A Mathematical Modelling Approach for Understanding the 2014 -2016 Ebola Virus Outbreak in Sierra Leone and Liberia

Thesis · October 2018

DOI: 10.13140/RG.2.2.31531.59685

2 authors:

Annika Brundyn
New York University
1 PUBLICATION **0** CITATIONS

SEE PROFILE

Sheetal Silal
University of Cape Town
52 PUBLICATIONS **557** CITATIONS

SEE PROFILE

A Mathematical Modelling Approach for Understanding the 2014 - 2016 Ebola Virus Outbreak in Sierra Leone and Liberia

Annika Brundyn

Supervisor: Dr Sheetal Silal

Department of Statistics, University of Cape Town

October 2018

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Ebola virus disease (EVD) has captured the imagination of the public and experts alike. This fascination is partly due to its overall rare occurrence and typically very few outbreaks, the severe, often graphic symptoms associated with EVD and the extremely high case-fatality rates, ranging anywhere between 25% and 90% [13]. Moreover, the identity of the natural Ebola virus reservoir remains unknown. This lack of knowledge means that the effect of novel Ebola virus introductions into human populations cannot easily be discerned or predicted, let alone prevented, adding to the enigma of the virus in the public's eye.

Health officials have been aware of Ebola Virus since 1976, but its history of sparse small outbreaks led to scepticism of its potential to cause large-scale damage. The death of a toddler in South-Eastern Guinea, believed to have been infected by bats in December 2013, evolved to be the largest Ebola Virus Disease outbreak documented in history. Spanning 2014 to 2016, the outbreak predominantly affected the West-African countries of Guinea, Liberia and Sierra Leone. Over 28,600 cases, and over 11,300 deaths resulting from Ebola were reported during the outbreak. In comparison, there were 2,427 reported cases and 1,597 deaths in all other known cases and outbreaks of Ebola since 1976 combined [13].

In June 2018, sporadic cases were identified and confirmed to be EVD in the Democratic Republic of Congo. As of August 29, 2018, a total of 116 EVD cases including 77 deaths have been recorded there. The nature of the virus makes it unlikely that it will ever exert such a tragic effect in developed countries. Factors frequently identified as contributing to the unprecedented scale of the outbreak include the slow identification of the virus and initial response by authorities, high population mobility across porous borders and into dense urban areas, damaged public health systems, severe shortages of health care workers, cultural beliefs and behavioural practices [64]. In summary, the threat of Ebola has not subsided, and the virus continues to prey on the damaged infrastructure of countries whose resources are already strained as they continue to control various other infectious diseases such as HIV, Tuberculosis and Malaria. Further investigation of the poorly-understood mechanisms driving the sporadic reappearance of the disease may contribute to more efficient control of Ebola outbreaks in future.

Furthermore, the challenges associated with containing the 2014 - 2016 outbreak provided crucial lessons applicable to disease management in general. The outbreak is said to have "demonstrated the lack of international capacity to respond to a severe, sustained, and geographically dispersed public health crisis" [64]. Additionally, it provided a desperate reminder of the gravity of contextual consideration in public health response. With the exponential advances in medicine and the growing

capability to provide accessible health care, we neglect the importance of individual and community buy-in. Our faith and high regard in science, does not always translate to useful solutions on the ground. For these reasons, Ebola Virus Disease provides a valuable case study in understanding the dynamics of unfamiliar infectious diseases, and determining appropriate response in highly contextual environments.

Mathematical modelling has become an integral tool for aiding our understanding of the dynamics of infectious diseases, and its application has helped decision-makers to investigate potential outcomes and strategies. An important benefit of mathematical modelling is the ability to enact exogenous change on the modelled system in order to predict the impact of interventions without committing any real resources; a key benefit in a setting with scarce resources and time constraints.

This study aims to use mathematical models to simulate Ebola Virus Disease transmission in order to better understand the key forces that govern an epidemic and assess the impact of control measures on disease outcomes. Of the three countries, significantly more cases were recorded in Sierra Leone (>14,100) and Liberia (>10,600) than in Guinea (>3,800). Consequently, due to constraints in time and computational power, this analysis exclusively focused on modelling EVD in Sierra Leone and Liberia. Explicitly, the research objectives for this study are:

- To develop a model for the transmission dynamics of EVD, as it presented in Sierra Leone and Liberia during 2014 - 2016, and estimate the key parameters that drove it's behaviour from data.
- To estimate the basic reproductive number as a measure of outbreak severity.
- To assess the impact of intervention and control measures, and the timing thereof.
- To explore the role of culture and context in shaping the course of an epidemic.

In answering these questions surrounding the dynamics of EVD transmission, two deterministic mathematical models with compartmental structures are presented. The first model captures many of the more nuanced features of the Ebola virus but proved to be too complex to uniquely estimate the multiple unknown parameters from the limited data. Therefore, to preserve model identifiability a second, simpler, model was developed. This model extends from the traditional Susceptible-Infectious-Recovered (SIR) framework to include the essential features of EVD with the least number of unknown parameters. The known latency period of EVD is incorporated through the inclusion of an exposed compartment. Additionally, the gravity of death in EVD transmission and the importance of burial practice in West Africa motivated the inclusion of two death-related states: one in which a deceased victim of EVD is still infectious and another, in which they are safely buried or cremated and no longer capable of infecting others. This model subsequently provides the basis for assessing the inclusion of intervention measures.

Chapter 2 presents an extensive literature review on the Ebola virus history, biology and control, as well as a detailed description of the 2014 - 2016 West Africa outbreak, with a focus on Sierra Leone and Liberia. Existing research and applications in the mathematical modelling of Ebola virus are discussed. Chapter 3 describes the recorded case data available for the outbreak as provided by the Centre for Disease Control (CDC). This is the secondary dataset that is used throughout the study to inform the mathematical models. Chapter 4 forms an extensive discussion of the sequential development of compartmental disease models, ordinary differential equation modelling (the main methodology employed in the mathematical models) and use of the base reproductive number. Chapter 5 presents parameter estimates, model predictions and parameter sensitivity for both of the final models fitted to the epidemic data. Additionally, the impact and timing of both

intervention measures deployed and alternative solutions are discussed. Chapter 6 synthesises the obtained model results with insights from literature and other sources, in a summary of the lessons provided by the 2014 - 2016 West Africa outbreak of EVD.

Chapter 2

Literature Review

2.1 History of Ebola Virus Disease

Ebola Virus Disease (EVD) was first identified in 1976, when two simultaneous outbreaks of fatal haemorrhagic fever were recorded in different parts of Central Africa. The first outbreak occurred in Yambuku, Democratic Republic of Congo (DRC, formerly Zaire) in a village near the Ebola River; from which the virus takes its name. The second outbreak occurred in what is now Nzara, South Sudan, approximately 850 km away [71]. Owing to the proximity of the two outbreaks, public health officials initially assumed that they were a single event associated with an infected person who travelled between the two locations. However, they were subsequently found to be caused by two genetically distinct strains of the virus, which were later named: Zaire ebolavirus and Sudan ebolavirus [17]. Viral and epidemiological research suggests that Ebola virus existed before these two outbreaks were recorded [13].

The causative agent of Ebola virus is an RNA virus of the family Filoviridae, which includes three genera: Cuevavirus, Marburgvirus, and Ebolavirus [22]. Within the genus Ebolavirus, five strains have been identified: Zaire, Sudan, Bundibugyo, Ivory Coast (Taï Forest) and Reston. Reston ebolavirus has only been observed to infect non-human primates. Since its discovery in 1976, 29 sporadic outbreaks or case reports of EVD have been reported in predominantly African countries [17]. Zaire ebolavirus, Bundibugyo ebolavirus, and Sudan ebolavirus are the three species of Ebola virus responsible for the larger outbreaks in Africa. Zaire ebolavirus was associated with the 2014-2016 outbreak in West Africa, the most severe Ebola outbreak to date with over 28,600 cases identified. Prior to this, outbreaks and cases were predominantly limited to rural communities in Sudan, Democratic Republic of Congo, Republic of Congo, Gabon and Uganda. Most of these outbreaks were small in size with just seven outbreaks involving more than 100 cases. The largest of the outbreaks before 2014 occurred in Uganda in 2000–2001 with 425 cases and 224 deaths (Sudan strain) and DRC (Kikwit) in 1995 with 315 cases and 250 deaths (Zaire strain) [17, 13]. Bundibugyo ebolavirus, discovered in 2007, was associated with two outbreaks, one in DRC and the other on the border of DRC and Uganda. Taï Forest ebolavirus, the only other Ebola virus discovered in West Africa, was the cause of a single case identified in Côte d'Ivoire [13].

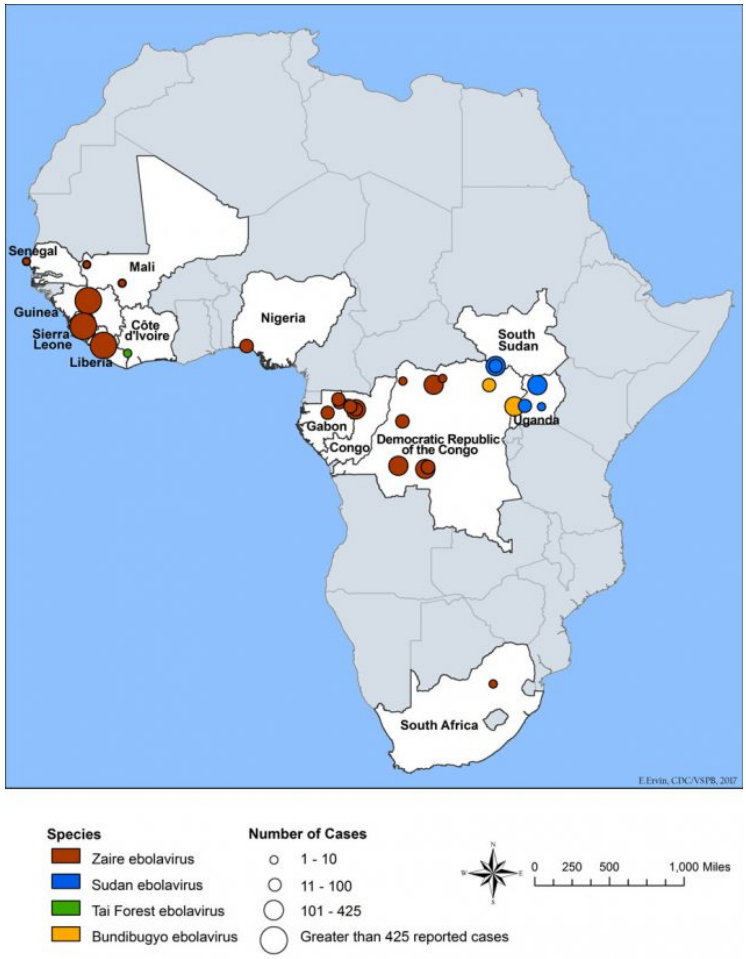

Figure 2.1: Ebola Virus Outbreaks by Species and Size, since 1976 [13])

2.2 Ebola Virus Disease

Ebola virus disease (EVD), formerly known as Ebola haemorrhagic fever, is a severe and deadly disease when presented in humans. Outbreaks from the same Ebola virus strain have resulted in varying case fatality estimates, but there is consensus that the Zaire ebolavirus is the most virulent among the five species and historically, also the most common. Since 1976, it has caused multiple outbreaks in Africa, with fatality rates ranging between 25% - 90% [27, 13, 71]. Despite the considerable international attention the 2014 - 2016 outbreak in West Africa received, the pathogenesis of the disease remains remarkably poorly understood.

2.2.1 Virus Transmission

It has not been established exactly where the Ebola virus originates. However, based on the nature of similar viruses, it is believed that the virus is animal-borne. Evidence strongly implicates bats as the natural reservoir hosts (source animal) for ebolaviruses [71]. Bats carrying the virus can transmit it to other animals, like apes, monkeys, duikers, porcupines and humans [13]. Scientists

continue to search for conclusive evidence of the bat's role in the transmission of Ebola. However, one explanation is that when bats drop partially eaten fruits and pulp, land mammals such as gorillas and duikers feed on these fallen fruits and are consequently infected [22]. This possible chain of events could serve as an indirect means of transmission from the natural host to other animal populations [74].

It is believed that infection from the animal population to humans occurs through close contact with the blood, secretions, organs, or other bodily fluids of infected animals, such as a fruit bats or non-human primates. This is called a spillover event. Factors like population growth, encroachment into forested areas, and direct interaction with wildlife (such as bush meat consumption) may have encouraged this type of contact and led to the spillover of the Ebola virus [13].

The virus spreads amongst people when human-to-human transmission occurs through direct contact with broken skin such as wounds, or mucous membranes (in the eyes, nose, mouth or vagina, for example) with the following:

- blood or body fluids (including but not limited to urine, saliva, sweat, faeces, vomit, breast milk, and semen) of a person who is sick with - or has died from - Ebola,
- objects (like needles and syringes), surfaces, and materials (e.g. bedding, clothing) that have been contaminated with body fluids from a person who is sick with Ebola, or the body of a person who has died from Ebola,
- infected fruit bats or primates (apes and monkeys, e.g. consumption of bushmeat)
- possibly from contact with semen from a man who has recovered from Ebola (for example, by having oral, vaginal, or anal sex) [13, 71]

Ebola virus is not an airborne disease, it requires an intimate level of direct contact with an infected individual who is almost certainly symptomatic, in order for the virus to spread. In general, transmission is unlikely to occur during the incubation period and the transmissibility increases with the duration of disease [2]. Health-care workers have frequently been infected while treating patients with suspected or confirmed EVD. Caring for EVD patients often requires very close contact and when control precautions are not strictly practised, the virus will almost certainly infect a caretaker.

During the 2014 - 2016 outbreak, a Liberian man infected with EVD travelled to the United States and managed to infect two Dallas nurses that cared for him prior to his death, despite stringent control measures. It is believed that an inconspicuous control breach such as incorrect removal of the protective gear resulted in their infection [75]. If transmission is possible in a developed well-functioning hospital despite the nurses wearing protective masks, gowns, shields and gloves - as well as working in a bleach washed environment - it is clear that there is great danger associated with patients being cared for in the family home. It is estimated that 74% of transmission occurred between close relatives during the 2014 - 2016 outbreak [13].

People remain infectious as long as their bodily fluids contain the virus and the viral load increases as the infection progresses, making direct contact with the bodies of those who died from EVD one of the most dangerous – and effective – methods of transmission [27]. Burial rituals and practices are another highly significant factor, and intimate ceremonies in many African countries (where religion and belief in an afterlife are common) pose another serious potential risk for transmission and infection.

Ebola virus can remain in certain bodily fluids after a person has recovered from the infection [13]. While there is no known risk of getting EVD through casual contact with an Ebola survivor,

the virus can possibly remain in immunologically privileged sites of the body for several months following acute infection. These are sites where viruses and pathogens, like the Ebola virus, are shielded from the survivor's immune system, even after being cleared elsewhere in the body. These areas include most notably the semen of males, as well as the interior of the eyes, and the central nervous system, particularly in the cerebrospinal fluid. It is not yet established whether the virus does in fact spread through sex post-recovery, or exactly how long it takes to clear the virus, but theoretically it is possible and the World Health Organisation (WHO) advises to abstain from sex or to use condoms for at least three months after the patient has recovered. EVD survivors are subjected to the regular testing of bodily fluids after recovery to ensure the virus is eventually fully cleared [13]. The virus can also survive for several hours on dry surfaces such as doorknobs and counter tops, and at least several days in bodily fluids, like blood, at room temperature.

2.2.2 Natural history and Symptom Development

Ebola is a severe acute viral illness often characterised by the sudden onset of common flu-like symptoms such as fever, fatigue, joint and muscular pain, sore throat, and headaches [71]. Initial symptoms typically appear between 8 and 10 days after exposure to the virus, but the incubation period can span 2 to 21 days [13]. Humans are not infectious until they develop symptoms. Typically vomiting, diarrhoea, and a rash follow as symptoms worsen, along with decreased function of the liver and kidneys. At this stage, in some cases people begin to bleed both internally and externally (e.g. bleeding from the eyes, nose, gums and mouth or blood in the stools). Ebola virus disease was initially named Ebola haemorrhagic fever due to first descriptions of Ebola in 1976 stating that up to 75% of patients had experienced haemorrhagic manifestations. Subsequently however, bleeding has not been noted as often and is not considered a universal feature of the illness [41]. That being said, these complications are common causes of death in fatal cases [27].

The most common signs and symptoms reported in West Africa during the 2014 - 2016 outbreak (beginning from symptom onset to the time the case was detected) include: fever (87%), fatigue (76%), vomiting (68%), diarrhoea (66%), and loss of appetite (65%) [13].

Patients with fatal disease usually develop more severe clinical signs early on during infection and die typically between days 6 and 16 of complications including multi organ failure and septic shock (mean of 7.5 days from symptom onset to death during the 2014 - 2016 outbreak in West Africa) [13]. In non-fatal cases, patients may have fever for several days and improve, typically around day 6. Normally symptoms dissipate 9.4 days after symptom onset in survivors.

There was speculation about the possibility of sub-clinical (asymptomatic) EVD exposure and infection in non-survivors. A study done by Glynn and colleagues [29] that followed up on the contacts of people with proven EVD in Sierra Leone showed that asymptomatic Ebola virus infections occur extremely rarely, even when individuals have direct contact to others infected with Ebola virus. This result is in line with the observation that individuals infected with Ebola virus typically experience grave and often lethal symptoms [36].

2.2.3 Diagnosis

EVD can be difficult to initially identify even in a well-functioning health system because its early symptoms closely mimic those of other common illnesses, including malaria, dengue fever, typhoid fever, viral illness, meningitis and gastroenteritis [71, 23, 22]. Thus, in an already weakened health system, the task of quickly but correctly identifying and isolating Ebola patients before laboratory test results are available, is particularly challenging. This fact can result in missed opportunities to isolate infectious patients (through incomplete screening sensitivity) and expose non-Ebola patients to nosocomial infection (through incomplete specificity) [46].

Confirmation that symptoms are caused by Ebola virus infection are made using the following diagnostic methods: [71]

- antibody-capture enzyme-linked immunosorbent assay (ELISA)
- antigen-capture detection tests
- serum neutralisation test
- reverse transcriptase-polymerase chain reaction (RT-PCR) assay
- electron microscopy
- virus isolation by cell culture.

Currently, the most common laboratory test to identify EVD relies on a reverse transcription PCR [46, 60], which is not a rapid point-of-care (POC) test but instead requires substantial laboratory infrastructure. In the West Africa outbreak, patient blood samples were typically sent to off-site laboratories set up through the international response. Although the test itself can be done in hours, the round trip from a health facility to the laboratory often took over three days, especially during the peak of the epidemic [60]. Although a few rapid POC EVD diagnostics were developed and field-tested during this outbreak, they are not yet ready for widespread commercial use [46].

In lieu of rapid POC EVD tests to identify EVD-positive cases, a standardised EVD case definition from the WHO was used during the epidemic as the primary tool for initially identifying potential EVD patients [66]. Because false negatives for EVD put patients and their communities at great risk, this case definition is broad (high sensitivity/low specificity). A broad case definition is also useful for epidemic surveillance [46]. The case definitions will be presented in more detail in Chapter 3.1. Patients meeting these case definitions, based on broad symptom and/or exposure criteria, were sent to holding centres for EVD testing and isolation. However, the broad case definition meant that negative and positive EVD patients were mixed together, often for days, until their test results were available and treatment facilities had beds for the positive patients. Although some holding centres tried to separate suspect patients based on wet (i.e. diarrhoea or vomiting) versus dry symptoms, this crude separation can expose Ebola-negative patients, particularly those with wet symptoms, to a higher risk of nosocomial infection [46].

2.2.4 Treatment and Vaccines

There is as yet no proven treatment or licensed vaccine. The course of treatment for infected patients primarily involves supportive care: providing symptomatic relief and ensuring rehydra-

tion while the body fights the infection. Intravenous fluids, antibiotics, and oxygen are usually employed. Treatment may also include the use of medications to control fever, help the blood clot, and maintain blood pressure. A range of potential treatments including blood products, immune therapies, and drug therapies are currently being evaluated, but these were not available during the outbreak [71]. Although fatality rates remained high, there is some evidence indicating that supportive care was effective in improving survival.

In a small study of 391 suspected, probable, or confirmed EVD cases in Bong County, Liberia by Weppelmann *et. al* [62], a 74% reduction in the risk of short term mortality was observed for patients hospitalised, compared to those not given medical intervention and after adjusting for age. The authors attribute this increase in survival to the fact that since less than 5% of the cases in the sample experienced unexplained bleeding and haemorrhage, hospitalisation likely helped to prevent intra-vascular volume depletion and replacement of electrolytes, which greatly reduces the potential for complication arising from hypovolemic shock.

A number of experimental vaccines have been in development since the late 1970s [39] but because EVD outbreaks are rare, and have, until 2014, been controlled quickly, commercial vaccine manufacturers demonstrated little urgency in advancing vaccines through clinical trials. That changed in 2014, with several vaccines previously tested only on animals being fast-tracked into Phase One clinical trials [58].

One such experimental vaccine known as rVSV-ZEBOV was introduced in a ring vaccination strategy - despite being unlicensed - in March 2016 when a flare-up of EVD was reported in Guinea. This was the first time that an Ebola vaccine had been used during an outbreak setting outside of a clinical trial [30]. The vaccine is reported to have been used and further investigated during the ongoing outbreak in the Democratic Republic of Congo, which has resulted in 87 cases between 1 - 16 August 2018 [70, 69]. The WHO has reported rVSV-ZEBOV to have been 100% effective in a ring vaccination protocol in a trial in Guinea during 2015. However, this effectiveness remains disputed given the lack of proper clinical trials [44].

Evidence shows that individuals who recover from Ebola infection develop antibodies that last for at least 10 years, possibly longer. It is not known if people who recover are immune for life or if they can become infected with a different species of Ebola. Some survivors have developed long-term complications, such as joint and vision problems [13].

2.2.5 Tools Against Ebola

In the absence of a licensed vaccine or a specific treatment during the large 2014-2016 outbreak, control and prevention strategies were focused on case management and minimizing direct contact with the infectious. This predominantly involved contact tracing, travel restrictions and quarantine, teams to facilitate safe burials and collaborating with local authorities to encourage cooperation and behavioural changes through education.

2.3 Time Line of the 2014-2016 West Africa Outbreak

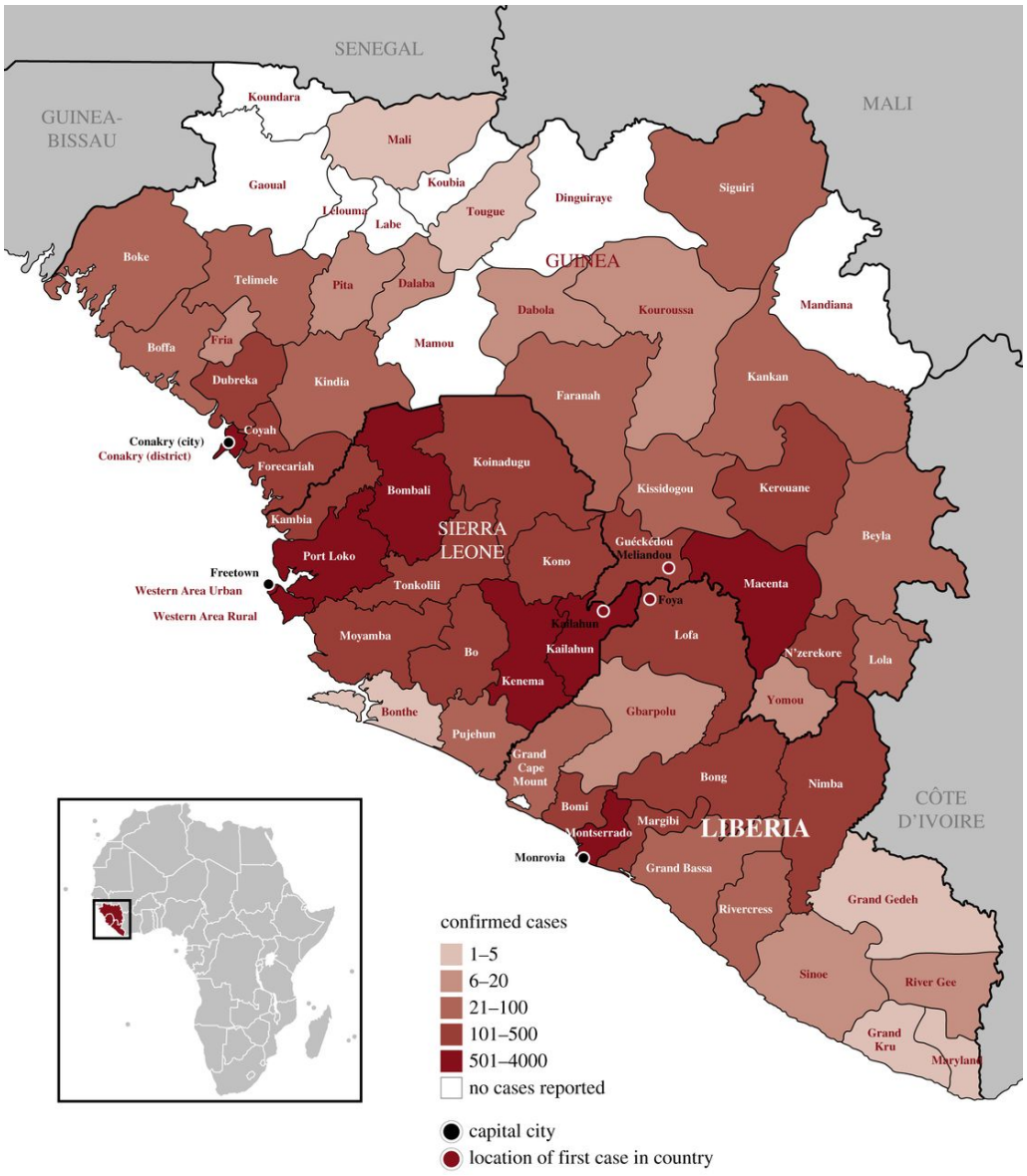

Figure 2.2: Geographic district-level map of total confirmed case in Guinea, Sierra Leone and Liberia during 2014 - 2016 outbreak [17], adapted from [68]

In December 2013, it was reported that an 18-month-old boy from Meliandou village within the Guéckédou prefecture located in South-Eastern Guinea (Figure 2.2) - bordering both Sierra Leone and Liberia - died after having been infected by bats [13, 21]. The toddler is suspected to have been the index case. Retrospective analyses traced the initial transmission chain from those infected by the toddler; this is provided in Appendix A.1. After five additional cases of fatal diarrhoea occurred in that area, an official medical alert was issued on January 24, 2014, to the district health officials. The Ebola virus soon spread to Guinea's capital city of Conakry, and on March 13, 2014, the Ministry of Health in Guinea issued an alert for an unidentified illness. Shortly after, the Pasteur Institute in France confirmed the illness as EVD caused by Zaire ebolavirus. On March 23, 2014, with 49 confirmed cases and 29 deaths, WHO officially declared an outbreak of EVD.

The identification of these early cases marked the beginning of the West Africa Ebola epidemic, the largest in history [13].

Initial efforts to control the outbreak were considered to be succeeding when in late April, a dip in reported cases in Guinea gave hope that the epidemic was beginning to subside and could be confined to largely one country. That hope was abandoned as the virus crossed international borders into neighbouring countries, and the initially small number of confirmed cases in bordering parts of Liberia and Sierra Leone rose sharply during May 2014 [21]. Weak surveillance systems and poor public health infrastructure contributed to the difficulty surrounding the containment of this outbreak and by July 2014, the outbreak had spread to the capitals of all three countries. This was the first time EVD extended out from more isolated, rural areas and into densely populated urban centres, providing an unprecedented opportunity for transmission [13].

On August 9, 2014, the World Health Organization declared the deteriorating situation in West Africa to be a Public Health Emergency of International Concern (PHEIC), which is designated only for events with a risk of potential international spread or that require a coordinated international response [68].

A country needs to be 42 days without any new cases to be declared Ebola-free by WHO [71]. The epidemic time frame and the end of the outbreak differed between the countries affected. Liberia was first declared Ebola-free in May 2014. However, additional cases appeared and the peak of transmission only occurred between August - September where Liberia was reporting between 300 - 400 new cases every week. Again, the epidemic seemed to abate and the outbreak was declared over on May 9, 2015 - only to re-emerge seven weeks later when a 17-year-old man died from the disease and additional cases followed. The same happened in September 2015. On January 14, 2016, WHO declared Liberia Ebola-free and no additional cases have been detected since.

After an initial declaration in November 2015, Sierra Leone announced it was Ebola-free on March 7, 2016. A preliminary statement in December 2015 was retracted when additional cases were discovered in March and April, and Guinea was finally declared Ebola-free in June 2016.

Guinea, Liberia and Sierra Leone were by far the worst hit countries. However, over the duration of the epidemic, 36 confirmed cases were reported from Italy, Mali, Nigeria, Senegal, Spain, the United Kingdom and the United States [13]. Two and a half years after the first case was discovered, the outbreak ended with 28,616 recorded cases and 11,325 deaths.

