

MATHEMATICAL MODELING OF THE 2014/2015 EBOLA EPIDEMIC IN WEST AFRICA

JEFF BARTLETT*, JAMES DEVINNEY†, AND ERIC PUDLOWSKI‡

Faculty mentor: WOLFGANG BANGERTH¹

Abstract. Accurately predicting the future number of infected patients and deaths during an epidemic is an important step towards efficiently allocating resources – doctors, nurses, hospital beds, public awareness campaigns, or foreign aid – in combating a disease that spreads through a community. In this paper, we develop mathematical models for the Ebola epidemic that started in West Africa in 2014 and that has infected more than 20,000 people so far.

To this end, we create a discrete time, age structured model using assumptions based on publicly available data from Sierra Leone and West Africa. We show that with reasonable assumptions, we provide an accurate fit to the reported number of infections and deaths due to Ebola in Sierra Leone. The close fit to past data provides hope that the model can also serve as a prediction for the future and we verify this in an appendix for data that became available after we created our model.

AMS subject classifications. 97M60, 97M10, 34K60

1. Introduction. On March 25th, 2014 the World Health Organization (WHO) reported an outbreak of Ebola fever in the southeastern part of Guinea in West Africa [7]. Prior to this announcement, Guinea had experienced 86 suspected cases and a total of 59 deaths linked to an unidentified virus. The French branch of Doctors without Borders identified Zaire Ebola virus as the cause of these deaths, a strain of the Ebola virus characterized by a hemorrhagic fever and high mortality rate. Within the next few days, additional cases spread to Liberia and Sierra Leone, a total of 26 confirmed as Ebola. This number rose to 100 cases in a month, and in 4 months to 909 confirmed cases – more than any other Ebola outbreak ever before. At the time of writing, there were almost 25,000 reported cases of infection, and almost 10,000 deaths; both numbers are probably significant underestimates of reality.

While the infection has remained contained almost exclusively to the three countries mentioned above, a report at the end of August 2014 by the Centers for Disease Control and Prevention (CDC) in the United States confirmed Ebola cases in 13 countries [7]. Some of these countries saw transmission within their borders, but ultimately all controlled their cases through rigorous quarantine and tracking of all individuals who may have come into contact with patients. Yet, this expansion into several outside countries (such as the United States) supported the public perception that the virus was leaving West Africa, and the epidemic quickly received significant global media attention.

For infections that spread to such large numbers of people, mathematical models can be used to predict the further progression of an epidemic. This provides insight into both the number of infected patients one can expect in the future, as well as how many of those will in fact die. Such models are useful because they can help

*Department of Electrical Engineering, Texas A&M University, College Station, TX 77843-3368, USA (jbartlett12026@tamu.edu).

†Department of Mathematics, Texas A&M University, College Station, TX 77843-3368, USA (jdevinney11@tamu.edu).

‡Department of Mathematics, Texas A&M University, College Station, TX 77843-3368, USA (Eric.Pudlowski@gmail.com).

¹Department of Mathematics, Texas A&M University, College Station, TX 77843-3368, USA (bangerth@math.tamu.edu).

allocate resources – doctors, medication, mobile clinics, money, or international aid – in anticipation of what is expected to happen and where it will have the biggest impact. Since almost all decisions in health care, in particular in developing countries like those in West Africa, are made in an environment of scarcity, making the best of available resources based on models is clearly useful. Furthermore, models allow us to identify *targets for interventions*. For example, epidemiological models have parameters for transmission rates from one patient to uninfected others, with an associated mortality rate. A model then allows us to ask questions such as “By how much do we have to reduce transmission rates to stop the epidemic?” or “By how much do we have to improve patient care to limit the total number of deaths to X ?” The answer to such questions can then again be matched with available resources: Do we spend the available money on building better clinics, on isolation wards, or on public awareness campaigns through local media?

This paper presents a mathematical model for the spread of Ebola in West Africa during the 2014/2015 epidemic. The model is based on publicly available data concerning the Ebola virus and its biological nature. Parameters describing human interaction and transmission coincide with the World Health Organization’s data about patient and death numbers during the first several months of the epidemic. We show that our model can be used to accurately track reported patient numbers and deaths; that it identifies a change in how the affected countries responded to the disease in late 2014; and how it can be used to define targets where interventions are necessary to bring the epidemic under control. In an appendix, we also show that our model’s predictions beyond the data points we used in its development, closely match actual case numbers encountered since then.

The remainder of the paper is organized as follows: In Section 2, we outline the biological and epidemiological background of Ebola. Section 3 introduces the “traditional” approach towards modeling infectious diseases and its shortfalls. We then introduce our age structured, discrete-time model in Section 4, along with its parameters and our estimates for these parameters. We compare this model against available data in Section 5. It will turn out that the model fits the initial stages of the epidemic well, but deviates significantly beyond November 2014. Sections 6 and 7 therefore introduce a model that incorporates changes in individual behavior and compares the resulting predictions with data, respectively. Section 8 provides conclusions. An appendix performs a retrospective analysis with the data that has become available after we finished developing our models and fitting its parameters.

2. Ebola: Motivation and Background. Ebola has garnered public interest mostly because of its low survival rate compared with most other modern diseases (between 30 and 70% of infected patients die within 2–3 weeks [23]) and graphic symptoms: patients start to bleed both internally and externally, including bloody stool and vomit. In many cases, the loss of fluids ultimately leads to death. Unlike other diseases endemic in West Africa such as Malaria or Yellow Fever that require a vector such as mosquitoes, Ebola is also readily spread from person to person: the CDC states that the Ebola virus is spread through “direct contact with the blood or body fluids” such as saliva, sweat, semen, and mucus [19]. Ebola is referred to as a “stable virus” with a constant and low mutation rate [15, 19], i.e., it is not likely to mutate into a disease that is more contagious (for example, airborne). However, the virus changes enough to reinfect those who have already recovered from the virus [19]. For the purposes of our model, this means that those who have recovered from Ebola will eventually become susceptible to getting it again.

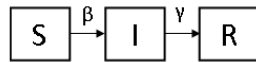


FIG. 3.1. *A traditional SIR Model*

In our mathematical model, we strive to accurately represent the clinical progression of an Ebola infection through incubation, infectiousness, and recovery or death. When a person contracts Ebola, they are in the incubation stage until the virus has built up enough to make them infectious. Once infectious, they are able to pass on the virus to anyone that comes into contact with their bodily fluids. Both the incubation and infectious periods are estimated by the CDC to be 6 days [21], though the model we will present below will use a more refined characterization of these days. Ebola kills roughly half the people it infects, and if after the average six day period of infectiousness the patient has survived Ebola, they start entering a period of recovery [6]. Recovery can take any period of time and varies on the immune system of the victim. The minimum recovery time is 21 days, and patients are kept for at least this amount of time after an initial suspicion of infection [6].

