学　位　論　文

Statistical modeling to estimate the protective effect of case isolation and reconstruct transmission dynamics during the 2014 Ebola virus disease epidemic

（２０１４年エボラウイルス病流行における個体情報を基にした感染者隔離の有効性評価）

２０２０年９月

北　海　道　大　学

陳逸軒

Yat Hin CHAN

学　位　論　文

Statistical modeling to estimate the protective effect of case isolation and reconstruct transmission dynamics during the 2014 Ebola virus disease epidemic

（２０１４年エボラウイルス病流行における個体情報を基にした感染者隔離の有効性評価）

２０２０年９月

北　海　道　大　学

陳逸軒

Yat Hin CHAN

Contents

List of Publication and Presentations

List of Publication

1. **Chan, Y.H**, and Nishiura, H. Estimating the protective effect of case isolation with transmission tree reconstruction during the Ebola outbreak in Nigeria, 2014. Journal of the Royal Society Interface. 17: 20200498 `http://doi.org/10.1098/rsif.2020.0498`

List of Presentations

1. **Chan, Y.H.**, and Nishiura, H. Data and dynamics reconstruction of Ebola virus disease outbreaks in Nigeria and Sierra Leone. Innovative Mathematical Modeling for the Analysis of Infectious Disease Data (IMAID) 2017, Institute of Statistical Mathematics, Japan.

2. **Chan, Y.H.**, and Nishiura, H. Reconstructing Ebola transmission in Nigeria and Sierra Leone using partially observed data. Japanese Society for Mathematical Biology (JSMB) 2017, Hokkaido University, Japan.

Summary

Background and Purpose The 2014 Ebola virus disease (EVD) epidemic caused more than 28,000 cases and 11,000 deaths internationally. The epidemic mainly affected West African countries, spreading initially from Guinea, then to Liberia and Sierra Leone. A minor outbreak also occurred in Nigeria. Multi-country epidemics led to an announcement of the public health emergency of international concern (PHEIC) and the transmission routes were contributed by human-to-human infections. Patients developed symptoms such as fever, vomiting or diarrhea, acting as infectious sources and transmitting the viruses to others by contact with bodily fluids. The mainstream of intervention was case isolation and contact tracing because vaccines were not available for the general public. This work aims to devise a statistical model to estimate the impact of case isolation using observational data. The model was applied to two countries, Nigeria and Sierra Leone. Empirically observed data from both countries did not offer a full spectrum of the epidemiological datasets for all individuals. Thus, a modeling framework was designed to reconstruct transmission and to analyze the dynamics in detail. Due to different methods of data reconstruction, the studies in two countries are presented separately in Chapters 1 and 2. The former was based on a dataset containing relatively less missing information whereas the latter required a more complicated approach due to larger parts of missing pieces.

Materials and Methods This work was based on two local outbreaks in Nigeria and Sierra Leone that involved 20 and 49 cases respectively, of which 20 and 44 were hospitalized. In particular, transmission trees that provide information on sources of infection and time events such as dates of symptom onset and hospitalization of each individual were analyzed. However, there were only partially observed details while some missing pieces require reconstruction. Missing dates were imputed using distributions of various intervals between time events. For example, an incubation period with the mean of 9 days was used to evaluate the timing of symptom onset from observation of exposure. Subsequently, a statistical model that quantifies the protective effect of case isolation as a reduction in secondary transmission rate was developed. Accordingly, the reproduction number representing infected cases per individual is reduced and the serial interval representing time lag of symptom onsets between primary and secondary cases is shortened. The modeling was combined with Bayesian statistical approaches to estimate parameters and to assess missing links in the transmission trees simultaneously. Two statistical methods, maximum likelihood estimation (MLE) and Markov chain Monte Carlo (MCMC), were adopted in the estimation. Furthermore, a sensitivity analysis was performed by considering different model assumptions and statistical methods. Alternative scenarios on the reproduction number, protective effect of unknown hospitalized cases and proposal distribution in MCMC were evaluated.

Results The time from symptom onset to hospitalization follows Weibull distributions in both outbreaks. The mean values were 4.0 and 5.0 days and SD were 2.3 and 2.5 days for the data from Nigeria and Sierra Leone respectively. The key time events of all individuals were probabilistically reconstructed by various time lags. The best-fitted models were found to be employing geometrically distributed secondary case numbers in both outbreaks while the serial intervals follow gamma and Weibull distribution for the data from Nigeria and Sierra Leone. The transmission dynamics were represented using mean and 95% credible interval (CI) of the corresponding posterior distributions of parameters. The protective effect of case isolation was 39.7% (95%CI: 2.4%-82.1%) and 23.9% (95%CI: 1.1%-69.5%) in the outbreaks in Nigeria and Sierra Leone, respectively. Most of the individual protective effect was similar to others sharing the same missing information patterns. Each missing link was represented by a list of possible candidates, which identified one and three most likely cases in two outbreaks respectively. The modeling demonstrated consequences in the absence of case isolation. The reproduction numbers at the beginning of two outbreaks were 10.0 (95%CI: 3.0-18.5) and 3.0 (95%CI: 1.2-9.1) respectively, which declined exponentially with the rates of 0.14 (95%CI: 0.07-0.23) and 0.019 (95%CI: 0.005-0.035) per day. The mean of unbiased serial intervals was 15.3 (95%CI: 14.2-16.6) and 12.6 (95%CI: 10.8-14.3) days. As the sensitivity analysis, the reproduction numbers were found to depend on the dynamical assumption in time (e.g. exponential decline). The protective effect of both known and unknown hospitalized cases was influenced by the variation. The serial intervals remained unchanged under alternative scenarios.

Discussion The statistical modeling estimated the protective effect of case isolation during the 2014 EVD outbreaks in Nigeria and Sierra Leone using partially observed contract tracing data. The model captures the dynamical features of EVD transmission using two main components, the reproduction number and serial interval. The former indicates the outbreak magnitude and the latter shows the speed of generation replacement. The estimates of serial intervals coincide with other studies of the two-week interval. The intervals were shortened in two outbreaks as a consequence of successful control. On the other hand, the reproduction numbers in two outbreaks were significantly different due to the index case as a super-spreader in Nigeria, whereas there was no super-spreading event during the outbreak in Sierra Leone. However, the protective effect and final sizes indicated that the interventions were more successful in Nigeria. The unknown hospitalization of the outbreak in Sierra Leone gave another explanation to the controlling impact. Overall, the major factor that lowered the protective effect is infections that occurred during isolation including transmission in hospitals. In addition, traditional burials with a series of practices during ceremonies made the isolation less effective in the early phase before the EVD burial teams conducting safe burials. Nevertheless, the major limitation of this model is the assumption on infectiousness starting from the time of symptom onset, which is valid for EVD but not broadly applicable to other infectious diseases.

Conclusion The modeling approach explicitly estimated the protective effect of case isolation and reconstructed missing information of transmission among all individuals involved. The statistical model was formulated to demonstrate that the case isolation during two local outbreaks of the 2014 EVD in Nigeria and Sierra Leone effectively reduced reproduction number and shortened the serial interval. Due to such local containment efforts in preventing secondary transmissions, affected countries were able to avoid the 2014 EVD becoming a global pandemic.

List of Abbreviations

AIC	Akaike information criterion
BIC	Bayesian information criterion
CDF	Cumulative distribution function
CFR	Case fatality rate
CI	Credible interval
DRC	Democratic Republic of the Congo
EHC	Ebola health center
EVD	Ebola virus disease
i.i.d.	Independent and identically distributed
MCMC	Markov chain Monte Carlo
MH	Metropolis-Hastings
MLE	Maximum likelihood estimation
PDF	Probability density function
PHEIC	Public health emergency of international concern
PHM	Posterior harmonic mean
PHU	Peripheral health unit
PMF	Probability mass function
SD	Standard deviation

Introduction

Ebola virus disease (EVD) is a viral hemorrhagic fever caused by the Ebola virus and generated the largest EVD epidemic in 2014. The disease spread widely from Guinea to Liberia, Sierra Leone, and Nigeria (WHO Ebola Response Team 2014). The infection routes were initially from wildlife to humans, which then expanded by human-to-human transmission (Ponce et al. 2019). Infected cases would be symptomatic and subsequently become infectious to transmit the viruses by bodily fluids, for example through physical contacts by series of practices during traditional burial ceremonies (Ajelli et al. 2015). The key interventions against EVD spreading were non-pharmaceutical measures including case isolation and contact tracing as vaccines were not yet ready. As infections were observed to be during the traditional burials, safe burials replaced to prevent further transmission in the later phase of the epidemic.

Many epidemiological modeling studies were conducted to evaluate the performance of both pharmaceutical and non-pharmaceutical interventions (Henao-Restrepo et al. 2015, 2017, Kucharski et al. 2015a,b, 2016, Pandey et al. 2014, Shen et al. 2015, Drake et al. 2015, Chowell et al. 2015, Fang et al. 2016, Lokuge et al. 2016, Merler et al. 2015, Yamin et al. 2015, Rivers et al. 2014, Webb et al. 2015, Camacho et al. 2015, Lewnard et al. 2014). However, all the modeling focused on population-level estimation and there is a lack of individual-level estimation. The models assumed a homogeneous population setting or consists of groups in heterogeneity. Characteristics of individuals were omitted in model formulation. In the context of case isolation, the timing of hospitalization since infection or symptom onset is different among patients. The individual impact would consequently different. Thus this is essential to infer a statistical model at an individual-level to estimate the impact of interventions.

This work aims to build a mathematical model to estimate the protective effect of case isolation given surveillance data including time events and source of infection of all patients. First, mathematical models can demonstrate various transmission scenarios that could not be observed in reality. For example, modeling can simulate transmission in the absence of case isolation to compare with the scenario in the presence of case isolation. As a result, the impact of case isolation can be evaluated. Second, individual data of each patient are essential in the analysis to evaluate the entire epidemiological characteristics and to identify how transmission occurred during the epidemic. However, there is no perfect method to collect data in the surveillance system and links between primary and secondary cases are not always identifiable by contact tracing. In most situations, this is almost impossible to observe every single detail of patients, especially before developing symptoms. Therefore, there are two key points in this work. One is to reconstruct data to reveal transmission during the epidemic and another one is to estimate model parameters to present transmission in the presence or absence of case isolation.

The model was based on the following mathematical modeling studies about data reconstruction and analysis of infectious diseases. Wallinga and Teunis (Wallinga & Teunis 2004) developed a

mathematical model to reconstruct transmission trees using dates of symptom onset of patients and applied on the 2003 SARS epidemics. Eichner and Dietz (Eichner & Dietz 2003) analyzed the 1967 smallpox epidemic in Abakaliki, Nigeria and reconstructed dates of exposure using dates of symptom onset of individuals to evaluate the impact of vaccination. Stockdale et al. (Stockdale et al. 2017) extended the study of Eichner and Dietz and replaced maximum likelihood estimation (MLE) by Bayesian statistical analysis with Markov chain Monte Carlo (MCMC) methods to avoid likelihood approximations.

By combining the above-mentioned approaches, the protective effect of case isolation can be estimated by reconstructing transmission trees and assessment of transmission dynamics to reveal the entire epidemiological characteristics of individuals. Precisely, the main model parameters include the protective effect of case isolation, reproduction number, and serial interval. The protective effect of case isolation represents a relative reduction in secondary transmission and quantifies how many infections could be avoided since patients were isolated. The reproduction number represents the magnitude of the epidemic whereas the serial interval represents the speed of generation replacement.

This mathematical modeling was applied to two datasets from Nigeria and Sierra Leone. The approaches are similar but different due to the completeness of reported data. Chapter 1 presents the study in Nigeria, of which the data were relatively complete with less missing information (Folarin et al. 2016, Shuaib et al. 2014). There was at most one time event requiring reconstruction of each individual while all individuals were hospitalized. Chapter 2 presents the study in Sierra Leone, of which the data contained more missing pieces and required a more complicated way for reconstruction and analysis (Ajelli et al. 2015). There were at most two time events requiring reconstruction and at most three in likelihood calculation. In addition, patients with unknown hospitalized information made the reconstruction and evaluation more challenging.

Chapter 1: Estimating the protective effect of case isolation with transmission tree reconstruction during the Ebola outbreak in Nigeria, 2014

Introduction

Ebola virus disease (EVD) is an acute severe human infection caused by the Ebola virus, which was first identified in Zaire in 1976 (Bowen et al. 1977, Johnson et al. 1977, Pattyn et al. 1977). Once infected, the incubation period (i.e. the time from exposure to the onset of illness) ranges from 2 to 21 days, with a mean of 9.1 days and standard deviation (SD) of 7.3 days (WHO Ebola Response Team 2014). Clinical signs and symptoms include fever, headache, fatigue, diarrhoea, vomiting, stomach pain, unexplained bleeding or bruising, and muscle pain. Due to its zoonotic nature, EVD is transmitted to the human population primarily from wildlife, notably fruit bats or primates (Ponce et al. 2019). Subsequently, human-to-human transmission occurs via direct contact with bodily fluids, with the handling of deceased victims (e.g. during burials) being a particularly significant channel for transmission. The largest ever epidemic of EVD occurred in West Africa during 2014-2016 and generated more than 28,600 confirmed, probable and suspected cases, with 11,325 deaths (CDC 2019). During the 2014-2016 epidemic, this highly fatal disease initially appeared in Guinea and then spread to neighbouring countries in West Africa, most notably Liberia and Sierra Leone. Besides the three countries most affected, small local outbreaks were reported in Nigeria and elsewhere (WHO Ebola Response Team 2014), with the Nigerian outbreak notable for being contained swiftly.

With regards to possible countermeasures against EVD, no effective vaccines were available during the early stage of the 2014-2016 outbreak, while various clinical studies were conducted during the later period of the epidemic (Henao-Restrepo et al. 2015, 2017, Kucharski et al. 2016). Thus, the mainstream interventions employed during the 2014-2016 epidemic were non-pharmaceutical, and for instance included interventions at healthcare settings relying on contact tracing and case isolation – i.e., restricting the movement of contacts of diagnosed and infectious individuals, respectively. Including analysis of early transmission dynamics, epidemiological modelling studies were performed to assess the effectiveness of public health interventions that encompass hospital admission, case isolation and contact tracing (Kucharski et al. 2015a,b, 2016, Pandey et al. 2014, Shen et al. 2015, Drake et al. 2015, Chowell et al. 2015, Fang et al. 2016, Lokuge et al. 2016, Merler et al. 2015, Yamin et al. 2015, Rivers et al. 2014, Webb et al. 2015, Camacho et al. 2015, Lewnard et al. 2014). Nigeria successfully contained its EVD outbreak by using contact tracing and case isolation, and published datasets on the outbreak in Lagos and Port Harcourt, Nigeria have documented the epidemiological dynamics in detail (Folarin et al. 2016, Shuaib et al. 2014).

Supporting the effectiveness of hospital admission and case isolation, a seminal paper by Kucharski

et al. (Kucharski et al. 2015a) has demonstrated that building a substantial number of EVD treatment centres helps curb the epidemiological dynamics of EVD, which actually became the mainstream of strategic global support for the control of EVD from 2014-2016. However, such evaluation has been restricted to the population level impact (i.e. the effectiveness of building provisional hospitals in reducing the incidence of EVD) (Pandey et al. 2014, Kucharski et al. 2015b, Lokuge et al. 2016, Merler et al. 2015, Lewnard et al. 2014) or to evaluations that measured the overall transmission risk at the population level (e.g. household secondary attack risk (Fang et al. 2016)). To fully understand the protective effect of case isolation, it is important to infer the effect of preventing secondary transmission at an individual level.

As part of the post-epidemic evaluation, here we devised a novel statistical model to assess the protective effect of case isolation against EVD transmission. Addressing the biased nature of observed data associated with interventions (e.g. shortened time interval between onset of illness for primary and secondary cases), the present study aims to quantify the protective effect of case isolation from the transmission network data for EVD. Specifically, we explore a well traced dataset on EVD from Nigeria in 2014.