The scope of this outbreak, both in terms of cases and geography, can be attributed to the unprecedented circulation of EVD into crowded urban areas, increased mobilisation across borders, and conflicts between key infection control practices and prevailing cultural and traditional practices in West Africa. Engaging local leaders in prevention programs and messaging, along with careful policy implementation at the national and global level, helped to eventually contain the spread of the virus and put an end to this outbreak.

It was decided to provide country specific time lines of key events as well as basic country statistics in 2014 prior to the outbreak for both Sierra Leone and Liberia, as this provided useful insight for understanding the similarities and differences between the outbreaks in the two countries. The time line for Sierra Leone is provided in Table 2.1 on page 21. The time line for Liberia is provided in Table 2.2 on page 22. Unless stated otherwise, the information relating to key events were obtained from WHO and CDC in the following sources: [67, 13, 68, 17]. Table 2.3 on page 23 provides simple summary statistics for both countries in 2014. For this table, all values were adapted from World Bank data for 2014 unless otherwise stated [43].

Sierra Leone	
30 March 2014	Several suspected cases reported from Sierra Leone (confirmed cases in Liberia and Guinea)
24 May 2014	Confirmed cases reported in Kenema, Sierra Leone. They are traced back to the funeral of a widely respected traditional healer from Kailahun who had contracted the disease after treating Ebola patients from across the border in Guinea.
26 May 2014	Outbreak confirmed.
11 June 2014	Borders with Liberia and Guinea, and a number of schools are closed.
July 2014	All schools closed (reopened April 2015).
15 July 2014	The Ministry of Health establishes an Emergency Operations Centre (EOC) at the WHO Country Office in Freetown.
September 2014	Quarantine restrictions instituted in high risk areas.
19 September 2014	The first three-day shut-down is launched and six days later, three additional districts are placed under quarantine. In total almost a third of the population is under lockdown.
16 October 2014	The EOC announces two Ebola cases in the far north of Sierra Leone, which marks the arrival of cases in every district in the country.
26 October 2014	Outbreak peaks in Sierra Leone.
Oct - Nov 2014	Curfews imposed in Freetown.
17 December 2014	Western Area Surge is officially launched. In partnership with WFP, UNDP, UNICEF, CDC and others, the surge is aimed at sourcing urgently needed supplies and equipment, encouraging community mobilisation, as well as surveillance and contact tracing.
13 February 2015	Hundreds of homes (approximately 700) in the capital are placed under quarantine for 21 days.
18 February 2015	Door-to-door searching for 'hidden' Ebola patients is launched. Spike in cases in Port Loko district, east of the capital, attributed to unsafe burials and patients being hidden from the authorities.
28 February 2015	New cases across the country prompt the reinstatement of the lifted ban.
27 March 2015	Second 3-day shut-down of around 6 million people.
12 June 2015	Curfews imposed in Port Loko and Kambia.
September 2015	Vaccine trial for front-line workers under way.
7 November 2015	Sierra Leone declared Ebola-free for the first time.
Nov - March 2016	Additional case clusters discovered.
7 March 2016	Sierra Leone declared Ebola-free for the last time. Liberia was already declared Ebola-free in January 2016. Guinea will only be Ebola-free in June 2016.

Table 2.1: Sierra Leone 2014/2016 Time line

Liberia	
30 March 2014	Outbreak confirmed by WHO.
17 June 2014	Liberia reports that Ebola has reached its capital, Monrovia.
July 2014	Two ETCs opened in Monrovia and Foya; government closes most border points and many schools (reopened February 2015).
20 July 2014	An airline passenger from Liberia introduces the virus into Lagos, Nigeria marking the first time that Ebola enters a new country via international air travel.
27 July 2014	Liberian President Ellen Johnson Sirleaf declares the closing of borders, with the exception of Roberts International Airport, where screening centres are added. The President also announces that football events are banned, all schools and universities are closed, and that the military is to be employed in quarantining the worst-affected communities.
1 August 2014	State of emergency declared, enhanced contact tracing and quarantining measures instituted.
4 August 2014	Liberian government orders all bodies of Ebola victims to be cremated.
19 August 2014	President declares a nationwide curfew and orders two communities to be completely quarantined, with no movement in or out of the areas.
20 August 2014	West Point protests occur and authorities clash with members of West Point neighbourhood in Monrovia, one of the communities put under quarantine [6].
Aug - Sept 2014	Additional ETCs built.
28 September 2014	Ebola outbreak peaks in Liberia.
25-29 October 2014	National reporting transitions from aggregate to case-based data which results in a peak in recorded cases.
29 October 2014	WHO reports the rate of infections in Liberia has slowed, due in part to changes in cultural mortuary practices.
13 November 2014	State of emergency lifted.
15 February 2015	Schools reopen after months of closing due to Ebola outbreak.
22 February 2015	Lifting of nationwide curfews imposed and re-opening of land border crossings.
9 May 2015	Liberia first declared Ebola-free.
June - Dec 2015	Additional clusters of cases detected.
14 January 2016	Liberia declared Ebola-free for the last time.

Table 2.2: Liberia 2014/2016 Time line

Country statistic	Sierra Leone	Liberia
Population	6.3 million	4.4 million
Rural Population (% of total)	60.4	50.7
Gross Domestic Product per capita (US$)	792.6	457.9
Capital City	Freetown	Monrovia
Physicians per 1000 people (as of 2010)	0.022	0.014
Total number of reported Ebola cases (WHO 2013 - 2016)	14124	10678
Total number of Ebola deaths (WHO 2014 - 2016)	3956	4810

Table 2.3: Country statistics for Sierra Leone and Liberia

2.4 Mathematical Modelling in Epidemiology

The use of mathematical models in epidemiology has become increasingly important for public health officials to understand the effects of disease spread amidst dense urban populations, for predicting the socio-economic effects of major epidemics, and for forecasting the effects of different intervention strategies [47]. A mathematical model will always be a simplification designed to obey the assumptions it is based on, and therefore will never perfectly capture a complex system. However, it provides a fast, cost-effective tool for understanding the complexity inherent in systems.

2.4.1 Compartmental Models in Epidemiology

The earliest account of mathematical modelling of spread of disease was carried out in 1766 by Daniel Bernoulli [11]. However, it was the emergence of compartmental models in the 1920s that established the basic foundations of the subject. One of the simplest and most fundamental of all epidemiological models is the so-called SIR model developed by Kermack and McKendrick in 1927 [33]. In this model, a population is divided into susceptible, infective and recovered individuals, with the functions S(t), I(t) and R(t) denoting their respective fractions in the populations at time t (measured, for example, in days). The evolution of these quantities is described by the differential equations: [33]

$$\frac{dS}{dt} = -\beta SI$$

$$\frac{dI}{dt} = \beta SI - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

where the derivatives dS/dt, dI/dt and dR/dt measure the rates of change of the quantities S(t), I(t), and R(t). The transmission parameter β is the average number of individuals that one

infected individual will infect per time unit, assuming that all contacts this individual makes are with susceptible individuals. Thus, a more highly infectious disease has a higher β. The number γ is the rate of recovery, so that $1/\gamma$ is the average time period during which an infected individual remains infectious. The product βS(t)*I(t) is the total infection rate, the fraction of the population that will be infected per unit time at time t [11]. To understand this, note that if a fraction I(t) of the population is currently infected, they would then infect a fraction βI(t) of the population per unit time if all of their contacts were with susceptible individuals, but as only a fraction S(t) of the population is currently susceptible, they will only infect βI(t) S(t) per unit time.

The ratio β/γ is also known as the basic reproductive number R_0, which is an important index for quantifying the transmission of pathogens. R_0 is defined as the expected number of secondary cases produced by a single (typical) infection in a completely susceptible population. It is important to note that R_0 is a dimensionless number and not a rate, which would have units of $time^{-1}$. Some authors incorrectly call R_0 the "basic reproductive rate" [32].

This model results in a fixed population N (S+I+R) where members of the population mix homogeneously (interact with one another to the same degree). There is no entry into or departure from the population as the dynamics of the disease are much faster than the time scale of birth and death processes; and hence the impact of these processes on the population can be ignored. Any inherent age, demographic and spatial structure is also ignored. There is no initial immunity as all 'members' of the susceptible population are equally likely to get infected. The model infers permanent immunity: once recovered, a second infection is impossible. The incubation period of the infectious agent is instantaneous, and the duration of infectivity is the same as the duration of the disease (one is infectious as long as one has the disease). Discrete individuals do not exist in the model, and it is assumed that individuals who reside in the compartments are identical and therefore variation among individuals is unimportant. Thus compartment models are described as population-level models. It is fractions of the population that flow between compartments and these movements are continuous. The rate of recovery λ is constant for each member of the population and hence the average duration of infectiousness (and in this case disease) is $1/\lambda$ [56].

The SIR model forms the basis for many extensions based on model requirements. Transitions between each of the compartments capture the overall dynamics of the epidemic, showing how the population as a whole gets infected and eventually recovers. These models capture aggregate behaviours over the whole population. Individuals and interactions between individuals are omitted, and so these models can be used with relatively little computational power and relatively small amounts of input. This allows for the models to capture large-scale (including country-wide/continent-wide pandemic) disease spread with relatively little effort [47].

2.4.2 Mathematical Modelling of Ebola

Several mathematical models have been developed to study the transmission dynamics of EVD, the majority of which were compartmental models.

The SEIR (susceptible-exposed-infectious-recovered) deterministic compartmental model structure is commonly used in the mathematical modelling of diseases with an incubation period, accounted for by the 'exposed' compartment. Several mathematical models have extended upon the SEIR structure in modelling Ebola virus. Many mathematical models for EVD are also focused on optimal control analysis and thus extensions frequently include various intervention measures [18].

Ahmad *et. al* proposed a SEIR-type model distinguishing between high risk and low risk susceptibles with the addition of hospitalisation, quarantine and vaccination controls [2]. Ivorra *et al.* proposed a general deterministic spatial temporal model with vital dynamics in [31] which was tested on the Guinea outbreak data. Models discriminated between early and late stage infections through the use of more than one infectious class such as in [4, 10]. Xia *et al.* distinguished between suspected and probable infectious cases [73].

Not all compartmental model structures assumed the latent period and excluded the exposed state. Njankou and Nyabadza used an SIHDR model to study the potential impact of limited hospital beds and hospitalisation constraints on the EVD outbreaks in Liberia and Sierra Leone [20]. Berge *et. al* developed a model to incorporate indirect transmission through a contaminated environment such as consumption of bush meat [9].

Some of these model extensions are summarised in Table 2.4 with the following compartmental definitions: S = susceptible, E = exposed, I = infected/infectious, R = recovered, H = hospitalised, V = vaccinated, D = dead, B = buried, F = funeral, P = Ebola virus pathogens in the environment.

Model	**Transmissions**	**Control**	**Reference**
SSEIHR	Community, hospital	Quarantine, vaccination	[2]
SEIH(DB)R	Hospital, dead, funeral	Quarantine, contact tracing, safe burial	[31]
SIHDR	Hospital, dead, funeral	Quarantine, contact tracing, safe burial	[20]
SEIIFR	Community, hospital, funeral	Isolation and safe burial, ETU	[4, 10]
SVEIIHFR	Community, hospital, funeral	Vaccination, media, isolation, quarantine, safe burial	[73]
SIPDR	Community, dead, contaminated environment	Education, safe burial	[9]
SEI(HF)R	Hospital, dead, funeral	Contact tracing, treatment, safe burial	[50]
SEIIHFR	Community, hospital, funeral	Media, isolation, quarantine, safe burial	[54]

Table 2.4: Ebola Transmission SIR/SEIR Variant Models and control measures investigated

Some studies did use alternative model structures to the traditional compartmental model. Bartlett *et al.* used a deterministic, discrete time, age structured model which separated people by the number of days into an infection [5]. Differences in district-level transmission dynamics were analysed using a generalised linear mixed effects model by Krauer and colleagues in [35].

Many of the interesting and complex models discussed were derived in theoretical mathematical settings, where implementation was restricted to numerical simulations. Fewer models were directly aimed at better understanding the Ebola virus dynamics and uniquely estimating the parameters that resulted in the outbreak. During the 2014 - 2016 outbreak, all efforts were necessarily focused on eradicating Ebola in West Africa to prevent further transmission and hence many of the models developed during this time were aimed at optimal control analysis. Given that even less data was available during the outbreak, scenario analysis was often limited to numerical simulation.

This paper is partly motivated by the limited research dedicated exclusively to the unique estimation and better understanding of the Ebola virus dynamics as observed in humans, as opposed to optimal control analysis and hypothetical intervention testing. The geographic location and the context of any future Ebola outbreak is uncertain, and the appropriate control measures will very much depend on this context as well as resources available at that time. In the large West Africa epidemic, state enforced cremation was successful in Liberia, but not a viable strategy available to Sierra Leone. It is unrealistic to assume general intervention strategies exist that are equally appropriate for all outbreaks. Consequently, this paper aims to supplement previous research by harnessing the full retrospective scope of data and information in providing more specific estimates for the dynamics of EVD and how it differed between Sierra Leone and Liberia, the two countries hardest hit by Ebola. Thorough understanding of the disease dynamics and how different factors interact, provides a good basis for tailoring the right response and preparing more specific optimal control models in the event of future out breaks.

Chapter 3

Data

3.1 Description

The data provides the cumulative EVD case counts and cumulative deaths recorded at different dates for the three countries: Sierra Leone, Liberia and Guinea. The data was compiled from situation reports prepared by the World Health Organisation (WHO) during the 2014-2016 outbreak but has been made publicly available by the Centre for Disease Control (CDC) [14].

During the outbreak, WHO recommended that mobile teams and health stations use three broad definitions in classifying an EVD case. Cases were recorded as either suspected, probable or confirmed and the respective definitions are provided in Table 3.1. WHO reported 'cases' to be the sum total of all three definitions. Due to the existing challenges in case management and detection, and the urgency to focus efforts on outbreak control and resource constraints, laboratory confirmation of an EVD infection was often non-viable.

3.2 Exploratory Data Analysis

Exploratory data analysis helps us to make sense of the data presented before diving straight into analysis. It facilitates the initial investigation into potential insights we could derive from the data. The plots of the raw cumulative case and death data for both Sierra Leone and Liberia are provided in figures 3.1 and 3.2, respectively. Despite the lack of independent variables present in this analysis, these epidemic curves are useful when supplemented with the historical information provided in the literature review of this paper.

Suspected	• Any person - alive or dead - who has or had EVD symptoms, as well as contact with a suspected, probable or confirmed EVD case, or a dead or sick animal. • Any person with sudden onset of high fever and at least three of the following EVD symptoms: headache, vomiting, anorexia, loss of appetite, lethargy, aching muscles or joints, breathing difficulties, diarrhoea, stomach pain, difficulty swallowing, or hiccups. • Any person with inexplicable bleeding, or who died suddenly from an unexplained case.
Probable	• Any suspected case evaluated by a clinician. • Any deceased suspected case (where it has not been possible to collect specimens for laboratory confirmation) having an epidemiological link with a confirmed case.
Confirmed	• Any suspected or probable that tests positive for EVD in laboratory testing.

Table 3.1: WHO definitions of EVD cases [66]

3.2.1 Sierra Leone

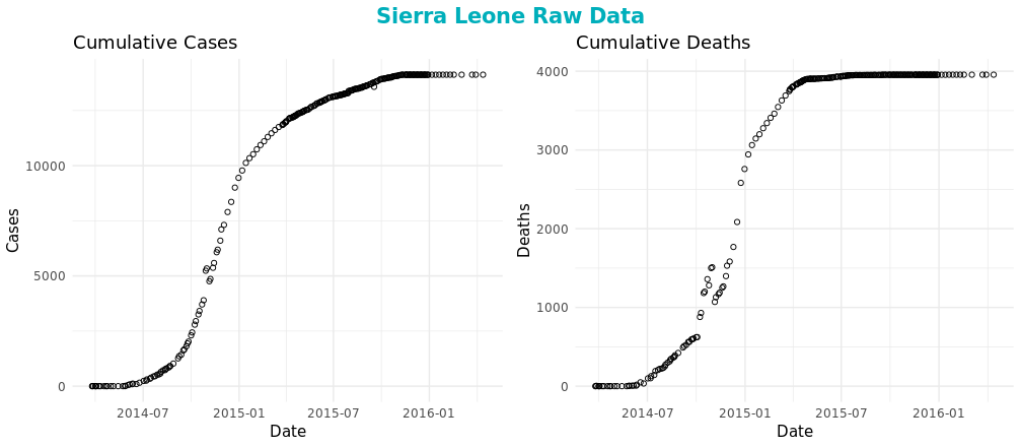

Figure 3.1: Raw case and death data in Sierra Leone

3.2.2 Liberia

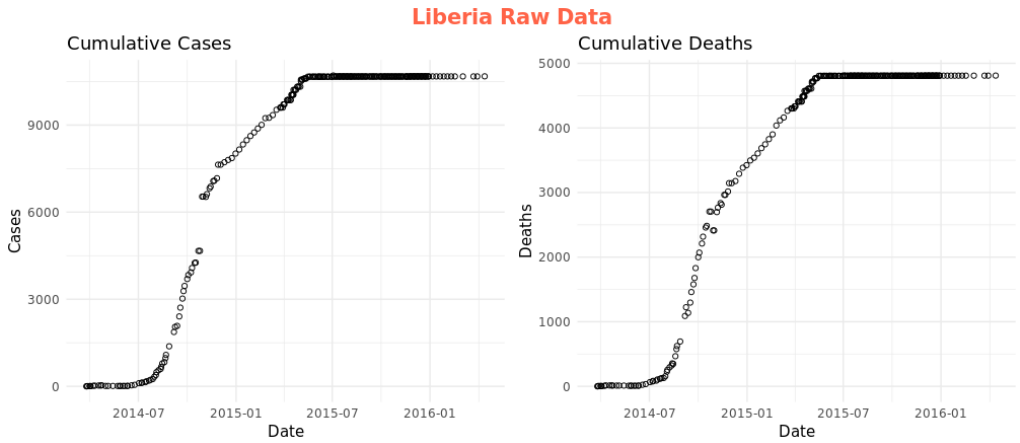

Figure 3.2: Raw case and death data in Liberia

The outbreak in Liberia was officially declared over in January 2016, but it is clear from Figure 3.2 that the epidemic curve flattens out from mid 2015 already, and was considered to be over as early as May 2015. Due to the detection of stray cases and small clusters, the country failed to achieve Ebola-free status for 42 consecutive days - WHO's requirement for an official end of an outbreak to be declared - until a year later (Table 2.2).

3.3 Data Cleaning

Discrepancy in the data is to be expected given the difficulty in detecting, confirming and recording cases during a crisis. There is evidence to suggest that both case data and death data were under-reported in the official records [45]. However, the documentation of cases is considered to have been better than for deaths and the available death data proved to be very noisy. This is in part due to the fact that burials could not be accurately quantified in the areas known to continue traditional burials [64], and efforts were aimed at tracking down live cases for quarantine and supportive care. Consequently, it was decided to smooth the death data by simply selecting observations slightly further apart that still capture the overall shape of the death curve but does not attempt to fit to all of the noise present in the daily data.

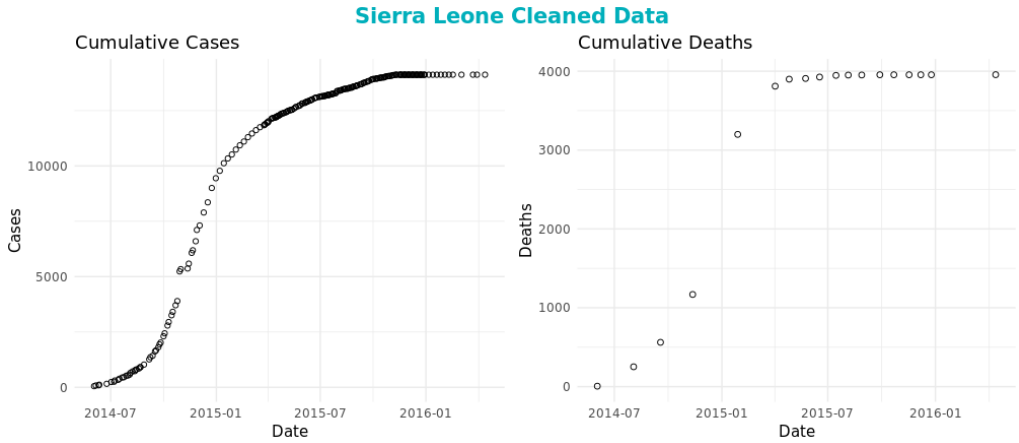

Figure 3.3: Sierra Leone clean data

For both Liberia and Sierra Leone, the data for the initial stage of the epidemic are available, however it is sparse and unreliable; often the number of deaths exceed the number of cases for a given recorded day. It was decided to consider data after the date by which at least 50 cases had been recorded for each country. This corresponded to the June 2, 2014 for Sierra Leone and June 24, 2014 for Liberia. Obvious errors such as duplicate date entries and sudden small dips in the cumulative cases were individually considered and either removed or corrected when the intended value was clear. Given the relatively few observations available for the outbreak, this was done in an effort to preserve as much data as possible. The plots for the cleaned datasets used in model fitting are provided in Figures 3.3 and 3.4.

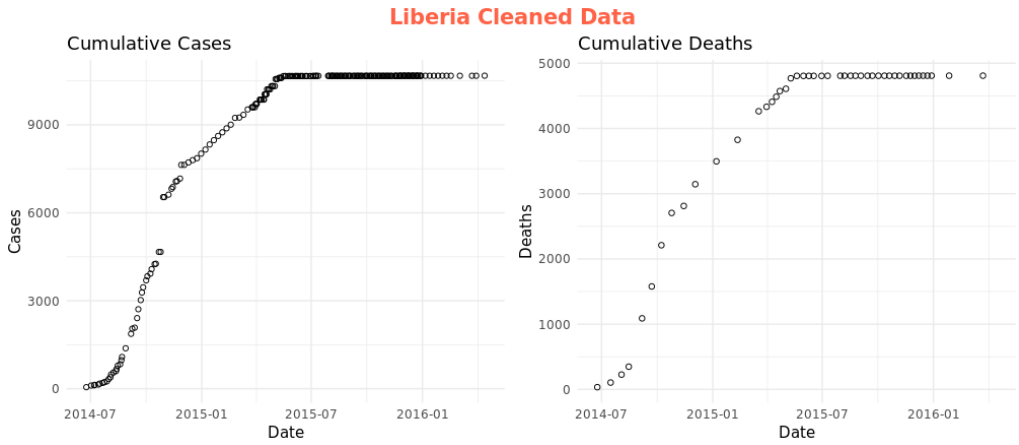

Figure 3.4: Liberia clean data

3.4 Limitations

Describing the mechanics of Ebola transmission, much like any other infectious disease, requires data on both treated and untreated cases, but data is often restricted to those who presented at healthcare facilities or were detected by surveillance authorities. Early in the epidemic, the CDC estimated an under-reporting factor of 2.5 [43] which certainly improved as contact tracing strengthened. Liberia reported near 100% effective case management towards the end of the

outbreak. Changes and improvement in case coverage and contact tracing could also have led to the spikes in the reported cases witnessed, as opposed to increases in the actual disease incidence. A related problem was the disjointed collection and recording of cases by different authorities. When closely looking at disease incidence in Sierra Leone, a dramatic variation in the number of cases reported from the end of October to the beginning of November 2014 is present. CDC reports that this jump is simply due to a change in data sources: prior to October 2014, the cumulative total numbers were derived from a combination of patient databases and country situation reports. Later, the revised approach used numbers compiled by the Ministries of Health and WHO country offices [14].

Despite these limitations, the data remains very valuable in better understanding the course of the outbreak.

Chapter 4

Methodology

The very first step in choosing to replicate a transmission pattern using a mathematical model is to decide on the class of model and structure for that model. This analysis made use of deterministic compartmental models to describe the population dynamics of EVD.

4.1 Model 1: The Full Model

The total population is divided into nine mutually exclusive compartments which are classified as: low risk susceptible (SL), high risk susceptible (SH), exposed (E), infected (I), hospitalised (H), recovered and infectious (R_1), recovered and not infectious (R_2), dead and infectious (D), dead and safely disposed (buried or cremated) (B). The proposed model provides an extension to the deterministic models presented in [2, 50]. Ahmad et al. [2] introduced the split between low risk and high risk susceptible individuals. Rivers et al. [50] explicitly included transmission resulting from contact with deceased individuals through a funeral state.

High risk susceptible refers to individuals with a higher rate (probability) of acquiring infection (i.e. $\psi_{HR} > 1$). Typically high risk susceptibles are considered to be women, children, health care workers and doctors. The rest of the population is included in the low risk susceptible section. All newborns are assumed to be low risk susceptible as there is no known vertical transmission of the infection [2].

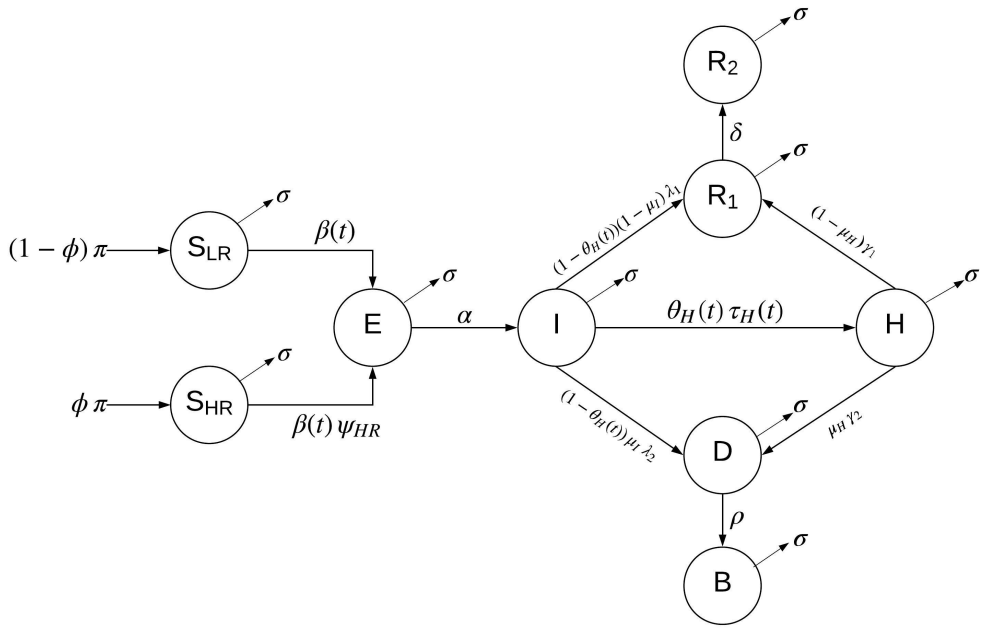

Figure 4.1: Model 1 flow diagram

Individuals are recruited into the susceptible populations at a rate π with probability ϕ of being a high risk individual. Deaths or exits due to reasons other than Ebola (from any state) are captured by the rate σ. After exposure to the Ebola virus, susceptibles become infected and can transmit the disease. Importantly, this model explicitly distinguishes between transmission through contact with infectious individuals (I), infectious corpses (D), hospitalised individuals (H) and recovered individuals (R_1). Generally speaking, the effective transmission parameter captures both the transmissibility (i.e. the probability of infection given contact) as well as the average rate of contact between susceptible and infectious individuals.

We would expect the transmissibility to differ depending on the state of the Ebola virus infection. For example, corpses are often more infectious than living patients and certainly more infectious than recovered individuals. The average rate of contact between susceptible and different infectious groups will also reasonably differ based on the typical type of contact (e.g. health care workers versus community) but also quite possibly by culture and country. Cultural burial practice will significantly influence the expected level of intimate contact between a diseased and living person. Hospitalised or quarantined patients may experience less direct contacts but often result in the infection of health care workers. All four rates are functions of time and are summarised by the single term $\beta(t)$ in the model flow diagram (Figure 4.1).

After exposure, infected individuals become symptomatic at a rate α. Once infected, individuals may be hospitalised with a time dependent probability $\theta_H(t)$ at rate $\tau(t)$. Fluctuating resource constraints are captured by the time dependent relationship in these two parameters. Non-hospitalised infectious individuals can recover from the disease at rate λ_1, or die from the infection at rate λ_2. The probability of death (or fatality rate) is denoted by μ_I. Hospitalised EVD patients may recover at a rate γ_1 or die from the disease at a rate γ_2 with fatality rate μ_H. There is some evidence to suggest that hospitalised EVD patients experience slightly better chances of survival than non-hospitalised cases, hence we have distinguished between the two fatality rates.

Given that viral persistence has been noted in immunologically privileged sites for several months

after acute infection, an EVD survivor first transitions into R_1 before transitioning to state R_2 at a rate δ, after which they are no longer infectious. Once fully recovered (R_2), the individual remains immune and cannot re-enter the susceptible population as EVD survivors are known to develop antibodies that last for at least 10 years [13].

If an infected individual dies, their corpse remains infectious until it is disposed of through safe burial or cremation which occurs at a rate $\rho(t)$. By allowing for the transmission rates, the probability of hospitalisation, the rate of hospitalisation and the rate of safe burial to vary with time, we can adequately capture the extreme behavioural changes observed and control measures implemented with the progression of the Ebola outbreak.