When modeling a disease such as Ebola, details such as the culture and tradition cannot be overlooked. Many West African societies follow a custom for funerals that involves a very personal encounter with the diseased corpses. Many times before the funeral, the sister of the deceased’s father is responsible for cleaning, bathing, and dressing the body [9]. The family members of the dead, after the ritual ceremony, will wash hands in a common bowl and either touch or kiss the face of the corpse [24]. Given that patients often show bleeding, vomiting and diarrhea, and that all of these body fluids are infectious, this practice clearly poses a health risk and has contributed to the spread of the disease; at the same time, it is a major target for intervention through public awareness campaigns to reduce the infection rate.

Such interventions also need to be cognizant of local culture. If the bodies of those who have perished of Ebola are left to be handled by professionals in hazmat suits with limited family contact, then this implies a separate and less meaningful burial than what many West Africans are accustomed to. There have even been reports of families breaking into contained morgues to retrieve the bodies of their dead relatives [20]. All opportunities for interventions are therefore constrained to what is acceptable in terms of regional culture.

Our model will not attempt to directly describe these human interactions; instead it will use statistical assumptions about how many other people each infected patient infects, and it is through this parameter that we can model changes in behavior.

3. “Traditional” mathematical epidemiology models. The “traditional” tools of mathematical epidemiology are a class of models that are commonly referred to as “SIR” since they categorize the overall population into three categories: **S**usceptible (don’t have the disease, but have potential of getting it), **I**nfected (infected with the disease), and **R**ecovering (don’t have the disease and don’t have the potential of getting it again). A graphical depiction of the simplest such model is shown in Fig. 3.1.

SIR models, at their core, model these interacting populations over time. The basic SIR model works by moving a number of people from the susceptible population into the infected population at a rate called the infection rate β . The infection rate is typically determined as a percentage of the people in the susceptible class times

the number of already infected people. Similarly, people move from the infected population into the recovered population at the recovery rate γ . The simplest SIR model, shown in Fig. 5.1, is also known as the Kermack-McKendrick Model.

SIR-type models can vary in complexity. Many can include traits like incubation periods, vaccines, the ability to return to the susceptible class, etc. They do so by adding additional classes of people (e.g., infected but not yet infectious) and by adding more arcs between classes (e.g., from recovered to susceptible), or allow classes to lose population due to death or gain population through births. An early example of application to Ebola can be found in [2], and SIR-type models for the Ebola outbreaks in Congo (1995) and Uganda (2000) are described in [12].

Despite their conceptual simplicity and their success in describing data, SIR models also have drawbacks. Among these is that the parameters in these models are often difficult to interpret and correlate with mechanistic descriptions of how a disease progresses. For example, SIR models do not have a coefficient that would indicate how many secondary infections each patient causes on average; rather, it has a parameter that indicates the *fraction* of the overall (susceptible) population each patient infects, even in cases where it is clear that patients can not possibly interact with the entire population or where the patient population is so small compared to the susceptible population that the size of the susceptible population should be of no relevance to the model. Secondly, the basic construction of all SIR models is through systems of ordinary differential equations in which certain *fractions* of the population of one class moves into another class each unit time. Consequently, a population can never completely die out, even if there is no further inflow; it can only exponentially decay. This is at odds with the actual progression of non-chronic diseases such as Ebola in which the immune system completely clears the body of viruses after a certain time. Consequently, if there are no new infections, we expect the number of infected persons to be zero after some time, not just a small and exponentially decaying number that could still infect others. Third, SIR models assume the presence of a finite susceptible population. This is useful in cases such as the flu where a significant fraction of the population gets infected, recovers, and is then protected from further infection (i.e., has moved to the recovered population). However, in cases such as the West Africa Ebola epidemic, the number of infected people is (and will likely remain) a vanishingly small fraction of the overall population. It is therefore neither necessary nor useful to explicitly track the size of the susceptible population since it is essentially constant and, furthermore, essentially infinite.

It is for these reasons that we desired to derive a more realistic age-structured model rather than the original SIR model or one of its variations. The model derived in the next section will not explicitly track the size of the susceptible population, and we will strive to define it in terms of parameters that can be explicitly correlated with epidemiologically traceable events such as one patient infecting another person.

4. An improved, age structured model.

4.1. The basic model. Given the observation at the end of the previous section, we decided to utilize a discrete-time, age structured model instead. Here, “discrete-time” means that we only consider populations in each class on a day-by-day basis, rather than the continuous time models implied by differential equations. “Age structured” means that we will consider separate classes for people who are 1, 2, . . . days into an infection. This allows us to easily model that patients do not show the same infectiousness throughout their disease, and that the death rate is also different depending on the stage of the disease. Furthermore, since each patient moves from the,

for example, “infected 10 days ago” either to the “infected 11 days ago” class or the “dead” class, over the course of one day, and that we only track patients up to their 21st day of infection, we obtain a model that after 21 days of no new infections has no patients left – as desired. In contrast to the SIR model, we do not separately track susceptible and recovered classes; this is appropriate in the current case since the total number of infections (a few ten thousands) is so small compared to the total population of the three West African countries that the susceptible class always contains the entire population for all practical purposes, and the recovered class is negligible for all practical purposes.

We can describe a model that satisfies these requirements through the following set of equations where $I_k(t)$ denotes the number of people who are k days into their disease on day t (e.g., counted from June 4th, 2014, or any other arbitrarily chosen date):

$$I_1(t + 1) = \sum_{k=1}^{21} \beta_k I_k(t), \quad (4.1)$$

$$I_2(t + 1) = (1 - \delta_1) I_1(t), \quad (4.2)$$

$$I_3(t + 1) = (1 - \delta_2) I_2(t), \quad (4.3)$$

⋮

⋮

$$I_{21}(t + 1) = (1 - \delta_{20}) I_{20}(t), \quad (4.4)$$

$$D(t + 1) = D(t) + \sum_{i=k}^{21} \delta_k I_k(t), \quad (4.5)$$

$$C(t + 1) = C(t) + I_1(t). \quad (4.6)$$

Equation (4.1) describes the number of people who are newly infected on day $t + 1$. Since they must have been infected by someone else who was already infected on the previous day t , this equation involves all of the variables $I_k(t)$. The numbers β_k indicate the number of susceptible people each patient infects on day k of her disease.