Materials and Methods

Description of the EVD outbreak data from Nigeria

Two published datasets from the Nigerian EVD outbreak were used (Folarin et al. 2016, Shuaib et al. 2014). The outbreak involved 20 cases in total, including one imported case, with all cases being confirmed and reported during 17 July 2014 to 20 October 2014. After a rapid detection of the index case, the government established an Emergency Operations Centre with an Incident Management System to coordinate quick response and decision making using previous experience dealing with a polio outbreak in 2012 (Shuaib et al. 2014). The index case potentially exposed 72 people and resulted in 894 traced contacts, of which 891 were followed up (Shuaib et al. 2014). Due to concerted control efforts, the proportion of cases hospitalized in Nigeria was 100% and here we aim to estimate the protective effect of the hospitalization. Based on contact tracing, dates of onset of illness and hospitalization were recorded, and links between infectors and infectees were established. Throughout this article, the source of infection (i.e. infectors) is denoted as $\mathbf{v}$. Additionally, exposure date is written as $\mathbf{t}^e$, date of onset of illness is $\mathbf{t}^s$, hospitalization date is $\mathbf{t}^h$, and date of death is $\mathbf{t}^d$.

In the present study, we combined the datasets from two independent publications (Folarin et al. 2016, Shuaib et al. 2014). In principle, the data we used adhered to Folarin et al. (Folarin et al. 2016), but we referred to another study (Shuaib et al. 2014) due to partially missing information. The next section describes the details of our procedure for reconstruction of the original data. Even with the merger of the data, the dataset was incomplete, specifically the dates $\mathbf{t}$ and sources of infection $\mathbf{v}$. There was only one case where the source of infection was unknown. For the only case with an unknown infector, we summarized a list of six possible infectors. On the other hand, there were several cases in the outbreak for which dates were missing, in particular, dates of onset of illness $\mathbf{t}^s$ and hospitalization $\mathbf{t}^h$, making it necessary to impute dates. Specifically, there were 10 cases for which both dates were known and 10 cases for which either date was missing and hence dates had to be imputed.

Data merging

The date of onset of illness for Case 1 was not described by Folarin et al. (Folarin et al. 2016), while another study specified it as 17 July 2014 (Shuaib et al. 2014), a date referred to throughout our study as Day 0. Dates of exposure and onset of illness for Case 12 were identical in the original data (Folarin et al. 2016). This is technically impossible, and so we chose to treat this date as the date of exposure rather than that of the onset of illness. We made this decision because no onset of illness event was recorded for 23 July 2014 in any other dataset (Shuaib et al. 2014), and also because the most likely dates of onset for Case 12 would have been from 30 July to 10 August 2014, based on transmission tree data in (Shuaib et al. 2014), which would agree closely with the known incubation period distribution with mean 9.1 days and standard devia-

tion (SD) 7.3 days (WHO Ebola Response Team 2014). By overlaying observed branches in the presence of three generations, something seen among cases infected in Port Harcourt (Folarin et al. 2016, Shuaib et al. 2014), the dates of onset of illness for Cases 18 and 19 were identified as 31 August 2014 and 29 July 2014, respectively. Based on visual inspection of the transmission tree (Shuaib et al. 2014), the date of onset of illness for Case 17 was identified as 2 August 2014 by tracing links connected to Case 13. Similarly, the date of onset of illness for Case 20 was considered to be 16 August 2014, and moreover, Case 20 was considered to have been infected by one of the first generation cases (i.e. cases infected by the index case). Thus, the possible primary cases of Case 20 are Cases 5, 6, 7, 10, 11 or 12, but it was not possible to uniquely identify any one of these on the transmission tree (Shuaib et al. 2014). The resulting dataset can be found in the author's website (https://github.com/imlouischan/ebola-ng).

Definition of time events and time intervals

Here we reconstruct the epidemic tree using the partially observed dataset, applying similar methods to those published elsewhere (Eichner & Dietz 2003, Stockdale et al. 2017), and subsequently estimate the protective effect of case isolation. For this reason, we firstly define several time events and intervals that would be central for reconstruction. Let i serve as an index of individuals. We have four different time events, namely, dates of exposure t_i^e, onset of illness t_i^s, hospitalization t_i^h and death t_i^d. As it is usually the case, these timelines are incomplete (see Fig 1), but knowing the dates of onset of illness t_i^s and hospitalization t_i^h is essential to evaluate the protective effect of case isolation. For each case, there was at least one observed time event. From these time events, there are two types of time lag distributions, i.e., the incubation period τ^{es} and the time from onset of illness to hospitalization τ^{sh}. We calculate $\tau^{xy} = t^y - t^x$, where the subscripts x and y represent e, s or h. Then, missing events are estimated using time lag distributions $f^l(\tau)$, where τ is time lag and l is a time lag that bridges the observed and missing dates for each individual. To impute missing dates, we prioritized the use of the incubation period, because parameter estimates of the distribution were derived from a published study with a large sample size (WHO Ebola Response Team 2014). For example, for a case i, while the date of exposure t_i^e is observed and the onset of illness t_i^s is missing, the date of onset of illness $\hat{t}_i^s = t_i^e + \tau^{es}$ is quantified from the date of exposure t_i^e and the estimated incubation period τ^{es}, rather than working backwards and estimating t_i^s from hospitalization t_i^h. Four patterns of time event reconstrction are summarized in Table 1. Moreover, theoretical boundaries for quantifying missing dates are described in the following section.

Incubation period distribution

As for the incubation period distribution, we employed the independently and identically distributed incubation period for another large-scale outbreak from 2014-2016, assuming that it followed a gamma distribution with mean 9.1 days and SD 7.3 days (WHO Ebola Response

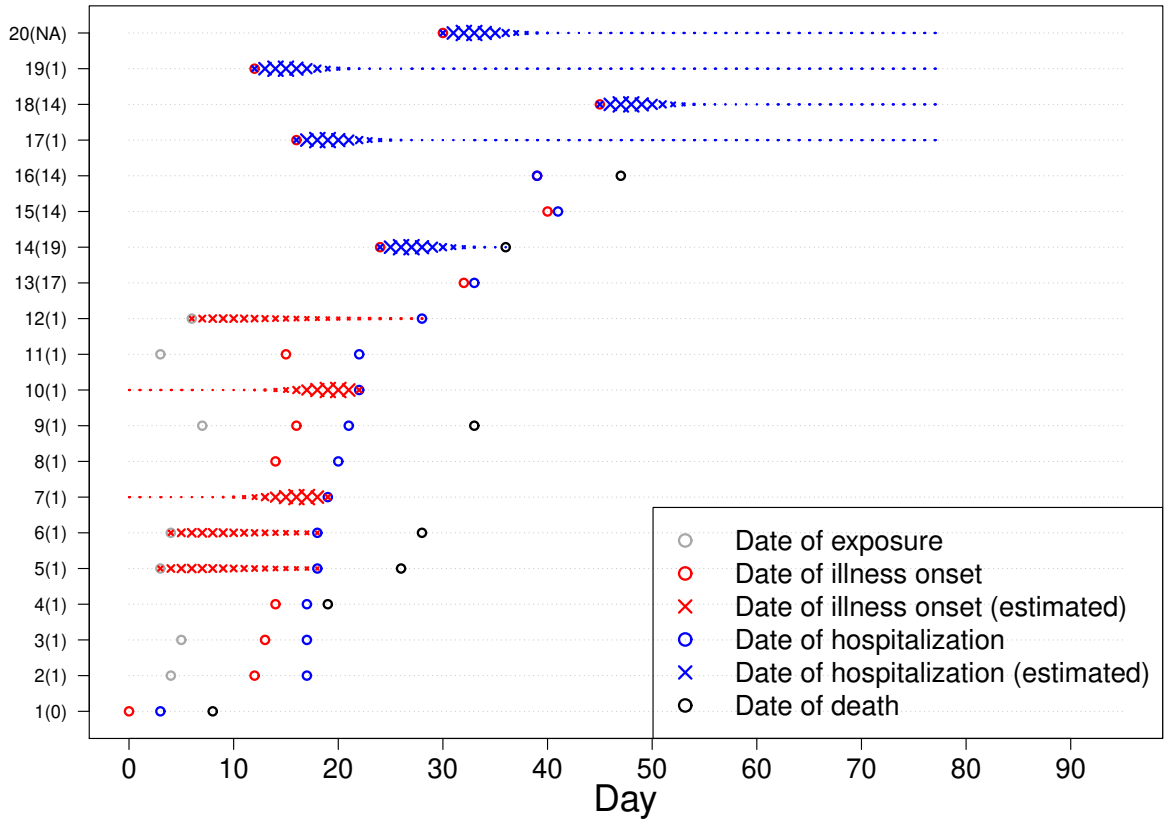


Figure 1: **The reconstructed timelines consisting of observed time events and the probability of reconstructed (missing) time events (n=20 cases).** The observed time events are denoted by unfilled circles. The probabilities of reconstructed time events are measured by cross size. The grey, red, blue, and black symbols denote the dates of exposure, onset of illness, hospitalization, and death, respectively. The numbers shown on the y-axis represent case identity numbers (and associated infectors, where NA means the infector is missing).

Table 1: **Theoretical patterns of observed and missing time events.** Each case i is categorized into four different patterns for reconstructing the dates of the two important events – i.e., onset of illness and hospitalization. Four time events, the dates of exposure t_i^e, onset of illness t_i^s, hospitalization t_i^h, and death t_i^d are observed O, missing M or trivial (i.e. either is ok) $-$. The reconstructed dates of onset of illness $\hat{t}_i^s$ and hospitalization $\hat{t}_i^h$ are estimated using time lags depending on observed patterns. The last column shows the count of cases that fell into the specified pattern.

Pattern	t_i^e		t_i^s		t_i^h	t_i^d	$\hat{t}_i^s$	$\hat{t}_i^h$	Observed
$i \in D^{(1)}$	$-$		O		O	$-$	t_i^s	t_i^h	10/20
$i \in D^{(2)}$	$-$		O	$\rightarrow$	M	$-$	t_i^s	$t_i^s + \tau^{sh}$	5/20
$i \in D^{(3)}$	O	$\rightarrow$	M		O	$-$	$t_i^e + \tau^{es}$	t_i^h	3/20
$i \in D^{(4)}$	M		M	$\leftarrow$	O	$-$	$t_i^h - \tau^{sh}$	t_i^h	2/20

Team 2014) (Fig 2A). Due to the discrete nature of our observed data, the distribution was then transformed from continuous to discrete by $f(\tau) = F(\tau+1) - F(\tau)$ for $\tau = 0, 1, 2, ...$, where $f(\tau)$ and $F(\tau)$ represent the probability mass function (PMF) and cumulative distribution function (CDF). The sample mean of the incubation period observed in Nigeria was 9.3 days, similar to the estimate from the abovementioned larger sample study (WHO Ebola Response Team 2014), while the SD was only 1.9 days, with this smaller latter value perhaps reflecting the small sample size in Nigeria. Fig 2A illustrates the mismatch between two datasets.

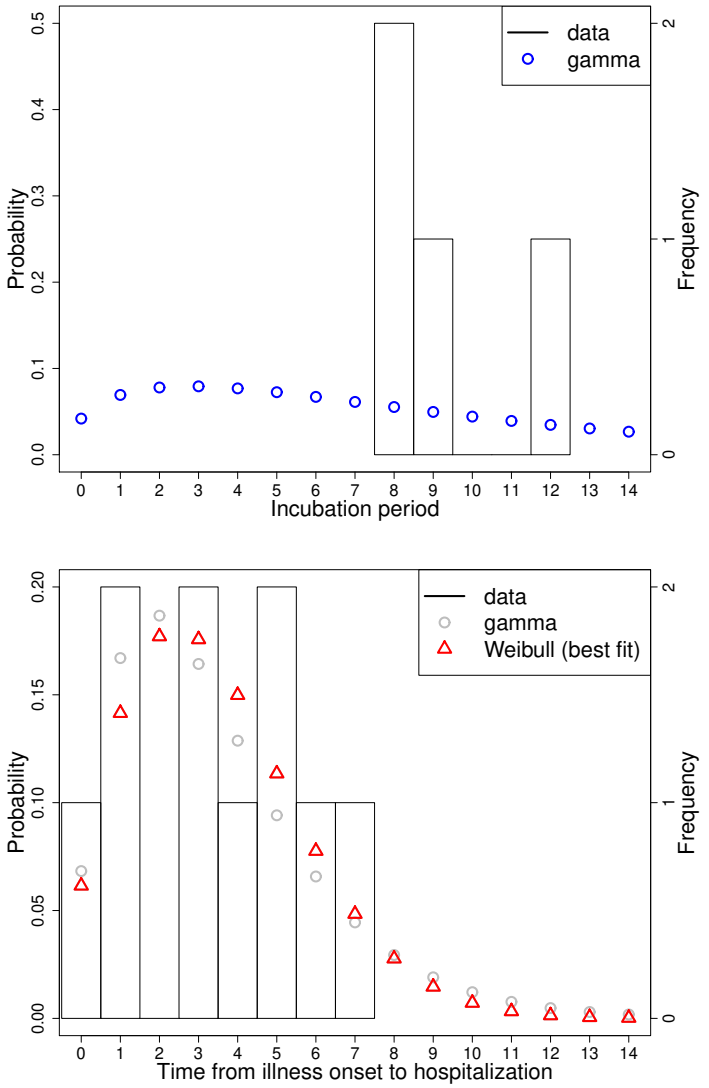


Figure 2: The probability mass functions (PMF) of (A) the incubation period and (B) the time from onset of illness to hospitalization. The observed data are shown in bars. The unfilled blue circles represent a published estimate (WHO Ebola Response Team 2014). The unfilled circles and unfilled triangles represent the predicted (fitted) values by employing two distributions, gamma and Weibull. The red and grey symbols show the best and second-best fit results, respectively. (A) The mean incubation period (with SD) was estimated as 9.1 (7.3) days. (B) The mean time from onset of illness to hospitalization (with SD) was estimated to be 4.0 (2.3) days.

Boundaries for missing date of an event

Missing dates were quantified within intervals between observed dates and the date of emptying isolation wards and exiting follow-up on 2 October 2014 (Day77), denoted as t_{exit} (Shuaib et al. 2014). The outbreak ended and the country was declared Ebola free by local authorities and WHO on 20 October 2014 (Day95), denoted as t_{end} (Folarin et al. 2016). For a case i, the dates of exposure, onset of illness, hospitalization and death are aligned as the following sequence, i.e. $t_i^e \leq t_i^s \leq t_i^h \leq t_i^d \leq t_{exit} < t_{end}$. Because we assume that human Ebola cases are not infectious before the onset of illness, the date of exposure of a case i follows the date of onset of illness of its infector v_i, i.e. $t_{v_i}^s \leq t_i^e$. Using these two inequalities, we write boundaries of missing events. The upper bound of the date of hospitalization t_i^h is $\bar{B}(t_i^h) = \min(t_i^d, t_{exit})$. Similarly, the lower and upper bounds of the date of onset of illness t_i^s are $\underline{B}(t_i^s) = \max(t_i^e, t_{v_i^n}^s, t_{v_i^n}^e)$ and $\bar{B}(t_i^s) = \min(t_i^h, t_i^d, \mathbf{t}_{u_i^n}, t_{exit})$, respectively. v_i^n represents all the 'ancestors' of case i and $\mathbf{t}_{u_i^n}$ represents all observed time events of the 'descendants' of case i. Note that the lower bound of the date of hospitalization t_i^h is the same as the lower bound of the date of onset of illness t_i^s if the latter is missing.