The compartments and parameters driving the movement between these compartments for Model 1 is summarised in Figure 4.1, Table 4.1 and Table 4.2. Note that the total population size is not necessarily assumed constant in this model.

Variable	Description
S_{LR}(t)	Population of low risk susceptible individuals
S_{HR}(t)	Population of high risk susceptible individuals
E(t)	Population of latent individuals
I(t)	Population of infectious individuals
H(t)	Population of hospitalised individuals
R_1(t)	Population of infectious recovered individuals
R_2(t)	Population of fully recovered individuals
D(t)	Population of dead but still infectious individuals
B(t)	Population of safely buried/cremated individuals

Table 4.1: Summary of Model 1 compartments

Variable	Description
π	Natural birth rate / recruitment rate
ϕ	Probability of a high risk individual
σ	Natural death rate (not due to Ebola)
$\beta_I(t)$	Time dependent transmission rate from infectious individuals
$\beta_D(t)$	Time dependent transmission rate from dead individuals
$\beta_H(t)$	Time dependent transmission rate from hospitalised individuals
$\beta_R(t)$	Time dependent transmission rate from recovered individuals
ψ_{HR}	Modification parameter for infection rate of high risk susceptibles
α	Rate at which latent individuals become infectious
$\theta_H(t)$	Time dependent probability of an infected individual being hospitalised
$\tau_H(t)$	Time dependent hospitalisation rate for infected individuals
μ_I	Fatality rate for non-hospitalised infected individuals
μ_H	Fatality rate for hospitalised infected individuals
λ_1	Recovery rate of infected individuals
λ_2	Disease-induced death rate of infected individuals
γ_1	Recovery rate of hospitalised individuals
γ_2	Disease-induced death rate of infected individuals
δ	Rate at which recovered individuals become not infectious
ρ	Rate at which infectious corpses are safely buried or disposed of

Table 4.2: Summary of Model 1 parameters

4.1.1 Feasibility of Model 1

Identifiability issues were a major concern in this analysis due to the lack of credible data and the wide ranging estimates for parameters presented in the literature. The lack of consistent estimates in understanding the clinical course of the disease motivated estimating those parameters from the data. However a lack of data restricts the number of parameters that can be estimated without resulting in identifiability issues. That is, for complex models consisting of many parameters and various compartments, there were several parameter sets that provided an equally good model fit, i.e. minimized the error sum of squares (SSE) in the least squares estimation procedure implemented.

Model 1 was developed without the consideration of data limitations and presented the 'ideal' approach. However to maintain model integrity and avoid identifiability issues, Model 1 was repeatedly simplified until a simpler model - which grouped multiple effects whilst preserving only the key characteristics of the disease - would provide more certain parameter estimates. Model 2, below, restricts the number of parameters and compartments in order to balance the complexity of the models versus the actual observations of the 2014 - 2016 epidemic.

4.2 Two Models to Analyse the Impact of Intervention

4.2.1 Model structure

Two models are used in combination to estimate the effect of intervention. The first model does not account for any intervention. The second model expands upon the first model by introducing the effect of control measures and estimating the aggregated effect of these measures.

Model without Intervention:

This model does not include for any methods of intervention. We formulate a deterministic model of six distinct compartments of susceptible (S), exposed (E), infectious (I), recovered (R), dead and infectious (D), and dead and safely disposed (buried or cremated) (B). After exposure to Ebola virus, susceptibles become infected and can transmit the disease. After c contacts with susceptibles, an infectious individual can transmit the disease with a probability τ. Thus, cτ is the effective transmission rate of the disease. For transmission through contact with infectious living individuals, this is captured by the term β_I. Similarly, for contact with the deceased but still infectious this is captured by β_D.

After exposure, infected individuals become symptomatic at a rate α. Once symptomatic, the individual becomes infectious. Infectious individuals recover at a rate $lambda_1$ with probability $(1-\mu)$ or they die from the disease at a rate $lambda_2$ with probability μ. Hence μ represents the fatality rate for the disease. Corpses are disposed at a rate ρ through burial or cremation, after which they are no longer infectious.

The size of the total population, calculated by: $N = S + E + I + R + D + B$, is assumed constant because it did not vary considerably during the modelling time (approximately 18 months).

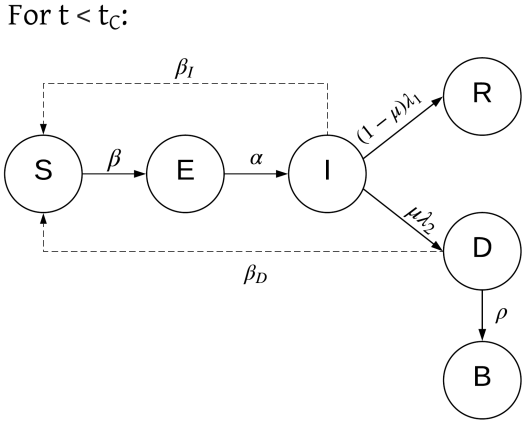

Figure 4.2: Model 2 with no intervention

Model with Intervention

We expand the model above by introducing interventions. The 2014 - 2016 West Africa outbreak was brought to an end by the serious efforts of various key players enforcing a multitude of control measures. This included hospitalisation, quarantine, the construction of many brand new Ebola Treatment Units (ETUs), travel bans and border controls, safe burial teams, contact tracing, national lock downs, closure of schools and public spaces, coupled with educational campaigns to inspire behavioural changes and better cooperation with authorities.

These interventions differ in many ways, however they all aimed at reducing the direct contact between susceptible and infectious individuals; whether dead or alive. Given that there is currently no cure or vaccine to provide immunity to the general susceptible community, isolation of Ebola patients in health care facilities or quarantine, and movement restrictions act to reduce contact with those that are living and infectious. Safe burial initiatives or state-enforced cremation, community surveillance and even mass-education aimed not only at the public but also at health care workers, act to reduce contact with infectious corpses. There was also a discernible behavioural shift that occurred a few months into the fast escalation of the 2014 - 2016 outbreak when public mistrust of foreign authorities subsided allowing for better cooperation in all three worst affected countries [76].

Since there is no indication in the literature that efforts were more focused on interventions that reduced contact with the living as opposed to efforts that reduced contact with corpses, these effected are assumed equal in order to restrict the number of parameters to be estimated. The various control measures and intervention effects discussed are simply introduced into the model through an aggregated intervention term (η) which reduces the effective transmission rates after some time (t_C) at which initial panic has subsided such that control measures are effectively implemented and cooperation between authorities and the public has been established.

This model is comparable to the model presented by Rivers et al. in [50]. However, this model chose to exclude the hospitalised state due to the lack of information available resulting in identifiability problems. Instead, the aggregated intervention term is similar to an approach implemented by Bartlett *et al.* [5].

The model remains exactly the same except that after some time t_C the effective transmission rates β_I and β_D are multiplied by a factor η, where $0 < \eta < 1$, which reduces both of the transmission rates.

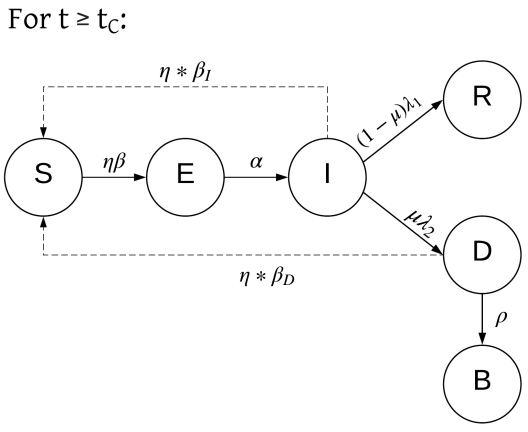

Figure 4.3: Model 2 with intervention

4.2.2 Model Equations

Without Intervention

$$\begin{aligned}
\frac{dS}{dt} &= -\beta_I \left(\frac{I}{N}\right) S - \beta_D \left(\frac{D}{N}\right) S \\
\frac{dE}{dt} &= \beta_I \left(\frac{I}{N}\right) S + \beta_D \left(\frac{D}{N}\right) S - \alpha E \\
\frac{dI}{dt} &= \alpha E - (1-\mu)\lambda_1 I - \mu\lambda_2 I \\
\frac{dR}{dt} &= (1-\mu)\lambda_1 I \\
\frac{dD}{dt} &= \mu\lambda_2 I - \rho D \\
\frac{dB}{dt} &= \rho D
\end{aligned} \tag{4.1}$$

With Intervention

$$\eta_t = \begin{cases} 1 & \text{if } t < t_C \\ \eta & \text{if } t \geq t_C \end{cases} \tag{4.2}$$

$$\begin{aligned}
\frac{dS}{dt} &= -\beta_I \eta_t \left(\frac{I}{N}\right) S - \beta_D \eta_t \left(\frac{D}{N}\right) S \\
\frac{dE}{dt} &= \beta_I \eta_t \left(\frac{I}{N}\right) S + \beta_D \eta_t \left(\frac{D}{N}\right) S - \alpha E \\
\frac{dI}{dt} &= \alpha E - (1-\mu)\lambda_1 I - \mu\lambda_2 I \\
\frac{dR}{dt} &= (1-\mu)\lambda_1 I \\
\frac{dD}{dt} &= \mu\lambda_2 I - \rho D \\
\frac{dB}{dt} &= \rho D
\end{aligned} \tag{4.3}$$

Variable	Description
S(t)	Population of susceptible individuals
E(t)	Population of exposed latent individuals
I(t)	Population of infectious individuals
R(t)	Population of recovered individuals
D(t)	Population of dead and infectious individuals
B(t)	Population of safely disposed dead individuals (burial/cremation)

Table 4.3: Description of model 2 compartments

4.2.3 Model Assumptions

It is necessary to highlight the assumptions implicit in the choice of the structure for Model 2 to be used in parameter estimation, as well as to provide contextual motivation for these assumptions.

Compartmental epidemiological models use the transitions between each of the compartments to capture the overall dynamics of the epidemic, showing how the population as a whole becomes infected and eventually recovers. It captures aggregate behaviours over the whole population. Individuals and interactions between individuals are omitted, and so these models capture large-scale (such as country-wide) disease spread with relative ease [47].

Discrete individuals are not considered, rather aggregated population behaviour and movements between compartments form the focus. Aggregating population behaviour requires the assumption that all members are homogeneous and differ only with regards to their disease state. Members are also assumed to mix homogeneously and without discrimination. Individuals within each compartment are hence assumed to be identical, and variation among individuals is unimportant [56]. As such, we have ignored any inherent age, demographic and spatial structure in the population; therefore this is clearly an unlikely representation of individuals within any country. However, given the fact that cases were reported from all 16 districts in Sierra Leone, and 14 out of 15 districts in Liberia - as well as the scarcity of district level data - this aggregation to the national level was deemed necessary.

A fixed population size (N) not allowing for entry or departure into the model was assumed. Natural changes to the population size such as births or deaths from causes other than Ebola were ignored. This assumption implies that the dynamics of EVD are faster than the time scale of natural birth and death. Given that the entire outbreak in all three countries occurred over a period of approximately two years and that a small proportion of total population were affected, it is reasonable to assume that the population size did not change significantly over the course of the epidemic.

There is uncertainty as to whether recovered individuals are afforded life-long immunity to the virus or whether infection could occur with a different species, but research shows that those who do recover develop antibodies that last for at least 10 years [13]. Given that the 2014-2016 outbreak occurred in a relatively short time frame and that the epidemic was caused exclusively by the Zaire ebolavirus, the model infers permanent immunity. As such, survivors cannot re-enter the susceptible population once recovered.

There is no known risk of becoming infected with EVD through casual contact with a survivor although, as previously discussed, the virus takes longer to clear from immunologically privileged sites, thereby making transmission through sexual intercourse post-recovery possible. This motivated the inclusion of two recovered states in the full Model 1. However as yet, there is little conclusive evidence of this viral persistence and clinicians are required to be overly precautious in preventing the spread of infection until clearer answers are obtained.

Typically the symptoms of infected EVD patients associated with the expelling of bodily fluids (such as through vomiting and bleeding) are significant contributing factors in the ease of transmission. Hence, transmission occurring after a patient's recovery was not a key characteristic of Ebola needing to be estimated. Additionally, evidence suggests that asymptomatic presentation of EVD infection of non-survivors is extremely unlikely and hence this was not factored into the model [36].

The model assumes all dead individuals first pass through the infectious (D) state before being safely disposed of (B). This is attributed to two effects primarily. The first is the strong cultural significance of burial in West Africa, while the second effect is that corpses remain extremely infectious; even when severe control measures are applied in their disposal, infection remains possible.

4.2.4 Basic Reproductive Number

The basic reproductive number, R_0 briefly introduced in section 2.4.1 is a fundamental concept in mathematical biology. It is a threshold parameter, intended to quantify the spread of disease by estimating the average number of secondary infections produced by the introduction of a single (typical) infection into an otherwise wholly susceptible population. It thereby provides a measure of initial disease spread or the invasion strength of an epidemic. In a well-mixed homogeneous population, an epidemic can only take off if $R_0 > 1$, in which case, initial infections will grow exponentially and the disease will persist. Whereas if $R_0 < 1$, the disease cannot successfully invade and will die out in the long run. If $R_0 = 1$, the disease becomes endemic.

The basic reproductive number for the baseline scenario was calculated in a similar manner as described in [38] by evaluation of the steady state conditions in order to obtain the following expression:

$$R_0 = \frac{\beta_I}{\lambda_1 + \mu(\lambda_2 - \lambda_1)} + \frac{\beta_D \mu \lambda_2}{\rho\big(\lambda_1 + \mu(\lambda_2 - \lambda_1)\big)}$$

From the above, we can see that R_0 is broken into two components, representing the respective contributions of community (infectious class) and deceased individuals in transmission. These values are calculated for the start of an epidemic to establish its potential for persistence, but clearly this will change if intervention measures are instituted. We take the value R_C to represent the base reproductive number from the time that control measures are first introduced (i.e. t_C).

With intervention ($t = t_C$):

$$R_C = \eta * \left(\frac{\beta_I}{\lambda_1 + \mu(\lambda_2 - \lambda_1)} + \frac{\beta_D \mu \lambda_2}{\rho\big(\lambda_1 + \mu(\lambda_2 - \lambda_1)\big)} \right)$$

There are various methods to derive a R_0 threshold from mathematical models, and different values may be obtained depending on the method used. This may be dealt with by interpreting the obtained values purely as a threshold indicative of the disease persistence in a specific population and not as the true value of R_0. As these values do not uniquely estimate the true R_0 for a disease, caution should be applied in their interpretation as they are not generally comparable in absolute terms and especially not for different diseases [37]. Consequently, it was decided not to compare the calculated R_0 values in this analysis to those of other infectious diseases, but rather to assess the extent to which these values agree with previous calculations for the 2014-2016 Ebola outbreak and in the assessment of the interventions.

4.3 Data Fitting

Deterministic models were fit to the full outbreak data for both Sierra Leone and Liberia using least squares optimisation. Parameters were estimated using the 'L-BFGB-B' optimisation algorithm, a quasi-Newton method which allows for box constraints [12]. The first 150 observations of reported cases were given one-half of the weight in the model fitting to preferentially fit the more recent case data. Smoothing of the death data implied reducing the number of data points and hence the weights of the remaining observations were increased by a factor of 1.5 to ensure an adequate fit to both the case and the death data. The optimiser was constrained to plausible parameter values as indicated in the literature, such as an upper bound of 20 days for infection duration, 4 days for disposal of an infectious corpse and 0 to 1 for probabilities or proportions. In order to ensure global optima during optimisation, the procedure was run with over 500 initial value seeds.

Preserving model identifiability implied that not all parameters could be estimated from the data but instead were empirically assumed from prior studies. These estimates may be improved upon in future but for the time being, uncertainty and sensitivity analysis allows for the identification of the critical parameters most influencing EVD dynamics and for which, correct estimation is of greater importance.

All models were implemented in R v3.4.0 [49].

Chapter 5

Results

5.1 Parameter Estimation

Given the wide range of estimates provided in the literature regarding the epidemiological features of the Ebola virus, it was challenging to select any set of fixed parameters that provided an impressive fit to the data. There is also reason to believe that the dynamics of the Ebola Virus Disease outbreak was considerably different in Sierra Leone, Liberia and Guinea; and hence would result in different disease parameters [65, 67]. It was decided to fix the parameters inherent to EVD to estimates obtained from the literature, and to estimate all the parameters that have reason to vary with country.

Scientists could not find any evidence that the Zaire strain presented in the 2014 - 2016 West Africa outbreak had mutated since previous outbreaks and hence, we assume the reported average incubation period, and average duration of infection before recovery, or death, as fixed when modelling the disease in both Sierra Leone and Liberia.

Despite being the longest, largest and deadliest outbreak witnessed, the case fatality rate reported was more volatile and generally lower for the 2014 - 2016 outbreak compared to any of the previous 24 outbreaks. A population of particular interest for ongoing monitoring and public health surveillance is comprised of more than 17,000 survivors: Ebola patients who successfully recovered from their illness [55]. Historically, the fatality rates associated with outbreaks has varied drastically ranging anywhere between 25%-90%. Crude estimates for the fatality rates experienced during the 2014-2016 outbreak (total recorded deaths/total recorded cases) are calculated as 39.5% for the total outbreak, 67% for Guinea, 45% for Liberia and 28% for Sierra Leone. Of the 881 total recorded cases amongst health care workers and the average case-fatality rates were 51% in Guinea, 50.8% in Liberia and 72% in Sierra Leone. The discrepancy between the general fatality rates and those for health care workers could be possibly attributed to better documentation and follow-up regarding the cases of health care workers, or the advanced infections that they are exposed to within the healthcare setting.

On the other hand, information provided to clinicians by the Centre for Disease Control and Prevention (CDC) stated the case fatality proportion among patients with a known outcome to be 70%, 61% for hospitalised patients and ranging between 37-74% in Ebola Treatment Units [13]. Case fatality may be underestimated due to the difficulty in accurately recording deaths during a crisis, but it may also be a consequence of using broad case definitions, such that low specificity

resulted in the presentation of many false positives. In summary, there is uncertainty regarding the true fatality rate and hence it was decided to estimate this parameter from the available data.

Cultural practices influence the frequency of close contact, burial procedures and the attitudes and behaviours of individuals within a country. Political situation, international relations, geographical context, and infrastructure also determine how control measures are enforced. Thus transmission, burial and all intervention parameters were estimated from the data. Plausible ranges for each parameter, obtained from the literature, were implemented as box constraints and are provided in Table 5.1 along with the source. Where a point estimate was available, it is indicated in brackets next to the range.

Symbol	Description	Feasible range	Source
$1/\alpha$	Time from exposure to symptom onset	8 - 12 days (10)	[13], [71]
$1/\lambda_1$	Time from symptom onset to recovery	6 - 16 days (9.4)	[59]
$1/\lambda_2$	Time from symptom onset to death	6 - 16 days (7.5)	[13]
μ	Fatality rate	0.25 - 0.75	[13]
$1/\rho$	Time from death to safe disposal of body	8 hours - 5 days	[50]
t_C	Time at which control measures are effectively enforced	SL: Aug-Dec 14; LIB: Aug-Dec 14	[51, 68]
β_I	Effective contact rate with live infectious people	0.03 - 0.3	[50]
β_D	Effective contact rate with infectious corpses	0.1 - 0.9	[50]
η	Factor to decrease $beta_I$ and $beta_D$ for $t \geq t_C$	0 - 1	Not in literature

Table 5.1: Plausible parameter ranges obtainable from literature

The assumed total population size was informed by the estimates used by the CDC in 2014 [43]. The cleaned data starts from the date when there are at least 50 cases and so the starting value for the cumulative cases is taken to be 50 for Sierra Leone and 51 for Liberia. For both models it was assumed that there were 80 individuals in the exposed class by this date. The official CDC records show that the cumulative deaths on the respective dates were 6 for Sierra Leone and 34 for Liberia. It was assumed that if deaths had already been documented in official records, enough time would have passed for the corpse to have been appropriately disposed of, and hence these deaths fall into the B (buried/cremated) class. It was assumed there were half as many deaths that had occurred recently enough for the bodies to still be infectious and not safely disposed of (state D) in Sierra Leone (i.e. 3 deaths) and the same number of recent deaths assumed for Liberia. Given that it was very early on in the epidemic, it was assumed that there were no survivors in either country yet as most of the early infections in 2014 were fatal [17].

Variable	Sierra Leone	Liberia
N	6,092,000	4,294,000
S(0)	6,091,870	4,293,869
E(0)	80	80
I(0)	41	14
R(0)	0	0
D(0)	3	3
B(0)	6	34

Table 5.2: Summary of initial conditions

5.2 Results

5.2.1 Model Without Intervention

Initially, a SEIRDB model was fit without the explicit inclusion of any form of intervention and using a range of plausible parameter estimates obtained from epidemiological studies and existing mathematical models. The model flow diagram can be seen in Figure 4.2.

Interestingly, these parameters could not provide a suitable fit to the epidemic curve and always resulted in far more severe predictions for the outbreak. Similar results were obtained by Bartlett *et. al* [5]. This could be indicative of two effects: the first being the under-recording of EVD cases, and the second being the necessity of intervention and the various control measures enforced in ending the epidemic. The first explanation remains plausible but as the outbreak progressed, surveillance, case detection and management became cardinal to officially eliminating Ebola. The extent of under-reporting required in order to make the no intervention model fit the epidemic curve could possibly be expected in the first few weeks of the epidemic but is improbable to have lasted for the entire duration.

5.2.2 Model With Intervention

Fitting both the case and death data to the intervention model for both Sierra Leone and Liberia resulted in the parameter estimates shown in Table 5.3. Results will be discussed in section 6.1.

Symbol	Description	Sierra Leone	Liberia	Source
α	1/Latency period	0.1	0.1	From literature
λ_1	1/Time from symptom onset to recovery	0.1064	0.1064	From literature
λ_2	1/Time from symptom onset to death	0.1333	0.1333	From literature
μ	Fatality rate	0.2523	0.3941	Estimated
ρ	1/Time from death to safe disposal of body	0.4029	1.3994	Estimated
t_C	Time at which control measures are implemented	170 19 Dec. 2014	103 4 Nov. 2014	Estimated
β_I	Effective contact rate with infectious people	0.0909	0.1746	Estimated
β_D	Effective contact rate with infectious corpses	0.7173	0.3434	Estimated
η	Factor to decrease β_I and β_D for $t \geq t_C$	0.6195	0.5016	Estimated

Table 5.3: Final parameter estimates for both Sierra Leone and Liberia

It is worth noting that the estimated case fatality rates for both Sierra Leone and Liberia appear to be on the lower end of the range typically associated with EVD. Initially, a model was fit using only case data and ignoring the death data. In this case, the model still achieved an acceptable fit to the case data but the estimated case fatality rates were significantly higher (nearly 70% in both countries). Upon further analysis of the plot for the predicted deaths however, deaths appeared to be significantly over-estimated when compared to the actual data. It was decided to perform this analysis using all the available data (cases and deaths) and given the difficulty in accurately inferring the under-reporting present in the data, it was decided to use the data as is. However, this demonstrates the uncertainty of the data, the uncertainty in the case fatality rates associated with EVD and presents an area which this analysis can be improved upon.

5.3 Model Analysis

5.3.1 Goodness of Fit

Sierra Leone

Figures 5.1 and 5.2 show the fit of the model predictions to the cumulative case and death data respectively. The model provides an impressive fit to the cumulative case data and adequately captures the shape of the death data. Plots for the predicted susceptible, exposed, recovered and dead but infectious compartments are provided in Figure 5.3.

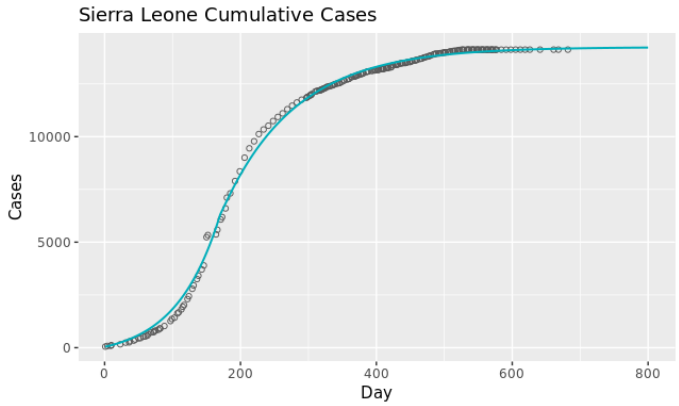

Figure 5.1: Model fit to Sierra Leone cumulative case data

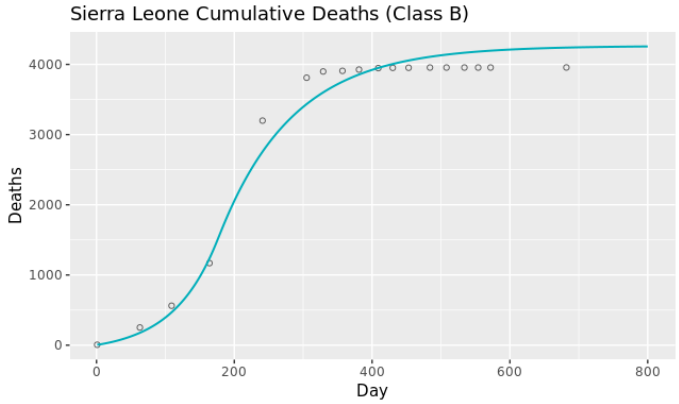

Figure 5.2: Model fit to Sierra Leone cumulative death data

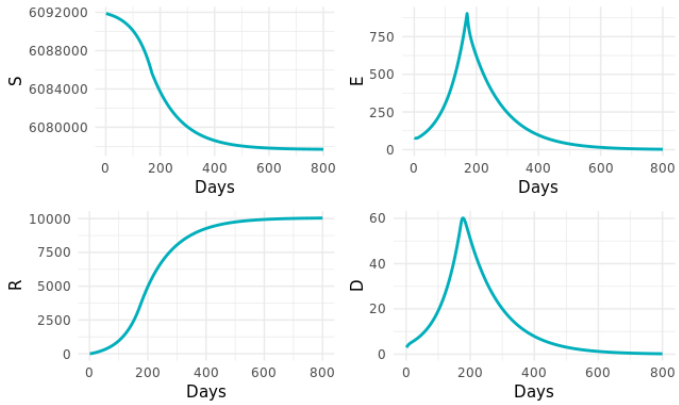

Figure 5.3: Susceptible, exposed, recovered and dead and infectious compartments for Sierra Leone

Liberia

Similarly, a satisfactory model fit to Liberian case and death data was accomplished.

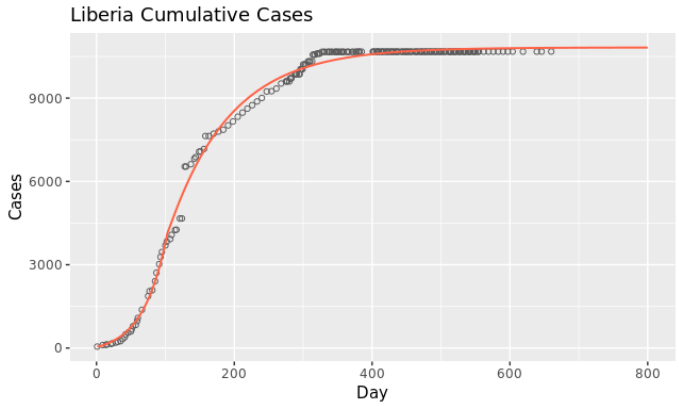

Figure 5.4: Model fit to Liberia cumulative case data

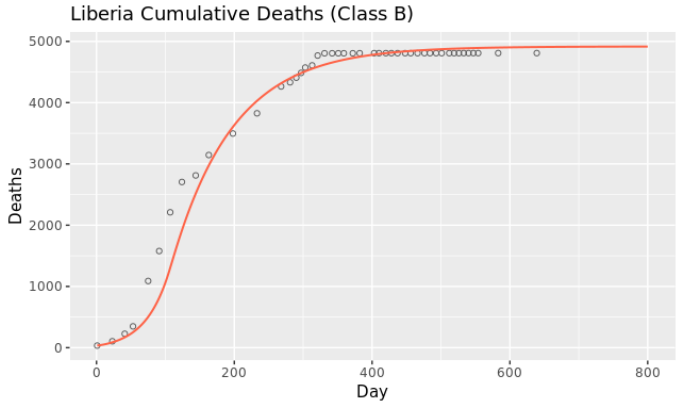

Figure 5.5: Model fit to Liberia cumulative death data

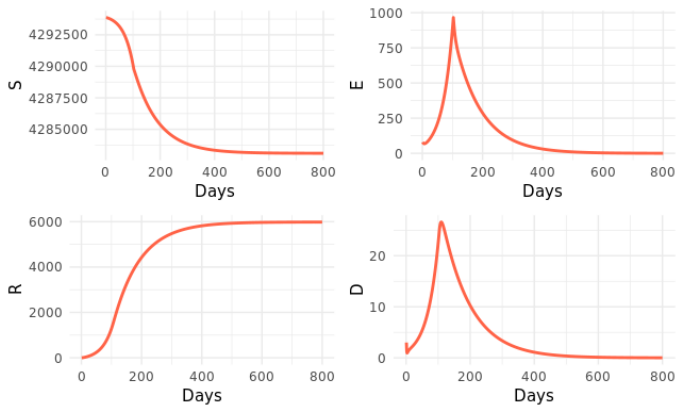

Figure 5.6: Susceptible, exposed, recovered and dead and infectious compartments for Liberia

5.3.2 Basic Reproductive Number

The baseline R_0 value for Sierra Leone (prior to intervention) was calculated as approximately 1.333, while R_0 for Liberia was 1.603, as provided in Table 5.6. Given that both values are greater than the threshold of 1, this indicates that infection will be able to spread in a population. These

estimates are comparable to those obtained by Khan and colleagues [34] where the resulting values were: 1.492 for Sierra Leone, and 1.757 for Liberia. Nearly all existing mathematical models have estimated R_0 values in the range of: 1.11 - 2.5 [3, 50, 73, 26, 61, 28] with the majority of estimates between 1.5 - 2. The R_C values after including the aggregated intervention parameter (η) is less than 1 for both countries.