Equations (4.2) through (4.4) describe that the number of infected people $k + 1$ days into their infection on day $t + 1$ equals the number of people k days into their infection on day t , minus that fraction δ_k of patients who die on their k th day into the infection.

The last two equations, (4.5) and (4.6), track the overall number of deaths and infection cases. In particular, the total number of deaths up to day $t + 1$, $D(t + 1)$, equals the total number of deaths up to day t , $D(t)$, plus the fraction δ_k of those patients who died on the k th day of their infection, accumulated over all 21 days on which we track these populations. Likewise, the total number of cases up to day $t + 1$, $C(t + 1)$ equals the total number of cases up to day t , $C(t)$, plus the number of new infections $I_1(t)$ encountered the previous day. We track $D(t)$, $C(t)$ because these numbers are reported by the WHO in their periodic briefings on the epidemic, and they allow us to compare the results of our models with reality.

Before discussing numerical values for the various parameters in this model, it is worth pointing out that δ_k denotes the *fraction* of people who die on day k of their infection. These fractions δ_k are of course between zero and one. Further, ours is a *deterministic model* since it predicts a precise number of people in each of the categories

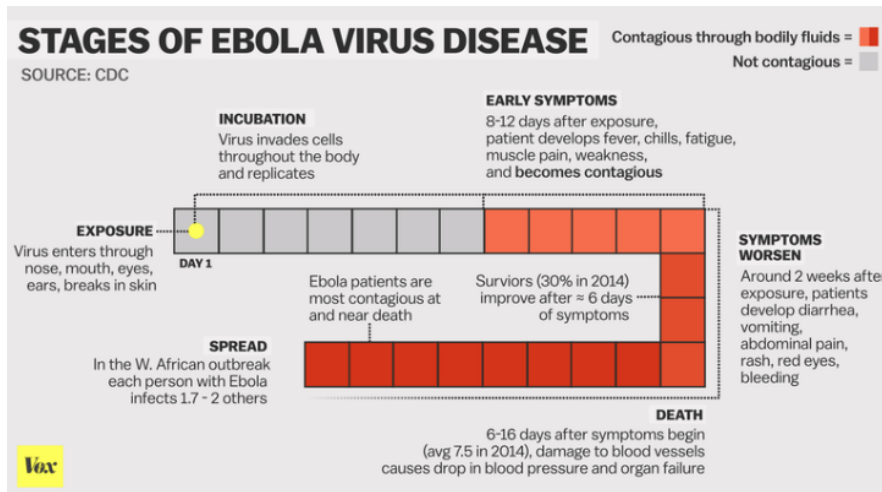


FIG. 4.1. *Stages of Ebola Virus (reproduced from [3]).*

for each day if the various parameters are known, but the parameters themselves are *statistical averages*: obviously, not every patient infects the same number of others, and no fractions of individuals die. Rather, the constants β_k and δ_k are averages over entire populations and it is clear that a model such as this can only be accurate if the number of patients and deaths is large enough to justify considering average outcomes. This is clearly the case for the situation we consider here.

4.2. Parameters in the model. For concrete predictions from this model, we need values for the various parameters β_k and δ_k as well as an initial number of infections. The values we use are informed by sources such as the CDC [14] and were then adjusted in ways that allowed us to fit the number of infections and deaths.

The model we have drawn up in the previous section has a large number of parameters – too many for practical purposes. We therefore need to strive to use insights about the disease progression to reduce the many parameters to a smaller number that are truly independent. To this end, we consider sources such as the illustration on the “Stages of Ebola” published by the CDC and shown in Figure 4.1. There, the CDC outlays the expected day-by-day breakdown of an Ebola infection. It suggests a length of incubation of 23 days. However we will go with 21 days as the CDC itself uses that number to model an Ebola outbreak Epidemic [5]. A 21 day period is also used by the World Health Organization.

Figure 4.1 suggests an incubation period before the symptoms begin and the carrier of the virus becomes contagious. One expects that after a number of days of showing symptoms, the viral load in a patient decreases again (see, for example, [8]) but this does not imply that patients immediately get better again or become less infectious, since the health effects of the virus are cumulative over a number of days. Consequently, we can reasonably assume that patient’s mortality and infectiousness continues to increase throughout the 21-day period.

From using the above considerations and those suggested by Figure 4.1, we decided to lump the infection and death parameters of our model into four categories each of sizes similar to those in the figure. Specifically, we assume that the variables β_k are constant between days $k = 1$ and 8, between days 9 and 13, between days 14 and

16, and between days 17 and 21. For the δ_k , we choose the intervals as $k = 1, \dots, 10$, $k = 11, \dots, 13$, $k = 14, \dots, 16$, and $k = 17, \dots, 21$. We choose the first group longer for the death coefficients δ_k based on the assumption that patients do not die immediately as they become infectious to others.² In summary, these assumptions are compatible with the general progression of the disease through four phases, as shown in Fig. 4.1, and reduce the number of parameters we have to determine from 63 to 9 (four groups each for β_k, δ_k , plus the initial population of diseased patients) – a much more manageable number not as prone to overfitting as the original model.

Using this lower-dimensional model, we fix the initial time of our model by choosing $t = 0$ to be the date June 4, 2014. We can then determine the values of the 9 independent parameters in our model by using a least squares fit to the actual data shown in the next section. Using the data points for the first 160 days of reported cases and deaths in Sierra Leone, i.e., until November 11, 2014, our fit yields the following parameters for the infection coefficients:

$$\begin{array}{lll}
 \beta_1 = 0 & \beta_2 = 0 & \beta_3 = 0 \\
 \beta_4 = 0 & \beta_5 = 0 & \beta_6 = 0 \\
 \beta_7 = 0 & \beta_8 = 0 & \beta_9 = 0.092 \\
 \beta_{10} = 0.092 & \beta_{11} = 0.092 & \beta_{12} = 0.092 \\
 \beta_{13} = 0.092 & \beta_{14} = 0.122 & \beta_{15} = 0.122 \\
 \beta_{16} = 0.122 & \beta_{17} = 0.153 & \beta_{18} = 0.153 \\
 \beta_{19} = 0.153 & \beta_{20} = 0.153 & \beta_{21} = 0.153.
 \end{array}$$