Model descriptions

Our modelling approach employs the so-called renewal equation,

$$j(t) = \int_0^\infty A(t,\tau) j(t-\tau) d\tau, \tag{1}$$

where $j(t)$ is the incidence (or the transient number of new infections) at calendar time t and $A(t,\tau)$ is the transmission rate per case at calendar time t and the infection-age (i.e., the time since infection) τ, which may satisfy $A(t,\tau) = R(t)s(\tau)$, where $R(t)$ is the instantaneous reproduction number and $s(\tau)$ is the probability density function (PDF) of the serial interval (Fraser 2007, Chowell & Nishiura 2008). Case isolation was performed during the course of the Nigerian epidemic. Throughout this study, we took the time of hospitalization as the time of case isolation, as the two occurred simultaneously in Nigeria. Let ε denote the relative reduction in the rate of secondary transmission in an isolated individual, quantified as a fixed parameter between 0 and 1, mirroring the protective effect of case isolation, where $\varepsilon = 1$ represents perfect isolation. Only during case isolation is the transmission rate assumed to be reduced to $(1-\varepsilon)A(\tau,t)$ per isolated case, i.e.,

$$\hat{A}(\tau,t) = \begin{cases} A(\tau,t) & \text{without isolation} \\ (1-\varepsilon)A(\tau,t) & \text{with isolation} \end{cases} \tag{2}$$

The time lag from onset of illness to hospitalization $\tau_i^{sh} = t_i^h - t_i^s$ varies by individual, and thus the reduction in the number of secondary transmissions at an individual level is described

as

$$\varepsilon_i = \left(1 - S(\tau_i^{sh})\right)\varepsilon, \tag{3}$$

where $S(\tau)$ is the cumulative distribution function (CDF) of the serial interval. We specifically consider ε_i for an individual i, because secondary transmissions averted by case isolation vary with the time delay τ_i^{sh} from onset of illness for the primary case. If a case was isolated immediately after the onset of illness, $\tau_i^{sh} = 0$ then the individual effect is maximized, i.e., $\varepsilon_i = \varepsilon$ ($\varepsilon_i < \varepsilon$ otherwise). Case isolation results in two different consequences for the transmission dynamics that are governed by the renewal equation. First, case isolation reduces secondary transmissions, i.e.,

$$\hat{R}(t_i) = R(t_i)(1 - \varepsilon_i) \tag{4}$$

$$= R(t_i)\left(\int_0^{\tau_i^{sh}} s(\tau)d\tau + (1-\varepsilon)\int_{\tau_i^{sh}}^{\infty} s(\tau)d\tau\right) \tag{5}$$

where τ_i^{sh} represents the disease-age (i.e. the time since onset of illness) at which the primary case i was isolated. Second, isolation shortens the serial interval distribution, i.e.,

$$\hat{s}(\tau_i) = \begin{cases} \frac{1}{1-\varepsilon_{v_i}} s(\tau_i) & \text{without isolation} \\ \frac{1-\varepsilon}{1-\varepsilon_{v_i}} s(\tau_i) & \text{with isolation} \end{cases} \tag{6}$$

$$= \begin{cases} \frac{s(\tau_i)}{\int_0^{\tau_{v_i}^{sh}} s(\tau)d\tau + (1-\varepsilon)\int_{\tau_{v_i}^{sh}}^{\infty} s(\tau)d\tau} & \text{without isolation} \\ \frac{(1-\varepsilon)s(\tau_i)}{\int_0^{\tau_{v_i}^{sh}} s(\tau)d\tau + (1-\varepsilon)\int_{\tau_{v_i}^{sh}}^{\infty} s(\tau)d\tau} & \text{with isolation} \end{cases} \tag{7}$$

where ε_{v_i} is the reduction effect of the primary case v_i. Hereafter, we refer to $s(\tau)$ in the absence of case isolation as the unbiased serial interval, while the observed time interval $\hat{s}(\tau)$ in the presence of case isolation is referred to as the biased serial interval. Using the reduced reproduction number $\hat{R}(t)$ and the distribution of shortened serial interval $\hat{s}(\tau)$, we intend to estimate the protective effect of case isolation ε.

Source of infection and partial reconstruction of the transmission tree

In terms of the transmission tree, the source of infection was unknown for one case, while for all other cases it was known. For the case with an unknown source of infection (Case 20), six possible candidates for the primary case existed, and the probabilistic reconstruction was conducted (Wallinga & Teunis 2004, Hens et al. 2012).

Unobserved infectors can be estimated using a transmission probability originally introduced by Wallinga and Teunis (Wallinga & Teunis 2004). The possible infectors are those with earlier onset of illness and the transmission probabilities of other cases are zero. This method has been

extended into a partial network reconstruction given contact information that restricts a list of possible infectors with positive transmission probabilities and others are zero (Hens et al. 2012). The transmission probability is written as

$$p_{ij}(\theta) = \frac{\hat{s}(\tau_{ij}^{ss};\theta)}{\sum_{k\in w_i}\hat{s}(\tau_{ik}^{ss};\theta)}, \tag{8}$$

where $\tau_{ik}^{ss} = t_i^s - t_k^s$ is the serial interval of case i and its possible infector k. w_i is a list of possible infectors of case i. The list consists of 6 infectors, namely Cases 5, 6, 7, 10, 11 and 12.

Likelihood functions

We first suppose that all data $\mathbf{D}$ are observed. In this model, the likelihood function denoted as $L(\theta|\mathbf{D})$ consists of two parts, number of observed secondary transmissions $L_r(\theta|\mathbf{D})$ and length of observed serial intervals $L_s(\theta|\mathbf{D})$.

$$L(\theta|\mathbf{D}) = L_r(\theta|\mathbf{D})L_s(\theta|\mathbf{D}) \tag{9}$$

The former consists of all reported cases, while the latter consists of only non-imported cases. The observed secondary case number r_i produced by each case i is assumed to follow a distribution p with the same mean value as the reproduction number $\hat{R}(t_i^s)$ at onset of illness. Three alternative distributions exist, geometric, Poisson and negative binomial. The likelihood of observing secondary case numbers is written as

$$L_r(\theta|\mathbf{D}) = \prod_{i\in\mathbf{D}} p(r_i;\theta). \tag{10}$$

We assume an exponential decreasing dynamic in the calendar time as $R(t) = R_0 e^{-\delta t}$, where R_0 is the reproduction number at time zero and δ is the exponential rate. The simple exponential decline in $R(t)$ has been empirically shown to fit better than the constant reproduction number in other outbreak case studies (Cauchemez et al. 2016, Nishiura et al. 2006). The observed serial intervals are assumed to follow a distribution $\hat{s}$, in which the unbiased distribution $s(\tau)$ follows two alternative distributions, gamma and Weibull. The likelihood is written as

$$L_s(\theta|\mathbf{D}) = \prod_{i\in\mathbf{D}\setminus\mathbf{D}^1} \hat{s}(\tau_i^{ss};\theta), \tag{11}$$

where $\tau_i^{ss} = t_i^s - t_{v_i}^s$ is the serial interval from the onset of illness in the primary case v_i to that in the secondary case i. Due to the assumption of a closed population, all cases except imported cases $\mathbf{D}^1$ are included in calculating the serial intervals.

None of the time events were fully observed, but instead were estimated using the maximum likelihood estimation (MLE). The next section provides further details on the reconstruction of

time events.

Reconstruction of time events

The observed dataset contained some cases for which the important time events, namely onset of illness and hospitalization, were only partially observed. The likelihood of those events was probabilistically described (Fig 1), and each event probability was considered at an individual level using convolution equations (Tables 2, 3, 4, 5). Specifically, we considered two different likelihood functions, L_r and L_s, where L_r accounts for the probability of observing a certain number of secondary transmissions per single primary case and L_s describes the probability of observing a time interval between primary and secondary cases. Theoretically, the former likelihood L_r is categorized into four types (Table 2), matching the timeline reconstruction (Table 1). On the other hand, the latter likelihood L_s is further categorized into six types due to each likelihood consisting of secondary case i and associated primary case v_i. For the description of L_s, we specifically need the date of onset of illness in both i and v_i, but only need the date of hospitalization for v_i. The observed dates are assumed to be independent and identically distributed (i.i.d.). The time from onset of illness to hospitalization is assumed to have the following two alternative distributions: Gamma and Weibull, which are discretized as the incubation period. The distribution parameters were estimated using the maximum likelihood method (Fig 2B). The computation was performed using the general-purpose optimization method `stats::optim` in R version 3.4.3 (Kite-Eating Tree) on Mac OSX. The alternative distribution is selected by Akaike information criterion (AIC).

Table 2: **Categories of the likelihood of secondary case L_r.** O, M and $-$ represent observed time events, missing time events, and cases for which there are both observed and missing time events, respectively. $\underline{B}(t_i^s) = \max(t_i^e, t_{v_i^n}^s, t_{v_i^n}^e)$ is the lower bound of t_i^s. $\bar{B}(t_i^s) = \min(t_i^h, t_i^d, t_{u_i^n}^e, t_{u_i^n}^s, t_{u_i^n}^h, t_{u_i^n}^d, t_{exit})$ is the upper bound of t_i^s. $\bar{B}(t_i^h) = \min(t_i^d, t_{exit})$ is the upper bound of t_i^s. v_i^n and u_i^n represent all the 'ancestors' and 'descendants' of i, respectively. t_{exit} represents the time when the isolation wards were empty.

$i \in$	t_i^e	t_i^s	t_i^h	t_i^d	$L_r(\theta \mid D)$
D^{-OO-}	$-$	O	O	$-$	$\prod_{i \in D^{-OO-}} p(t_i^s, t_i^h - t_i^s)$
D^{-OM-}	$-$	O	$\rightarrow$ M	$-$	$\prod_{i \in D^{-OM-}} \frac{\sum_{\tau^{sh}=0}^{\bar{B}(t_i^h)-t_i^s} p(t_i^s, \tau^{sh}) f^{sh}(\tau^{sh})}{\sum_{\tau^{sh}=0}^{\bar{B}(t_i^h)-t_i^s} f^{sh}(\tau^{sh})}$
D^{OMO-}	O $\rightarrow$	M	O	$-$	$\prod_{i \in D^{OMO-}} \frac{\sum_{\tau^{es}=0}^{\bar{B}(t_i^s)-t_i^e} p(t_i^e+\tau^{es}, t_i^h-t_i^e-\tau^{es}) f^{es}(\tau^{es})}{\sum_{\tau^{es}=0}^{\bar{B}(t_i^s)-t_i^e} f^{es}(\tau^{es})}$
D^{MMO-}	M	M $\leftarrow$	O	$-$	$\prod_{i \in D^{MMO-}} \frac{\sum_{\tau^{sh}=t_i^h-\bar{B}(t_i^s)}^{t_i^h-\underline{B}(t_i^s)} p(t_i^h-\tau^{sh}, \tau^{sh}) f^{sh}(\tau^{sh})}{\sum_{\tau^{sh}=t_i^h-\bar{B}(t_i^s)}^{t_i^h-\underline{B}(t_i^s)} f^{sh}(\tau^{sh})}$

Table 3: **Categories of the likelihood of serial interval L_s.** The 1st category corresponds to the pattern of $(v_i \in D^{-OO-}) \cap (i \in D^{-O--})$. O, M and $-$ represent observed time events, missing time events, and cases for which there are both observed and missing time events, respectively. $\underline{B}(t_i^s) = \max(t_i^e, t_{v_i^n}^s, t_{v_i^n}^e)$ is the lower bound of t_i^s. $\bar{B}(t_i^s) = \min(t_i^h, t_i^d, t_{u_i^n}^e, t_{u_i^n}^s, t_{u_i^n}^h, t_{u_i^n}^d, t_{exit})$ is the upper bound of t_i^s. $\bar{B}(t_i^h) = \min(t_i^d, t_{exit})$ is the upper bound of t_i^s. v_i^n and u_i^n represent all the 'ancestors' and 'descendants' of i, respectively. t_{exit} represents the time when the isolation wards were empty.

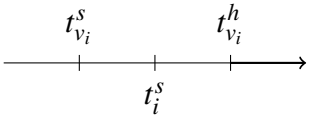
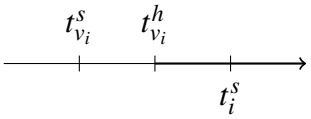
$L_s(\theta \mid D)$	1D timeline
$\prod_{\substack{v_i \in D^{-OO-} \\ i \in D^{-O--}}} \hat{s}(t_i^s - t_{v_i}^s, t_{v_i}^h - t_{v_i}^s)$	
$= \prod_{\substack{v_i \in D^{-OO-} \\ i \in D^{-O--} \\ t_{v_i}^h > t_i^s}} \frac{s(t_i^s - t_{v_i}^s)}{1-\varepsilon\left[1-S(t_{v_i}^h - t_{v_i}^s)\right]}$	
$\prod_{\substack{v_i \in D^{-OO-} \\ i \in D^{-O--} \\ t_{v_i}^h \le t_i^s}} \frac{(1-\varepsilon)s(t_i^s - t_{v_i}^s)}{1-\varepsilon\left[1-S(t_{v_i}^h - t_{v_i}^s)\right]}$	

Table 4: **Categories of the likelihood of serial interval L_s, Cont.** The 2nd and 3rd categories correspond to the pattern of $(v_i \in D^{-OO-}) \cap (i \in D^{OM--})$ and $(v_i \in D^{-OO-}) \cap (i \in D^{MMO-})$, respectively.