Country	Estimated R_0	Estimated R_C
Sierra Leone	1.332548	0.8254536
Liberia	1.602782	0.803916

Table 5.4: Estimated R_0 values for Sierra Leone and Liberia

It is common practice in statistical modelling to split the data into training and testing sets in order to validate the trained model on the unseen test set. For an epidemic curve we could split the training and testing sets into time periods and apply a rolling window. This prevents over-fitting to the data. Given the very small sample size and the relatively short duration of the epidemic, it was decided to perform sensitivity analysis on the parameter estimates and compare these estimates to those observed in epidemiological studies. It is not feasible to validate the models using data from other countries as the outbreaks differed too significantly.

5.4 Sensitivity Analysis

Sensitivity analysis refers to the study of how the uncertainty in the output of a mathematical model can be apportioned to different sources of uncertainty in its inputs. The objectives are broadly to identify the critical inputs (parameters and initial conditions) of a model and to quantify the extent to which input uncertainty impacts model outcome(s). A detailed description of the history and methodology of uncertainty and sensitivity analysis is given in [52, 72]. Since our model is deterministic in nature, the only sources of uncertainty are the model parameters and the initial conditions, which will be examined in the sensitivity and uncertainty analysis.

The importance of the basic reproduction number and its interpretation in infectious disease modelling, necessitates sensitivity and uncertainty analysis of R_0. Several crucial model parameters (β_I, β_D, λ_1, λ_2, ρ and μ) determine the value of R_0. The basic reproductive numbers provided in Table 5.6 were calculated by the input of specific parameter values into the expression. However, factors such as natural variation, errors in measurements and lack of measuring techniques contribute towards the associated uncertainty of these model parameters, and consequently uncertainty in R_0 [2]. It is therefore preferable to produce confidence intervals and to analyse the distribution of R_0 as opposed to interpreting point estimates. Latin hypercube sampling (LHS) was implemented by randomly selecting 10,000 samples in studying the influence of these six parameters on R_0. Additionally, the univariate relationship between R_0 and each parameter was plotted.

Briefly, LHS belongs to the Monte Carlo class of sampling methods, introduced by McKay *et. al* [42]. LHS allows an unbiased estimate of the average model output, with the advantage that it requires fewer samples than simple random sampling to explore the entire range for each parameter [42]. We treat each parameter in the model as a random variable, distributed according to an appropriate probability distribution. LHS implements a stratified sampling without replacement technique, where the random parameter distributions are divided into N equal probability intervals,

which are then sampled. N represents the chosen sample size. The choice for N should be at least k+1, where k is the number of parameters varied. The sampling is done by randomly selecting values from each probability distribution. Each interval for each parameter is sampled exactly once (without replacement), so that the entire range for each parameter is explored. A matrix is generated (which we call the LHS matrix) that consists of N rows for the number of simulations (sample size) and of k columns corresponding to the number of varied parameters. N model solutions are then simulated, using every combination of parameter values. The model output of interest is collected for each model simulation.

Since we lack any concrete information regarding the probability distributions of the model parameters, we assume that our model parameters are either normally or uniformly distributed, although it is quite possible that in reality some parameters are skewed towards a particular value. Parameters with more certain average values reported in the literature were assumed to be normally distributed about the average value. The parameter values used in sensitivity analysis were informed by biologically plausible ranges obtained from the literature as well as observational studies done during the outbreak. These ranges and the corresponding distributional assumptions are provided in Table 5.5.

The baseline calculation for R_0 does not however depend on all the model parameters, and so we assess the sensitivity of the predicted cases to the uncertainty in each of the parameter values. This is useful in providing an indication of which estimates should be focused on most for accurate estimation. We identify crucial model parameters by computing partial rank correlation coefficients (PRCC) which is a measured impact of each input parameter on some decided output [72]. Here we select the output to be the number of cumulative cases predicted after 800 days with the model run. We also performed the same analysis for the number of cases after 50 days, 200 days, and 400 days in order to compare the variable importance at different stages of the virus outbreak. PRCC reduces the non-linearity effects by rearranging the data in ascending order, replacing the values with their ranks and then providing the measure of monotonicity after the removal of the linear effects of each model parameter keeping all other parameters constant [52]. Larger absolute values for the PRCC, denotes stronger correlation between the chosen parameters and the output. Again we implement LHS with 10,000 model simulations.

The Spearman rank correlation method was used in calculating the partial rank correlation coefficients. Spearman rank correlation is a non-parametric method, hence it does not carry any assumptions about the distribution of the data which was preferred in this situation. It assumes that the data must be at least ordinal and the scores on one variable must be monotonically related to the other variable, both these assumptions were met.

Parameter distributions

Parameter	Sierra Leone	Liberia
α	N($\mu = 0.1$, $\sigma = 0.0125$)	N($\mu = 0.1$, $\sigma = 0.0125$)
λ_1	N($\mu = 0.1$, $\sigma = 0.0225$)	N($\mu = 0.1$, $\sigma = 0.0225$)
λ_2	N($\mu = 0.1$, $\sigma = 0.0225$)	N($\mu = 0.1$, $\sigma = 0.0225$)
ρ	U(0.25, 1.5)	U(0.25, 1.5)
t_C	U(140, 200)	U(73, 133)
β_I	U(0.05, 0.16)	U(0.12, 0.22)
β_D	U(0.5, 0.8)	U(0.24, 0.42)
μ	U(0.2, 0.7)	U(0.2, 0.7)
η	U(0.3, 0.7)	U(0.3, 0.7)

Table 5.5: Parameter values used in sensitivity analysis for Sierra Leone and Liberia

5.4.1 Multivariate Sensitivity Analysis of R_0

Uncertainty analysis for the base reproductive number prior to intervention (R_0), consisted of running 10,000 random samples (obtained from the LHS structure) in order to analyse the distribution of the respective R_0 values. These distributions are presented in Figures 5.7 and 5.8. The distribution was calculated on 10,000 runs of the LHS structure and shows R_0 to be concentrated around a median value of 1.626 for Sierra Leone and 1.808 for Liberia. In both cases, the distributions are slightly skewed to the right due to some large R_0 values obtained. The distributions have a similar shape and correspond almost exactly to the range of estimates provided in the literature as discussed in section 5.3.2. In avoiding distributional assumptions, table 5.6 provides the median, and first and third quartile values as opposed to confidence intervals.

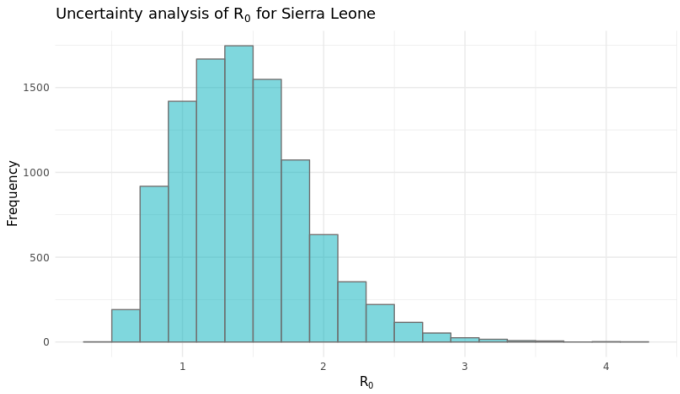

Figure 5.7: Uncertainty of R_0 for Sierra Leone

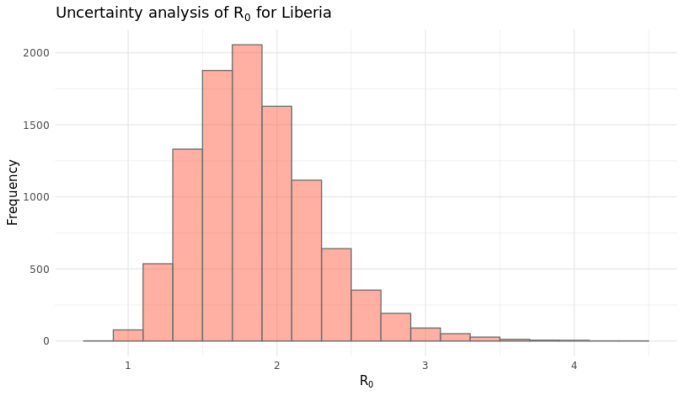

Figure 5.8: Uncertainty of R_0 for Liberia

Country	First Quartile	Median	Third Quartile
Sierra Leone	1.295	1.626	2.009
Liberia	1.563	1.808	2.099

Table 5.6: Estimated R_0 median, first and third quartile for Sierra Leone and Liberia

5.4.2 Univariate Relationships Between R_0 and Key Parameters

The univariate relationship between R_0 and each of the parameters that determine its value was analysed by varying parameters one at a time whilst fixing all other parameters at their estimated values, and observing the changes that occur in R_0. The relationships between R_0 and each of the parameters β_I, $beta_D$, λ_1 and ρ has an obvious interpretation for both Sierra Leone and Liberia.

The parameters μ and λ_2 indicate relationships of opposite sign in Sierra Leone compared to Liberia. For Sierra Leone, the relationship between both λ_2 and R_0, and μ and R_0 are positive and nearly linear. A possible explanation for this is that Sierra Leone was estimated to have a large β_D and a relatively long time until safe disposal of infected corpses (i.e. small ρ), hence when we fix these parameters, higher fatality rates (larger μ) result in more contacts with infectious corpses and increase R_0. Similarly, shorter infectious periods leading to death (larger λ_2) result in faster transitions to the dead and infectious state - where patients are more infectious than when alive - and they remain in this state for relatively long, thereby increasing R_0.

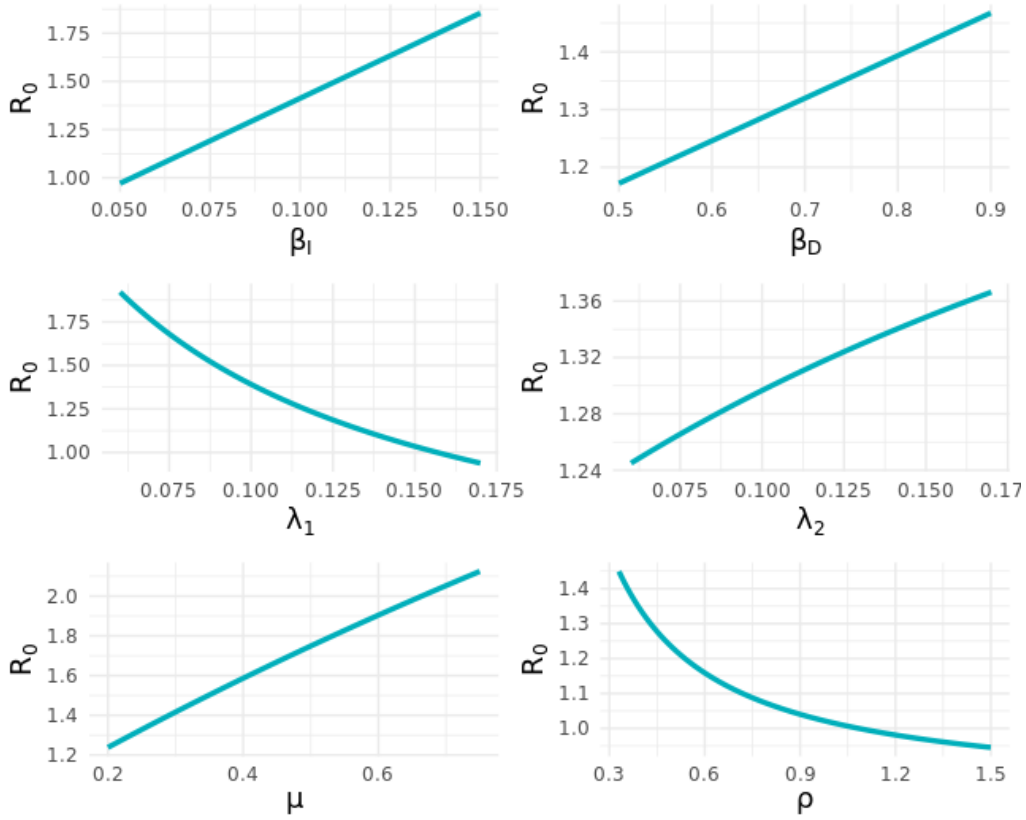

Figure 5.9: Univariate relationships between R_0 and key parameters in Sierra Leone

In contrast, Liberia demonstrates a negative relationship between λ_2 and R_0, and μ and R_0. This may be attributed to the fact that in Liberia contact with dead bodies was considered to be significantly lower, and the delay between death and disposal of a corpse significantly shorter. Infection occurring within live members of the community was therefore of greater importance in Liberia than in Sierra Leone. When infectious patients die faster (increased λ_2) or are more likely to die (increased μ), and there are very effective mechanisms in place to efficiently dispose of corpses, then this removal of infection allows for less transmission to occur between members of the community, thereby reducing R_0.

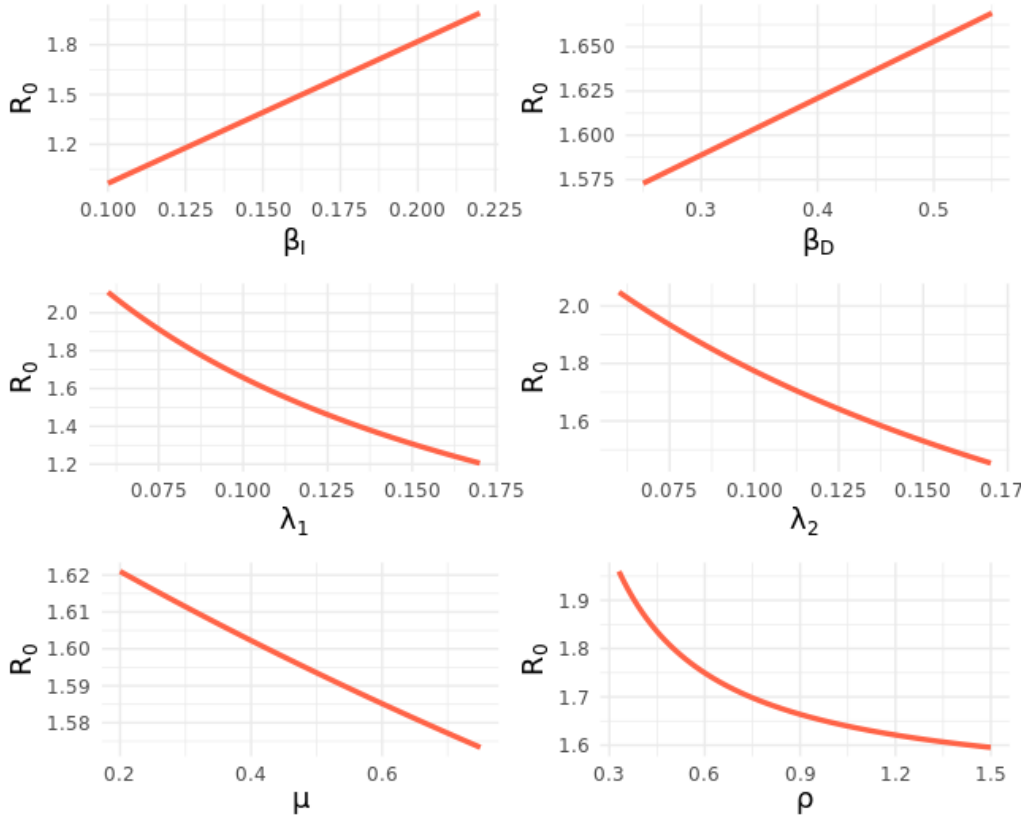

Figure 5.10: Univariate relationships between R_0 and key parameters in Liberia

5.4.3 Univariate Sensitivity Analysis

On the basis of the 10,000 independent samples generated using all 9 model parameters, we are able to determine the relative univariate importance of variables in model outcomes using partial rank correlation coefficients (PRCC). Larger absolute values for the PRCC, denotes stronger correlation between the selected parameter and the model outcome being investigated. We can also infer whether the relationship between the selected variable and the model outcome is positive or negative (as we have assumed a monotonic relationship) from the direction of the PRCC sign.

Initially the cumulative cases predicted for day 800 was used as the model outcome to assess parameter uncertainty, and the partial rank correlation coefficients are plotted in Figure 5.11 for Sierra Leone and Figure 5.12 for Liberia.

In both cases, the variables β_I, λ_1 and ρ had the largest |PRCC| values, they were very influential in determining the total cumulative cases at day 800 in the model. This makes sense as these values define the key parameters relating to transmission and, consequently, influence the overall size of the epidemic. Additionally, the direction of the relationships - as indicated by the sign of the PRCC - is intuitively interpretable. All else remaining constant, the larger the effective transmission rate in the community (β_I), the larger the number of cumulative cases at day 800. The rate of recovery from infection (λ_1) is the inverse of the length of the infectious period. Hence, the larger λ_1, the shorter the infectious period, the less transmission occurs, and the total number of cases reduces. Similarly, the inverse of ρ corresponds to the average length of time between

death and the safe disposal of a body so that it is no longer infectious. The larger ρ, the shorter the period in which post-mortem transmission can occur and hence less cases are produced.

In both cases, the latent period, captured by α was of relatively low importance. Given that individuals are not infectious in the latent period, it has little influence on the number of cases produced.

The parameters that capture post-mortem transmission is captured by μ, ρ and β_D. All three of these parameters were of greater relative importance in the model for Sierra Leone than for Liberia. This could be attributed to the differing time to disposal of corpses observed in both the model estimates as well as in the literature. It was estimated that the disposal of infected corpses in Sierra Leone took approximately 2.5 days on average, where as it took less than a day (average of 8.5 hours) in Liberia. When patients are buried or cremated shortly after death, as in Liberia, the post-mortem transmission rate becomes of lesser importance as individuals are buried faster than they can infect others. This was a key strategy in Liberia's containment of Ebola, within five months of the start of the outbreak, Liberian President Ellen Johnson Sirleaf ordered that all bodies of people killed by the Ebola virus be cremated. Additionally, we note λ_2 has greater relative importance in the Liberia model. A plausible explanation for this is that when infectious corpses are removed quickly, transmission within the community is of greater significance than post-mortem transmission.

The intervention parameters t_C and η are relatively influential in the cumulative cases recorded 800 days after the first 50 cases, but more so in Liberia than in Sierra Leone. In both cases η is more influential than t_C. The extent of intervention efforts is of greater significance than the exact time at which intervention occurs.

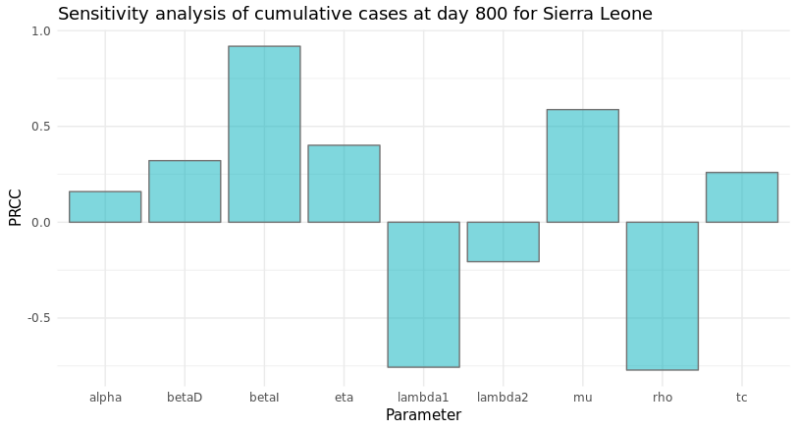

Figure 5.11: Sensitivity analysis of cumulative cases at day 800 for Sierra Leone

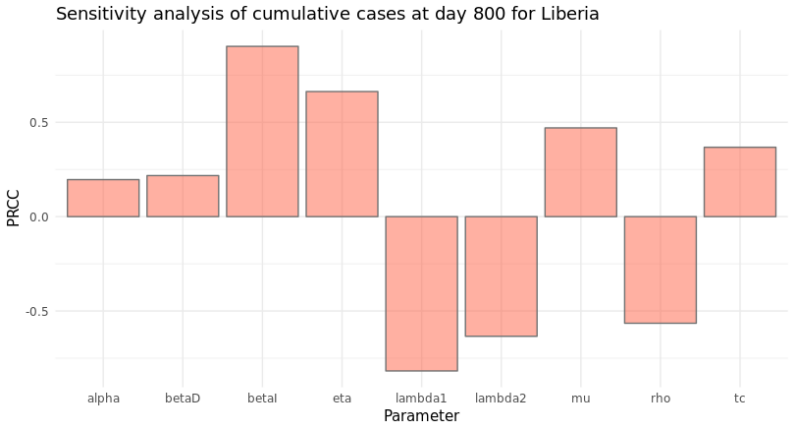

Figure 5.12: Sensitivity analysis of cumulative cases at day 800 for Liberia

The same analysis was repeated to measure the influence of parameter uncertainty on the cases predicted at day 50, day 200, and day 400, in addition to the cases at day 800. This allows for the comparison of changes in the relative importance of variables at different stages of the outbreak.

For both Sierra Leone (Figure 5.13) and Liberia (Figure 5.14), the start of the outbreak (day 50) is characterised by the inherent disease dynamics; intervention parameters are of little importance. The latency period is of greatest importance at day 50, and decreases as the outbreak progresses. When there are still few infections in the population at the start of an epidemic, the latent period and the delay in the onset of symptoms determines the speed at which the disease is able to spread.

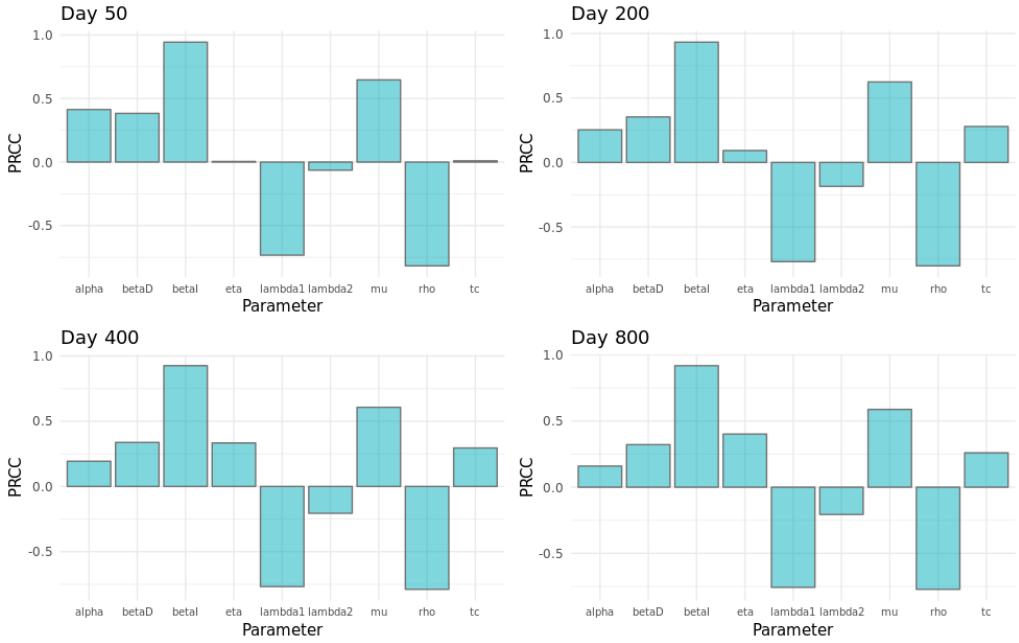

Figure 5.13: Sensitivity of parameters at different days in Sierra Leone

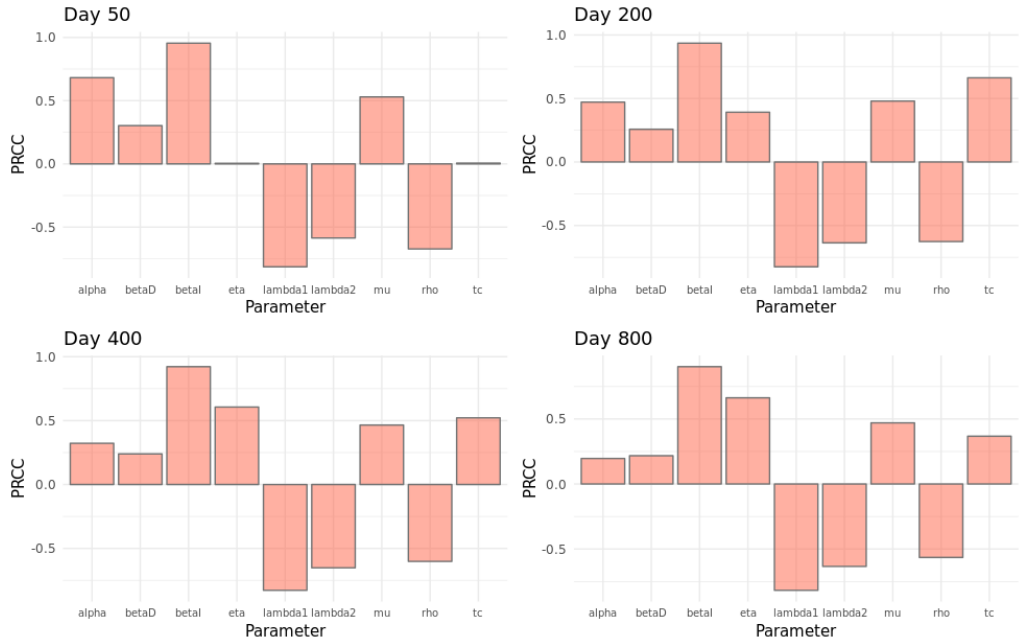

Figure 5.14: Sensitivity of parameters at different days in Liberia

5.5 Assessing the Impact and Timing of Intervention

In reflecting on the lessons of the 2014 - 2016 Ebola outbreak, the Centre for Disease Control (CDC) identified slow initial response as a primary factor in the unprecedented scale of the epidemic. Hence, it was decided to investigate the effect of earlier intervention on disease outcome [8].

The starting time of effective intervention efforts, t_C, was estimated to be 170 days for Sierra Leone (December 19, 2014) and 103 days (November 4, 2014) for Liberia. Figures 5.15 and 5.16 indicate the predicted cumulative cases if the estimated starting t_C was one month, two months or three months earlier in Sierra Leone and Liberia respectively. Additionally, two dashed vertical lines are drawn on each of the plots. The dark blue line represents the time t_C at which intervention controls are effectively implemented. The red dashed line represents the first time at which the case incidence is less than one, and is taken to be the time at which an epidemic has slowed down enough to be considered 'over'. Note that the time of intervention could equivalently be expressed as a threshold of the number of cases before intervention efforts and behavioural changes are realised (i.e. the corresponding y-value for the dashed blue line in each plot).