The corresponding values for the death coefficients are then

$$\begin{array}{lll}
 \delta_1 = 0 & \delta_2 = 0 & \delta_3 = 0 \\
 \delta_4 = 0 & \delta_5 = 0 & \delta_6 = 0 \\
 \delta_7 = 0 & \delta_8 = 0 & \delta_9 = 0 \\
 \delta_{10} = 0 & \delta_{11} = 0.0128 & \delta_{12} = 0.0128 \\
 \delta_{13} = 0.0128 & \delta_{14} = 0.0231 & \delta_{15} = 0.0231 \\
 \delta_{16} = 0.0231 & \delta_{17} = 0.032 & \delta_{18} = 0.032 \\
 \delta_{19} = 0.032 & \delta_{20} = 0.032 & \delta_{21} = 0.032.
 \end{array}$$

Our model is completed by stating that the infection started with 5.6 people in each of the classes $I_1(0), \dots, I_9(0)$, also determined by the least squares procedure, with the remaining categories $I_k, k = 10, \dots, 21$ empty. We also chose $D(0) = C(0) = 0$. We will be using these values to model the epidemic only in Sierra Leone, though small variations (such as the start date and initial numbers of patients) would also allow adaptation of the model to the other countries.

Using the above information, we can compute the number of people every infected patient infects himself, a number often referenced as the *reproductive number* or R_0 , and used to describe how easily a disease spreads [17]. In particular, if $R_0 > 1$, then the epidemic expands, whereas if $R_0 < 1$ then it will die out naturally. The goal of controlling an epidemic is to bring R_0 below one.

In the current context, we can compute it as follows: On the first day of their infection, every patient infects β_1 others; the probability of survival to day two of

²In reality, the length of any of these groups is certainly debatable as patient histories are highly variable. See, for example, [25, page 91].

TABLE 4.1
 R_0 for various common diseases (reproduced from [17]).

Disease	R_0
AIDS/HIV	2–5
Smallpox	3–5
Measles	16–18
Malaria	> 100

their infection is $(1 - \delta_1)$, where they infect another β_2 others; probability of survival to day three is $(1 - \delta_1)(1 - \delta_2)$ where they infect β_3 others; and so on. In formulas, the total number of infections from each patient is then, statistically speaking,

$$R_0 = \beta_1 + \beta_2(1 - \delta_1) + \beta_3(1 - \delta_1)(1 - \delta_2) + \dots = \sum_{k=1}^{21} \beta_k \prod_{j=1}^{k-1} (1 - \delta_j).$$

With our parameter values, we arrive at $R_0 = 1.47$. This value can be compared to that of other common diseases (see Table 4.1) that, by and large, have larger R_0 and consequently have historically also infected much larger numbers of patients. The value we compute for Ebola’s R_0 is roughly consistent with that reported in the literature. For example, Althaus [1] estimates $R_0 \approx 2.52$ for the very early phases of the epidemic in Sierra Leone; on the other hand, Hunt [16] estimates $R_0 \approx 1.51$. (Hunt notes that in the earliest phases of the epidemic, significantly before August 2014, his method estimates $R_0 \approx 2.26$, indicating that the value has dropped over time and thereby resolving the discrepancy between the results of Althaus and Hunt.)³

In the same way as above we can compute the likelihood with which a patient dies from Ebola, called the mortality M . Clearly, this likelihood is the sum over days k of the probability of the event characterized of (i) remaining in the infected class until day k and (ii) dying on day k . Using similar reasoning as above, this leads to the following expression:

$$M = \sum_{k=1}^{21} \delta_k \prod_{j=1}^{k-1} (1 - \delta_j).$$

With the values we have chosen, this formula equates to $M = 0.24$. This value seems at odds with reported data that report the total number of deaths to be approximately 40% of cases. (For example, the number of reported cases provided by the CDC in the area of Sierra Leone, Liberia, and Guinea total 23,539, with the number of deaths at 9,541 [10]. This gives a death rate of approximately 40%.) However, the definition for M we use above is the likelihood for any infected patient to die (accumulated over the entire infection history of the patient), whereas the reported numbers introduce various time delays into the calculation: for example, in the best situation, cases and deaths are reported immediately, but because the death of a patient typically trails the infection by several days, death cases should not be compared with patient numbers of the same day, but with those of a few days earlier. In practice, of course, reports are often incomplete and delayed, and sometimes erroneous (as is

³For comparison, Chowell *et al.* [12] estimate $R_0 \approx 1.83$ during the initial, exponential-growth phase of the earlier 1995 Ebola outbreak in Congo, and $R_0 \approx 1.34$ for an outbreak in Uganda in 2000, in both cases before any interventions are put in place.

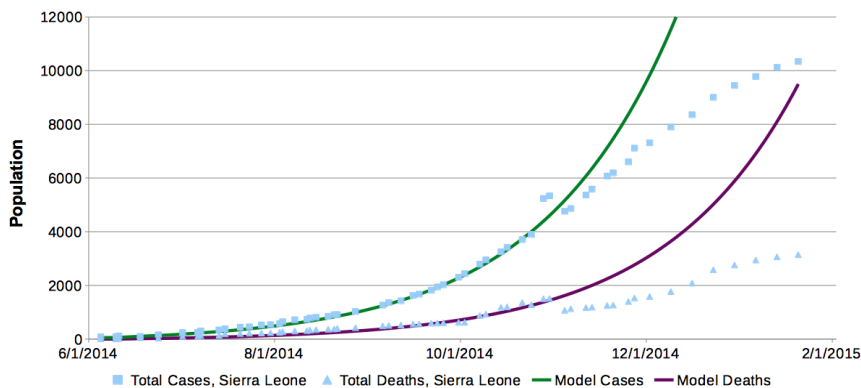


FIG. 5.1. Model fitted to the first 160 days and actual infected people and deaths over time.

clearly seen in the data we present below where the number of cumulative cases since the epidemic started occasionally *drops* compared to the previous week – reflecting a retroactive correction of data). Finally, the data reported is not that of patients and deaths during a given day, but accumulated over time; taking the ratio of these two numbers would then, at best, provide an *average* value for M , not an *instantaneous* one as used in our definition above. In any case, while these differing definitions of mortality prevent us from completely reconciling the value for M we compute above with those discussed in the literature, the ultimate success of our model does not hinge on computing this one value, but in accurately modeling the number of cases and deaths over time, as will be shown in the next sections.