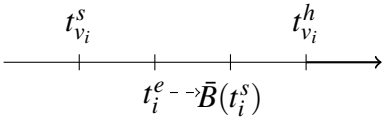
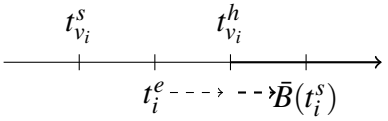
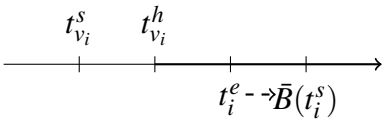
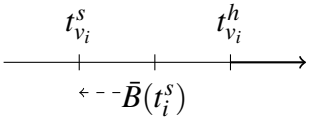
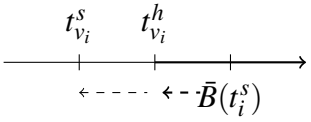
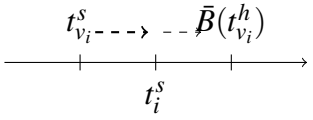
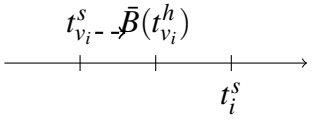
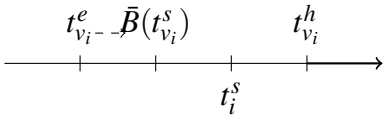
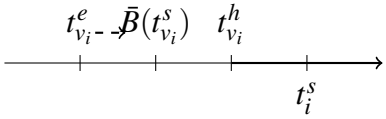
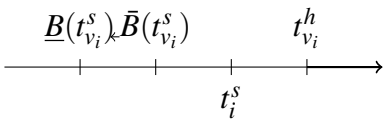
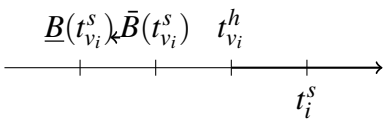
$L_s(\theta \mid D)$	1D timeline
$\prod_{\substack{v_i \in D^{-OO-} \\ i \in D^{OM--}}} \frac{\sum_{\tau^{es}=0}^{\bar{B}(t_i^s)-t_i^e} \hat{s}(t_i^e+\tau^{es}-t_{v_i}^s, t_{v_i}^h-t_{v_i}^s) f^{es}(\tau^{es})}{\sum_{\tau^{es}=0}^{\bar{B}(t_i^s)-t_i^e} f^{es}(\tau^{es})}$	
$= \prod_{\substack{v_i \in D^{-OO-} \\ i \in D^{OM--} \\ t_{v_i}^h > \bar{B}(t_i^s)}} \frac{\sum_{\tau^{es}=0}^{\bar{B}(t_i^s)-t_i^e} \frac{s(t_i^e+\tau^{es}-t_{v_i}^s)}{1-\varepsilon\left[1-S(t_{v_i}^h-t_{v_i}^s)\right]} f^{es}(\tau^{es})}{\sum_{\tau^{es}=0}^{\bar{B}(t_i^s)-t_i^e} f^{es}(\tau^{es})}$	
$\prod_{\substack{v_i \in D^{-OO-} \\ i \in D^{OM--} \\ t_i^e < t_{v_i}^h \le \bar{B}(t_i^s)}} \frac{\sum_{\tau^{es}=0}^{t_{v_i}^h-t_i^e} \frac{s(t_i^e+\tau^{es}-t_{v_i}^s)}{1-\varepsilon\left[1-S(t_{v_i}^h-t_{v_i}^s)\right]} f^{es}(\tau^{es}) + \sum_{\tau^{es}=t_{v_i}^h-t_i^e}^{\bar{B}(t_i^s)-t_i^e} \frac{(1-\varepsilon)s(t_i^e+\tau^{es}-t_{v_i}^s)}{1-\varepsilon\left[1-S(t_{v_i}^h-t_{v_i}^s)\right]} f^{es}(\tau^{es})}{\sum_{\tau^{es}=0}^{\bar{B}(t_i^s)-t_i^e} f^{es}(\tau^{es})}$	
$\prod_{\substack{v_i \in D^{-OO-} \\ i \in D^{OM--} \\ t_{v_i}^h \le t_i^e}} \frac{\sum_{\tau^{es}=0}^{\bar{B}(t_i^s)-t_i^e} \frac{(1-\varepsilon)s(t_i^e+\tau^{es}-t_{v_i}^s)}{1-\varepsilon\left[1-S(t_{v_i}^h-t_{v_i}^s)\right]} f^{es}(\tau^{es})}{\sum_{\tau^{es}=0}^{\bar{B}(t_i^s)-t_i^e} f^{es}(\tau^{es})}$	
$\prod_{\substack{v_i \in D^{-OO-} \\ i \in D^{MMO-}}} \frac{\sum_{\tau^{sh}=t_i^h-\bar{B}(t_i^s)}^{t_i^h-t_{v_i}^s} \hat{s}(t_i^h-\tau^{sh}-t_{v_i}^s, t_{v_i}^h-t_{v_i}^s) f^{sh}(\tau^{sh})}{\sum_{\tau^{sh}=t_i^h-\bar{B}(t_i^s)}^{t_i^h-t_{v_i}^s} f^{sh}(\tau^{sh})}$	
$= \prod_{\substack{v_i \in D^{-OO-} \\ i \in D^{MMO-} \\ t_{v_i}^h > \bar{B}(t_i^s)}} \frac{\sum_{\tau^{sh}=t_i^h-\bar{B}(t_i^s)}^{t_i^h-t_{v_i}^s} \frac{s(t_i^h-\tau^{sh}-t_{v_i}^s)}{1-\varepsilon\left[1-S(t_{v_i}^h-t_{v_i}^s)\right]} f^{sh}(\tau^{sh})}{\sum_{\tau^{sh}=t_i^h-\bar{B}(t_i^s)}^{t_i^h-t_{v_i}^s} f^{sh}(\tau^{sh})}$	
$\prod_{\substack{v_i \in D^{-OO-} \\ i \in D^{MMO-} \\ t_{v_i}^h \le \bar{B}(t_i^s)}} \frac{\sum_{\tau^{sh}=t_i^h-\bar{B}(t_i^s)}^{t_i^h-t_{v_i}^h} \frac{(1-\varepsilon)s(t_i^h-\tau^{sh}-t_{v_i}^s)}{1-\varepsilon\left[1-S(t_{v_i}^h-t_{v_i}^s)\right]} f^{sh}(\tau^{sh}) + \sum_{\tau^{sh}=t_i^h-t_{v_i}^h}^{t_i^h-t_{v_i}^s} \frac{s(t_i^h-\tau^{sh}-t_{v_i}^s)}{1-\varepsilon\left[1-S(t_{v_i}^h-t_{v_i}^s)\right]} f^{sh}(\tau^{sh})}{\sum_{\tau^{sh}=t_i^h-\bar{B}(t_i^s)}^{t_i^h-t_{v_i}^s} f^{sh}(\tau^{sh})}$	

Table 5: **Categories of the likelihood of serial interval L_s, Cont.** The 4th, 5th and 6th categories correspond to the pattern of $(v_i \in D^{-OM-}) \cap (i \in D^{-O--})$, $(v_i \in D^{OMO-}) \cap (i \in D^{-O--})$ and $(v_i \in D^{MMO-}) \cap (i \in D^{-O--})$, respectively.

$L_s(\theta \mid D)$	1D timeline
$\prod_{\substack{v_i \in D^{-OM-} \\ i \in D^{-O--}}} \frac{\sum_{\tau^{sh}=0}^{\bar{B}(t^h_{v_i})-t^s_{v_i}} \hat{s}(t^s_i - t^s_{v_i}, \tau^{sh}) f^{sh}(\tau^{sh})}{\sum_{\tau^{sh}=0}^{\bar{B}(t^h_{v_i})-t^s_{v_i}} f^{sh}(\tau^{sh})}$	
$= \prod_{\substack{v_i \in D^{-OM-} \\ i \in D^{-O--} \\ t^s_i < \bar{B}(t^h_{v_i})}} \frac{\sum_{\tau^{sh}=0}^{t^s_i - t^s_{v_i}} \frac{(1-\varepsilon) s(t^s_i - t^s_{v_i})}{1-\varepsilon\left[1-S(\tau^{sh})\right]} f^{sh}(\tau^{sh}) + \sum_{\tau^{sh}=t^s_i - t^s_{v_i}}^{\bar{B}(t^h_{v_i})-t^s_{v_i}} \frac{s(t^s_i - t^s_{v_i})}{1-\varepsilon\left[1-S(\tau^{sh})\right]} f^{sh}(\tau^{sh})}{\sum_{\tau^{sh}=0}^{\bar{B}(t^h_{v_i})-t^s_{v_i}} f^{sh}(\tau^{sh})}$	
$\prod_{\substack{v_i \in D^{-OM-} \\ i \in D^{-O--} \\ t^s_i \geq \bar{B}(t^h_{v_i})}} \frac{\sum_{\tau^{sh}=0}^{\bar{B}(t^s_{v_i})-t^s_{v_i}} \frac{(1-\varepsilon) s(t^s_i - t^s_{v_i})}{1-\varepsilon\left[1-S(\tau^{sh})\right]} f^{sh}(\tau^{sh})}{\sum_{\tau^{sh}=0}^{\bar{B}(t^h_{v_i})-t^s_{v_i}} f^{sh}(\tau^{sh})}$	
$\prod_{\substack{v_i \in D^{OMO-} \\ i \in D^{-O--}}} \frac{\sum_{\tau^{es}=0}^{\bar{B}(t^s_{v_i})-t^e_{v_i}} \hat{s}(t^s_i - t^e_{v_i} - \tau^{es}, t^h_{v_i} - t^e_{v_i} - \tau^{es}) f^{es}(\tau^{es})}{\sum_{\tau^{es}=0}^{\bar{B}(t^s_{v_i})-t^e_{v_i}} f^{es}(\tau^{es})}$	
$= \prod_{\substack{v_i \in D^{OMO-} \\ i \in D^{-O--} \\ t^s_i < t^h_{v_i}}} \frac{\sum_{\tau^{es}=0}^{\bar{B}(t^s_{v_i})-t^e_{v_i}} \frac{s(t^s_i - t^e_{v_i} - \tau^{es})}{1-\varepsilon\left[1-S(t^h_{v_i} - t^e_{v_i} - \tau^{es})\right]} f^{es}(\tau^{es})}{\sum_{\tau^{es}=0}^{\bar{B}(t^s_{v_i})-t^e_{v_i}} f^{es}(\tau^{es})}$	
$\prod_{\substack{v_i \in D^{OMO-} \\ i \in D^{-O--} \\ t^s_i \geq t^h_{v_i}}} \frac{\sum_{\tau^{es}=0}^{\bar{B}(t^s_{v_i})-t^e_{v_i}} \frac{(1-\varepsilon) s(t^s_i - t^e_{v_i} - \tau^{es})}{1-\varepsilon\left[1-S(t^h_{v_i} - t^e_{v_i} - \tau^{es})\right]} f^{es}(\tau^{es})}{\sum_{\tau^{es}=0}^{\bar{B}(t^s_{v_i})-t^e_{v_i}} f^{es}(\tau^{es})}$	
$\prod_{\substack{v_i \in D^{MMO-} \\ i \in D^{-O--}}} \frac{\sum_{\tau^{sh}=t^h_{v_i} - \bar{B}(t^s_{v_i})}^{t^h_{v_i} - \underline{B}(t^s_{v_i})} \hat{s}(t^s_i - t^h_{v_i} + \tau^{sh}, \tau^{sh}) f^{sh}(\tau^{sh})}{\sum_{\tau^{sh}=t^h_{v_i} - \bar{B}(t^s_{v_i})}^{t^h_{v_i} - \underline{B}(t^s_{v_i})} f^{sh}(\tau^{sh})}$	
$= \prod_{\substack{v_i \in D^{MMO-} \\ i \in D^{-O--} \\ t^s_i < t^h_{v_i}}} \frac{\sum_{\tau^{sh}=t^s_i - \underline{B}(t^s_{v_i})}^{t^h_{v_i} - \underline{B}(t^s_{v_i})} \frac{s(t^s_i - t^h_{v_i} + \tau^{sh})}{1-\varepsilon\left[1-S(\tau^{sh})\right]} f^{sh}(\tau^{sh})}{\sum_{\tau^{sh}=t^s_i - \underline{B}(t^s_{v_i})}^{t^h_{v_i} - \underline{B}(t^s_{v_i})} f^{sh}(\tau^{sh})}$	
$\prod_{\substack{v_i \in D^{MMO-} \\ i \in D^{-O--} \\ t^s_i \geq t^h_{v_i}}} \frac{\sum_{\tau^{sh}=t^h_{v_i} - \bar{B}(t^s_{v_i})}^{t^h_{v_i} - \underline{B}(t^s_{v_i})} \frac{(1-\varepsilon) s(t^s_i - t^h_{v_i} + \tau^{sh})}{1-\varepsilon\left[1-S(\tau^{sh})\right]} f^{sh}(\tau^{sh})}{\sum_{\tau^{sh}=t^h_{v_i} - \bar{B}(t^s_{v_i})}^{t^h_{v_i} - \underline{B}(t^s_{v_i})} f^{sh}(\tau^{sh})}$	

Bayesian parameter estimation

Bayesian inference with Markov chain Monte Carlo (MCMC) was adopted to obtain parameter estimates. This approach resembles the data augmentation strategy used elsewhere to estimate the source of infection during MERS-CoV transmission (Cauchemez et al. 2016). The prior distribution was assumed to be non-informative and flat. Due to the difficulties in simultaneous direct sampling of the posterior distribution of parameters and the source of infection, the distribution was approximated using the Metropolis-Hastings Markov chain Monte Carlo (MH-MCMC) algorithm. In each MH-MCMC iteration step, parameters were updated and then the source of infection was sampled given the proposed parameters. To improve mixing, the proposed distribution of the infector was not flat as assumed for the prior distribution of other parameters, but was weighted according to the transmission probability of each of the six listed possible infectors.

The simulation using the MH-MCMC algorithm was performed by 1,000,000 iterations with thinning of every 500th sample to reduce autocorrelation. Despite visually converged chains suggesting that the samples are well-mixed, the sampling was replicated in four independent chains with a burn-in length of half in each chain (Gelman et al. 2013). A total of 4,000 samples represented the posterior distribution. The algorithm was implemented in R version 3.4.3 (Kite-Eating Tree) on Mac OSX and R version 3.4.4 (Someone to Lean On) on Linux (Available from: https://github.com/imlouischan/ebola-ng).

Model selection and alternative time-dependent assumption

Regarding the combinations of alternative distributions used in observing the number of secondary cases (i.e. Poisson, Geometric or Negative binomial) and unbiased serial interval (i.e. Gamma and Weibull), six possible combinations existed. The best combination was selected according to three criteria, namely, Akaike information criterion (AIC), Bayesian information criterion (BIC) and model posterior distribution. We obtain a better model when AIC or BIC are lower. AIC and BIC are point estimators that take the best sample among 4,000. In contrast, we obtain a better model with a higher model posterior. Determining the model posterior given a uniform prior is as simple as calculating the weight of marginal likelihoods. The marginal likelihoods are calculated using the posterior harmonic mean (PHM) estimator (Vyshemirsky & Girolami 2007).

Besides comparing six possible combinations for L_r and L_s, we also evaluated the impact of time dependence in the reproduction number on our estimate of the protective effect of case isolation. That is, identifying the best fitted combination, we have additionally compared the reproduction number with and without exponential decline as a function of time. In the time-independent model, we set $\delta = 0$.

Results

The time from onset of illness to hospitalization was fitted better by Weibull than gamma distribution (see Fig 2B), yielding estimates of the mean (and SD) at 4.0 (2.3) days, used to reconstruct the timeline. Fig 1 shows the reconstructed timelines consisting of both the observed time events and the probabilistically reconstructed (missing) time events.

Employing Bayesian inference with MCMC, all three criteria (i.e. AIC, BIC and model posterior) indicated that the best model matches our assumptions, namely being that in which the number of secondary cases per primary cases is geometrically distributed and the unbiased serial interval is gamma-distributed (Table 6). The corresponding model posterior is 52.0%. There was one unobserved branch of the transmission tree, and the possible primary case associated with this branch was probabilistically identified and statistically ranked by the posterior distribution (Fig 3A). The most likely primary case of Case 20 was Case 11 with probability 53.1%, followed by Case 7 with 26.3% and Case 10 with 6.3%.

Table 6: **Model comparison employing different distributions.** Six possible models are compared using three different criterion values. AIC and BIC were calculated using the maximum values of log-likelihood $\ln(\hat{L})$ among MCMC samples. Model posteriors were calculated using marginal likelihoods.

Secondary cases	Serial interval	$\ln(\hat{L})$	AIC	BIC	Posterior (%)
Geometric	Gamma	-65.3	142.6	148.6	52.0
Geometric	Weibull	-67.3	146.6	152.6	19.7
Poisson	Gamma	-67.9	147.9	153.9	4.5
Poisson	Weibull	-70.1	152.1	158.1	0.8
Negative binomial	Gamma	-64.4	142.8	149.8	20.7
Negative binomial	Weibull	-66.4	146.8	153.7	2.5

The individual protective effect of case isolation ε_i and the protective effect without delay in case isolation ε are summarized in (Fig 3B). The mean of ε was estimated as 39.7% (95% credible interval (CI): 2.4%-82.1%). Smaller estimates of the individual protective effect were observed in Cases 5, 6 and 12, for which the date of onset of illness was reconstructed using the incubation period distribution (Fig 3B), while the individual effect was similarly distributed to ε (i.e. without time delay).

Given the posterior shape and scale parameters of gamma distribution, Fig 3C shows the unbiased serial interval distribution (i.e. in the absence of case isolation). The mean and SD were 15.3 (95%CI: 14.2-16.6) and 2.3 (95%CI: 1.6-3.5) days. In comparison with the biased sample mean, there was a lag of 0.5 day due to effective case isolation (see Fig 4). The distribution of the secondary case number per primary case, the mean of which is the reproduction number at time zero, is shown in Fig 3D. Assuming that the reproduction number declined over time, the mean at time zero and the rate of exponential decline were estimated to be 10.0 (95%CI: 3.0-18.5) and 0.14 (95%CI: 0.07-0.23), respectively.

