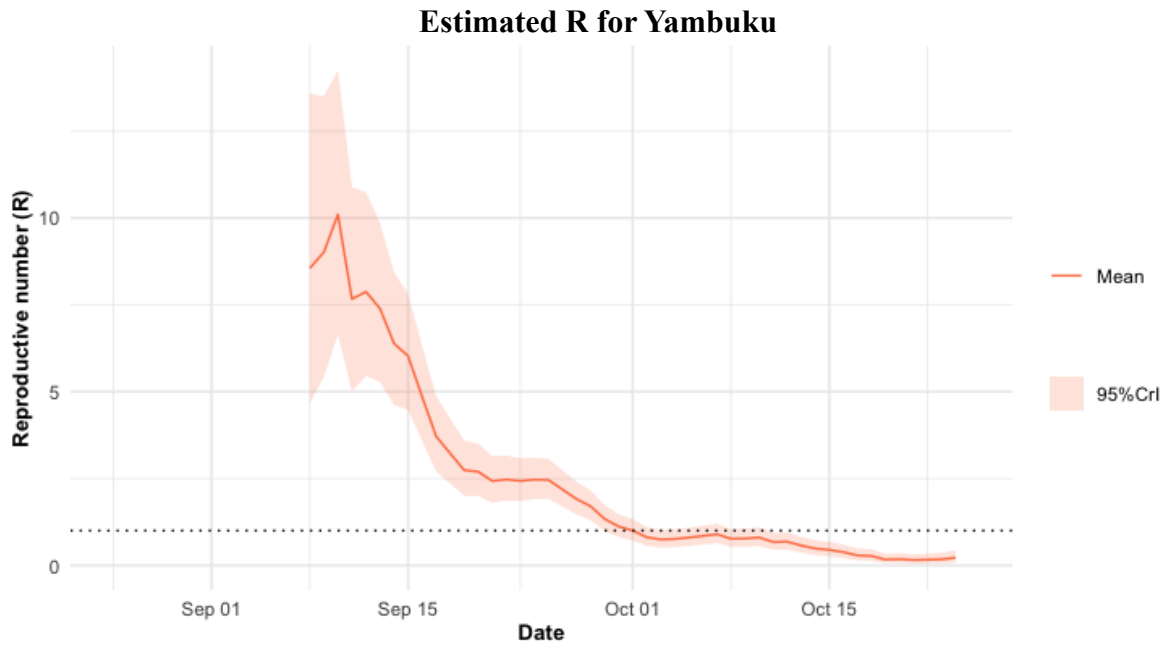
Looking at the epidemic curve for Sierra Leone, when divided by gender, we can see that women and men were infected at a much more equal rate. In fact, on some dates, the number of cases of males was higher than that of females, which was not the case during the Yambuku outbreak. This makes sense given that the ratio of female to male workers in certain hospital staff positions was more balanced in 2014-2015 than it was in 1976. It is also worth noting that around three months into the Sierra Leone outbreak, there were cases where the gender of the patient was not recorded. This could be due to the rapid spread of the disease, as the peak number of cases on the graph is where grey “NA” responses appear. Thus, epidemiologists may have been unable to keep up reporting the rapidly increasing cases.



When looking at the sex ratio for the Yambuku outbreak, the majority of the peaks in the graph indicate that women were the ones primarily contracting Ebola compared to men. While the peaks represent the ratio of women infected with Ebola during the short-lived outbreak, we can also see that the ratio rarely, if ever, fell below 1 (which indicates more men contracting the disease than women). This graph supports the epidemic curve's claims that the outbreak began in a clinic due to syringe use, and that the majority of nurses working there were female, making them more exposed and vulnerable to contracting the disease.