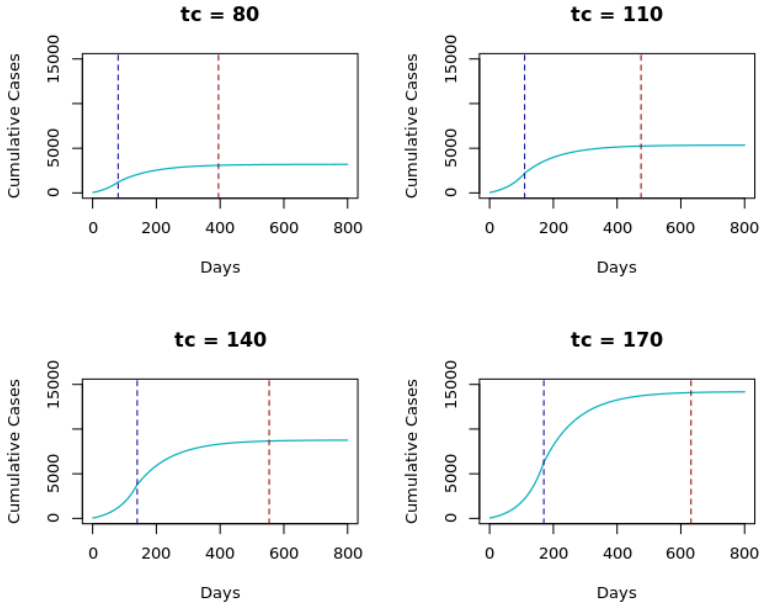

Figure 5.15: Comparing 1, 2 and 3 month earlier intervention in Sierra Leone

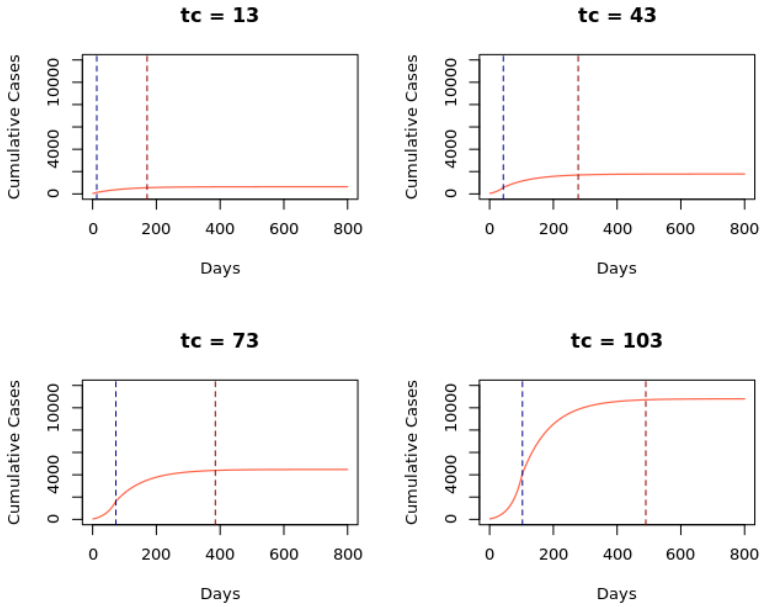

Figure 5.16: Comparing 1, 2 and 3 month earlier intervention in Liberia

These plots can be extended to express the relationship between timing and disease duration more generally. In Figure 5.17 we look at the total duration of the epidemic for a range of different threshold values expressed in terms of the number of cumulative cases before intervention is enforced. In this case, the disease duration is defined as the time from which at least 50 cumulative cases have been identified until the first time that case incidence is less than 1. Keeping η fixed at its estimated value, a t_C value of 30 days (1 month) corresponds to a threshold of 321 cases in Sierra Leone and 338 cases in Liberia. Practically, intervention measures are unlikely to be effectively enforced any earlier than within a month. The estimated intervention time of 170 days

in Sierra Leone, corresponds to the maximum threshold of 6088 cases in Figure 5.17. In Liberia, the estimated t_C value corresponds to a maximum threshold of 4143 cases in Figure 5.18.

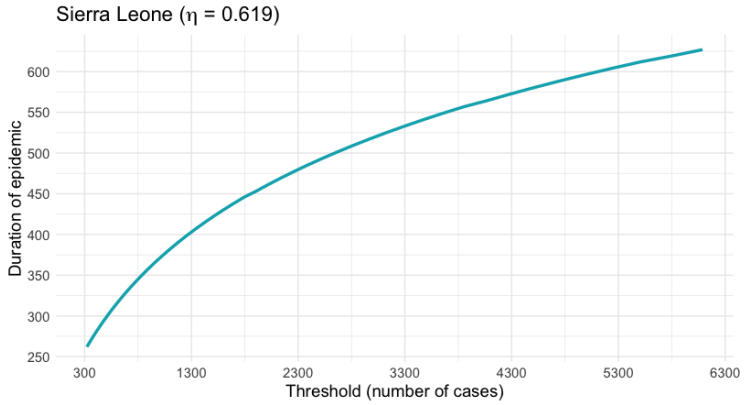

Figure 5.17: Relationship between intervention threshold and epidemic duration in Sierra Leone

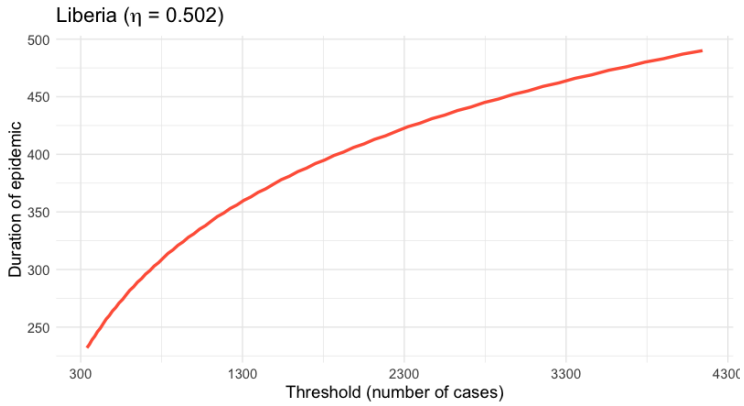

Figure 5.18: Relationship between intervention threshold and epidemic duration in Liberia

The above figures demonstrate that even with a quick initial response, stopping EVD remains a cumbersome and time-consuming matter, easily lasting for over 300 days. The argument remains that timing must be matched by more severe intervention efforts and control strategies; decreasing the η parameter in other words. The η factor acts to reduce the number of direct contacts occurring within the population through aggregating various controls such as quarantine, safe burials, behavioural changes, and case tracking. However, each of these direct measures are by default time-consuming and past experience has taught us that humans cannot suddenly alter their behaviour and maintain these changes for extended periods, even when faced with urgent threats [24]. Therefore, assuming control measures will be strong enough and individuals will adapt fast enough to more than halve their daily number of contacts in a few weeks (or even months) seems overly-optimistic. This is not to say that control strategies could not have been improved or that sooner implementation would have been futile, but rather that this was an exceptionally complicated task to achieve without treatment or vaccines available.

5.6 Vaccination as a Strategy for Epidemic Prevention

Vaccination has greatly reduced the burden of infectious diseases worldwide, allowing for the near eradication of many old diseases. In the case of the Ebola outbreak, an effective vaccine would compensate for the difficulty in reducing the large number of direct contacts between individuals daily and, provided the vaccine is readily available, can act much faster in eliminating disease. Promising results in the early trials of an Ebola vaccine in the Democratic Republic of Congo (DRC) has motivated assessing the impact of such a vaccine in the case of future outbreak.

While vaccination prevents the infection of an individual by ensuring their personal immunity, it offers the additional benefit of providing indirect protection to those who are not vaccinated. When control measures act to reduce the number of direct contacts, this needs to achieved between all individuals in an entire population, whereas vaccination only needs to occur in a certain proportion of the population in order to be effective. In epidemiological modelling, vaccination is often accompanied by this concept of 'herd immunity'. It refers to a form of indirect protection from an infectious disease that occurs when a large percentage of a population has become immune to an infection, thereby providing a measure of protection for individuals who are not immune. Put simply, the greater the proportion of individuals in a community who are immune, the smaller the probability that those who are not immune will come into contact with an infectious individual, thereby ensuring indirect protection [56].

The term 'herd immunity' has existed for almost a century but only gained popularity in recent decades with the increasing use of vaccines, discussions of disease eradication, and analyses of the costs and benefits of vaccination programs [25]. It's importance in epidemiological research escalated with the recognition, by Smith in 1970 and Dietz in 1975, of a simple threshold theorem - that if immunity (i.e. successful vaccination) were delivered at random and if members of a population mixed at random, such that on average each individual contacted R_0 individuals in a manner sufficient to transmit the infection, then incidence of the infection would decline if the proportion immune exceeded $(R_0 - 1)/R_0$, or $(1 - 1/R_0)$ [25, 28].

Herd immunity is established when the prevalence of protected persons (I) is higher than the herd immunity threshold (I_C). When this occurs, Ebola virus transmission is blocked within the given population. However, when prevalence is lower than the threshold, the number of infections will grow exponentially, thus spreading the virus. Using the mentioned variables, the critical proportion of protected individuals needed to establish herd immunity in a completely susceptible community can be determined from the equation: $I_C = 1 - (1/R_0)$ [28].

Suppose that the rZEBOV vaccine is 100% effective as reported from the small ring vaccination trial and that it infers permanent immunity. The HIT threshold theorem further requires the assumptions of: (1) random vaccination within the population, (2) homogenous mixing of persons within the population, (3) homogeneous distribution of vaccine-induced protected and infected persons within the population, and (4) a fully susceptible population [28]. Under the above assumptions, it is simple to calculate a crude estimate for the herd immunity threshold for the EVD outbreaks in Sierra Leone and Liberia, as well as an interval estimate obtained from the sensitivity analysis of R_0 in section 5.4.1 (as presented in Table 5.7).

Country	HIT Interval (Median)
Sierra Leone	23% - 50% (38%)
Liberia	36% - 52% (47%)

Table 5.7: Herd immunity thresholds for Sierra Leone and Liberia

Thus, if a 100% effective vaccine is licensed - and if approximately 38% of the Sierra Leonean population and 47% of the Liberian population were vaccinated prior to any future outbreak - in theory, no EVD epidemic will occur in these populations. The ranges provided above correspond with previous findings on required thresholds for various EVD outbreaks [28]. The assumption of 100% efficacy is not a requirement for the vaccination model, it just implies that greater proportions of the population will need to be vaccinated in order to prevent an epidemic if it is not met. The HIT threshold theorem is easily adjusted for vaccine effectiveness [28]. Let E be the proportion of individuals that are immune after receiving the vaccine (vaccine efficacy), then the critical vaccination coverage (V_C) needed to establish herd immunity can be determined by dividing the herd immunity threshold (I_C), by the level of vaccine effectiveness (E): $V_C = I_C/E = [1-1/R_0]/E$. The relationship between the critical vaccination coverage required (V_C) and vaccine effectiveness (E) is explored in Figure 5.19 for outbreaks varying in severity, captured by varying levels of R_0 as experienced in previous Ebola outbreaks.

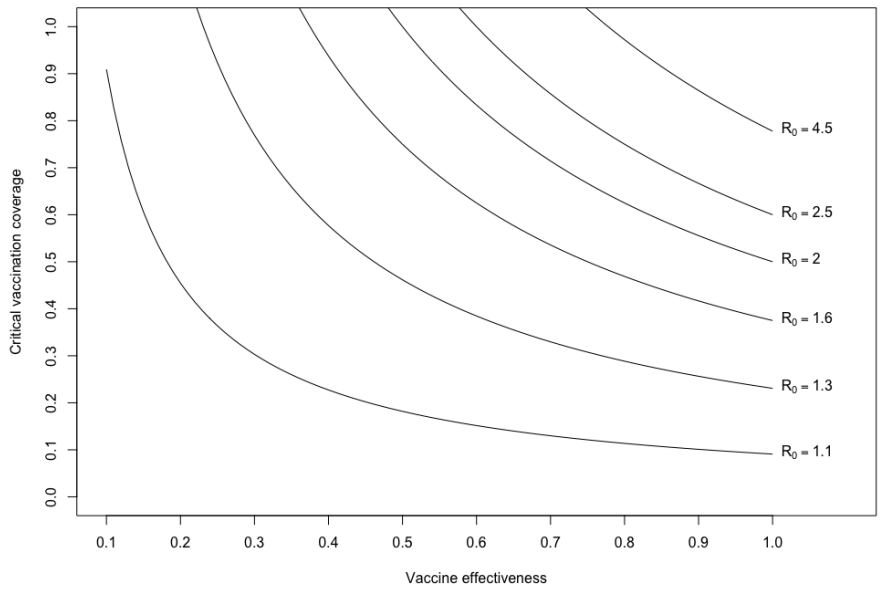

Figure 5.19: Critical vaccination coverage (%) needed to provide herd immunity against varying Ebola viruses (R_0) and variable vaccine efficacy (%), adjusted from [28]

A model in which vaccination is the only control strategy was developed in order to analyse epidemic duration and severity if a highly effective vaccine existed, was readily available, and community behaviour remain unchanged in both Sierra Leone and Liberia. In this model it is assumed that no one in the population is vaccinated at the beginning of an outbreak, but intense vaccination programmes roll out immediately after the first five cases are recorded. The aggregated intervention, terms represented by η and t_C were replaced by a daily rate of vaccination assuming 100% efficacy (ω) into a vaccinated and immune compartment (V), all other estimated parameters

maintained their values. The flow diagram for this model is provided in Figure 5.20 along with the model equations (5.1).

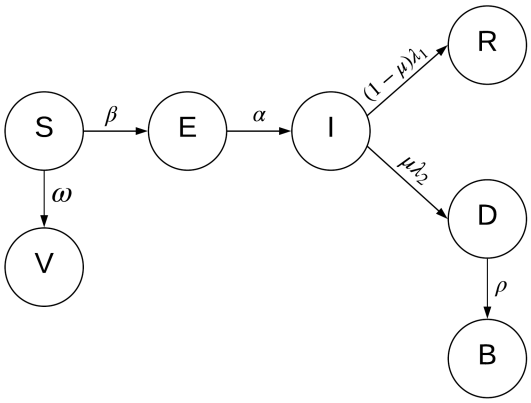

Figure 5.20: Model with vaccination

$$
\begin{aligned}
\frac{dS}{dt} &= -\beta_I \left(\frac{I}{N}\right) S - \beta_D \left(\frac{D}{N}\right) S - \omega S \\
\frac{dV}{dt} &= \omega S \\
\frac{dE}{dt} &= \beta_I \left(\frac{I}{N}\right) S + \beta_D \left(\frac{D}{N}\right) S - \alpha E \\
\frac{dI}{dt} &= \alpha E - (1-\mu)\lambda_1 I - \mu\lambda_2 I \\
\frac{dR}{dt} &= (1-\mu)\lambda_1 I \\
\frac{dD}{dt} &= \mu\lambda_2 I - \rho D \\
\frac{dB}{dt} &= \rho D
\end{aligned}
\tag{5.1}
$$

Figures 5.21 and 5.22 show the epidemic curves for varying levels of vaccination assuming vaccination is started once five cases have been detected in Sierra Leone and Liberia respectively. Instead of the disease duration on the x-axis however, the proportion of the total population vaccinated (V/N) is used, given that we have a constant vaccination rate. The initial population sizes were assumed as in Table 5.2. In Sierra Leone we notice that a daily vaccination rate of 0.001 would result in less than 1000 cases in total and 0.0015 would nearly prevent an epidemic happening at all. However in Liberia, it is estimated that slightly over double the rate (0.0025) is required to significantly slow the epidemic and bring the total cases below 1000.

Existence of a vaccine coupled with the infrastructure required for distribution provides an incredibly powerful tool in the elimination of infectious disease.

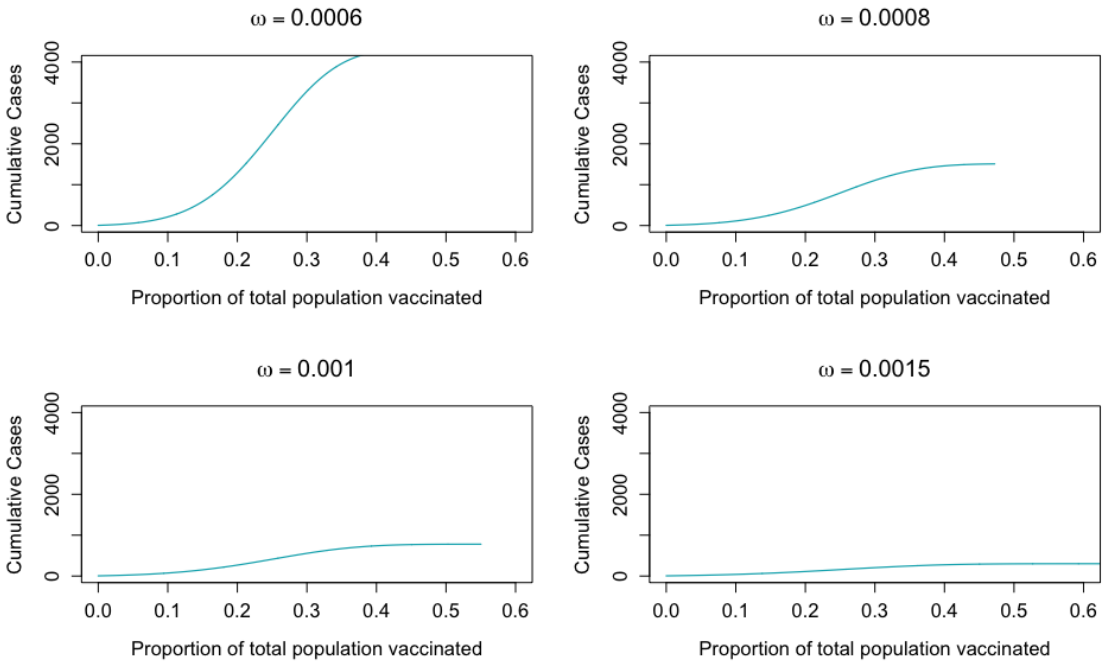

Figure 5.21: Cumulative cases for different fixed number of vaccinations per day (with Sierra Leone parameters)

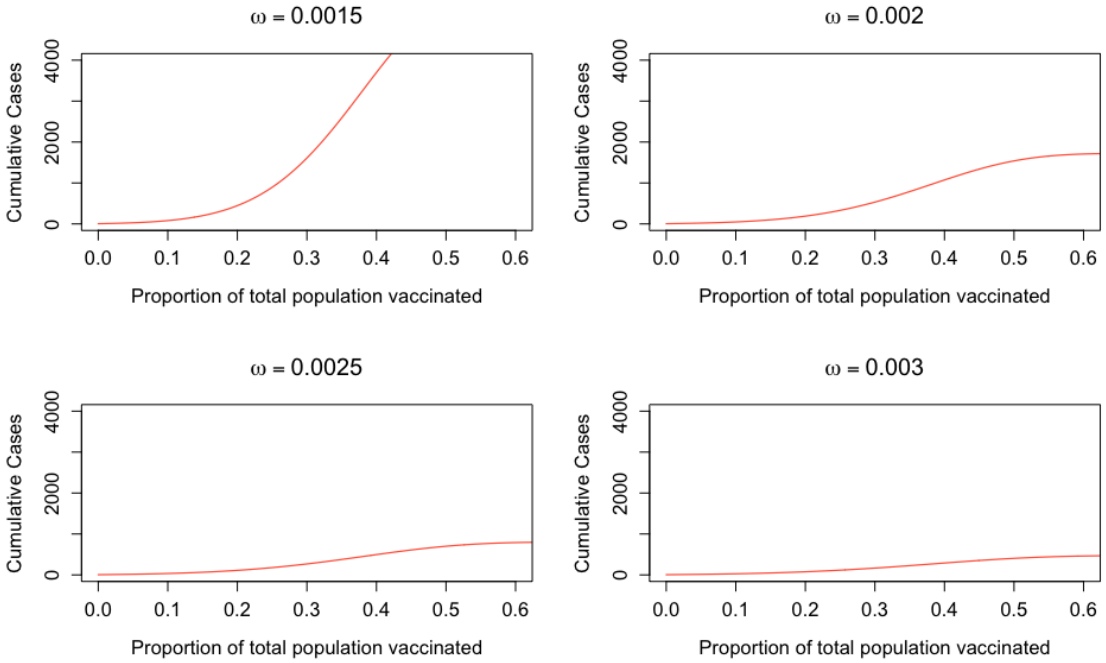

Figure 5.22: Cumulative cases for different fixed number of vaccinations per day (with Liberia parameters)

Chapter 6

Discussion and Conclusions

6.1 Mathematical Models of EVD

Mathematical modelling is an integral tool aiding our understanding of the dynamics of infectious diseases and their application has helped decision makers to investigate potential outcomes and strategies, especially in time- and resource-constrained situations.

The rare occurrence of Ebola Virus Disease (EVD) and the frightful symptoms, which often result in death, has engrossed the public and experts alike. Ebola was neither a new nor unfamiliar disease, health officials have been aware of its fatal consequences since 1976, however its history of sparse small outbreaks made many sceptical of its potential to cause large-scale damage. The unprecedented escalation of the virus in 2014, spurred much of the research that determined our fundamental understanding of EVD.

Thus, the mathematical modelling and research into the dynamics of EVD did not carry an established history for researchers to expand upon with the sudden onset of the West Africa outbreak. During 2014, the acceleration of infections and extremely high fatality associated with initial cases required officials and researchers alike to direct their immediate efforts at investigating the control measures that could prevent further escalation of the disease. Consequently, many of the mathematical models developed between 2014 and 2016 are focused on optimal control analysis and hypothetical scenario testing.

In reality, many of the effective controls identified in modelling were either inappropriate in context or unattainable. Vaccine and treatment development could not be achieved fast enough. Many response measures implemented through foreign aid did not take into consideration cultural context and were heavily undermined by community resistance in all three countries [64].

The subsided urgency of eradicating the Ebola virus and the full retrospective scope of data available, shifted the focus of this research away from optimal control analysis and towards determining a more conclusive understanding of the basic dynamics underlying EVD. Simple mathematical models were favoured over complex models in order to preserve model identifiability and the unique estimation of parameters that determine EVD dynamics.

To this end, compartmental models extending from the traditional Susceptible-Infectious-Recovered (SIR) framework were used in estimating and distinguishing between the effects present during the 2014 - 2016 EVD outbreak in Sierra Leone and Liberia, the two worst affected countries. A deter-

ministic six compartment model (SEIDBR) was developed to incorporate some of the key features of EVD: a latency period, infection resulting from close direct contact but distinguishing between contact with infectious patients and infectious corpses.

Initially the model was fit without the explicit inclusion of intervention and using a range of plausible parameter estimates obtained from epidemiological studies, as well as existing mathematical models. Interestingly, these parameters could not provide a suitable fit to the epidemic curve and always resulted in far more severe predictions for the outbreak. This could be indicative of two effects: the first being the under-recording of EVD cases and the second being the necessity of intervention and the various control measures enforced in ending the epidemic. The first explanation remains plausible but as the outbreak progressed, surveillance, case detection and management became cardinal to officially eliminating Ebola. The extent of under-reporting required in order to make the no intervention model fit the epidemic curve could possibly be expected in the first few weeks of the epidemic but is improbable to have lasted for the entire duration.

Thus, the 'no-intervention' model was extended to include the aggregated effect of various control measures and strategies implemented, as well as to incorporate death data obtained during the outbreak. It is assumed that after some critical time (referred to as t_C), various control measures act as levers to essentially reduce these two effects. Quarantine, hospitalisation, banning large social events, the closing of borders, case tracking, and educational media campaigns are all measures instituted to reduce contact with living infectious EVD patients. The roll-out of safe burial teams, state enforced cremation, disinfection of homes and again cultivating cultural awareness of the role of funerals and burial ceremonies in the spread of EVD are all measures acting on reducing contact with the infectious corpses of victims. These effects are aggregated by the crude addition of a term η ($0 < \eta < 1$) that reduces the transmission rates from infectious patients and corpses once drastic intervention efforts are introduced after time t_C.

Parameters inherent to Ebola virus that have remained relatively consistent since 1976 were fixed at estimates obtained from the literature. Parameters that did not show consistency in reporting - and that were reasonably expected to vary within the context of different countries - were estimated by fitting the model to recorded case data using a least squares optimisation approach. Sensitivity analysis was performed to identify critical model parameters, and to quantify the impact of input parameter uncertainty on the value of an output.

Parameter estimation and analysis was repeated for both Sierra Leone and Liberia independently. This comparison highlighted the importance of disease context: an identical virus in two neighbouring countries could result in rather different estimates, clearly influenced by factors specific to culture and environment.

Anthropologists have long been aware of the importance of funeral and burial rites in West Africa, yet the role of burial and post-mortem infections significantly differed in Sierra Leone and Liberia. The time period from death until the safe disposal of a corpse was estimated to be approximately 2,5 days in Sierra Leone but under a day (8,5 hours) in Liberia. The longer the time before safe disposal of a body, the more the expected contacts with a corpse is expected to be and consequently this alters the effective transmission parameter. The effective transmission rate resulting from deceased EVD patients in Sierra Leone was nearly double the rate experienced in Liberia. Conversely, the effective transmission rate resulting from living EVD patients in Liberia was nearly double that in Sierra Leone. By ensuring the fast removal of infectious corpses early on in the epidemic, Liberia managed to control EVD faster and prevent more EVD related deaths than Sierra Leone, despite the countries reporting similar weekly case incidences during the peak of the outbreak. This corroborates the findings published by the WHO in praise of Liberia's response to the 2014 - 2016

EVD outbreak [67]. Notably, Liberian President Sirleaf acted on early warnings from international agencies and went against culture and tradition by ordering that the bodies of Ebola victims be cremated and not buried.

Many social factors such as political climate, history and general public trust in authorities can completely undermine intervention efforts. One factor that may have contributed to the success of reducing post-mortem infection in Liberia that is not often mentioned, is religion. A demographic survey conducted in 2013, indicated that 78.2% of Sierra Leonean respondents were Muslim and 21.2% Christian [57]. This corresponds with estimates derived by the Pew Research Centre in 2015: Muslim: 78%; Christian: 20.9% [48]. In contrast, according to a 2013 survey, 84.2% of the Liberian population practices Christianity and 11.8% practice Islamic faith [19]. Islamic law strictly forbids cremation. Burial rites require a ritual washing ceremony including bathing to be performed by another Muslim shortly after death and typically followed by a gathering of the Muslim community around the deceased to offer prayers [15]. Traditionally, Christianity opposed cremation however in recent years, many Christians have elected to be cremated and most churches have changed their stance to acknowledge that there are valid sanitary, economic and social reasons for cremation but maintain that burial better demonstrates reverence for the deceased [16]. The consequences of religious beliefs and principles provide an interesting frame of reference when attempting to compare the successes of different countries.

The base reproductive number (R_0) is arguably the single most reported measure in epidemiological modelling. The estimated value of R_0 in Sierra Leone was 1.33, and 1.60 in Liberia. Despite various limitations in the interpretation and comparison of R_0 to other studies - and in other diseases where its derivation may differ - it provides a threshold for determining whether an epidemic will take off. Both estimated values exceed this threshold of 1 and hence the disease will persist in both populations. Inclusion of the intervention terms estimated, ensured that R_C, the effective reproductive after controls are enforced, was less than 1.

Slow initial response by international and local authorities was often cited as a key contributor to the unprecedented scale of the 2014 - 2016 Ebola outbreak [64]. Further analysis into the timing of intervention reinforced a rather obvious lesson: controls acting on reducing the direct personal contact between individuals, or that rely of large scale behavioural changes from large groups, is almost inevitably time-consuming. In nations filled with millions of people, individual-level controls are practically cumbersome and humans are very slow to adapt their behaviour even in the face of urgent threat.

Promising results in a ring vaccination trial of rVSV-ZEBOV in Guinea [70] and the deployment of this vaccine in the contained ongoing 2018 outbreak in the Democratic Republic of the Congo, motivated the analysis of vaccination as a control strategy to EVD. It is not necessary for an entire population to be vaccinated for a disease to die out. If a certain proportion of the population above the herd immunity threshold are vaccinated early on in an outbreak, this indirectly provides protection to those that are not vaccinated. The herd immunity threshold obtained for Sierra Leone was 38% and 47% for Liberia. In this case, no epidemic would occur in either country given the presence of a 100% effective vaccine.

6.2 Limitations and Future Work

A limitation of the models derived is the simplifying assumption of a closed, fixed population aggregated to behaviour on the national level. Metapopulation models allowing for population mobility permit a more realistic contact structure as district-level transmission differed and porous borders significantly contributed to how the epidemic spread [64]. The model structure and results derived above could provide a basis to such a model, but these extensions will bring with it significant challenges in obtaining additional data sources at the district-level and computational challenges in model development given the many districts and large geographical area affected by EVD.

In reality, human behaviour and intervention efforts will fluctuate over the duration of an outbreak and thus the assumption of static rates is a limitation of this model. This model could be improved by deriving time dependent contact rates and burial rates; as well as quantifying the extent of under-reporting.

Future work in this area also includes adding an economic cost component to the existing models. Overall, more than $3.611 billion (USD) was spent to fight the epidemic by December 2015 [13]. It is worth investigating whether unnecessary costs were incurred in inefficient implementation. Ebola also resulted in severe indirect costs on the provision of health care services in Guinea, Liberia and Sierra Leone. Redirecting limited health care services and resources to Ebola caused major setbacks in the treatment and control of other serious diseases, including HIV, tuberculosis and malaria, affecting much larger proportions of populations than Ebola. Frameworks for the determination of the appropriate control measures in a local context and evaluation of the economic costs (direct and indirect) involved remain invaluable research, particularly with new interventions such as vaccines on the horizon.

6.3 The Impact of Culture and Tradition on an Epidemic

Throughout this paper there has been made reference to the many ways in which culture, tradition and religion may direct the outcomes of disease. Although this is certainly not unique to Ebola, the 2014 - 2016 West Africa outbreak brought about a desperate reminder of the importance of context in global health.

Adherence to ancestral funeral and burial sites has repeatedly been singled out, both in this paper as well as in the overall literature, as fuelling large explosions of new cases. Guinea's Ministry of Health, indicated that 60% of cases could be linked to traditional burial and funeral practices. WHO staff in Sierra Leone similarly estimated that 80% of cases were linked to these practices [64].

The level of intimacy present in these ceremonies are difficult to imagine in Western environments. The funeral of a single healer in Kenema, Sierra Leone is estimated to have resulted in as many as 365 Ebola deaths [63]. Burial rites were reinforced by a number of secret societies; some mourners bathe in or anoint others with rinse water from the washing of corpses. Understudies of socially prominent members of these secret societies have been known to sleep near a highly infectious corpse for several nights, believing that doing so allows the transfer of powers [40].