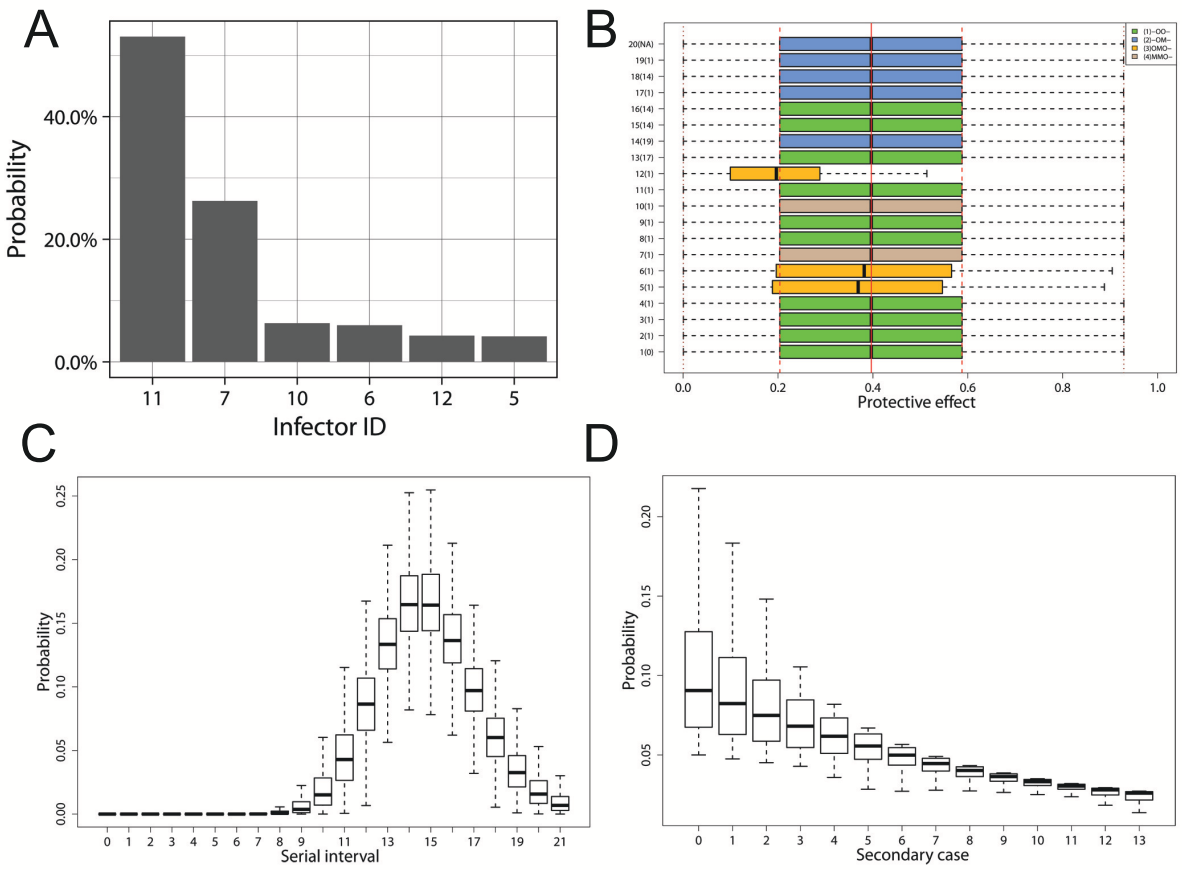


Figure 3: **The posterior distribution of estimated parameters.** (A) The posterior distribution of the primary case of Case 20. Case 11 was considered the most likely primary case with a probability of 53.1%, and was followed by Case 7 with 26.3% and Case 10 with 6.3%. (B) The posterior distribution of the individual protective effect of case isolation, ε_i, for each case. The green, yellow, blue and brown boxes show four observed patterns. O represents observed, M indicates missing and $-$ indicates unused (i.e. observed or missing), and the order of events follows the dates of exposure, onset of illness, hospitalization and death. The third pattern in yellow showed significantly lower individual protective effect than the other three patterns. The red solid, dashed and dotted lines show the mean, interquartile range, and minimum and maximum of the protective effect of case isolation without delay, ε. (C) The posterior distribution of unbiased serial intervals $s(\tau)$ with mean at 15.3 (95%CI: 14.2-16.6) and SD at 2.3 (95%CI: 1.6-3.5) days. The unbiased distribution represents the situation in the absence of case isolation. The shape and scale parameters of the gamma distribution were estimated as 45.0 (95%CI: 19.6-92.0) and SD at 0.3 (95%CI: 0.2-0.8), respectively. (D) The posterior distribution of the secondary case number per primary case $p(r)$, given the mean as the reproduction number at time zero, which was estimated to be 10.0 (95%CI: 3.0-18.5). The distributions shown in the figure were derived from the best fit model that employs geometric and gamma distributions, respectively, to describe the distributions of the observed number of secondary cases per primary case and the unbiased serial interval. The posterior distributions are shown using black boxes, ranging from the lower to the upper quartiles along with the median.

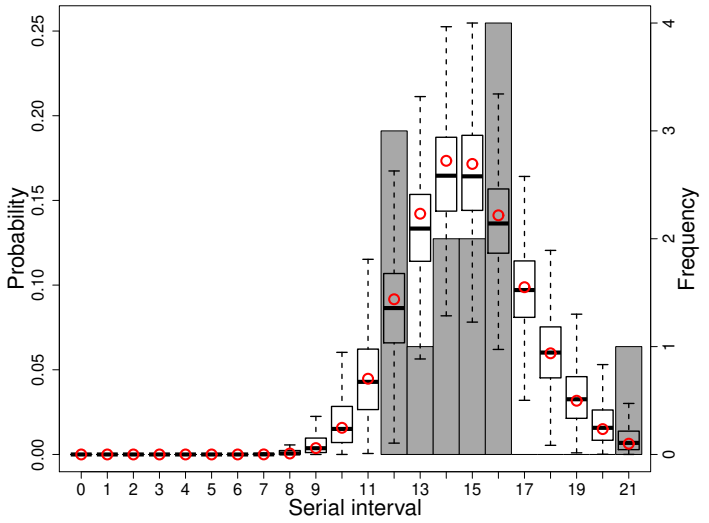


Figure 4: **The serial interval distribution.** The unbiased distribution $s(\tau)$ (i.e. in the absence of case isolation) is represented by using boxes as shown in Fig 3C. The mean and SD were 15.3 (95%CI: 14.2-16.6) and 2.3 (95%CI: 1.6-3.5) days. The corresponding biased distribution $\hat{s}(\tau)$ (i.e. in the presence of case isolation) is not shown due to various individual time from illness onset to hospitalization. As an alternative, the frequency of observed serial interval (i.e. biased / in the presence of case isolation) is shown as histogram using gray bars. The sample mean and SD were 14.8 and 2.5 days. The red circles represent the MLE fitted values by employing gamma distribution, similar to Fig 2B. The mean and SD were 15.3 and 2.3 days.

Fig 5 shows pairwise distributions of the parameter posterior and the marginals. There are two interesting pairs of parameters. First, the shape and scale parameters of unbiased serial interval distribution are strongly negatively correlated and exhibit a banana shaped distribution. Second, a positive correlation was observed between the reproduction number at time zero R_0 and the protective effect without time delay ε, with the correlation coefficient being 0.418.

Sensitivity analysis was performed to examine the dependence of the estimated protective effect of case isolation without time delay ε to variable assumption on the reproduction number. Specifically, we compared two scenarios, involving time-dependent and constant reproduction numbers, using $R(t) = R_0 e^{-\delta t}$, where we assumed $\delta > 0$ for the time-dependent model and $\delta = 0$ for the constant model. Table 7 and Fig 6 show the comparison of the posterior parameter distribution between these two model scenarios. Using two models, the shape and scale parameters of the unbiased serial interval and the identified primary case of Case 20 were similarly distributed, but exert a strong negative effect on the distributions of the reproduction number R_0 and protective effect of case isolation without time delay ε. With $\delta = 0$, the constant reproduction number was estimated as 3.2 (95%CI: 0.8-16.4), and the mean protective effect of case isolation was 67.2% (95%CI: 3.7%-94.2%). Nevertheless, all three model comparison criteria indicated that $\delta > 0$ fits better than $\delta = 0$. The lower panel of Fig 6 illustrates the estimated reproduction number over time, of which the declining mean crosses the criticality of 1 on Day 16.

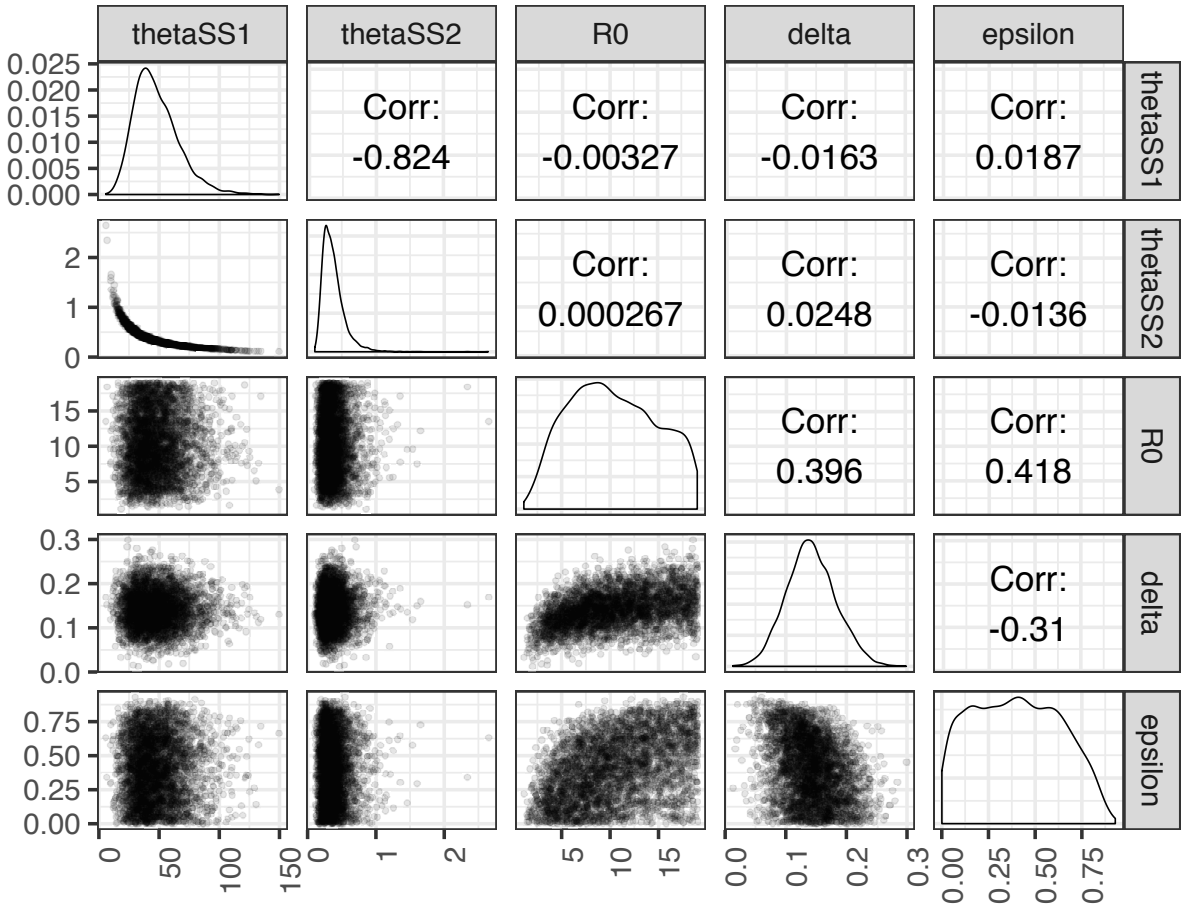


Figure 5: **The pairwise distributions of estimated parameters.** The posterior was derived from the best fit model that employs geometric and gamma distributions, respectively, to describe the distributions of the observed number of secondary cases per primary case and the unbiased serial interval. The parameters include the shape (thetaSS1) and scale (thetaSS2) of the unbiased serial interval, the reproduction number at time zero (R0), the exponentially decreasing rate of secondary transmission (delta) and the protective effect of case isolation without time delay (epsilon). The diagonal plots show the marginal distributions. The lower diagonal reflects dependence between parameters by scatter plots. The upper diagonal indicates correlation between parameters.

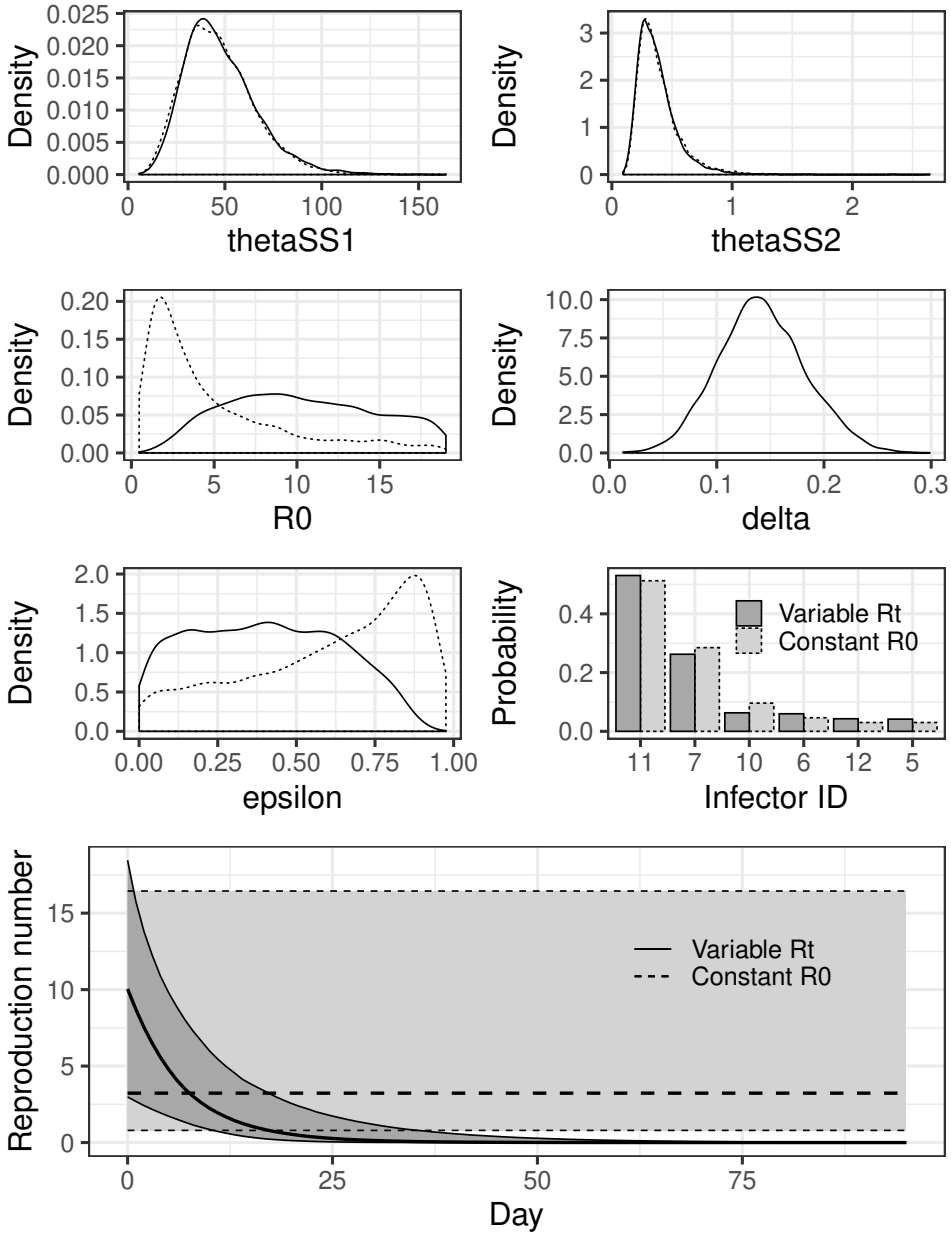


Figure 6: **The marginal parameter posterior distributions of two different model scenarios.** The solid and dashed lines show the model scenario in which the reproduction number decreased exponentially and remained constant over time, respectively. The parameters include the shape (thetaSS1) and scale (thetaSS2) parameters of unbiased serial interval, the reproduction number at time zero (R0), the exponentially decreasing rate of the reproduction number (delta), the protective effect of case isolation without time delay (epsilon) and an identified primary case of Case 20. The lower panel shows the estimated reproduction number during the outbreak. The thicker lines show the mean and the gray regions represent the 95%CI.