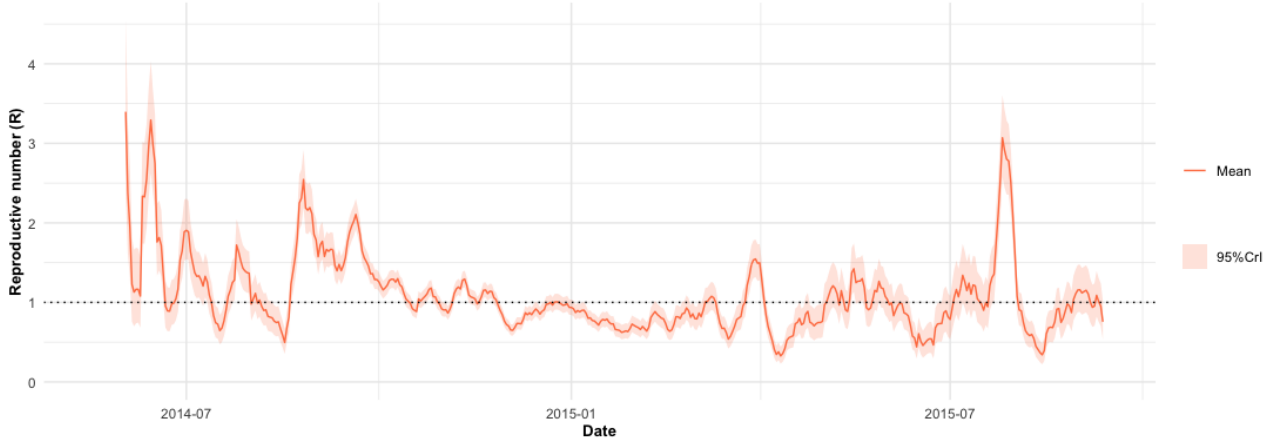


The sex ratio incidence of Sierra Leone's outbreak shows that at first, women contracted the disease at a much higher rate than men. However, the sex ratio of cases seems to change over the course of the outbreak after 4 months, the sex ratio moved closer to 1, indicating the ratio of men to women contracting Ebola is much more evenly balanced. In fact, in contrast to Yambuku's outbreak, Sierra Leone had more men contracting Ebola than women, as evidenced by the ratio falling below 1 twice. This is in line with the gender epidemic curve of the Sierra Leone outbreak.



The estimated reproductive number (R) tells us how infectious a disease is over the course of the outbreak. In Yambuku, the estimate of R was well above 1 for about 2 weeks, but then fell below 1 on October 1st. The high estimated R-value seen at first at around 8, indicates that the disease spread quickly throughout the community, alluding to the outbreak crisis and epidemic the city experienced. The gradual decline from mid-September to October could be explained by the fact that the Yambuku outbreak occurred in a small town with a small population. As a result, the majority of people would have contracted the disease and either died from it or developed a degree of immunity to slow its spread. Furthermore, many of the Ebola cases in Yambuku were contracted through syringe use rather than person-to-person contact. This could also influence the potential of transmission, as the first great spike in the graph could be attributed to syringe use, and the decline could be attributed to the closure of the hospital as well as the quarantine measures that were implemented. This is because, with these preventative measures in place, the spread of Ebola was more controlled and lessened, resulting the R-value falling below 1.

Estimated R for Sierra Leone



In Sierra Leone, however, there was a lot more variation in the estimated R-value of the spread of Ebola. We can see that there were spikes showing high transmission rates in the early days of the outbreak, which gradually tapered off to an R-value of less than one as the case incidence stabilised. However, once it was below 1, there were regular spikes in Ebola transmission and spread, particularly in October 2014, April 2015, and August 2015. This could be due to Ebola variant mutations causing an increase in the number of cases. Furthermore, the changing reproductive number could be due to the fluctuation of transmission rates through contact with body fluids, which is how the outbreak in Sierra Leone spread.

References

2014-2016 Ebola Outbreak Distribution in West Africa. (2021). Retrieved 1 May 2022, from https://www.cdc.gov/vhf/ebola/history/2014-2016-outbreak/distribution-map.html.

Ebola case counts. (2020). Retrieved 1 May 2022, from https://www.cdc.gov/vhf/ebola/history/2014-2016-outbreak/case-counts.html.

Ebola outbreak 2014-2016 - West Africa. Who.int. Retrieved 1 May 2022, from https://www.who.int/emergencies/situations/ebola-outbreak-2014-2016-West-Africa.

Forna, A., Dorigatti, I., Nouvellet, P., & Donnelly, C. (2020). Spatiotemporal variability in case fatality ratios for the 2013–2016 Ebola epidemic in West Africa. *International Journal Of Infectious Diseases*, *93*, 48-55. https://doi.org/10.1016/j.ijid.2020.01.046.