Equally unfamiliar were the response measures, like disinfecting houses, setting up barriers and fever checks, and the invasion by foreigners dressed in outlandish hazmat suits, who took people to hospitals or barricaded tent-like wards from which few returned [40].

Ebola preyed on another subtler deep-seated cultural trait: compassion. In West Africa, the virus spread through the networks that bind societies together in a culture that stresses compassionate care for the ill and ceremonial care for their bodies if they die [1]. Some doctors are thought to have become infected when they rushed, unprotected, to aid patients who collapsed in waiting rooms or on the grounds outside a hospital [64].

As several experts have noted [53, 1, 64, 7], when technical interventions cross purposes with entrenched cultural practices, culture always wins. Control efforts must work within the culture, not against it.

6.4 Conclusion

This study has designed a model and made use of all available data in estimating the parameters determining the underlying mechanisms of the Ebola Virus Disease as it presented in West Africa between 2014 - 2016. Simple model structures capturing the essential features of the disease's natural history allowed for the preservation of identifiability. The models produced sensible estimates of parameters, as well as of the base reproductive number, allowing for a narrower focus of the plausible parameter values than those currently obtainable from epidemiological studies and clinical reports. The model predictions also demonstrated an acceptable fit to the epidemic data. Broad analysis of interventions highlighted the importance of timing and the required effectiveness and coverage required of potential vaccination strategies for the fast elimination of Ebola Virus in future. The models developed provide tools with which to assess the impact of other proposed interventions in the case of containing future outbreak.

The nature of the Ebola virus makes it is unlikely that it will ever exert such a tragic effect in developed countries. However, the virus has preyed on the damaged infrastructure in countries that are in dire need of these resources. Analysis into the key forces that drove its unprecedented scale between 2014 - 2016 and the impact of various intervention efforts provide a strong foundation for determining appropriate responses in the specific context of any future outbreak. In this manner, mathematical models can form an integral part of the research, planning and evaluation of elimination-focused strategies so that Ebola elimination is achieved in a faster, more focused, and more cost effective manner in future.

Bibliography

[1] F. B. Agusto, M. I. Teboh-Ewungkem, and A. B. Gumel. Mathematical assessment of the effect of traditional beliefs and customs on the transmission dynamics of the 2014 ebola outbreaks. *BMC Medicine*, 13:96, 04 2015.

[2] M. D. Ahmad, M. Usman, A. Khan, and M. Imran. Optimal control analysis of ebola disease with control strategies of quarantine and vaccination. *Infectious Diseases of Poverty*, 5:72, July 2016. https://doi.org/10.1186/s40249-016-0161-6.

[3] C. L. Althaus. Estimating the reproduction number of ebola virus (ebov) during the 2014 outbreak in west africa. *PLOS Currents Outbreaks*, 09 2014. http://currents.plos.org/outbreaks/article/estimating-the-reproduction-number-of-zaire-ebolavirus-ebov-during-the-2014-outbreak-in-west-

[4] A. Barbarossa, Maria Vittoria ans Denes, G. Kiss, Y. Nakata, et al. Transmission dynamics and final epidemic size of ebola virus disease outbreaks with varying interventions. *PLOS ONE*, 10(7):1–21, 07 2015.

[5] J. Bartlett, D. J., and E. Pudlowski. Mathematical modeling of the 2014/2015 ebola epidemic in west africa. 9, 2015.

[6] BBC. Ebola crisis: Liberia police fire at monrovia protests, 08 2014. https://www.bbc.com/news/world-africa-28879471.

[7] S. Bedrosian, C. Young, S. L., et al. Lessons of risk communication and health promotion — west africa and united states. *Morbidity and Mortality Weekly Report*, 65:68–74, 2016.

[8] B. Bell, I. Damon, D. Jernigan, et al. Overview, control strategies, and lessons learned in the cdc response to the 2014–2016 ebola epidemic. *Morbidity and Mortality Weekly Report (CDC)*, 65(3):4–11, 07 2016.

[9] T. Berge, J.-S. Lubuma, G. Moremedi, N. Morris, and R. Kondera-Shava. A simple mathematical model for ebola in africa. *Journal of Biological Dynamics*, 11(1):42–74, 2017.

[10] E. Bonyah, K. Badu, and S. K. Asiedu-Addo. Optimal control application to an ebola model. *Asian Pacific Journal of Tropical Biomedicine*, 6(4):283–289, 11 2015.

[11] F. Brauer, P. van den Driessche, C. Bauch, J. Wu, C. Castillo-Chavez, D. Earn, Z. Feng, M. Lewis, J. Li, et al. *Mathematical Epidemiology*. Lecture Notes in Mathematics. Springer Berlin Heidelberg, 2008.

[12] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu. A limited-memory algorithm for bound constrained optimization. *SIAM Journal on Scientific and Statistical Computing*, 16:1190–1208, 1994.

[13] Centre for Disease Control. Ebola virus disease, 2017. https://www.cdc.gov/vhf/ebola/index.html.

[14] Centre for Disease Control. Ebola virus disease (evd) case counts, 2017. https://www.cdc.gov/vhf/ebola/history/2014-2016-outbreak/case-counts.html.

[15] M. Cheraghi, S. Payne, and M. Salsali. Spiritual aspects of end-of-life care for muslim patients: Experiences from iran. 11:468–74, 10 2005.

[16] CNN. Vatican issues guidelines on cremation, says no to scattering ashes, 10 2016. https://edition.cnn.com/2016/10/25/europe/cremation-vatican-scattering/.

[17] C. Coltart, B. Lindsey, et al. The ebola outbreak, 2013–2016: old lessons for new epidemics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1721), 2017.

[18] P. M. Datilo and Z. Ismail. A review on west african ebola epidemic dynamics models. *International Journal of Epidemiologic Research*, 3(4):386–401, 2016. http://ijer.skums.ac.ir/article_23154.html.

[19] DHS. Liberia: Demographic and health survey, 2013, 2013. https://dhsprogram.com/pubs/pdf/FR291/FR291.pdf.

[20] S. D. Djiomba-Njankou and F. Nyabadza. Modelling the potential impact of limited hospital beds on ebola virus disease dynamics. *Mathematical Methods in the Applied Sciences*, pages 1 – 17, 2018. https://onlinelibrary.wiley.com/doi/abs/10.1002/mma.4789.

[21] J. Farrar and P. Piot. The ebola emergency — immediate action, ongoing strategy. *New England Journal of Medicine*, 371(16):1545–1546, 2014. https://doi.org/10.1056/NEJMe1411471.

[22] H. Feldmann and T. Geisbert. https://www.thelancet.com/journals/lancet/article/piis0140-6736(10)60667-8/fulltext. *The Lancet*, 377(9768):849–862, 12 2010.

[23] H. Feldmann and T. Geisbert. https://www.thelancet.com/journals/lancet/article/piis0140-6736(10)60667-8/fulltext. *The Lancet*, 377(9768):849–862, 12 2010.

[24] E. P. Fenichel, C. Castillo-Chavez, M. G. Ceddia, G. Chowell, P. A. G. Parra, G. J. Hickling, G. Holloway, R. Horan, B. Morin, C. Perrings, M. Springborn, L. Velazquez, and C. Villalobos. Adaptive human behavior in epidemiological models. *Proceedings of the National Academy of Sciences*, 108(15):6306–6311, 2011.

[25] P. Fine, K. Eames, and D. L. Heymann. "herd immunity": A rough guide. *Clinical Infectious Diseases*, 52(7):911–916, 2011.

[26] D. Fisman, E. Khoo, and A. Tuite. Early epidemic dynamics of the west african 2014 ebola outbreak: Estimates derived with a simple two-parameter model. *PLOS Current Outbreaks*, 09 2014.

[27] H. Ghazanfar et al. Ebola, the killer virus. *Infectious Diseases of Poverty*, 4(15), 2015.

[28] K. Gittings and K. L. Matson. Establishing herd immunity against ebola through vaccination. *Vaccine*, 34(24):2644 – 2647, 2016.

[29] J. Glynn et al. Asymptomatic infection and unrecognised ebola virus disease in ebola-affected households in sierra leone: a cross-sectional study using a new non-invasive assay for antibodies to ebola virus. *The Lancet*, 17(6):645–653, 02 2017. https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(17)30111-1/fulltext.

[30] P.-S. Gsell, A. Camacho, et al. Ring vaccination with rvsv-zebov under expanded access in response to an outbreak of ebola virus disease in guinea, 2016: an operational and vaccine safety report. *The Lancet*, 17(12):1276–1284, 12 2017. https://www.thelancet.com/journals/laninf/article/PIIS1473-3099%2817%2930541-8/fulltext.

[31] B. Ivorra, D. Ngom, and A. Ramos. Be-codis: A mathematical model to predict the risk of human diseases spread between countries—validation and application to the 2014–2015 ebola virus disease epidemic. *Bulletin of Mathematical Biology*, 77(9):1668 – 1704, 09 2015.

[32] J. H. Jones. Notes on r0, 05 2007. https://web.stanford.edu/~jhj1/teachingdocs/Jones-on-R0.pdf.

[33] W. Kermack and A. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 115(772):700–721, 1927.

[34] A. Khan, M. Naveed, et al. Estimating the basic reproductive ratio for the ebola outbreak in liberia and sierra leone. *Infectious Diseases of Poverty*, 4(13), 02 2015.

[35] F. Krauer, S. Gsteiger, N. Low, C. H. Hansen, and C. L. Althaus. Heterogeneity in district-level transmission of ebola virus disease during the 2013-2015 epidemic in west africa. *PLOS Neglected Tropical Diseases*, 10(7):1–14, 07 2016.

[36] J. Kuhn and S. Bavari. Asymptomatic ebola virus infections—myth or reality? *The Lancet*, 17(6):570–571, 02 2017. https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(17)30110-X/fulltext.

[37] J. Li, D. Blakeley, and R. J. Smith. The failure of r0. *Computational and Mathematical Methods in Medicine*, 2011, 2011.

[38] A. Lloyd. Introduction to epidemiological modeling: Basic models and their properties, 01 2017.

[39] H. W. Lupton, R. D. Lambert, D. L. Bumgardner, J. B. Moe, and G. A. Eddy. Inactivated vaccine for ebola virus efficacious in guineapig model. *The Lancet*, 316(8207):1294 – 1295, 1980. http://www.sciencedirect.com/science/article/pii/S0140673680923521.

[40] A. Maxmen. How the fight against ebola tested a culture's traditions. *National Geographic*, 01 2015. https://news.nationalgeographic.com/2015/01/150130-ebola-virus-outbreak-epidemic-sierra-leone-funerals/.

[41] A. McElroy. Understanding bleeding in ebola virus disease. *Clinical advances in hematology and oncology*, 13:29–31, 2015.

[42] M. D. McKay, R. J. Beckman, and W. J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245, 1979. http://www.jstor.org/stable/1268522.

[43] M. I. Meltzer, C. Y. Atkins, S. Santibanez, B. Knust, et al. Estimating the future number of cases in the ebola epidemic — liberia and sierra leone, 2014–2015. *CDC Morbidity and Mortality Weekly Report (MMWR)*, 63(3), 09 2014. https://www.cdc.gov/mmwr/preview/mmwrhtml/su6303a1.htm.

[44] W. Metzger and S. Vivas-Martinez. Questionable efficacy of the rvsv-zebov ebola vaccine. *The Lancet*, 391(10125):1021, 03 2018. https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(18)30560-9/fulltext.

[45] K. Owada, T. Eckmanns, et al. Epidemiological data management during an outbreak of ebola virus disease: Key issues and observations from sierra leone. *Frontiers in Public Health*, 4(163), 08 2016.

[46] S. Oza et al. Symptom- and laboratory-based ebola risk scores to differentiate likely ebola infections. *Emerging Infectious Diseases*, 23(11):1792–1799, 2017. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5652431/.

[47] O. Ozmen, J. J. Nutaro, L. L. Pullum, and A. Ramanathan. Analyzing the impact of modeling choices and assumptions in compartmental epidemiological models. *SAGE: Simulation*, 92(5):459–472, 04 2016.

[48] Pew Research Centre. The future of world religions: Population growth projections, 2010-2050, 04 2015. http://www.pewforum.org/files/2015/03/PF_15.04.02_ProjectionsFullReport.pdf.

[49] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018. https://www.R-project.org/.

[50] C. M. Rivers, E. T. Lofgren, M. Marathe, S. Eubank, and B. L. Lewis. Modeling, the impact of interventions on an epidemic of ebola in sierra leone and liberia. *PLOS Current Outbreaks*, Oct 2014. http://currents.plos.org/outbreaks/article/obk-14-0043-modeling-the-impact-of-interventions-on-an-epidemic-of-ebola-in-sierra-leone-and-

[51] E. Ross, P. Angelides, et al. Sierra leone's response to the ebola outbreak: Management strategies and key responder experiences, 03 2017. https://www.chathamhouse.org/publication/sierra-leones-response-ebola-outbreak-management-strategies-key-responder#.

[52] M. Sanchez and S. Blower. Uncertainty and sensitivity analysis of the basic reproductive rate, tuberculosis as an example. *American Journal of Epidemiology*, 145(12):1127 – 37, 1996. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.129.9834&rep=rep1&type=pdf.

[53] N. Sharareh, N. Sabounchi, H. Sayama, and R. MacDonald. The ebola crisis and the corresponding public behavior: A system dynamics approach. *PLOS Currents Outbreaks*, 1, 11 2016.

[54] M. Shen, Y. Xiao, and L. Rong. Modeling the effect of comprehensive interventions on ebola virus transmission. *Scientific Reports*, 5, 07 2015.

[55] J. Shultz, Z. Espinel, M. Espinola, and A. Rechkemmer. Distinguishing epidemiological features of the 2013–2016 west africa ebola virus disease outbreak. *Disaster Health*, 3(3):78–88, 08 2016. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5314891/.

[56] S. Silal. Mathematical modelling of infectious diseases, 04 2018.

[57] Statistics Sierra Leone (SSL). Sierra leone demographic and health survey 2013, 10 2014. https://dhsprogram.com/pubs/pdf/PR42/PR42.pdf.

[58] The History of Vaccines. Ebola virus disease and ebola vaccines, 01 2018. https://www.historyofvaccines.org/content/articles/ebola-virus-disease-and-ebola-vaccines.

[59] G. Velásquez, O. Aibana, E. Ling, et al. Time from infection to disease and infectiousness for ebola virus disease, a systematic review. *Clinical Infectious Diseases: An Official Publication*

of the Infectious Diseases Society of America, 61(7):1135 – 1140, 10 2015. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4560911/.

[60] N. F. Walker, C. S. Brown, D. Youkee, P. Baker, N. Williams, et al. Evaluation of a point-of-care blood test for identification of ebola virus disease at ebola holding units, western area, sierra leone, january to february 2015. *Eurosurveillance*, 20(12), 2015.

[61] X.-S. Wang and L. Zhong. Ebola outbreak in west africa: Real-time estimation and multiple-wave prediction. *Mathematical Biosciences and Engineering*, 12(5):1055–1063, 10 2015.

[62] T. A. Weppelmann, B. Donewell, U. Haque, W. Hu, R. J. S. Magalhaes, M. Lubogo, L. Godbless, S. Shabani, J. Maeda, H. Temba, T. C. Malibiche, N. Berhanu, W. Zhang, and L. Bawo. Determinants of patient survival during the 2014 ebola virus disease outbreak in bong county, liberia. *Global Health Research and Policy*, 1(1), Jun 2016. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5675064/.

[63] WHO. Sierra leone: a traditional healer and a funeral, 2014. http://www.who.int/csr/disease/ebola/ebola-6-months/sierra-leone/en/.

[64] WHO. Factors that contributed to undetected spread of the ebola virus and impeded rapid containment, 01 2015. http://www.who.int/csr/disease/ebola/one-year-report/factors/en/.

[65] A. Wilkinson and J. Fairhead. Comparison of social resistance to ebola response in sierra leone and guinea suggests explanations lie in political configurations not culture. *Critical Public Health*, 27(1):14–27, 2017. https://www.tandfonline.com/doi/full/10.1080/09581596.2016.1252034.

[66] World Health Organisation. Case definition recommendations for ebola or marburg virus diseases, 08 2014. http://www.who.int/csr/resources/publications/ebola/ebola-case-definition-contact-en.pdf.

[67] World Health Organisation. The ebola outbreak in liberia is over, 05 2015. http://www.who.int/mediacentre/news/statements/2015/liberia-ends-ebola/en/.

[68] World Health Organisation. Key events in the who response to the ebola outbreak, 2015. http://www.who.int/csr/disease/ebola/one-year-report/who-response/en/.

[69] World Health Organisation. Ebola situation reports: Democratic republic of the congo, 08 2018. http://www.who.int/ebola/situation-reports/drc-2018/en/.

[70] World Health Organisation. Ebola vaccination begins in north kivu, 08 2018. http://www.who.int/news-room/detail/08-08-2018-ebola-vaccination-begins-in-north-kivu.

[71] World Health Organisation. Ebola virus disease, 2018. http://www.who.int/en/news-room/fact-sheets/detail/ebola-virus-disease.

[72] J. Wu, R. Dhingra, M. Gambhir, and J. Remais. Sensitivity analysis of infectious disease models: methods, advances and their application. *Journal of the Royal Society Interface*, 10(86), 06 2013. http://rsif.royalsocietypublishing.org/content/10/86/20121018.

[73] Z.-Q. Xia, S.-F. Wang, S.-L. Li, L.-Y. Huang, W.-Y. Zhang, G.-Q. Sun, Z.-T. Gai, and Z. Jin. Modeling the transmission dynamics of ebola virus disease in liberia. *Nature: Scientific Reports*, 5, 09 2015. https://www.nature.com/articles/srep13857#t2.

[74] Xinyu Liu and Juhan Sonin and Emily Twaddell, 10 2014.

[75] H. Yan. American nurse with protective gear gets ebola - how could this happen? *CNN*, 10 2014. https://edition.cnn.com/2014/10/13/health/ebola-nurse-how-could-this-happen/index.html.

[76] T. Zolnikov. New targets for behaviour change in ebola outbreaks: Ideas for future interventions. *Intervention: Journal of Mental Health and Psycological Support in Conflicted Affected Areas*, 16(2):79–85, 2018.

Appendices

Appendix A

Additional information and figures

A.1 Initial transmission chain for the 2014 - 2016 outbreak

"Retrospective analyses traced the source of the outbreak to Meliandou village within the prefecture of Guéckédou, in the forested region of southeastern Guinea that borders both Sierra Leone and Liberia. The suspected index case was a two-year old child who fell ill on 2 December 2013 and died 4 days later [29]. A second epidemiological investigation confirmed the source village and index case, but the date of death of the index case was documented as the end of December 2013. Other family members rapidly became unwell and died between 13 December 2013 and 1 January 2014 (the index case's mother, sister, and grandmother). A village midwife who cared for the index case during his illness also fell ill—during her hospitalization in the nearest town, Guéckédou, she likely infected another healthcare worker (HCW) who was hospitalized in Macenta Hospital and is thought to have triggered the spread of the infection to a larger town. The midwife also had epidemiological links to cases in villages around Guéckédou prefecture (Dandou Pombo, Dawa and Gbandou Villages) between January and March 2014. The initial case fatality rate was 86% (12 of the 14 original patients with a known outcome died). Baize et al. reported this initial transmission chain in October 2014, and an adaptation of the initial transmission tree is shown in figure A.1" [17].

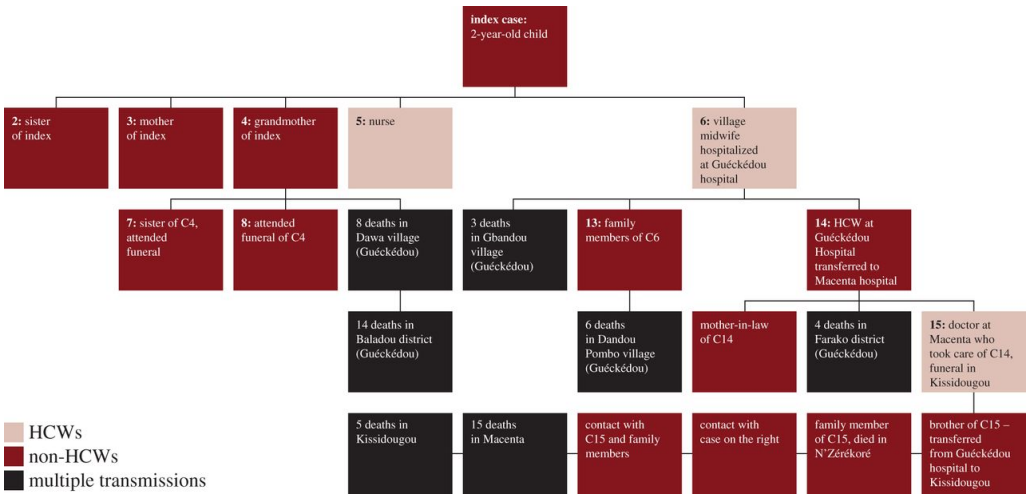

Figure A.1: Initial transmission chain in Guinea. HCW, healthcare worker. [17]

A.2 Assessing the impact and timing of intervention

The relationship between timing and cases avoided can be extended further. By calculating the total number of cases after 800 days for a sequence of plausible intervention times, we can infer the number of cases avoided for every day that intervention was implemented sooner in this model. This relationship between number of days sooner than the estimated intervention date and the number of cases are plotted for both countries in Figures A.2 and A.3.

Both graphs show a concave down, increasing shape, implying that at 'later' stages of the outbreak, one day of earlier action prevents many cases but this effect slows down if intervention occurs early in the outbreak.

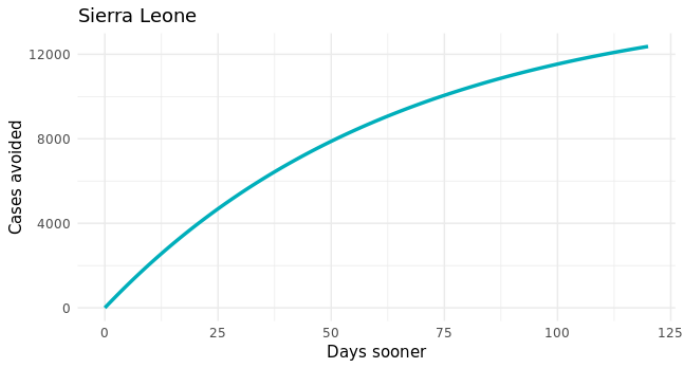

Figure A.2: Cases avoided for every day of earlier intervention Sierra Leone

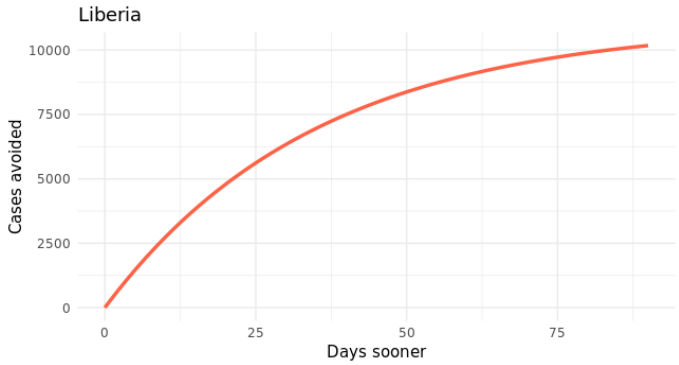

Figure A.3: Cases avoided for every day of earlier intervention Liberia

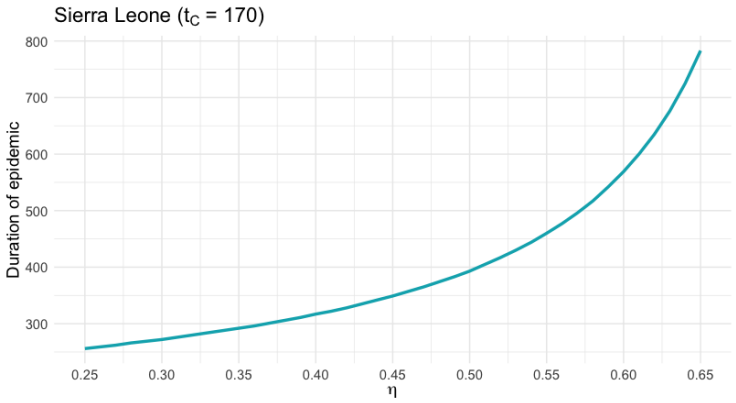

Figure A.4: Assessing the impact of intervention intensity in Sierra Leone

A.3 Vaccination model equations

$$
\begin{aligned}
\frac{dS}{dt} &= -\beta_I \left(\frac{I}{N}\right) S - \beta_D \left(\frac{D}{N}\right) S - \omega S \\
\frac{dV}{dt} &= \omega S \\
\frac{dE}{dt} &= \beta_I \left(\frac{I}{N}\right) S + \beta_D \left(\frac{D}{N}\right) S - \alpha E \\
\frac{dI}{dt} &= \alpha E - (1-\mu)\lambda_1 I - \mu\lambda_2 I \\
\frac{dR}{dt} &= (1-\mu)\lambda_1 I \\
\frac{dD}{dt} &= \mu\lambda_2 I - \rho D \\
\frac{dB}{dt} &= \rho D
\end{aligned}
\tag{A.1}
$$

Appendix B

Code

B.1 Data cleaning

B.1.1 Sierra Leone

```
###### Prepare Sierra Leone data ------

# clear environment and set working directory
rm(list = ls())
wd = "~/Ebola/final"
setwd(wd)

# install required libraries
library(ggplot2)
library(grid)
library(gridExtra)

# read in the raw data
full_data_raw = read.csv("CDCallcasecounts.csv", sep = ";")

full = full_data_raw
colnames(full)[1] = "Date"
full$Date = as.Date(full$Date, format = "%Y/%m/%d")
full = full[-266,] #remove empty final row

# only look at dataframe columns for Sierra Leone - column 6 and 7
sl.data.raw = full[,c(1,6,7)]

sl.cases = sl.data.raw

# remove any missing observations
sl.cases = na.omit(sl.cases)

# rename column headings
colnames(sl.cases)[2] = "Cases"
```

```
colnames(sl.cases)[3] = "Deaths"

# plot raw SL case data
sl_raw_cases = ggplot(sl.cases, aes(Date, Cases)) +
  geom_point(shape = 1) +
  ggtitle("Cumulative Cases") +
  theme_minimal()
#sl_raw_cases

# plot raw SL death data
sl_raw_deaths = ggplot(sl.cases, aes(Date, Deaths)) +
  geom_point(shape = 1) +
  ggtitle("Cumulative Deaths") +
  theme_minimal()
#sl_raw_deaths

# arrange plots of raw data in a grid format
title1=textGrob("Sierra Leone Raw Data", gp=gpar(fontface="bold", fontsize = 16, col = "#00AI
grid.arrange(sl_raw_cases, sl_raw_deaths, nrow = 1, ncol = 2,
             top = title1)

# remove duplicate dates
dup.dates = which( duplicated( sl.cases$Date))
sl.data = sl.cases[-dup.dates,]

# order data from earliest date to latest date
sl.data = sl.data[ order(sl.data$Date), ]

# start from date when cases are first > 50
row.sl = min(which(sl.data$Cases >= 50))
start.sl = sl.data$Date[row.sl]      # SL: starting date "2014-06-02", observation 18
sl.data = sl.data[c(row.sl:nrow(sl.data)),]

# convert dates to days after initial starting date
sl.data$Day = as.numeric(difftime(sl.data$Date, start.sl-1, units = "days"))
sl.data = sl.data[,c(1,4,2,3)]
rownames(sl.data) = 1:nrow(sl.data)

# identify position of dips in cumulative cases
sl.data$inc = c(NA, sl.data$Cases[2:nrow(sl.data)] - sl.data$Cases[1:(nrow(sl.data)-1)] )
neg.inc.sl = which(sl.data$inc < 0)
sl.data = sl.data[-c(5, 44, 45, 76, 141, 170),]
rownames(sl.data) = 1:nrow(sl.data)

# smooth the death data to only select every 14th point
sl.data$Deaths[-seq(1, nrow(sl.data), 14)] = NA

# plot non missing death indices
death_ind <- ! is.na(sl.data$Deaths)
plot(Deaths ~ Date, data = sl.data , subset = death_ind, type="l")

# plot cleaned SL case data
sl_clean_cases = ggplot(sl.data, aes(Date, Cases)) +
```

```
  geom_point(shape = 1) +
  ggtitle("Cumulative Cases") +
  theme_minimal()

# plot cleaned SL death data
sl_clean_deaths = ggplot(sl.data, aes(Date, Deaths)) +
  geom_point(shape = 1, size = 1.7) +
  ggtitle("Cumulative Deaths") +
  theme_minimal()

# arrange plots of cleaned data in grid format
title1=textGrob("Sierra Leone Cleaned Data", gp=gpar(fontface="bold", fontsize = 16, col = "
grid.arrange(sl_clean_cases, sl_clean_deaths, nrow = 1, ncol = 2,
             top = title1)

# save sierra leone data in rds and csv format
saveRDS(sl.data, file = "sl_data_23Sept.rds")
write.csv(sl.data, "sl_data_23Sept.csv")
```