Table 7: **The parameters estimated in two scenarios.** Scenario 1 and 2 represent that the reproduction number decreased exponentially and remained constant over time, respectively. The parameters include the protective effect of case isolation without time delay (epsilon), the reproduction number at time zero (R0), the exponentially decreasing rate of secondary transmission (delta) and the shape (thetaSS1) and scale (thetaSS2) of the unbiased serial interval. The corresponding mean and SD of the unbiased serial interval $s(\tau)$ are included.

-	Scenario 1	Scenario 2
Parameters	Mean (95%CI)	Mean (95%CI)
epsilon (%)	39.7 (2.4-82.1)	67.2 (3.7-94.2)
R0	10.0 (3.0-18.5)	3.2 (0.8-16.4)
delta (day^{-1})	0.14 (0.07-0.23)	0
thetaSS1	45.0 (19.5-92.0)	44.6 (18.2-88.2)
thetaSS2	0.3 (0.2-0.8)	0.3 (0.2-0.9)
$s(\tau)$ mean (day)	15.3 (14.2-16.6)	15.3 (14.2, 16.6)
$s(\tau)$ SD (day)	2.3 (1.6-3.5)	2.3 (1.6, 3.6)

Discussion

The present study estimated the protective effect of case isolation during the 2014 EVD outbreak in Nigeria, using partially observed contact tracing data for a total of 20 cases. We developed a statistical model that features the time interval between onset of illness for the primary case and the secondary case and also the number of secondary transmissions per primary case, while addressing problems of missing data for time events (e.g. date of hospitalization) and partially reconstructing the transmission dynamics. We have shown that the observed serial intervals are likely shorter than the unbiased serial intervals, and also that case isolation likely resulted in the observed number of secondary cases per single primary case being smaller than it would otherwise have been. Assuming that the effective reproduction number exponentially declines over time, the protective effect of case isolation in reducing secondary transmission was estimated at 39.7% (95%CI: 2.4%-82.1%), successfully quantifying the contribution of case isolation to the transmission dynamics in Nigeria during 2014. While theoretical aspects of study designs for determining vaccine effect have been broadly discussed (Halloran et al. 2010), to our knowledge, the present study is the first to apply the proposed model to EVD data with reconstruction of both time events and transmission network.

The most important take-home message from the present study is that the effectiveness of case isolation could be explicitly estimated from observational data, provided the transmission network is known (at least partially) and so too are the dates of onset of illness and/or hospitalization. The estimated protective effect ε is interpreted as a relative reduction of the secondary transmission rate, representing the transient reduction in the frequency of secondary transmission. Since the delay from onset of illness to case isolation varies by case, we have also estimated the protective effect of each individual ε_i, where its maximum value ε is considered as the protective effect without time delay. From the formulation used for this estimation, we have shown that the individual protective effect depends on the rapidity of case finding and public health response, as well as the individual observed patterns of data in our model. However, the estimated ε varied greatly according to model assumptions regarding the time dependence of the reproduction number. While the mean value of ε was 39.7%, with an exponentially declining reproduction number, the value with a stationary (i.e. time-independent) reproduction number reached 67.2%. In both assumptions, the uncertainty bound was very broad. Although all three model selection criteria favoured the exponentially declining reproduction number, the discrepancy here is caused by the existence of assumptions that are not fully supported, and thus, the protective effect of case isolation still involves a degree of uncertainty. In Nigeria, contact tracing and case isolation were the mainstream interventions in 2014, and in hospitals, standard precautions were practiced by healthcare professionals (Folarin et al. 2016, Shuaib et al. 2014). The present study indicates that, although the uncertainty bound remains wide, the practice of case isolation has successfully reduced secondary transmissions by 40%-67%.

Additionally, it should be noted that the proposed model also estimated the unbiased reproduc-

tion number, i.e., the reproduction number without case isolation. Assuming an exponential decline, the reproduction number at time zero was estimated at 10.0 (95%CI: 3.0-18.5), while under an assumption of a stationary transmissibility, the reproduction number at time zero was estimated at about one third of this level, i.e. 3.2 (95%CI: 0.8-16.4). The sample mean of observed (and biased) number of secondary transmissions was 0.9 (SD: 3.0). These cannot be directly compared against the basic reproduction number estimates, i.e., ranging from about 1.3 to 2.6, from a larger scale epidemic (Shen et al. 2015, Lewnard et al. 2014, Chowell & Nishiura 2014, Nishiura & Chowell 2015), but at least our estimates demonstrate that the unbiased reproduction number that adjusts the protective effect of case isolation can exceed the series of estimates derived from models that did not explicitly take into account the impact of case isolation.

We have also estimated the unbiased length of the serial interval, with the estimated mean at 15.3 (95%CI: 14.2-16.6) days and SD at 2.3 (95%CI: 1.6-3.5) days. The sample mean and SD of the serial interval in the presence of case isolation were 14.8 and 2.5 days, indicating that although the mean was not greatly shortened it was at least shortened by half a day (see Fig 4). The two-week interval resembled the published estimates (Chowell & Nishiura 2014, Nishiura & Chowell 2015). Two technical points should be further discussed. First, we adopted published estimates of the incubation period distribution from a larger epidemic (WHO Ebola Response Team 2014). The use has led to some differences among individuals in protective effects (i.e. Fig 3B) that reflect their observation patterns. The reconstruction of dates of the onset of illness using the incubation period (Type 3 in Table 1) and the reconstruction using the time from onset of illness to hospitalization (Types 2 and 4 in Table 1) have different characteristics due to the distributions from different samples and outbreaks. However, we believe that the way we addressed those distributions might be the best practice, simultaneously addressing issues of limited sample size and effectively using the available dataset. Second, individual variations existed in the number of secondary transmissions. In fact, the index case was the main patient who infected 13 individuals and became a super-spreader. The abovementioned big difference between time-dependent and time-independent reproduction numbers can be partially explained by the existence of this index case. It has also contributed to forming a heavy tail in the distribution of secondary transmissions.

The present study has various limitations that must be discussed. First, our model assumed that infectiousness starts from the date of onset of illness, which is not broadly applicable, e.g. HIV/AIDS. Second, strictly speaking, serial interval and generation time are used interchangeably in many studies, but the present study explicitly handled the time of onset of illness and strictly used the serial interval. Third, the serial interval distribution was assumed to be stationary, which means it is independent of calendar time throughout the outbreak. On the other hand, the reproduction number was allowed to vary with calendar time to capture the transmission dynamics. Strictly speaking, both would vary with time and infection age. Fourth, the application

of the model to this dataset from Nigeria was limited by a few missing pieces of information, namely in some cases either the date of onset of illness or hospitalization of each individual was not observed. The time event reconstruction was relatively simple and the reconstruction was limited to one dimension. However, surveillance data of individuals during unexpected outbreaks are often imperfectly collected and require complicated reconstruction. In the future we plan to extend our modelling approach by applying it to datasets with higher dimensional imputations.

In conclusion, the present study has shown that the case isolation of the 2014 EVD outbreak in Nigeria was effective, and the public health interventions by the Nigerian government should be commended for successfully controlling the outbreak. Accounting for the protective effect of case isolation, we demonstrate that the observed number of secondary cases per primary case and serial interval were, respectively, reduced and shortened.

Chapter 2: Assessment of case isolation and transmission dynamics: the 2014 Ebola virus disease outbreak in Pujehun District, Sierra Leone

Introduction

The largest Ebola virus disease (EVD) epidemic initially occurred in Guinea and spread to neighboring counties, Liberia and Sierra Leone (CDC 2019). Apart from the three most-affected countries, outbreaks also occurred in Nigeria, Democratic Republic of the Congo (DRC), Senegal and Mali (WHO Ebola Response Team 2014). The EVD caused 28,652 cases and 11,325 deaths worldwide in 2014-2016 (CDC 2019). In Pujehun District, Sierra Leone, a local outbreak occurred since 7th July 2014 (Day 0) as an index case developed symptoms (Ajelli et al. 2015). Due to effective intervention measures including contact tracing and case isolation, the final size was fortunately limited to 49 cases, including 9 imported cases (Ajelli et al. 2015). The district was declared Ebola-free by local authorities on 10th January 2015 (Day 187) (Ajelli et al. 2015).

It is now well-known that the outbreak was caused by the Ebola virus, which was first identified in Zaire, 1976 (Bowen et al. 1977, Johnson et al. 1977, Pattyn et al. 1977). However, at the beginning of the 2014 EVD epidemic, it was an unknown disease with an unknown transmission route. After the outbreak occurred, the characteristic of EVD was intensively investigated and it was found that patients would become infectious since the onset of symptoms including fever, headache, fatigue, diarrhea, vomiting, stomach pain, unexplained bleeding or bruising, and muscle pain. An infectious patient would then create a chain reaction due to human-to-human transmission via direct contact with bodily fluid, particularly with the handling of deceased patients during burials. As the vaccine was not available, isolation of cases was instead one of the most important management. A quick response to hospital admission was essentially associated with the incubation period (time from exposure to symptom onset), time from symptom onset to hospitalization, and serial interval (or generation time) between primary and secondary cases. All epidemiological information is useful for creating intervention strategies, and however, was barely known at the beginning of the outbreak.

Many epidemiological modeling studies were conducted to assess performance of non-pharmaceutical interventions (Pandey et al. 2014, Lewnard et al. 2014, Kucharski et al. 2015a,b, 2016, Lokuge et al. 2016, Merler et al. 2015, Shen et al. 2015, Drake et al. 2015, Chowell et al. 2015, Fang et al. 2016, Yamin et al. 2015, Rivers et al. 2014, Webb et al. 2015, Camacho et al. 2015). One of the main strategies supporting the control of EVD was building EVD treatment centers and the corresponding effectiveness was evaluated (Kucharski et al. 2015a,b, Pandey et al. 2014, Lokuge et al. 2016, Merler et al. 2015, Lewnard et al. 2014). However, such as evaluations focused on the population level, which was similar to measuring household secondary attack

risk (Fang et al. 2016). A study also evaluated the impact of intervention measures including the probability of hospitalization, the time delay of contact tracing, probability of safe burials and EVD bed capacity (Ajelli et al. 2015). However, there is a lack of studies on individual level assessing secondary transmission and evaluating the impact of case isolation during the 2014 EVD outbreak.

Here, we developed a statistical model to demonstrate the protective effect of case isolation during the local outbreak in Pujehun District, Sierra Leone. We revealed unbiased quantities including serial interval and reproduction number in the absence of control interventions, by removing the impact of relative reduction on transmission risk given case isolation. More importantly, missing data including individual time events such as symptom onset and hospitalization, as well as the source of infection were imputed using both maximum likelihood estimation (MLE) and Bayesian inference framework with Markov chain Monte Carlo (MCMC) sampling. Cases with unknown hospitalization were also evaluated and estimated with the protective effect.

Materials and Methods

Outbreak description

A dataset from the 2014 EVD outbreak in Pujehun District, Sierra Leone was reconstructed and published in a study (Ajelli et al. 2015). The data were obtained by analyzing hospital registers, contact tracing forms and interviews of healthcare workers, survived cases and relatives of deceased cases. The local outbreak involved in total 49 cases confirmed and reported from 7th July 2014 (Day 0), on which the index case developed symptoms, to 10th January 2015 (Day 187), on which the district was declared Ebola-free by local authorities. There were 18% (9/49) of imported cases and 82% (40/49) non-imported cases who were infected by the other cases found in the district. The hospitalization rate was 90% (44/49) and the case fatality rate (CFR) was 86% (42/49). The outbreak occurred in a geographically isolated and low population density region, mainly including two areas called Zimmi Town and Dumagbe Village. It demonstrated a unique example of successful control in a rural area. The healthcare system in the district consists of one hospital with a capacity of 87 beds and 74 peripheral health units (PHUs). Two Ebola health centers (EHCs) including one in the hospital, were operative during the outbreak and provided in total 20 beds for suspected EVD cases. The implemented intervention measures included case detection, isolation, contact tracing (including 250 contact tracers), safe burials, population awareness and compliance with intervention policies. One of the critical factors of human-to-human transmission at the initial stage of the outbreak was traditional burials, including practices of washing, touching or kissing dead bodies. Thus, hospitalized cases were safely buried after the arrival of EVD burial teams who conducted disinfection procedures for burial ceremonies.

In this article, we used two types of established data, individual time events and primary cases of infection. The individual time events are written as $\mathbf{t}$, of which five types including dates of exposure, symptom onset, hospitalization, death, and discharge for case i are written as $\mathbf{t}_i^e$, $\mathbf{t}_i^s$, $\mathbf{t}_i^h$, $\mathbf{t}_i^d$, and $\mathbf{t}_i^r$ respectively. Due to the existence of missing time events, the two key events, which are symptom onset and hospitalization, were essential for dynamics reconstruction and thus required an imputation. There are various patterns observed for the individual time events which require different forms of imputation using time lag distributions. Fig 7 shows timelines of all individuals. In addition, the primary cases of infection are denoted as $\mathbf{v}$. The transmission tree of this outbreak was partially reconstructed that 39 out of 40 non-imported cases were found with known infectors while Case 41 was the only case with missing information. The observed primary cases are shown on the y-axis of Fig 7. Note that Case 61 was infected by Cases 58 and 59, who were an identical pair but Case 58 is shown for simplicity.

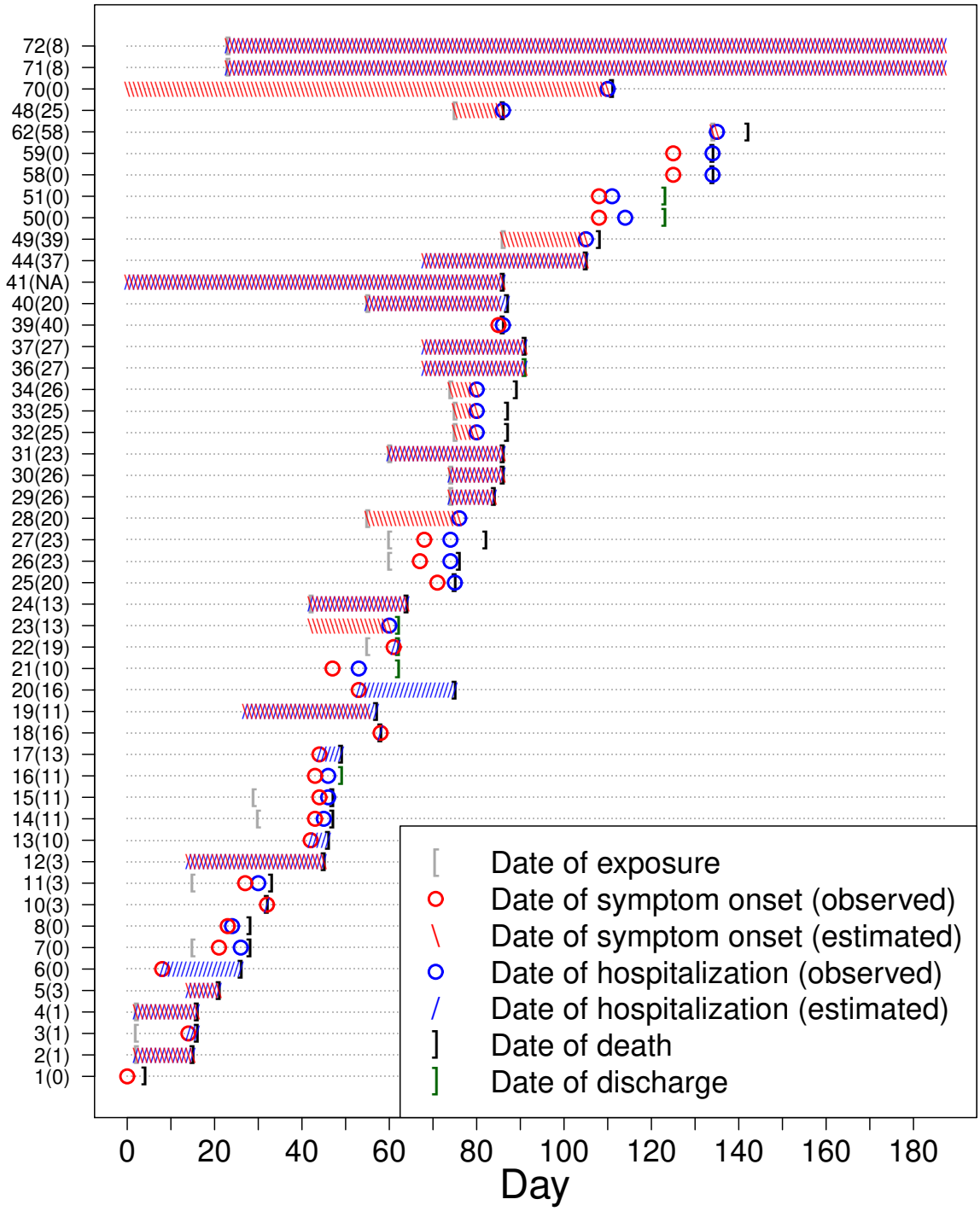


Figure 7: **The individual timelines.** The timelines consisting of observed time events and the possible missing time events of all 49 cases. The observed dates of exposure, death, and discharge are presented by gray, black, and green square brackets. The observed dates of symptom onset and hospitalization are presented by red and blue unfilled circles. The estimated missing time events are imputed within the ranges presented by slash symbols. The numbers shown on the y-axis represent case identities and their primary cases in parentheses, where 0 refers to nine imported cases and NA of Case 41 means missing.