5. Comparing predictions with reality in Sierra Leone. The quality of a model lies in whether it fits reality. Using the model we have outlined above along with the parameters and initial conditions discussed in the previous section, we can predict cumulative case numbers $C(t)$ and cumulative deaths $D(t)$ as a function of time, for the country of Sierra Leone. We show these in Fig. 5.1 along with the actual numbers provided by the CDC [4]. As can be seen, our model provides a quite reasonable fit to the data up until late October or early November 2014, i.e., for the first 160 days from June 4, 2014 ($t = 0$) for which we used data to fit.

After this time, the actual data shows a slow-down of the number of cases and deaths. This is generally attributed to changes in behavior in the affected communities that reduced the likelihood of secondary infections. This prompted us to limit the least squares fit for the parameters shown above to the first 160 days. We will discuss the causes and how this change in the epidemic can be incorporated into our model in the following sections. However, before moving on, we would like to consider what our model predicts what would have happened without the interventions that changed people’s behavior. Fig. 5.2 shows predictions from our model to one year after its start time. It suggests that, if the epidemic had remained unaddressed, there would have been more than 600,000 cases and more than 150,000 deaths by June of 2015. We know now that this will not likely happen, but predictions of similar magnitude could nevertheless be found in both the general press and in scientific publications [5, 13, 16, 18, 21].

6. Changing the course of an epidemic. With sufficient resources and well trained personnel, Ebola can be contained. For example, when Ebola cases started to

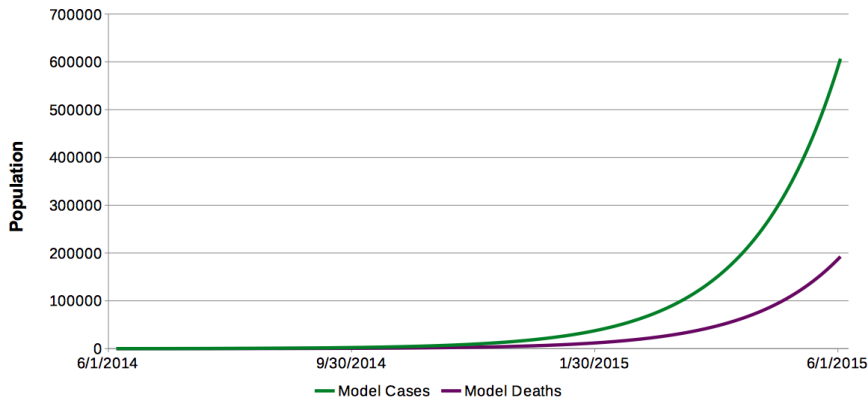


FIG. 5.2. Predictions of the model fitted to the first 160 days model over the course of 1 year.

appear in the United States and Europe, authorities were able to quickly corral it by determining who the few infected patients had come into contact with in order to prevent further infections by quarantining and monitoring suspected contacts. Patients also received care at a level that spared little expense.

This level of attention to every single patient and their contacts is not feasible in West Africa, despite the fact that since the WHO declared the Ebola outbreak an epidemic, the CDC, the U.S. military, other foreign governments, a plethora of non-governmental organizations, and the United Nations have set up treatment units, labs, and have facilitated the mobilization of the general population.

In order to avoid further progression of an epidemic, governments and organizations aiding in recovery must therefore work at a level that affects the *statistical probabilities* of spreading the disease or dying from it, rather than the *individual outcomes* of patients. Typical means to this end include advertising safe health practices and implementing guidelines and policies on dealing with the disease. Indeed, *information* is often the most valuable tool in dealing with epidemics.

Ebola can primarily be controlled by lowering the contact rate between infected and healthy individuals, for example through quarantine. There must be close interaction with a diseased body to induce infection; the virus is spread by bodily fluids such as mucus, semen, saliva, etc. Simple tactics such as covering one’s mouth when coughing or sneezing, washing hands, and avoiding unnecessary exchange of bodily fluids can ensure a slowed spread of disease. With safer funeral practices that respect ritualistic practices, the contact rate from dead bodies can be lowered. All of these behaviors can be indirectly affected by public awareness campaigns, within the cultural limits of a society as discussed above (an example of a public awareness poster produced by the CDC is shown in Fig. 6.1). They can also be directly affected; for example, by establishing over 140 teams of “burial expert”, and laying out a 17 page guideline for culturally respective burials for Ebola victims, one can expect to see a drop in Ebola related deaths [26]. As for live contact rate, policies as easy as informing people not to go near the sick or those who have had contact with Ebola can help. All of these factors will affect the β_k parameters we have used in our model that describe secondary infections. Finally, medical advances and better hospital treatment can help to reduce the fractions δ_k of those who die from the virus.

At the same time, obstacles to such interventions are numerous in developing countries. For example, in Sierra Leone, the literacy rate was 35 percent in 2013 and

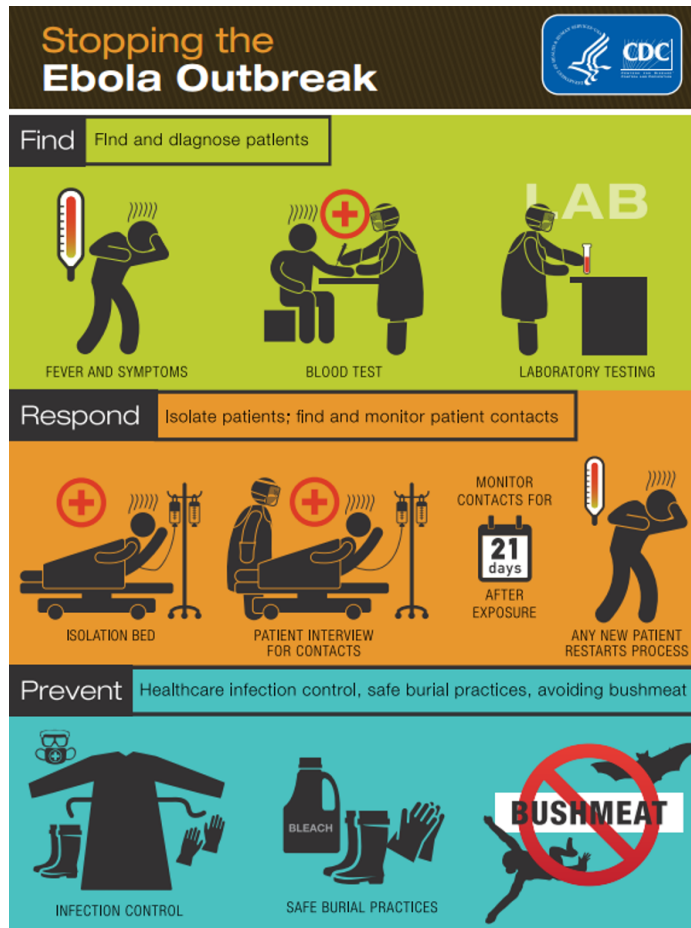


FIG. 6.1. Intervention poster used by CDC [6].

less than half the population was urbanized [11]. This means the population is both harder to reach and communicate with than in developed countries where medical warnings reach a large majority of the population and where, generally speaking, governments have a track record that allows the population to trust such government warnings.