B.1.2 Liberia

```
###### Prepare Liberia data ------

# clear environment and set working directory
rm(list = ls())
wd = "~/Liberia code"
setwd(wd)

# install required libraries
library(ggplot2)
library(grid)
library(gridExtra)

# read in the raw data
full_data_raw = read.csv("CDCallcasecounts.csv", sep = ";")

full = full_data_raw
colnames(full)[1] = "Date"
full$Date = as.Date(full$Date, format = "%Y/%m/%d")
full = full[-266,] #remove empty final row

# only look at dataframe columns for Liberia - column 4 and 5
lib.cases = full[,c(1,4,5)]

# remove any missing observations
lib.cases = na.omit(lib.cases)

# rename column headings
colnames(lib.cases)[2] = "Cases"
colnames(lib.cases)[3] = "Deaths"

# plot raw LIB case data
```

```
lib_raw_cases = ggplot(lib.cases, aes(Date, Cases)) +
  geom_point(shape = 1) +
  ggtitle("Cumulative Cases") +
  theme_minimal()

# plot raw LIB death data
lib_raw_deaths = ggplot(lib.cases, aes(Date, Deaths)) +
  geom_point(shape = 1) +
  ggtitle("Cumulative Deaths") +
  theme_minimal()

# arrange plots of raw data in a grid format
title1=textGrob("Liberia Raw Data", gp=gpar(fontface="bold", fontsize = 16, col = "tomato"))
grid.arrange(lib_raw_cases, lib_raw_deaths, nrow = 1, ncol = 2,
             top = title1)

# remove duplicate dates
dup.dates = which( duplicated( lib.cases$Date))
lib.data = lib.cases[-dup.dates,]

# order data from earliest date to latest date
lib.data = lib.data[ order(lib.data$Date), ]

# start from date when cases are first > 50
row.lib = min(which(lib.data$Cases >= 50))
start.lib = lib.data$Date[row.lib]
lib.data = lib.data[c(row.lib:nrow(lib.data)),]

# convert dates to days after initial starting date
lib.data$Day = as.numeric(difftime(lib.data$Date, start.lib-1, units = "days"))
lib.data = lib.data[,c(1,4,2,3)]
rownames(lib.data) = 1:nrow(lib.data)

# identify position of dips in cumulative cases
lib.data$inc = c(NA, lib.data$Cases[2:nrow(lib.data)] - lib.data$Cases[1:(nrow(lib.data)-1)]
lib.data[c(125, 126, 128, 129, 130, 132, 133, 134, 135), 2] = 10672
neg.inc = which(lib.data$inc < 0)
lib.data = lib.data[-c(39, 118),]
rownames(lib.data) = 1:nrow(lib.data)

#plot(Deaths ~ Date, data = lib.data, type="l")
#plot(Deaths ~ Date, data = lib.data , subset = seq(1,nrow(lib.data),5), type="l")
#plot(Cases ~ Date, data = lib.data , type="l")
#plot(inc ~ Date, data = lib.data , type="l")

# smooth the death data to only select every 5th point
lib.data$Deaths[-seq(1, nrow(lib.data), 5)] = NA

# plot non missing death indices
death_ind <- ! is.na(lib.data$Deaths)
plot(Deaths ~ Date, data = lib.data , subset = death_ind, type="l")

# delete a few inconsistent observations - clear errors
```

```
lib.data = lib.data[-c(123:133),]

# plot cleaned LIB case data
lib_clean_cases = ggplot(lib.data, aes(Date, Cases)) +
  geom_point(shape = 1) +
  ggtitle("Cumulative Cases") +
  theme_minimal()

# plot cleaned LIB death data
lib_clean_deaths = ggplot(lib.data, aes(Date, Deaths)) +
  geom_point(shape = 1, size = 1.6) +
  ggtitle("Cumulative Deaths") +
  theme_minimal()

# arrange plots of cleaned data in grid format
title1=textGrob("Liberia Cleaned Data", gp=gpar(fontface="bold", fontsize = 16, col = "tomato
grid.arrange(lib_clean_cases, lib_clean_deaths, nrow = 1, ncol = 2,
             top = title1)

# save liberia data in rds and csv format
saveRDS(lib.data, file = "lib_data_23Sept.rds")
write.csv(lib.data, "lib_data_23Sept.csv")
```

B.2 Model, parameter estimation and sensitivity analysis

B.2.1 Sierra Leone

```
###### Sierra Leone: Model fitting, parameter estimation ------

# clear environment and set working directory
rm(list = ls())
wd = "~/sierra leone code"
setwd(wd)

# install required libraries
library(deSolve)
library(gtools)    # for logit function
library(ggplot2)
library(gridExtra)

# read in data, create dates sequence
sl.data = readRDS("sl_data_23Sept.rds")
data.f = sl.data[,c(2:4)]
dates.seq = data.f$Day  # save dates for which case counts are available
death.dates.seq = data.f$Day[!is.na(data.f$Deaths)] # save dates for which death counts are
# Note: starting date: "2014-06-02" = day 1

# set initial values
InitPop = 6092000
E0 = 80
```

```
I0 = 41
R0 = 0
D0 = 3 #half?
B0 = 6
Inc0 = 50
S0 = InitPop - E0 - I0 - R0 - D0 - B0
start = c(S = S0, E = E0, I = I0, R = R0, D = D0, B = B0, Inc = Inc0)

# create model times vector
startday = 1
endday = 800   #Data stops at day 682
times = seq(startday, endday, 1)

#### Define functions -----

# SEIRDB function for estimation
seirdb.est = function(t, x, parms){
  with(as.list(c(parms,x)), {
    betaI = exp(logbetaI)          #effective contact rate with infectious people (alive)
    betaD = exp(logbetaD)          #effective contact rate with dead but infectious people
    alpha = exp(logalpha)          #1/latency period
    lambda1 = exp(loglambda1)      #1/period of infection to survival - still infectious
    lambda2 = exp(loglambda2)      #1/period of infection to death
    rho = exp(logrho)              #1/time to dispose of a body
    mu = inv.logit(logitmu)        #fatality rate
    eta = inv.logit(logiteta)      #factor to decrease betaI for t > tc
    tc = exp(logtc)                #time of intervention/control measures implemented

    if (t >= tc) {
      etat = eta
    } else {
      etat = 1
    }

    N = S + E + I + R + D

    dS = - betaI*etat*(I/N)*S - betaD*etat*(D/N)*S

    dE = betaI*etat*(I/N)*S + betaD*etat*(D/N)*S - alpha*E

    dI = alpha*E - (1 - mu)*(lambda1)*I - mu*(lambda2)*I

    dR = (1 - mu)*lambda1*I

    dD = mu*lambda2*I - rho*D

    dB = rho*D

    dInc = betaI*etat*(I/N)*S + betaD*etat*(D/N)*S

    output = c(dS, dE, dI, dR, dD, dB, dInc)
    list(output)
  })
```

```
}

# SEIRD function - without transformed parameters for estimation
seirdb = function(t, x, parms){
  with(as.list(c(parms,x)), {

    if (t >= tc) {
      etat = eta
      #rhot = rho2
    } else {
      etat = 1
      #rhot = rho1
    }

    N = S + E + I + R + D

    dS = - betaI*etat*(I/N)*S - betaD*etat*(D/N)*S

    dE = betaI*etat*(I/N)*S + betaD*etat*(D/N)*S - alpha*E

    dI = alpha*E - (1 - mu)*(lambda1)*I - mu*(lambda2)*I

    dR = (1 - mu)*lambda1*I

    dD = mu*lambda2*I - rho*D

    dB = rho*D

    dInc = betaI*etat*(I/N)*S + betaD*etat*(D/N)*S

    output = c(dS, dE, dI, dR, dD, dB, dInc)
    list(output)
  })
}


# Function for calculating sum of squared errors from case and death data
seirdb.sse = function(varparms, fixparms, times, start, data) {
  seirdb.lse = ode(times = times, y = start, func = seirdb.est, parms = c(varparms, fixparms)

  error.cum.cases = (seirdb.lse[dates.seq, 8] - data$Cases)^2
  error.cum.cases[-c(1:48)] = 2*error.cum.cases[-c(1:48)]   #values after 180 days is observ

  error.cum.deaths = (seirdb.lse[death.dates.seq, 7] - data$Deaths[data$Day %in% death.dates

  sse.cases = sum(error.cum.cases)
  sse.deaths = sum(error.cum.deaths)
  sse = sse.cases + 1.5*sse.deaths  #since death data is smoothed, weighted 1.5 times

  return(sse)
}

# Base reproductive number
```

```
r0.fn2 = function(estms){
  betaI = estms['betaI']
  betaD = estms['betaD']
  lambda1 = estms['lambda1']
  lambda2 = estms['lambda2']
  mu = estms['mu']
  rho = estms['rho']
  r0 = ( (betaI / (lambda1 + mu*(lambda2 - lambda1))) + ((betaD * mu * lambda2)/(rho*(lambda
  return(r0)
}

#### Estimate parameter values from data ------

# only need to run this the first time to initialize values for total error and estimates
min.err = 1000000000000
min.start = min.estms = NULL

# set number of iterations to run with different starting values
nsim = 500

# provide values for fixed parameters
fixparms = c(logalpha = log(1/10),
             loglambda1 = log(1/9.4),
             loglambda2 = log(1/7.5))

#start_time <- Sys.time()  #to measure run time

# create for loop to generate random starting values for variable parameters,
# optimize variable parameters using L-BFGS-B method
# calculate total error with these model parameters, if the total error is less
# than the current saved minimum error, save the estimated parameters as the
# best parameters ('min.estms')

for (i in 1:nsim) {

  varparms = c(logbetaI = log(runif(1, 0.15, 0.2)),
               logbetaD = log(runif(1, 0.1, 0.4)),
               logrho = log(runif(1, 0.25, 1.5)),
               logtc = log(runif(1, 60, 280)),
               logiteta = logit(runif(1, 0.01, 0.6)),
               logitmu = logit(runif(1, 0.25, 0.7)))

  sl.optim = optim(par = varparms, seirdb.sse, fixparms = fixparms, method = "L-BFGS-B",
                   times = times, start = start, data = data.f,
                   lower = c(-10, -10, log(0.2), -10, -10, logit(0.2)),
                   upper = c(log(2), log(2), log(1), log(300), logit(0.8), logit(0.8)))
  sl.sse = sl.optim$value

  if (sl.sse < min.err) {
    min.err = sl.sse
    min.start = varparms
    min.estms = sl.optim$par
  }
```

```
}

#end_time <- Sys.time()
#end_time - start_time


# back transform estimates from log/logit scale to original scale

all.estms = c( fixparms, min.estms)
all.estms = setNames(all.estms, c("alpha", "lambda1", "lambda2",
                                  "betaI", "betaD",  "rho", "tc", "eta", "mu"))
all.estms
estms.no.tr = all.estms

all.estms[c("alpha", "lambda1", "lambda2",
            "betaI", "betaD","rho", "tc")] = sapply(all.estms[c("alpha", "lambda1", "lambda2"
                                                               "betaI", "betaD","rho", "tc"
all.estms[c("eta", "mu") ] = sapply(all.estms[c("eta", "mu")], inv.logit)
all.estms

# save estimates in rds and csv format
saveRDS(all.estms, file = 'sl_final_24Sept.rds')
write.table(all.estms, file = "final_sl.csv")

model.estms = all.estms
#model.estms = readRDS('sl_final_24Aug.rds')
model.estms

# calculate r0 value
r0_sl = r0.fn2(model.estms)
r0_sl

# fit model using estimated parameters
model.f = ode(times = times, y = start, func = seirdb, parms = model.estms)

#### plot predicted behaviour of model compartments ------

model.f2 = as.data.frame(model.f)

# cumulative cases
pInc = ggplot(data.f, aes(x = Day, y = Cases)) +
  geom_point(shape = 1, color="gray35") +
  geom_line(data = model.f2, aes(x = time, y = Inc), col = "#00AFBB", size = 0.7) +
  ggtitle("Sierra Leone Cumulative Cases")
  #theme_minimal()
pInc

#deaths
pB = ggplot(data.f, aes(x = Day, y = Deaths)) +
  geom_point(shape = 1, color="gray35") +
  geom_line(data = model.f2, aes(x = time, y = B), col = "#00AFBB", size = 0.7) +
  ggtitle("Sierra Leone Cumulative Deaths (Class B)")
```

```
  #theme_minimal()
pB


#incidence
model.f3 = model.f2
model.f3$inc_predict = c(NA, model.f3[2:nrow(model.f3),8] - model.f3[1:(nrow(model.f3)-1),8]

pI = ggplot(sl.data, aes(x = Day, y = inc)) +
  geom_point(shape = 1, color="gray35") + ylab("Cases") +
  geom_line(data = model.f3, aes(x = time, y = inc_predict), col = "#00AFBB", size = 0.7) +
  ggtitle("Sierra Leone Case Incidence (Class I)")
  #theme_minimal()
pI

#grid.arrange(pInc, pB, pI, ncol = 1)

#S
pS = ggplot(model.f2, aes(x = time, y = S)) +
  geom_line(color = "#00AFBB", size = 1) +
  xlab("Days") + ylab("S") +
  theme_minimal()
  #theme(axis.text.x=element_blank())

#E
pE = ggplot(model.f2, aes(x = time, y = E)) +
  geom_line(color = "#00AFBB", size = 1) +
  xlab("Days") + ylab("E") +
  theme_minimal()

#R
pR = ggplot(model.f2, aes(x = time, y = R)) +
  geom_line(color = "#00AFBB", size = 1) +
  xlab("Days") + ylab("R") +
  theme_minimal()

#D
pD = ggplot(model.f2, aes(x = time, y = D)) +
  geom_line(color = "#00AFBB", size = 1) +
  xlab("Days") + ylab("D") +
  theme_minimal()

grid.arrange(pS, pE, pR, pD, nrow = 2, ncol = 2)

###### Sensitivity analysis for Sierra Leone ------

# set working directory, read in saved parameter estimates
wd = "~/Desktop/Project/Drafts/mathematical-modelling-spread/code/sl"
setwd(wd)
model.estms = read.csv("final_sl.csv", sep = " ", skip = 1, header = FALSE)
model.estms = setNames( model.estms$V2, model.estms$V1)  #turn df into named vector
model.estms
```

```
# install required libraries
library(deSolve)
library(lhs)
library(ppcor)
library(data.table)
library(ggplot2)

#### Defining functions -----

# base reproductive number
r0.calc = function(params){

  betaI = params['betaI']
  betaD = params['betaD']
  lambda1 = params['lambda1']
  lambda2 = params['lambda2']
  mu = params['mu']
  rho = params['rho']

  r0 = ( (betaI / (lambda1 + mu*(lambda2 - lambda1))) + ((betaD * mu * lambda2)/(rho*(lambda
  return(r0)
}

#### Univariate relationship between R0 and key parameters -----

## betaI
betaI_range = seq(0.05, 0.16, length.out = 500)
model.estms2 = model.estms
betaI_r0_range = NULL
for (i in 1:length(betaI_range)){
  model.estms2['betaI'] = betaI_range[i]
  betaI_r0_range = c(betaI_r0_range, r0.calc(model.estms2))
}
betaI_info = data.frame(betaI = betaI_range, r0 = betaI_r0_range)

r0_betaI = ggplot(betaI_info, aes(x = betaI, y = r0)) +
  geom_line(size = 1.2, colour = "#00AFBB") + xlab(expression(beta[I])) + ylab(expression('R
  theme_minimal() +
  theme( axis.title.x = element_text(size=rel(1.2)),
         axis.title.y = element_text(size=rel(1.2)))
  #theme(axis.title.y = element_text(angle=0))
#r0_betaI


## betaD
betaD_range = seq(0.5, 0.8, length.out = 500)
model.estms2 = model.estms
betaD_r0_range = NULL
for (i in 1:length(betaD_range)){
  model.estms2['betaD'] = betaD_range[i]
  betaD_r0_range = c(betaD_r0_range, r0.calc(model.estms2))
}
betaD_info = data.frame(betaD = betaD_range, r0 = betaD_r0_range)
```

```
r0_betaD = ggplot(betaD_info, aes(x = betaD, y = r0)) +
  geom_line(size = 1.2, colour = "#00AFBB") + xlab(expression(beta[D])) + ylab(expression('R
  theme_minimal() +
  theme( axis.title.x = element_text(size=rel(1.2)),
         axis.title.y = element_text(size=rel(1.2)))
#theme(axis.title.y = element_text(angle=0))
#r0_betaD


## lambda1
lambda1_range = seq(0.06, 0.17, length.out = 500)
model.estms2 = model.estms
lambda1_r0_range = NULL
for (i in 1:length(lambda1_range)){
  model.estms2['lambda1'] = lambda1_range[i]
  lambda1_r0_range = c(lambda1_r0_range, r0.calc(model.estms2))
}
lambda1_info = data.frame(lambda1 = lambda1_range, r0 = lambda1_r0_range)

r0_lambda1 = ggplot(lambda1_info, aes(x = lambda1, y = r0)) +
  geom_line(size = 1.2, colour = "#00AFBB") + xlab(expression(lambda[1])) + ylab(expression(
  theme_minimal() +
  theme( axis.title.x = element_text(size=rel(1.2)),
         axis.title.y = element_text(size=rel(1.2)))
#theme(axis.title.y = element_text(angle=0))
#r0_lambda1


## lambda2
lambda2_range = seq(0.06, 0.17, length.out = 500)
model.estms2 = model.estms
lambda2_r0_range = NULL
for (i in 1:length(lambda2_range)){
  model.estms2['lambda2'] = lambda2_range[i]
  lambda2_r0_range = c(lambda2_r0_range, r0.calc(model.estms2))
}
lambda2_info = data.frame(lambda2 = lambda2_range, r0 = lambda2_r0_range)

r0_lambda2 = ggplot(lambda2_info, aes(x = lambda2, y = r0)) +
  geom_line(size = 1.2, colour = "#00AFBB") + xlab(expression(lambda[2])) + ylab(expression(
  theme_minimal() +
  theme( axis.title.x = element_text(size=rel(1.2)),
         axis.title.y = element_text(size=rel(1.2)))
#theme(axis.title.y = element_text(angle=0))
#r0_lambda2


## mu
mu_range = seq(0.2, 0.7, length.out = 500)
model.estms2 = model.estms
mu_r0_range = NULL
for (i in 1:length(mu_range)){
```

```
  model.estms2['mu'] = mu_range[i]
  mu_r0_range = c(mu_r0_range, r0.calc(model.estms2))
}
mu_info = data.frame(mu = mu_range, r0 = mu_r0_range)

r0_mu = ggplot(mu_info, aes(x = mu, y = r0)) +
  geom_line(size = 1.2, colour = "#00AFBB") + xlab(expression(mu)) + ylab(expression('R'[0]))
  theme_minimal() +
  theme( axis.title.x = element_text(size=rel(1.2)),
         axis.title.y = element_text(size=rel(1.2)))
#theme(axis.title.y = element_text(angle=0))
r0_mu


## rho
rho_range = seq(0.33, 1.5, length.out = 500)
model.estms2 = model.estms
rho_r0_range = NULL
for (i in 1:length(rho_range)){
  model.estms2['rho'] = rho_range[i]
  rho_r0_range = c(rho_r0_range, r0.calc(model.estms2))
}
rho_info = data.frame(rho = rho_range, r0 = rho_r0_range)

r0_rho = ggplot(rho_info, aes(x = rho, y = r0)) +
  geom_line(size = 1.2, colour = "#00AFBB") + xlab(expression(rho)) + ylab(expression('R'[0]))
  theme_minimal() +
  theme( axis.title.x = element_text(size=rel(1.2)),
         axis.title.y = element_text(size=rel(1.2)))
#theme(axis.title.y = element_text(angle=0))
#r0_rho


grid.arrange(r0_betaI, r0_betaD, r0_lambda1, r0_lambda2, r0_mu, r0_rho, ncol = 2, nrow = 3)

#### Assessing impact of earlier intervention (changes in tC) ------

# 4 plots - 3 monthly changes
tc_range = c(80, 110, 140, 170)
model.estms2 = model.estms
par(mfrow = c(2,2))

for ( i in 1:length(tc_range)){
  model.estms2['tc'] = tc_range[i]
  fit = ode(times = times, y = start, func = seirdb, parms = model.estms2)
  diff_inc = c(NA, fit[2:nrow(fit),8] - fit[1:(nrow(fit)-1),8] )
  #peak = which.max(diff_inc) #which(diff_inc <= 0)[1]
  slow_down = which(diff_inc < 1)[1]
  plot(fit[,1], fit[,8], type = "l", ylim = c(0,15000), xlim = c(0,800),
       main = paste("tc =", tc_range[i]), xlab = "Days", ylab = "Cumulative Cases", col = "#
  abline(v = tc_range[i], lty = 2, col = "darkblue")
  abline(v = slow_down, lty = 2, col = "darkred")
  print(c(tc_range[i], slow_down, fit[800,8]))
```

```
}

# now look at relationship numerically
tc_diff_range = seq(0, 120, 0.25)
cases_avoided = NULL
total_cases = model.f[800,8]

for (i in 1:length(tc_diff_range)){
  model.estms2['tc'] = model.estms['tc'] - tc_diff_range[i]
  fit = ode(times = times, y = start, func = seirdb, parms = model.estms2)
  cases_avoided = c(cases_avoided, (total_cases - fit[800,8]))
}
df.cases_avoided = data.frame("days" = tc_diff_range, "cases" = cases_avoided)

tc_early = ggplot(df.cases_avoided, aes(x = days, y = cases, group = 1)) +
  geom_line(size = 1.2, color = "#00AFBB") + xlab("Days sooner") + ylab("Cases avoided") +
  ggtitle("Sierra Leone") +
  theme_minimal()

tc_early

#### Relationship between 'disease threshold' and duration of the epidemic -----
tc_range2 = seq(30, 170, 3)
model.estms2 = model.estms
threshold_vals = c()
duration_vals = c()

for (i in 1:length(tc_range2)){
  model.estms2['tc'] = tc_range2[i]
  fit = ode(times = times, y = start, func = seirdb, parms = model.estms2)
  diff_inc = c(NA, fit[2:nrow(fit),8] - fit[1:(nrow(fit)-1),8] )
  #peak = which.max(diff_inc) #which(diff_inc <= 0)[1]
  slow_down = which(diff_inc < 1)[1]

  threshold_vals = c(threshold_vals, fit[tc_range2[i], 8])

  duration_vals = c(duration_vals, (slow_down))
}
thresh_data = data.frame(threshold = threshold_vals, duration = duration_vals)

p_thresh = ggplot(thresh_data, aes(x = threshold, y = duration)) +
  geom_line(size = 1.2, color = "#00AFBB") + xlab("Threshold (number of cases)") + ylab("Dur
  ggtitle( expression( paste("Sierra Leone (", eta, " = 0.619)"))) +
  scale_x_continuous(breaks = seq(300, 6300, 1000)) +
  scale_y_continuous(minor_breaks = seq(0, 800, 25), breaks = seq(200, 800, 50)) +
  theme_minimal() +
  theme(text = element_text(size = 14))
p_thresh

#### Relationship between eta and duration of the epidemic keeping tc constant -----
eta_range2 = seq(0.25, 0.65, 0.01)
model.estms2 = model.estms
threshold_vals = c()
```

```
duration_vals = c()

for (i in 1:length(eta_range2)){
  model.estms2['eta'] = eta_range2[i]
  fit = ode(times = times, y = start, func = seirdb, parms = model.estms2)
  diff_inc = c(NA, fit[2:nrow(fit),8] - fit[1:(nrow(fit)-1),8] )
  #peak = which.max(diff_inc) #which(diff_inc <= 0)[1]
  slow_down = which(diff_inc < 1)[1]

  duration_vals = c(duration_vals, (slow_down)) #- model.estms2['tc']
}

thresh_data = data.frame(eta = eta_range2, duration = duration_vals)

p_eta = ggplot(thresh_data, aes(x = eta, y = duration)) +
  geom_line(size = 1.2, color = "#00AFBB") + xlab(expression(eta)) + ylab("Duration of epider
  ggtitle( expression(paste("Sierra Leone (", t[C], " = 170)"))) +
  scale_x_continuous(minor_breaks = seq(0.2, 0.65, 0.025), breaks = seq(0.2, 0.65, 0.05)) +
  scale_y_continuous(minor_breaks = seq(0, 800, 50), breaks = seq(0, 800, 100)) +
  theme_minimal() +
  theme(text = element_text(size = 14))
p_eta

#### Assessing changes in eta -----

#4 plots - 3 monthly changes
eta_range = c(0.45, 0.5, 0.55, 0.6)
model.estms2 = model.estms
par(mfrow = c(2,2))

for ( i in 1:length(eta_range)){
  model.estms2['eta'] = eta_range[i]
  fit = ode(times = times, y = start, func = seirdb, parms = model.estms2)
  diff_inc = c(NA, fit[2:nrow(fit),8] - fit[1:(nrow(fit)-1),8] )
  #peak = which.max(diff_inc) #which(diff_inc <= 0)[1]
  slow_down = which(diff_inc < 1)[1]
  plot(fit[,1], fit[,8], type = "l", ylim = c(0,15000), xlim = c(0,800),
       main = bquote(eta == .(eta_range[i])),
                     xlab = "Days", ylab = "Cumulative Cases", col = "#00AFBB", lwd = 1.2)
      #main = paste("eta =", eta_range[i]), xlab = "Days", ylab = "Cumulative Cases", col =
  abline(v = tc_range[i], lty = 2, col = "darkblue")
  abline(v = slow_down, lty = 2, col = "darkred")
  print(c(tc_range[i], slow_down, fit[800,8]))
}


#### Multivariate sensitivity analysis on R0 -----

n = 10000 # number to sample
k = 6 #number of variables

lhsp = randomLHS(n, k)
```

```
samples = data.frame( betaI = qunif( lhsp[,1], 0.06, 0.18),   #check these two
                      betaD = qunif( lhsp[,2], 0.5, 0.9),     #check
                      lambda1 = qnorm( lhsp[,3], 0.1, 0.0225),
                      lambda2 = qnorm( lhsp[,4], 0.1, 0.0225),
                      mu = qunif( lhsp[,5], 0.2, 0.7),
                      rho = qunif( lhsp[,6], 0.25, 1.5)
)

save.vals = NULL

for (i in 1:n) {
  #run = ode( times = times, y = start, func = seirdb, parms = samples[i,])
  save.vals[i] = r0.calc(samples[i,])
}

save.vals = unlist(save.vals)

save.vals2 = data.frame(values = save.vals)
ggplot(save.vals2, aes(values)) +
  geom_histogram(binwidth = 0.2, fill = "#00AFBB", color = "grey45", alpha = 0.5) +
  xlab(expression("R"[0])) + ylab("Frequency") +
  ggtitle(expression(paste("Uncertainty analysis of ", "R"[0], " for Sierra Leone"))) +
  theme_minimal()

t.test(save.vals)
summary(save.vals)

pcor_vals = c()

for ( i in 1:k){
  vals = pcor.test(x = samples[,i], y = save.vals, z = samples[,-i], method = "spearman")
  pcor_vals = c(pcor_vals, vals$estimate)
}

pcor_vals
pcor_vals = setNames(pcor_vals, colnames(samples))
barplot(pcor_vals, ylim = c(-1, 1))


#### Univariate sensitivity analysis on cumulative cases ----

# Using LHS
n = 10000 # number to sample
k = 9 #number of variables

lhsp = randomLHS(n, k)

samples = data.frame( alpha = qnorm( lhsp[,1], 0.1, 0.0125),
                      lambda1 = qnorm( lhsp[,2], 0.1, 0.0225),
                      lambda2 = qnorm( lhsp[,3], 0.1, 0.0225),
                      betaI = qunif( lhsp[,4], 0.05, 0.16),
                      betaD = qunif( lhsp[,5], 0.5, 0.8),
                      rho = qunif( lhsp[,6], 0.25, 1.5),
```

```
                        tc = qunif( lhsp[,7], 140, 200),
                        eta = qunif( lhsp[,8], 0.3, 0.7),
                        mu = qunif( lhsp[,9], 0.2, 0.7)
                        #E0 = qunif( lhsp[,9], 30, 70),
                        #I0 = qunif( lhsp[,10], 10, 40),
                        #D0 = qunif( lhsp[,11], 10, 30)
)

save.vals50 = NULL
save.vals200 = NULL
save.vals400 = NULL
save.vals800 = NULL

for (i in 1:n) {
  run = ode( times = times, y = start, func = seirdb, parms = samples[i,])
  save.vals50[i] = run[50, 8]
  save.vals200[i] = run[200,8]
  save.vals400[i] = run[400,8]
  save.vals800[i] = run[800,8]
}

save.vals50 = unlist(save.vals50)
save.vals200 = unlist(save.vals200)
save.vals400 = unlist(save.vals400)
save.vals800 = unlist(save.vals800)
#hist(save.vals, main = expression(paste("Uncertainty analysis of ", "R"[0], " for Liberia"))

pcor_vals50 = c()
pcor_vals200 = c()
pcor_vals400 = c()
pcor_vals800 = c()

for ( i in 1:k){
  vals50 = pcor.test(x = samples[,i], y = save.vals50, z = samples[,-i], method = "spearman"
  vals200 = pcor.test(x = samples[,i], y = save.vals200, z = samples[,-i], method = "spearman
  vals400 = pcor.test(x = samples[,i], y = save.vals400, z = samples[,-i], method = "spearman
  vals800 = pcor.test(x = samples[,i], y = save.vals800, z = samples[,-i], method = "spearman
  pcor_vals50 = c(pcor_vals50, vals50$estimate)
  pcor_vals200 = c(pcor_vals200, vals200$estimate)
  pcor_vals400 = c(pcor_vals400, vals400$estimate)
  pcor_vals800 = c(pcor_vals800, vals800$estimate)
}

#pcor_vals
pcor_vals50 = setNames(pcor_vals50, colnames(samples))
pcor_vals200 = setNames(pcor_vals200, colnames(samples))
pcor_vals400 = setNames(pcor_vals400, colnames(samples))
pcor_vals800 = setNames(pcor_vals800, colnames(samples))

barplot(pcor_vals800, ylim = c(-1, 1), cex.names = 0.8,
        ylab = "PRCC", main = "Sensitivity analysis of cumulative cases at day 800 for Sierra
        col = "#00AFBB")
```

```
vals800 = data.frame(pcor_vals800)
setDT(vals800, keep.rownames = TRUE)[]
colnames(vals800) = c("variable", "value")

# try do with ggplot
#need to put into long format
sens_bar800 = ggplot(vals800, aes(x = variable, y = value)) +
  geom_bar(stat = "identity", fill = "#00AFBB", col = "grey45", alpha = 0.5) +
  ylab("PRCC") + xlab("Parameter") + ggtitle("Sensitivity analysis of cumulative cases at day
  theme_minimal()

sens_bar800

#other 4 plots
vals50 = data.frame(pcor_vals50)
setDT(vals50, keep.rownames = TRUE)[]
colnames(vals50) = c("variable", "value")

sens_bar50 = ggplot(vals50, aes(x = variable, y = value)) +
  geom_bar(stat = "identity", fill = "#00AFBB", col = "grey45", alpha = 0.5) +
  ylab("PRCC") + xlab("Parameter") + ggtitle("Day 50") +
  theme_minimal() +
  theme(axis.text.x = element_text(size = 7.5))

#sens_bar50

vals200 = data.frame(pcor_vals200)
setDT(vals200, keep.rownames = TRUE)[]
colnames(vals200) = c("variable", "value")

sens_bar200 = ggplot(vals200, aes(x = variable, y = value)) +
  geom_bar(stat = "identity", fill = "#00AFBB", col = "grey45", alpha = 0.5) +
  ylab("PRCC") + xlab("Parameter") + ggtitle("Day 200") +
  theme_minimal() +
  theme(axis.text.x = element_text(size = 7.5))

#sens_bar200

vals400 = data.frame(pcor_vals400)
setDT(vals400, keep.rownames = TRUE)[]
colnames(vals400) = c("variable", "value")

sens_bar400 = ggplot(vals400, aes(x = variable, y = value)) +
  geom_bar(stat = "identity", fill = "#00AFBB", col = "grey45", alpha = 0.5) +
  ylab("PRCC") + xlab("Parameter") + ggtitle("Day 400") +
  theme_minimal() +
  theme(axis.text.x = element_text(size = 7.5))


sens_bar800.v2 = ggplot(vals800, aes(x = variable, y = value)) +
  geom_bar(stat = "identity", fill = "#00AFBB", col = "grey45", alpha = 0.5) +
  ylab("PRCC") + xlab("Parameter") + ggtitle("Day 800") +
  theme_minimal() +
```

```
  theme(axis.text.x = element_text(size = 7.5))


grid.arrange(sens_bar50, sens_bar200, sens_bar400, sens_bar800.v2, nrow = 2, ncol = 2)
```