Model formulation

The modeling approach of this study employs the renewal equation,

$$j(t) = \int_0^\infty A(t,\tau) j(t-\tau) d\tau, \tag{12}$$

where $j(t)$ is the incidence number at the time t and $A(t,\tau)$ is the transmission rate per individual at the time t with the infection age (i.e. time since infection) τ. The transmission rate may further be separated into two independent variables, reproduction number $R(t)$ and serial interval distribution $s(\tau)$, which satisfies $A(t,\tau) = R(t)s(\tau)$ (Fraser 2007, Chowell & Nishiura 2008).

In our modeling, we here assume that cases have been isolated since the time of hospitalization. For this local outbreak in Pujehun District, the hospitalization rate was 90% (44/49). Of all 49 cases, there are 24 cases with and 24 cases without observed dates of hospitalization, and the index case who is known to be non-isolated. Further, of those 24 cases without observed dates of hospitalization, there are 2 cases discharged. Therefore, there are 18 (out of 22) cases have been isolated, however, without observed dates of hospitalization or discharge. We take the hospitalization probability denoted as $p^h = 82\%$ (18/22) for those 22 unknown hospitalized cases; whereas the hospitalization probability is 1 (i.e. $p^h = 100\%$) for the other 26 cases.

During case isolation, the transmission rate is assumed to be reduced,

$$\hat{A}(t,\tau) = \begin{cases} A(t,\tau) & \text{without isolation} \\ (1-\varepsilon^h)A(t,\tau) & \text{with isolation} \end{cases}, \tag{13}$$

where ε^h is the relative reduction ranged from 0 to 1, quantifying the protective effect of case isolation. For the cases with known and unknown hospitalization, the protective effect is different depending on the hospitalization probability, $\varepsilon^h = 1-(1-\varepsilon)^{p^h}$ or,

$$\begin{cases} 1-(1-\varepsilon)^1 & = \varepsilon & \text{for 26 known hospitalized cases} \\ 1-(1-\varepsilon)^{0.82} & = \varepsilon^h & \text{for 22 unknown hospitalized cases} \end{cases}, \tag{14}$$

where ε is the protective effect of case isolation with an evidence (i.e. of cases with known hospitalization). The protective effect of unknown hospitalized cases is assumed to be a mean value over 22 individuals, and thus is lower than the one of known hospitalization. As an alternative, we also considered a scenario with the assumption of the protective effect of unknown hospitalization ε^h to be an independent variable without using the hospitalization probability p^h.

Moreover, the protective effect of individuals varies depending on the time from symptom onset

to hospitalization,

$$\varepsilon_i^h = \left[1 - S(\tau_i^{sh})\right]\varepsilon^h, \tag{15}$$

where S is the cumulative distribution function (CDF) of the serial interval as a function of the time from symptom onset to hospitalization τ_i^{sh} of case i. The individual protective effect is maximized only if a case was isolated immediately after symptom onset (i.e. $\tau_i^{sh} = 0$). Thus, the protective effect ε^h is also considered as the effect without time delay.

Consequently, the reproduction number and serial interval distribution in the renewal equation (Eq. 12) need a modification according to the protective effect of case isolation. First, the secondary transmission is reduced, and thus the reproduction number is

$$\hat{R}(t_i) = R(t_i)(1 - \varepsilon_i^h) \tag{16}$$

$$= R(t_i)\left(\int_0^{\tau_i^{sh}} s(\tau)d\tau + (1-\varepsilon^h)\int_{\tau_i^{sh}}^{\infty} s(\tau)d\tau\right) \tag{17}$$

where τ_i^{sh} is time from symptom onset to hospitalization representing the infection age at the time of isolation of the primary case i. Additionally, we assume an exponential declining transmissibility in time as $R(t) = R_0 e^{-\delta t}$, where R_0 is the reproduction number at time zero (Day 0) and δ is the exponential declining rate. The exponential dynamics has shown a better performance than a constant reproduction number in MERS and plague outbreaks (Cauchemez et al. 2016, Nishiura et al. 2006). As an alternative, we also considered a scenario with the assumption of a constant reproduction number. Second, the serial interval is shortened since isolation, and thus the corresponding distribution is

$$\hat{s}(\tau_i) = \begin{cases} \frac{1}{1-\varepsilon_{v_i}^h} s(\tau_i) & \text{without isolation} \\ \frac{1-\varepsilon}{1-\varepsilon_{v_i}^h} s(\tau_i) & \text{with isolation} \end{cases} \tag{18}$$

$$= \begin{cases} \frac{s(\tau_i)}{\int_0^{\tau_{v_i}^{sh}} s(\tau)d\tau + (1-\varepsilon^h)\int_{\tau_{v_i}^{sh}}^{\infty} s(\tau)d\tau} & \text{without isolation} \\ \frac{(1-\varepsilon)s(\tau_i)}{\int_0^{\tau_{v_i}^{sh}} s(\tau)d\tau + (1-\varepsilon^h)\int_{\tau_{v_i}^{sh}}^{\infty} s(\tau)d\tau} & \text{with isolation} \end{cases}, \tag{19}$$

where $\varepsilon_{v_i}^h$ is the protective effect of primary case v_i infected secondary case i. In the following, the above-defined quantities $\hat{R}$ and $\hat{s}$ are referred to biased, describing the observed data in the presence of case isolation; whereas the unbiased quantities R and s in the absence of case isolation. The two unbiased quantities and protective effect of case isolation ε were estimated simultaneously during Markov chain Monte Carlo (MCMC) sampling.

Parameter estimation

In this modeling, we first construct the likelihood function supposing that the data $\mathbf{D}$ are fully observed. The full likelihood function denoted as $L(\theta|\mathbf{D})$ consists of two parts,

$$L(\theta|\mathbf{D}) = L_r(\theta|\mathbf{D})L_s(\theta|\mathbf{D}), \tag{20}$$

where $L_r(\theta|\mathbf{D})$ is the likelihood of observed secondary case number and $L_s(\theta|\mathbf{D})$ is the likelihood of observed serial interval. First, the likelihood of observed secondary case number consists of all cases $i \in \mathbf{D}$ and is written as

$$L_r(\theta|\mathbf{D}) = \prod_{i\in\mathbf{D}} p(r_i;\theta), \tag{21}$$

where the observed secondary case number r_i generated by case i is assumed to be following a distribution p with mean as the reproduction number $\hat{R}(t_i^s)$ at the time of symptom onset. There are three alternative distributions, which are geometric, negative binomial and Poisson. In practice, the 49 cases $i \in \mathbf{D}$ were divided into three groups (the index case $i \in \mathbf{D}^1$, the known hospitalized case $i \in \mathbf{D}^h$ and the unknown hospitalized case $i \in \mathbf{D}\backslash\{\mathbf{D}^1, \mathbf{D}^h\}$), which are without protective effect, with the protective effect ε and ε^h, respectively. Second, the likelihood of observed serial interval consists of only non-imported cases $i \in \mathbf{D}\backslash\mathbf{D}^I$ and is written as

$$L_s(\theta|\mathbf{D}) = \prod_{i\in\mathbf{D}\backslash\mathbf{D}^I} \hat{s}(\tau_i^{ss};\theta), \tag{22}$$

where $\hat{s}$ is the biased distribution of serial interval in the presence of case isolation and $\tau_i^{ss} = t_i^s - t_{v_i}^s$ is the time between symptom onsets of the primary case v_i and the secondary case i. There are two alternative distributions, which are Weibull and gamma, for the unbiased serial interval s in the absence of case isolation. In practice, the 40 non-imported cases $i \in \mathbf{D}\backslash\mathbf{D}^I$ or $i|v_i \in \mathbf{D}$ were also divided into three groups (of the infector was the index case $i|v_i \in \mathbf{D}^1$, the known hospitalized case $i|v_i \in \mathbf{D}^h$ or the unknown hospitalized case $i|v_i \in \mathbf{D}\backslash\{\mathbf{D}^1, \mathbf{D}^h\}$).

While the two main time events, dates of symptom onset and hospitalization, are missing for some cases, it requires data imputation to assess transmission dynamics and quantify the protective effect of case isolation. The dates were imputed using six key time lag distributions, namely exposure-to-onset τ^{es} (i.e. incubation period), onset-to-hospitalization τ^{sh}, onset-to-death τ^{sd}, onset-to-discharge τ^{sr}, hospitalization-to-death τ^{hd} and hospitalization-to-discharge τ^{hr}. We assume that the incubation period was the same as that of the entire epidemics in West Africa and was gamma distributed with mean at 9.1 and SD at 7.3 days overall countries (WHO Ebola Response Team 2014). The other five distributions were estimated using the observed time events from the local outbreak with the maximum likelihood estimation (MLE) method. The alternative distribution, Weibull or gamma, was selected according to a higher likelihood

value. Due to the discrete nature of observed data, the distributions were transformed from continuous to discrete by $f(\tau) = F(\tau+1) - F(\tau)$ for $\tau = 0, 1, 2, ...$, where $f(\tau)$ represents the probability mass function (PMF) and $F(\tau)$ represents the cumulative distribution function (CDF) of Weibull or gamma distribution.

Given the estimated distributions, each individual is represented with either observed dates or possible missing dates with certain probabilities. The probabilities were normalized by the length of possible dates based on the observed pattern of time events. Table 8 shows the pattern of all categories imputing the key dates from this local outbreak. There are two types of categories, of which the first was employed for both dates of symptom onset and hospitalization and the second was employed for only the dates of symptom onset. The second part was designed for computational efficiency as the dates of hospitalization for the secondary case i in the likelihood of observed serial interval was not needed.

Table 8: **Categories of observed and missing individual time events.** Each case i is categorized based on observed patterns of individual time events. The categorization is designed for imputing the two important events (i.e. symptom onset and hospitalization). The upper part shows eight categories for the reconstruction of both dates of symptom onset and hospitalization; whereas the lower part shows five categories for the reconstruction of the dates of symptom onset. The time events, dates of exposure t_i^e, symptom onset t_i^s, hospitalization t_i^h, and death or discharge $t_i^{d,r}$ are observed O, missing M or trivial (i.e. either observed or missing) $-$. The last categories of both parts used the dates of discharge t_i^r instead of death t_i^d. The reconstructed dates of symptom onset $\hat{t}_i^s$ and hospitalization $\hat{t}_i^h$ were estimated using observed time events $t_i^{\cdot}$ and time lags $\tau^{\cdot}$. The last column shows the case counts of specified categories.

Pattern	t_i^e		t_i^s		t_i^h		$t_i^{d,r}$	$\hat{t}_i^s$	$\hat{t}_i^h$	Observed
$i \in D^{(1)}$	$-$		O		O		$-$	t_i^s	t_i^h	15/48
$i \in D^{(2)}$	$-$		O	$\rightarrow$	M		$-$	t_i^s	$t_i^s + \tau^{sh}$	8/48
$i \in D^{(3)}$	O	$\rightarrow$	M		O		$-$	$t_i^e + \tau^{es}$	t_i^h	7/48
$i \in D^{(4)}$	M		M	$\leftarrow$	O		$-$	$t_i^h - \tau^{sh}$	t_i^h	2/48
$i \in D^{(5)}$	O	$\rightarrow$	M		M	$\leftarrow$	O	$t_i^e + \tau^{es}$	$t_i^d - \tau^{hd}$	7/48
$i \in D^{(6)}$	O	$\rightarrow$	M	$\rightarrow$	M		M	$t_i^e + \tau^{es}$	$t_i^e + \tau^{es} + \tau^{sh}$	2/48
$i \in D^{(7)}$	M		M	$\leftarrow$	M	$\leftarrow$	O	$t_i^d - \tau^{hd} - \tau^{sh}$	$t_i^d - \tau^{hd}$	6/48
$i \in D^{(8)}$	M		M	$\leftarrow$	M	$\leftarrow$	O	$t_i^r - \tau^{hr} - \tau^{sh}$	$t_i^r - \tau^{hr}$	1/48
$i \in D^{(I)}$	$-$		O		$-$		$-$	t_i^s		24/49
$i \in D^{(II)}$	O	$\rightarrow$	M		$-$		$-$	$t_i^e + \tilde{\tau}^{es}$		16/49
$i \in D^{(III)}$	M		M	$\leftarrow$	O		$-$	$t_i^h - \tilde{\tau}^{sh}$		2/49
$i \in D^{(IV)}$	M		M	$\leftarrow$	$\leftarrow$	$\leftarrow$	O	$t_i^d - \tilde{\tau}^{sd}$		6/49
$i \in D^{(V)}$	M		M	$\leftarrow$	$\leftarrow$	$\leftarrow$	O	$t_i^r - \tilde{\tau}^{sr}$		1/49

The model parameters were estimated using Metropolis-Hastings MCMC (MH-MCMC) with Bayesian inference. At each iteration, all parameters, except the primary case, are first proposed using a non-informative flat prior assumption. Then, the possible primary case is proposed using

the transmission probability of case i infected by case j,

$$p_{ij}(\theta) = \frac{\hat{s}(\tau_{ij}^{ss};\theta)}{\sum_{k\in w_i}\hat{s}(\tau_{ik}^{ss};\theta)}, \tag{23}$$

where $\tau_{ik}^{ss} = t_i^s - t_k^s$ is the biased serial interval between secondary case i and possible primary case k, who is on the possible infector list w_i. The transmission probability was initially proposed in the study of the 2003 SARS outbreak (Wallinga & Teunis 2004) and was also used in partial network reconstruction of the 2009 influenza pandemics (Hens et al. 2012). The transmission probability (or the serial interval) is calculated based on the proposed parameters in the first step. The possible infector list from this local outbreak consists of 41 cases with possible earlier dates of symptom onset, however, except Cases 71 and 72 of low likelihoods to improve computational efficiency. As an alternative, we also considered a scenario with flat proposal distribution to reduce computation cost. As the transmission probability requires a heavy computation on serial intervals with data imputation, the flat proposal would significantly accelerate the process.

The sampling process was performed by 100,000 iterations with thinning of every 100th sample. The sampling was replicated in four independent chains of each 1,000 thinned samples, and thus posterior distribution is represented by using in total 4,000 samples. The algorithm was implemented in R version 3.4.3 (Kite-Eating Tree) on Mac OSX and R version 3.4.4 (Someone to Lean On) on Linux (Available from: https://github.com/imlouischan/ebola-sl).