It is generally understood that during the current epidemic, governments and international organizations required time to understand and analyze the situation before large-scale efforts to stop the epidemic were initiated. Consequently, the number of cases increased largely unchecked for the first several months and relief efforts ramped up only in the fall of 2014. The effects of this could be seen with the change in behavior of the epidemic from unchecked growth to a tapering off, as shown in Fig. 5.1. We will account for this in our model in the next section.

7. A model that accounts for interventions. As mentioned above, the model predictions shown in Fig. 5.1 conform well to the data for the months July through November 2014, i.e., for the data points we used in our fit. However, after that there appears to be a slow-down of infections in the data reported by the CDC but not reflected in our model. We attribute this to changes in behavior resulting from

interventions such as those discussed in the previous section and that reduced the rate of infection.

In reality, such changes in behavior do not happen overnight. However, for simplicity and in the absence of more concrete data on the efficacy of the public awareness campaigns and changes in medical procedures, let us assume that indeed the infection rates dropped instantaneously on the 160th day (November 11, 2014) and investigate how this affects our model predictions. To this end, we use the final values $I_k(160), C(160), D(160)$ of the model discussed above as the initial values for a model with the same structure but different values of the parameters for the time after $t = 160$.

We can fit this second model to the data points after that day and obtain a different set of parameters:

$$\begin{array}{lll}
 \beta_1 = 0 & \beta_2 = 0 & \beta_3 = 0 \\
 \beta_4 = 0 & \beta_5 = 0 & \beta_6 = 0 \\
 \beta_7 = 0 & \beta_8 = 0 & \beta_9 = 0.0412 \\
 \beta_{10} = 0.0412 & \beta_{11} = 0.0412 & \beta_{12} = 0.0412 \\
 \beta_{13} = 0.0412 & \beta_{14} = 0.0623 & \beta_{15} = 0.0623 \\
 \beta_{16} = 0.0623 & \beta_{17} = 0.0902 & \beta_{18} = 0.0902 \\
 \beta_{19} = 0.0902 & \beta_{20} = 0.0902 & \beta_{21} = 0.0902
 \end{array}$$

and

$$\begin{array}{lll}
 \delta_1 = 0 & \delta_2 = 0 & \delta_3 = 0 \\
 \delta_4 = 0 & \delta_5 = 0 & \delta_6 = 0 \\
 \delta_7 = 0 & \delta_8 = 0 & \delta_9 = 0 \\
 \delta_{10} = 0 & \delta_{11} = 0.0025 & \delta_{12} = 0.0025 \\
 \delta_{13} = 0.0025 & \delta_{14} = 0.010 & \delta_{15} = 0.010 \\
 \delta_{16} = 0.010 & \delta_{17} = 0.020 & \delta_{18} = 0.020 \\
 \delta_{19} = 0.020 & \delta_{20} = 0.020 & \delta_{21} = 0.020
 \end{array}$$

Just like before, the values were determined by a least squares fit to the actual data after day 160.

Using these parameters, we can recompute the value for R_0 and obtain $R_0 = 0.82$ and $M = 0.13$ after day 160. As discussed above, a value of R_0 below one implies that the epidemic will eventually die down, and this is of course what we see in the data. This can also be compared with reproductive numbers R_0 after interventions started in other Ebola outbreaks; for example, [12] estimate that after interventions started, R_0 dropped to 0.51 in the 1995 Congo outbreak, and 0.66 for the 2000 Uganda epidemic. Our value is larger but can be explained by the fact that the Congo and Uganda outbreaks were much smaller and consequently much more resources per affected person could be brought to bear on containing the epidemic than in late 2014 in West Africa.

With the values stated above, we can re-run the model to obtain predictions beyond the original 160 days and, in particular, about the future. Results are shown in Fig. 7.1. We see that, compared to the original model in Fig. 5.1, the model now fits the data well not only until October or November 2014, but in fact for the entire time

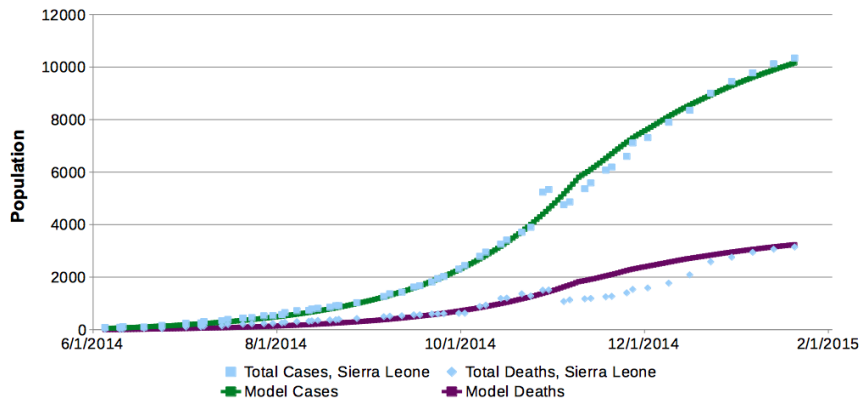


FIG. 7.1. Composite model and actual infected people and deaths over time.

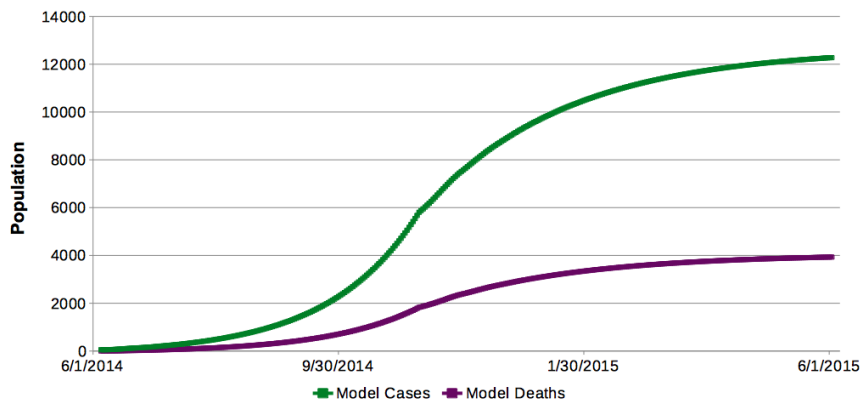


FIG. 7.2. Predictions by the composite model over the course of 1 year.

period for which we have data reported by the CDC. Similarly, running our model for an entire year (see Fig. 7.2), we can see the cumulative number of infections leveling off over the course of 2015, as should be expected for a reproductive number R_0 less than one.