B.2.2 Liberia

```
###### Liberia: Model fitting, parameter estimation ------

# clear environment and set working directory
rm(list = ls())
wd = "~/Liberia code"
setwd(wd)

# install required libraries
library(deSolve)
library(gtools)    # for logit function
library(ggplot2)
library(gridExtra)

# read in data, create dates sequence
lib.data = readRDS("lib_data_23Sept.rds")
data.f = lib.data[,c(2:4)]
dates.seq = data.f$Day # save dates for which case counts are available
death.dates.seq = data.f$Day[!is.na(data.f$Deaths)] # save dates for which death counts are

# starting date: "2014-06-02" = day 1

# set initial values
InitPop = 4294000
E0 = 80
I0 = 14
R0 = 0
D0 = 3
B0 = 34
Inc0 = 51
S0 = InitPop - E0 - I0 - R0 - D0 - B0
start = c(S = S0, E = E0, I = I0, R = R0, D = D0, B = B0, Inc = Inc0)

# create model times vector
startday = 1
endday = 800   # data stops at day
times = seq(startday, endday, 1)

#### Define functions -----

# SEIRDB function for estimation
seirdb.est = function(t, x, parms){
  with(as.list(c(parms,x)), {
    betaI = exp(logbetaI)          #effective contact rate with infectious people (alive)
    betaD = exp(logbetaD)          #effective contact rate with dead but infectious people
    alpha = exp(logalpha)          #1/latency period
```

```
    lambda1 = exp(loglambda1)       #1/period of infection to survival - still infectious
    lambda2 = exp(loglambda2)       #1/period of infection to death
    rho = exp(logrho)               #1/time to dispose of a body
    mu = inv.logit(logitmu)         #fatality rate
    eta = inv.logit(logiteta)       #factor to decrease betaI for t > tc
    tc = exp(logtc)                 #time of intervention/control measures implemented

    if (t >= tc) {
      etat = eta
    } else {
      etat = 1
    }

    N = S + E + I + R + D

    dS = - betaI*etat*(I/N)*S - betaD*etat*(D/N)*S

    dE = betaI*etat*(I/N)*S + betaD*etat*(D/N)*S - alpha*E

    dI = alpha*E - (1 - mu)*(lambda1)*I - mu*(lambda2)*I

    dR = (1 - mu)*lambda1*I

    dD = mu*lambda2*I - rho*D

    dB = rho*D

    dInc = betaI*etat*(I/N)*S + betaD*etat*(D/N)*S

    output = c(dS, dE, dI, dR, dD, dB, dInc)
    list(output)
  })
}

# SEIRD function - without transformed parameters for estimation
seirdb = function(t, x, parms){
  with(as.list(c(parms,x)), {

    if (t >= tc) {
      etat = eta
      #rhot = rho2
    } else {
      etat = 1
      #rhot = rho1
    }

    N = S + E + I + R + D

    dS = - betaI*etat*(I/N)*S - betaD*etat*(D/N)*S

    dE = betaI*etat*(I/N)*S + betaD*etat*(D/N)*S - alpha*E

    dI = alpha*E - (1 - mu)*(lambda1)*I - mu*(lambda2)*I
```

```
    dR = (1 - mu)*lambda1*I

    dD = mu*lambda2*I - rho*D

    dB = rho*D

    dInc = betaI*etat*(I/N)*S + betaD*etat*(D/N)*S

    output = c(dS, dE, dI, dR, dD, dB, dInc)
    list(output)
  })
}


# Function for calculating sum of squared errors from case and death data
seirdb.sse = function(varparms, fixparms, times, start, data) {
  seirdb.lse = ode(times = times, y = start, func = seirdb.est, parms = c(varparms, fixparms)

  error.cum.cases = (seirdb.lse[dates.seq, 8] - data$Cases)^2
  error.cum.cases[-c(1:48)] = 2*error.cum.cases[-c(1:48)]     #values after 180 days is obse

  error.cum.deaths = (seirdb.lse[death.dates.seq, 7] - data$Deaths[data$Day %in% death.dates

  sse.cases = sum(error.cum.cases)
  sse.deaths = sum(error.cum.deaths)
  sse = sse.cases + 1.5*sse.deaths

  return(sse)
}

# Base reproductive number
r0.fn2 = function(estms){
  betaI = estms['betaI']
  betaD = estms['betaD']
  lambda1 = estms['lambda1']
  lambda2 = estms['lambda2']
  mu = estms['mu']
  rho = estms['rho']
  r0 = ( (betaI / (lambda1 + mu*(lambda2 - lambda1))) + ((betaD * mu * lambda2)/(rho*(lambda
  return(r0)
}

#### Estimate parameter values from data ------

# only need to run this the first time to initialize values for total error and estimates
min.err = 1000000000000
min.start = min.estms = NULL

# set number of iterations to run with different starting values
nsim = 500

# provide values for fixed parameters
```

```
fixparms = c(logalpha = log(1/10),
             loglambda1 = log(1/9.4),
             loglambda2 = log(1/7.5))

#start_time <- Sys.time()  #to measure run time

# create for loop to generate random starting values for variable parameters,
# optimize variable parameters using L-BFGS-B method
# calculate total error with these model parameters, if the total error is less
# than the current saved minimum error, save the estimated parameters as the
# best parameters ('min.estms')

for (i in 1:nsim) {

  varparms = c(logbetaI = log(runif(1, 0.15, 0.2)),
               logbetaD = log(runif(1, 0.1, 0.4)),
               logrho = log(runif(1, 0.25, 1.5)),
               logtc = log(runif(1, 40, 120)),
               logiteta = logit(runif(1, 0.01, 0.6)),
               logitmu = logit(runif(1, 0.25, 0.7)))

  sl.optim = optim(par = varparms, seirdb.sse, fixparms = fixparms, method = "L-BFGS-B",
                   times = times, start = start, data = data.f,
                   lower = c(-10, -10, log(0.2), -10, -10, logit(0.2)),
                   upper = c(log(2), log(2), log(1.5), log(300), logit(0.8), logit(0.8)))
  sl.sse = sl.optim$value

  if (sl.sse < min.err) {
    min.err = sl.sse
    min.start = varparms
    min.estms = sl.optim$par
  }
}

#end_time <- Sys.time()
#end_time - start_time


# back transform estimates from log/logit scale to original scale

all.estms = c( fixparms, min.estms)
all.estms = setNames(all.estms, c("alpha", "lambda1", "lambda2",
                                  "betaI", "betaD",  "rho", "tc", "eta", "mu"))
all.estms
estms.no.tr = all.estms

all.estms[c("alpha", "lambda1", "lambda2",
            "betaI", "betaD","rho", "tc")] = sapply(all.estms[c("alpha", "lambda1", "lambda2"
                                                                "betaI", "betaD","rho", "tc"
all.estms[c("eta", "mu") ] = sapply(all.estms[c("eta", "mu")], inv.logit)
all.estms

# save estimates in rds and csv format
```

```
saveRDS(all.estms, file = 'lib_final_24Sept.rds')
write.table(all.estms, file = "final_lib.csv")

model.estms = all.estms
model.estms

# calculate r0 value
r0_lib = r0.fn2(model.estms)
r0_lib

# fit model using estimated parameters
model.f = ode(times = times, y = start, func = seirdb, parms = model.estms)

#### plot predicted behaviour of model compartments ------

model.f2 = as.data.frame(model.f)

# cumulative cases
pInc = ggplot(data.f, aes(x = Day, y = Cases)) +
  geom_point(shape = 1, color="gray35") +
  geom_line(data = model.f2, aes(x = time, y = Inc), col = "tomato", size = 0.7) +
  ggtitle("Liberia Cumulative Cases")
#theme_minimal()
pInc

#deaths
pB = ggplot(data.f, aes(x = Day, y = Deaths)) +
  geom_point(shape = 1, color="gray35") +
  geom_line(data = model.f2, aes(x = time, y = B), col = "tomato", size = 0.7) +
  ggtitle("Liberia Cumulative Deaths (Class B)")
#theme_minimal()
pB

#incidence
pI = ggplot(lib.data, aes(x = Day, y = inc)) +
  geom_point(shape = 1, color="gray35")  + ylab("Cases") +
  geom_line(data = model.f2, aes(x = time, y = I), col = "tomato", size = 0.7) +
  ggtitle("Liberia Case Incidence (Class I)")
#theme_minimal()
pI

#grid.arrange(pInc, pB, pI, ncol = 1)

#S
pS = ggplot(model.f2, aes(x = time, y = S)) +
  geom_line(color = "tomato", size = 1) +
  xlab("Days") + ylab("S") +
  theme_minimal()
#theme(axis.text.x=element_blank())

#E
pE = ggplot(model.f2, aes(x = time, y = E)) +
  geom_line(color = "tomato", size = 1) +
```

```
  xlab("Days") + ylab("E") +
  theme_minimal()

#R
pR = ggplot(model.f2, aes(x = time, y = R)) +
  geom_line(color = "tomato", size = 1) +
  xlab("Days") + ylab("R") +
  theme_minimal()

#D
pD = ggplot(model.f2, aes(x = time, y = D)) +
  geom_line(color = "tomato", size = 1) +
  xlab("Days") + ylab("D") +
  theme_minimal()

grid.arrange(pS, pE, pR, pD, nrow = 2, ncol = 2)

###### Sensitivity analysis for Liberia ------

# set working directory, read in saved parameter estimates
wd = "~/Desktop/Project/Drafts/mathematical-modelling-spread/code/lib"
setwd(wd)
model.estms = read.csv("final_lib.csv", sep = " ", skip = 1, header = FALSE)
model.estms = setNames( model.estms$V2, model.estms$V1)  #turn df into named vector
model.estms

# install required libraries
library(deSolve)
library(lhs)
library(ppcor)
library(data.table)
library(ggplot2)

#### Defining functions -----

# base reproductive number
r0.calc = function(params){

  betaI = params['betaI']
  betaD = params['betaD']
  lambda1 = params['lambda1']
  lambda2 = params['lambda2']
  mu = params['mu']
  rho = params['rho']

  r0 = ( (betaI / (lambda1 + mu*(lambda2 - lambda1))) + ((betaD * mu * lambda2)/(rho*(lambda
  return(r0)
}

#### Univariate relationship between R0 and key parameters -----

## betaI
betaI_range = seq(0.12, 0.22, length.out = 500)
```

```
model.estms2 = model.estms
betaI_r0_range = NULL
for (i in 1:length(betaI_range)){
  model.estms2['betaI'] = betaI_range[i]
  betaI_r0_range = c(betaI_r0_range, r0.calc(model.estms2))
}
betaI_info = data.frame(betaI = betaI_range, r0 = betaI_r0_range)

r0_betaI = ggplot(betaI_info, aes(x = betaI, y = r0)) +
  geom_line(size = 1.2, colour = "tomato") + xlab(expression(beta[I])) + ylab(expression('R'
  theme_minimal() +
  theme( axis.title.x = element_text(size=rel(1.2)),
         axis.title.y = element_text(size=rel(1.2)))
#theme(axis.title.y = element_text(angle=0))
#r0_betaI


## betaD
betaD_range = seq(0.24, 0.42, length.out = 500)
model.estms2 = model.estms
betaD_r0_range = NULL
for (i in 1:length(betaD_range)){
  model.estms2['betaD'] = betaD_range[i]
  betaD_r0_range = c(betaD_r0_range, r0.calc(model.estms2))
}
betaD_info = data.frame(betaD = betaD_range, r0 = betaD_r0_range)

r0_betaD = ggplot(betaD_info, aes(x = betaD, y = r0)) +
  geom_line(size = 1.2, colour = "tomato") + xlab(expression(beta[D])) + ylab(expression('R'
  theme_minimal() +
  theme( axis.title.x = element_text(size=rel(1.2)),
         axis.title.y = element_text(size=rel(1.2)))
#theme(axis.title.y = element_text(angle=0))
#r0_betaD


## lambda1
lambda1_range = seq(0.06, 0.17, length.out = 500)
model.estms2 = model.estms
lambda1_r0_range = NULL
for (i in 1:length(lambda1_range)){
  model.estms2['lambda1'] = lambda1_range[i]
  lambda1_r0_range = c(lambda1_r0_range, r0.calc(model.estms2))
}
lambda1_info = data.frame(lambda1 = lambda1_range, r0 = lambda1_r0_range)

r0_lambda1 = ggplot(lambda1_info, aes(x = lambda1, y = r0)) +
  geom_line(size = 1.2, colour = "tomato") + xlab(expression(lambda[1])) + ylab(expression('I
  theme_minimal() +
  theme( axis.title.x = element_text(size=rel(1.2)),
         axis.title.y = element_text(size=rel(1.2)))
#theme(axis.title.y = element_text(angle=0))
#r0_lambda1
```

```
## lambda2
lambda2_range = seq(0.06, 0.17, length.out = 500)
model.estms2 = model.estms
lambda2_r0_range = NULL
for (i in 1:length(lambda2_range)){
  model.estms2['lambda2'] = lambda2_range[i]
  lambda2_r0_range = c(lambda2_r0_range, r0.calc(model.estms2))
}
lambda2_info = data.frame(lambda2 = lambda2_range, r0 = lambda2_r0_range)

r0_lambda2 = ggplot(lambda2_info, aes(x = lambda2, y = r0)) +
  geom_line(size = 1.2, colour = "tomato") + xlab(expression(lambda[2])) + ylab(expression('R
  theme_minimal() +
  theme( axis.title.x = element_text(size=rel(1.2)),
         axis.title.y = element_text(size=rel(1.2)))
#theme(axis.title.y = element_text(angle=0))
#r0_lambda2


## mu
mu_range = seq(0.2, 0.7, length.out = 500)
model.estms2 = model.estms
mu_r0_range = NULL
for (i in 1:length(mu_range)){
  model.estms2['mu'] = mu_range[i]
  mu_r0_range = c(mu_r0_range, r0.calc(model.estms2))
}
mu_info = data.frame(mu = mu_range, r0 = mu_r0_range)

r0_mu = ggplot(mu_info, aes(x = mu, y = r0)) +
  geom_line(size = 1.2, colour = "tomato") + xlab(expression(mu)) + ylab(expression('R'[0]))
  theme_minimal() +
  theme( axis.title.x = element_text(size=rel(1.2)),
         axis.title.y = element_text(size=rel(1.2)))
#theme(axis.title.y = element_text(angle=0))
r0_mu


## rho
rho_range = seq(0.33, 1.5, length.out = 500)
model.estms2 = model.estms
rho_r0_range = NULL
for (i in 1:length(rho_range)){
  model.estms2['rho'] = rho_range[i]
  rho_r0_range = c(rho_r0_range, r0.calc(model.estms2))
}
rho_info = data.frame(rho = rho_range, r0 = rho_r0_range)

r0_rho = ggplot(rho_info, aes(x = rho, y = r0)) +
  geom_line(size = 1.2, colour = "tomato") + xlab(expression(rho)) + ylab(expression('R'[0]))
  theme_minimal() +
```

```
  theme( axis.title.x = element_text(size=rel(1.2)),
         axis.title.y = element_text(size=rel(1.2)))
#theme(axis.title.y = element_text(angle=0))
#r0_rho


grid.arrange(r0_betaI, r0_betaD, r0_lambda1, r0_lambda2, r0_mu, r0_rho, ncol = 2, nrow = 3)

#### Assessing impact of earlier intervention (changes in tC) -----

#4 plots - 3 monthly changes
tc_range = c(13, 43, 73, 103)
model.estms2 = model.estms
par(mfrow = c(2,2))

for ( i in 1:length(tc_range)){
  model.estms2['tc'] = tc_range[i]
  fit = ode(times = times, y = start, func = seirdb, parms = model.estms2)
  diff_inc = c(NA, fit[2:nrow(fit),8] - fit[1:(nrow(fit)-1),8] )
  #peak = which.max(diff_inc) #which(diff_inc <= 0)[1]
  slow_down = which(diff_inc < 1)[1]
  plot(fit[,1], fit[,8], type = "l", ylim = c(0,12000), xlim = c(0,800),
       main = paste("tc =", tc_range[i]), xlab = "Days", ylab = "Cumulative Cases", col = "tc
  abline(v = tc_range[i], lty = 2, col = "darkblue")
  abline(v = slow_down, lty = 2, col = "darkred")
  print(c(tc_range[i], slow_down, fit[800,8]))
}

# now look at relationship numerically
tc_diff_range = seq(0, 90, 0.25)
cases_avoided = NULL
total_cases = model.f[800,8]

for (i in 1:length(tc_diff_range)){
  model.estms2['tc'] = model.estms['tc'] - tc_diff_range[i]
  fit = ode(times = times, y = start, func = seirdb, parms = model.estms2)
  cases_avoided = c(cases_avoided, (total_cases - fit[800,8]))
}
df.cases_avoided = data.frame("days" = tc_diff_range, "cases" = cases_avoided)

tc_early = ggplot(df.cases_avoided, aes(x = days, y = cases, group = 1)) +
  geom_line(size = 1.2, color = "tomato") + xlab("Days sooner") + ylab("Cases avoided") +
  ggtitle("Liberia") +
  theme_minimal()

tc_early

#### Relationship between 'disease threshold' and duration of the epidemic -----
tc_range2 = seq(30, 103, 1)
model.estms2 = model.estms
threshold_vals = c()
duration_vals = c()
```

```
for (i in 1:length(tc_range2)){
  model.estms2['tc'] = tc_range2[i]
  fit = ode(times = times, y = start, func = seirdb, parms = model.estms2)
  diff_inc = c(NA, fit[2:nrow(fit),8] - fit[1:(nrow(fit)-1),8] )
  #peak = which.max(diff_inc) #which(diff_inc <= 0)[1]
  slow_down = which(diff_inc < 1)[1]

  threshold_vals = c(threshold_vals, fit[tc_range2[i], 8])

  duration_vals = c(duration_vals, (slow_down))
}
thresh_data = data.frame(threshold = threshold_vals, duration = duration_vals)

p_thresh = ggplot(thresh_data, aes(x = threshold, y = duration)) +
  geom_line(size = 1.2, color = "tomato") + xlab("Threshold (number of cases)") + ylab("Dura
  ggtitle( expression( paste("Liberia (", eta, " = 0.502)"))) +
  scale_x_continuous(breaks = seq(300, 6300, 1000)) +
  scale_y_continuous(minor_breaks = seq(0, 800, 25), breaks = seq(200, 800, 50)) +
  theme_minimal() +
  theme(text = element_text(size = 14))
p_thresh


#### Multivariate sensitivity analysis on R0 ------

n = 10000# number to sample
k = 6 #number of variables

lhsp = randomLHS(n, k)

samples = data.frame( betaI = qunif( lhsp[,1], 0.12, 0.2),   #check these two
                      betaD = qunif( lhsp[,2], 0.24, 0.41),     #check
                      lambda1 = qnorm( lhsp[,3], 0.1, 0.0225),
                      lambda2 = qnorm( lhsp[,4], 0.1, 0.0225),
                      mu = qunif( lhsp[,5], 0.2, 0.7),
                      rho = qunif( lhsp[,6], 0.25, 1.5)
)

save.vals = NULL

for (i in 1:n) {
  #run = ode( times = times, y = start, func = seirdb, parms = samples[i,])
  save.vals[i] = r0.calc(samples[i,])
}

save.vals = unlist(save.vals)

save.vals2 = data.frame(values = save.vals)
ggplot(save.vals2, aes(values)) +
  geom_histogram(binwidth = 0.2, fill = "tomato", color = "grey45", alpha = 0.5) +
  xlab(expression("R"[0])) + ylab("Frequency") +
  ggtitle(expression(paste("Uncertainty analysis of ", "R"[0], " for Liberia"))) +
  theme_minimal()
```

```
t.test(save.vals)
summary(save.vals)

#hist(save.vals, main = expression(paste("Uncertainty analysis of ", "R"[0], " for Sierra Le

pcor_vals = c()

for ( i in 1:k){
  vals = pcor.test(x = samples[,i], y = save.vals, z = samples[,-i], method = "spearman")
  pcor_vals = c(pcor_vals, vals$estimate)
}

pcor_vals
pcor_vals = setNames(pcor_vals, colnames(samples))
barplot(pcor_vals, ylim = c(-1, 1))



#### Univariate sensitivity analysis on cumulative cases ------

# Using LHS
n = 10000 # number to sample
k = 9 #number of variables

lhsp = randomLHS(n, k)

samples = data.frame( alpha = qnorm( lhsp[,1], 0.1, 0.0125),
                      lambda1 = qnorm( lhsp[,2], 0.1, 0.0225),
                      lambda2 = qnorm( lhsp[,3], 0.1, 0.0225),
                      betaI = qunif( lhsp[,4], 0.12, 0.22),
                      betaD = qunif( lhsp[,5], 0.24, 0.42),
                      rho = qunif( lhsp[,6], 0.25, 1.5),
                      tc = qunif( lhsp[,7], 73, 133),
                      eta = qunif( lhsp[,8], 0.3, 0.7),
                      mu = qunif( lhsp[,9], 0.2, 0.7)
                      #E0 = qunif( lhsp[,9], 30, 70),
                      #I0 = qunif( lhsp[,10], 10, 40),
                      #D0 = qunif( lhsp[,11], 10, 30)
)

save.vals50 = NULL
save.vals200 = NULL
save.vals400 = NULL
save.vals800 = NULL

for (i in 1:n) {
  run = ode( times = times, y = start, func = seirdb, parms = samples[i,])
  save.vals50[i] = run[50, 8]
  save.vals200[i] = run[200,8]
  save.vals400[i] = run[400,8]
  save.vals800[i] = run[800,8]
}
```

```
save.vals50 = unlist(save.vals50)
save.vals200 = unlist(save.vals200)
save.vals400 = unlist(save.vals400)
save.vals800 = unlist(save.vals800)
#hist(save.vals, main = expression(paste("Uncertainty analysis of ", "R"[0], " for Liberia"))

pcor_vals50 = c()
pcor_vals200 = c()
pcor_vals400 = c()
pcor_vals800 = c()

for ( i in 1:k){
  vals50 = pcor.test(x = samples[,i], y = save.vals50, z = samples[,-i], method = "spearman"
  vals200 = pcor.test(x = samples[,i], y = save.vals200, z = samples[,-i], method = "spearman
  vals400 = pcor.test(x = samples[,i], y = save.vals400, z = samples[,-i], method = "spearman
  vals800 = pcor.test(x = samples[,i], y = save.vals800, z = samples[,-i], method = "spearman
  pcor_vals50 = c(pcor_vals50, vals50$estimate)
  pcor_vals200 = c(pcor_vals200, vals200$estimate)
  pcor_vals400 = c(pcor_vals400, vals400$estimate)
  pcor_vals800 = c(pcor_vals800, vals800$estimate)
}

#pcor_vals
pcor_vals50 = setNames(pcor_vals50, colnames(samples))
pcor_vals200 = setNames(pcor_vals200, colnames(samples))
pcor_vals400 = setNames(pcor_vals400, colnames(samples))
pcor_vals800 = setNames(pcor_vals800, colnames(samples))

vals800 = data.frame(pcor_vals800)
setDT(vals800, keep.rownames = TRUE)[]
colnames(vals800) = c("variable", "value")

# try do with ggplot
#need to put into long format
sens_bar800 = ggplot(vals800, aes(x = variable, y = value)) +
  geom_bar(stat = "identity", fill = "tomato", col = "grey45", alpha = 0.5) +
  ylab("PRCC") + xlab("Parameter") + ggtitle("Sensitivity analysis of cumulative cases at day
  theme_minimal()

sens_bar800

#other 4 plots
vals50 = data.frame(pcor_vals50)
setDT(vals50, keep.rownames = TRUE)[]
colnames(vals50) = c("variable", "value")

sens_bar50 = ggplot(vals50, aes(x = variable, y = value)) +
  geom_bar(stat = "identity", fill = "tomato", col = "grey45", alpha = 0.5) +
  ylab("PRCC") + xlab("Parameter") + ggtitle("Day 50") +
  theme_minimal() +
  theme(axis.text.x = element_text(size = 7.5))
```

```
#sens_bar50

vals200 = data.frame(pcor_vals200)
setDT(vals200, keep.rownames = TRUE)[]
colnames(vals200) = c("variable", "value")

sens_bar200 = ggplot(vals200, aes(x = variable, y = value)) +
  geom_bar(stat = "identity", fill = "tomato", col = "grey45", alpha = 0.5) +
  ylab("PRCC") + xlab("Parameter") + ggtitle("Day 200") +
  theme_minimal() +
  theme(axis.text.x = element_text(size = 7.5))

#sens_bar200

vals400 = data.frame(pcor_vals400)
setDT(vals400, keep.rownames = TRUE)[]
colnames(vals400) = c("variable", "value")

sens_bar400 = ggplot(vals400, aes(x = variable, y = value)) +
  geom_bar(stat = "identity", fill = "tomato", col = "grey45", alpha = 0.5) +
  ylab("PRCC") + xlab("Parameter") + ggtitle("Day 400") +
  theme_minimal() +
  theme(axis.text.x = element_text(size = 7.5))


sens_bar800.v2 = ggplot(vals800, aes(x = variable, y = value)) +
  geom_bar(stat = "identity", fill = "tomato", col = "grey45", alpha = 0.5) +
  ylab("PRCC") + xlab("Parameter") + ggtitle("Day 800") +
  theme_minimal() +
  theme(axis.text.x = element_text(size = 7.5))


grid.arrange(sens_bar50, sens_bar200, sens_bar400, sens_bar800.v2, nrow = 2, ncol = 2)
```