We can further analyze the predictions of our model by looking not at *cumulative* case numbers, but at *new* infections. In our model, this is represented by the variable $I_1(t)$. These are shown in Fig. 7.3. For the days up to our 160th day, our model predicts a steadily increasing number of infections. Following the time when we assume behaviors changed, new infections drastically drop and are predicted to continue dropping beyond the timeframe for which we have actual data. Very similar graphs have indeed been reported from Liberia, for example, with a peak at approximately the same time as in our model [21, 22].⁴

⁴The data that is available for Sierra Leone and that we attempt to fit here does not allow for a direct comparison with Fig. 7.3 as cases are not reported in daily or weekly increments but only as cumulative cases that, in each week, often retroactively include much older cases that were not – but should have been – reported in previous weeks; or, in some cases, retroactively reduce the number of cases but without updating previous weeks’ numbers – implying a *negative* number of new cases. This can, for example, be seen in the “noisiness” of the data shown in Fig. 5.1.

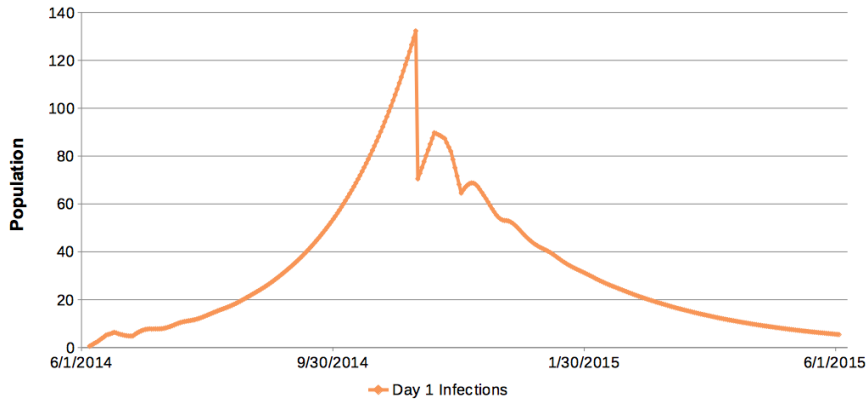


FIG. 7.3. Composite model predictions for new infections $I_1(t)$ over 1 year.

8. Conclusion. In this paper, we have introduced a model that intended to describe the spread of the current Ebola epidemic in West Africa. The model uses an age structured approach in which we track each patient throughout the entire infection history from being infected to complete recovery or death.

Through comparison with actual data that has been reported on the epidemic in Sierra Leone, we showed that our model can accurately describe reality quantitatively, and this provides the basis to make forecasts for the future development of an epidemic. It also allows to investigate questions such as “What would happen if we were able to reduce the number of secondary infections by 20%?” or “What would happen if we were able to reduce the mortality by 20%?”. In both of these cases, one could run the model we have presented with different values for the parameters β_k, δ_k starting at the current time and compare the outcomes with predictions in which the parameters are kept unchanged. Such simulations of “What if?” scenarios help in deciding where limited resources to combat an epidemic are best used, but such scenarios are only useful if the underlying model is an accurate representation of the epidemic itself. The results we have shown here demonstrate that our model provides such a base.

Appendix A. A retrospective with more data.

Since writing this paper, new data has been published by the CDC and we were intrigued to compare the accuracy of our model using data only up to January 21, 2015. Fig. A.1 therefore shows our model prediction against the data that has since become available. We can see that the model accurately predicted the total number of cases with an error on the order of 2%. On the other hand, the actual number of deaths from Ebola did not increase as much as the model predicted.

The number of deaths and cases are linked by the parameters δ_k . Since our model accurately predicted the number of cases but not of deaths, the implication of this deviation is that either the death rate has dropped since early 2015, or alternatively that infections are now more accurately reported than they were before. In either case, we can conclude that the predictions we obtained from our model were quite close to what in fact happened. Looking even beyond this data, by August 17, 2015, the number of cases and deaths in Sierra Leone stood at 13,489 and 3,952, respectively – well within the range of predictions that can be obtained with the help of our model.

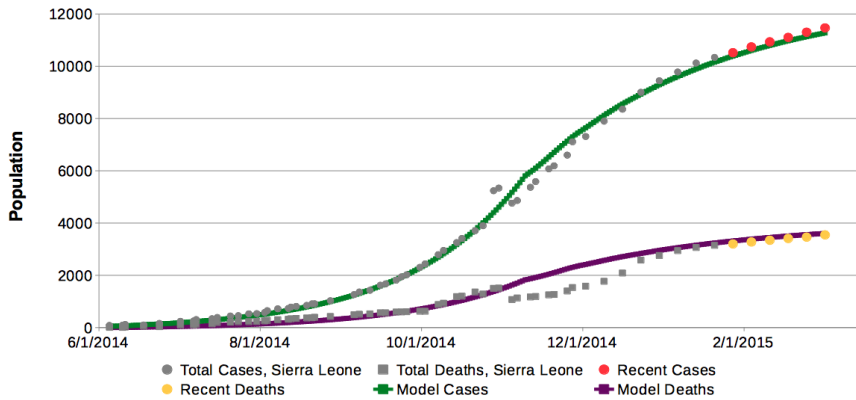


FIG. A.1. Composite model as shown in Fig. 7.1, but with additional data points (highlighted) that were not available when the model was fitted.

REFERENCES

- [1] C. L. ALTHAUS, *Estimating the reproduction number of Ebola Virus (EBOV) during the 2014 outbreak in West Africa*, PLoS Currents, 1 (2014).
- [2] J. ASTACIO, D. BRIERE, M. GUILLÉN, J. MARTÍNEZ, F. RODRÍGUEZ, AND N. VALENZUELA-CAMPOS, *Mathematical models to study the outbreaks of Ebola*, Tech. Report Biometrics Unit Technical report BU-1365-M, retrieved from <https://ecommons.library.cornell.edu/bitstream/1813/31962/1/BU-1365-M.pdf>, Cornell University, 1996.
- [3] J. BELLUZ, *9 questions you were too afraid to ask about Ebola*, retrieved from <http://www.vox.com/2014/10/9/6905347/too-afraid-to-ask-about-ebola-virus-outbreak-symptoms>, 2014.
- [4] CENTERS FOR DISEASE CONTROL AND PREVENTION, *Cumulative Reported Cases Graphs*, retrieved from <http://www.cdc.gov/vhf/Ebola/outbreaks/2014-westafrica/cumulative-cases-graphs.html>, 2014.
- [5] —, *Estimating the Future Number of Cases in the Ebola Epidemic – Liberia and Sierra Leone, 2014–2015*, retrieved from <http://www.cdc.gov/mmwr/preview/mmwrhtml/su6303a1.htm>, 2014.
- [6] —, *Infographics and Illustrations*, retrieved from <http://www.cdc.gov/vhf/Ebola/resources/infographics.html>, 2014.
- [7] —, *Previous updates: 2014 West Africa outbreak*. retrieved from <http://www.cdc.gov/vhf/Ebola/outbreaks/2014-westafrica/previous-updates.html>, 2014.
- [8] —, *Review of Human-to-Human Transmission of Ebola Virus*, retrieved from <http://www.cdc.gov/vhf/ebola/transmission/human-transmission.html>, October 2014.
- [9] —, *Transmission*. retrieved from Ebola, retrieved from <http://www.cdc.gov/vhf/Ebola/transmission/index.html>, 2014.
- [10] —, *2014 Ebola Outbreak in West Africa - Case Counts*, retrieved from <http://www.cdc.gov/vhf/ebola/outbreaks/2014-west-africa/case-counts.html>, 2015.
- [11] CENTRAL INTELLIGENCE AGENCY, *The World Factbook*, retrieved from <https://www.cia.gov/library/publications/the-worldfactbook/geos/sl.html>, 2014.
- [12] G. CHOWELL, N. W. HENGARTNER, C. CASTILLO-CHAVEZ, P. W. FENIMORE, AND J. M. HYMAN, *The basic reproductive number of Ebola and the effects of public health measures: The cases of Congo and Uganda*, Journal of Theoretical Biology, 229 (2004), pp. 119–126.
- [13] D. FISMAN, E. KHOO, AND A. TUIITE, *Early Epidemic Dynamics of the West African 2014 Ebola Outbreak: Estimates Derived with a Simple Two-Parameter Model*, PLoS Currents, 6 (2014).
- [14] CENTERS FOR DISEASE CONTROL AND PREVENTION, *2014 Ebola Outbreak in West Africa - Reported Cases Graphs*, retrieved from <http://www.cdc.gov/vhf/ebola/outbreaks/2014-west-africa/cumulative-cases-graphs.html>, 2015.
- [15] T. HOENEN, D. SAFRONETZ, A. GROSETH, K. R. WOLLENBERG, O. A. KOITA, B. DIARRA, I. S. FALL, F. C. HAIDARA, F. DIALLO, M. SANOGO, Y. S. SARRO, A. KONE, A. C. G. TOGO, A. TRAORE, M. KODIO, A. DOSSEH, K. ROSENKE, E. DE WIT, F. FELDMANN, H. EBIHARA,

- V. J. MUNSTER, K. C. ZOON, H. FELDMANN, AND S. SOW, *Mutation rate and genotype variation of Ebola virus from Mali case sequences*, *Science*, 348 (2015), pp. 117–119.
- [16] A. G. HUNT, *Exponential growth in Ebola outbreak since May 14, 2014*, *Complexity*, 20 (2014), pp. 8–11.
- [17] M. KEELING, *The Mathematics of Diseases*, retrieved from <http://plus.maths.org/content/os/issue14/features/diseases/index>, 2011.
- [18] K. KUPFERSCHMIDT, *Estimating the Ebola epidemic*, *Science*, 345 (2014), p. 1108.
- [19] R. MACINTYRE, *Reinfection with Ebola – is it possible and should health worker survivors use PPE?*, retrieved from <https://sphcm.med.unsw.edu.au/infectious-diseases-blog/reinfection-ebola-%E2%80%93-93-it-possible-and-should-health-worker-survivors-use-ppe>, University of New Wales School of Public Health and Community Medicine, (2014).
- [20] B. MANANAVIRE, *Family claims bad handling of ‘Ebola’ corpse*, retrieved from <http://nehandaradio.com/2015/02/12/family-claims-bad-handling-ebola-corpse/>, Nehanda Radio, (2015).
- [21] M. MELTZER ET AL., *Estimating the Future Number of Cases in the Ebola Epidemic – Liberia and Sierra Leone, 2014–2015*, retrieved from <http://www.cdc.gov/mmwr/preview/mmwrhtml/su6303a1.htm>, *Morbidity and Mortality Weekly Report*, 63 (2014), pp. 1–14.
- [22] LIBERIAN MINISTRY OF HEALTH, *Liberia Ebola daily sitrep no. 283 for 22th February 2015*, retrieved from <http://www.mohsw.gov.lr/documents/Sitrep%20283%20Feb%2022th%202015.pdf>, 2015.
- [23] J. OLU-MAMMA, *Ebola Survival Improving in Sierra Leone*, retrieved from <http://www.cbsnews.com/news/ebola-survival-improving-in-sierra-leone/>, CBSNews.com, (2014).
- [24] A. PANDEY, K. E. ATKINS, J. MEDLOCK, N. WENZEL, J. P. TOWNSEND, J. E. CHILDS, T. G. NYENSWAH, M. L. NDEFFO-MBAH, AND A. P. GALVANI, *Strategies for Containing Ebola in West Africa*, *Science*, 346 (2014), pp. 991–995.
- [25] S. R. PATTYN, ed., *Ebola Virus Haemorrhagic Fever. Proceedings of an International Colloquium on Ebola Virus Infection and Other Haemorrhagic Fevers held in Antwerp, Belgium, 6-8 December, 1977*, Elsevier/North-Holland Biomedical Press, 1978.
- [26] WORLD HEALTH ORGANIZATION, *Sierra Leone: A traditional healer and a funeral*, retrieved from <http://www.who.int/csr/disease/Ebola/ebola-6-months/sierra-leone/en/>, (2014).