Model selection and scenario comparison

There are six combinations of alternative distributions of secondary case numbers (i.e. geometric, negative binomial and Poisson) and serial interval (i.e. Weibull and gamma). The combinations were ranked according to three criteria, Akaike information criterion (AIC), Bayesian information criterion (BIC) and model posterior distribution. A better combination is obtained while a lower AIC, a lower BIC, or a higher model posterior. AIC and BIC are point estimates taking the best samples among 4,000 whereas the model posterior considers the entire distribution represented by 4,000 samples and was calculated using the weight of marginal likelihoods given a uniform model prior assumption (i.e. same prior among all combinations). The marginal likelihoods were calculated using the posterior harmonic mean (PHM) estimator (Vyshemirsky & Girolami 2007).

In addition to the baseline scenario, we considered alternative scenarios to investigate sensitivity to the assumptions including time dependence of the reproduction number, dependence of the protective effect of unknown hospitalized cases and the informative proposal distribution. The reproduction number can alternatively remain constant over time instead of exponential declining. In other words, the reproduction number (as well as transmission rate) is time-independent, i.e. $\delta = 0$ and the number of estimated parameters would -1. The protective effect of unknown

hospitalized cases ε^h can be alternatively sampled as an independent parameter rather than being a function of the protective effect of known hospitalization ε and hospitalization probability p^h. In this scenario, the number of estimated parameters would $+1$. The proposal distribution during sampling can be alternatively flat and non-informative rather than using the transmission probability p_{ij}. The number of estimated parameters would remain unchanged in this alternative.

Results

Five time lag distributions used in imputation were the best fitted by Weibull distribution over gamma distribution using MLE. The corresponding estimated means of onset-to-hospitalization, onset-to-death, onset-to-discharge, hospitalization-to-death and hospitalization-to-discharge were 5.0, 7.1, 10.8, 3.6 and 7.5 days; and the standard deviations (SD) were 2.5, 5.8, 6.3, 3.5 and 3.8 days, respectively. The estimates of five time lag distributions are summarized in Table 9 and both estimated distributions and observed data are plotted in Fig 8. As a result, Fig 7 shows the reconstructed timeline of each individual, using both observed and ranges of imputed individual time events. The corresponding probabilities of imputed time events of individuals are shown in Fig 9.

Table 9: **Maximum likelihood estimation (MLE) of time lag distributions.** All five distributions were selected among Weibull or gamma distribution using MLE. The estimates of mean and SD were calculated using shape and scale parameters.

Time lag distribution	Best model	mean	SD
onset-to-hospitalization (τ^{sh})	Weibull	5.0	2.5
onset-to-death (τ^{sd})	Weibull	7.1	5.8
onset-to-discharge (τ^{sr})	Weibull	10.8	6.3
hospitalization-to-death (τ^{hd})	Weibull	3.6	3.5
hospitalization-to-discharge (τ^{hr})	Weibull	7.5	3.8

In the baseline scenario, the model combining geometrically distributed secondary case number and Weibull distributed serial interval was the best supported by all three criteria. The corresponding AIC and BIC were the lowest at 373.6 and 384.9 and the model posterior was 83.1%. Table 10 shows the selection criteria of all six combinations. The second-best model was the one of secondary case numbers following negative binomial distribution and serial intervals following Weibull distribution. It was estimated with 16.8% model posterior and the highest log-likelihood value of -180.1. Hereafter, rather than employing model averaging, the following results are based on this best-ranked model.

Fig 10 shows the posterior distribution of primary case infected Case 41. The top three possible infectors were Case 25, 27 and 26 with probabilities 16.6%, 15.8% and 14.5%, respectively. Fig 11 shows the individual protective effect of case isolation. In general, the individual protective effect was higher for cases with known hospitalization than those without. Cases 40 and 31 were the two with the lowest individual protective effect and belong to the 5th category (see Table 8). The protective effect of known hospitalization as a maximum value was estimated with the mean as 23.9% (95%CI: 1.1%-69.5%) and the one of unknown hospitalization was 20.0% (95%CI: 0.9%-62.2%). Fig 12 A and B shows both unbiased distributions of the secondary case number and serial interval in the absence of isolation. The mean number of secondary cases per primary case at time zero was estimated as 3.0 (95%CI: 1.2-9.1), while the exponential rate of decline over time was estimated to be 0.019 (95%CI: 0.005-0.035). The mean and SD of the

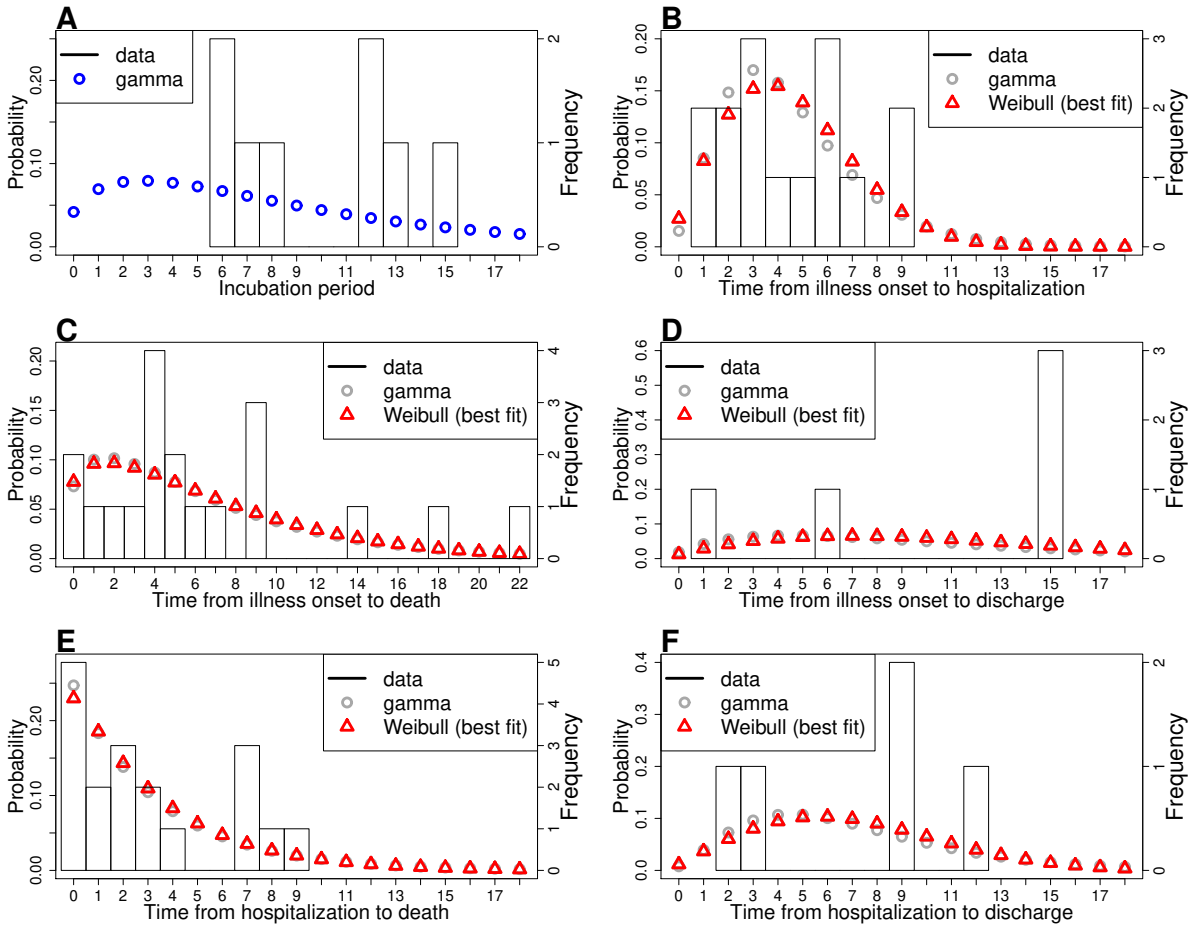


Figure 8: **The estimated time lag distributions.** A: The probability mass function (PMF) of the incubation period (time from exposure to symptom onset). The unfilled blue circles represent published estimates employing gamma distribution, which however is unmatched by sample variance. The observed frequencies of data are shown in bars. B: The PMF of time from symptom onset to hospitalization. C: The PMF of time from symptom onset to death. D: The PMF of time from symptom onset to discharge. E: The PMF of time from hospitalization to death. F: The PMF of time from hospitalization to discharge. The unfilled red triangles and gray circles represent maximum likelihood estimates employing Weibull and gamma distributions, respectively. All five distributions were best fitted with higher likelihood values using Weibull distribution.

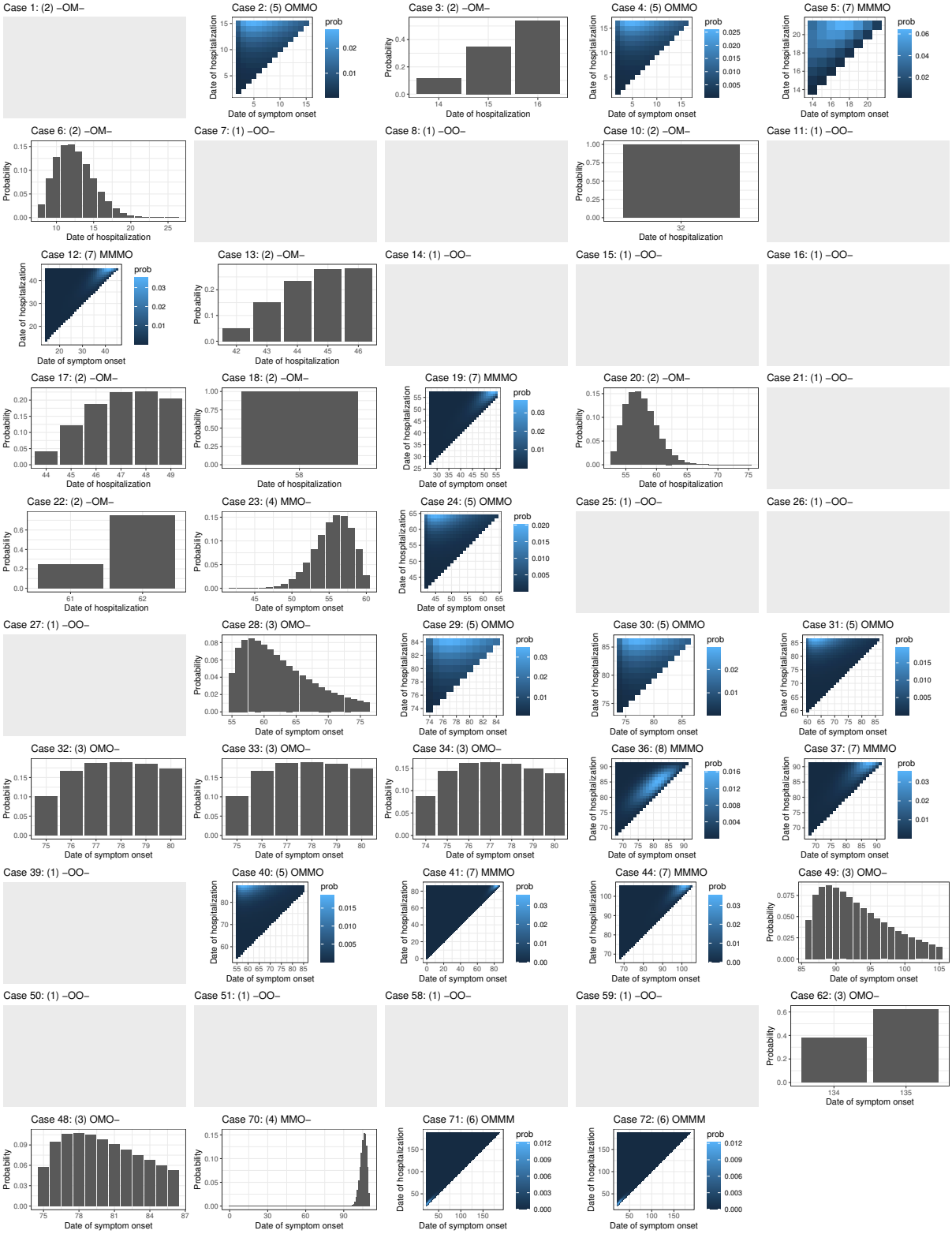


Figure 9: **The probability of imputed time events (dates of symptom onset and hospitalization) of all individuals.** All 49 individuals are categorized into 8 (shown in parentheses, see the upper part of Table 1). The 1st category: no data imputation is required for the cases with both observed dates of symptom onset and hospitalization. The 2nd, 3rd and 4th categories: the 1D probability of either date of symptom onset or hospitalization are shown. The 5th-8th categories: the 2D probability of both dates of symptom onset and hospitalization are shown. The lighter and darker blue color represents a higher and lower probability, respectively.

Table 10: **Model selection of alternative distributions.** Six alternative models were compared using three criteria, AIC, BIC and model posteriors. AIC and BIC were calculated using the point estimate of log-likelihood $\ln(\hat{L})$. All criteria indicated that the best model was the one combining secondary case numbers following geometric distribution and serial intervals following Weibull distribution. The second-best model, of which secondary case number followed negative binomial distribution and serial interval followed Weibull distribution, was estimated with the highest log-likelihood value.

Secondary case	Serial interval	$\ln(\hat{L})$	AIC	BIC	Posterior (%)
Geometric	Weibull	-180.8	373.6	384.9	83.1
Negative binomial	Weibull	-180.1	374.3	387.5	16.8
Poisson	Weibull	-186.5	384.9	396.3	0.2
Geometric	Gamma	-183.0	378.0	389.4	0.0
Negative binomial	Gamma	-182.3	378.6	391.9	0.0
Poisson	Gamma	-188.6	389.2	400.6	0.0

unbiased serial interval were estimated as 12.6 (95%CI: 10.8-14.3) and 4.0 (95%CI: 3.1-5.4) days. Fig 13 shows the comparison to the biased serial interval with observed data, of which the sample mean and SD were 13.7 and 4.5 days. The direct fitted estimates in the presence of case isolation employed Weibull distribution using MLE, which was the same as estimating the key time lag distributions. The biased mean and SD were 14.1 and 3.9 days. Fig 14 shows the marginal and pairwise posterior distribution of parameters. Two parameters, reproduction number at time zero (R0) and exponential rate (delta), were positively correlated with the correlation coefficient as 0.687.

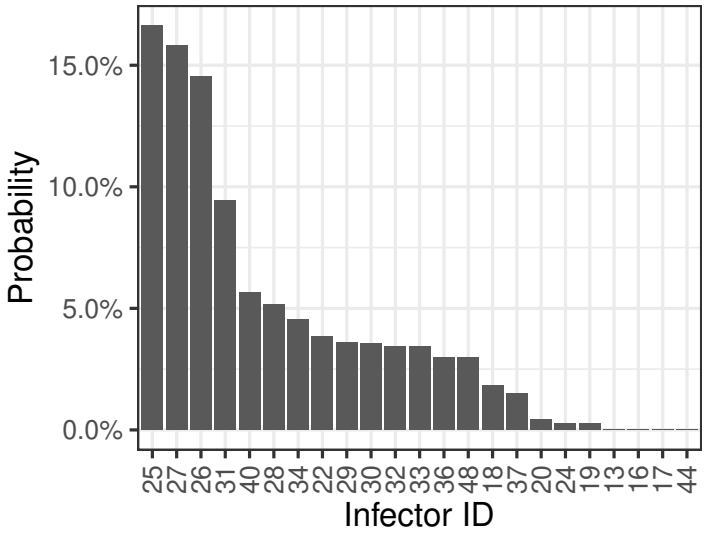


Figure 10: **The posterior distribution of the primary case of Case 41.** The top three were Case 25, 27 and 26 with probabilities 16.6%, 15.8% and 14.5%, respectively.

Fig 15 and Table 11 show the comparison between alternative scenarios on the assumptions of time-dependence of reproduction number and dependence of the protective effect of unknown hospitalization. The shape and scale parameters of the unbiased serial interval distribution and

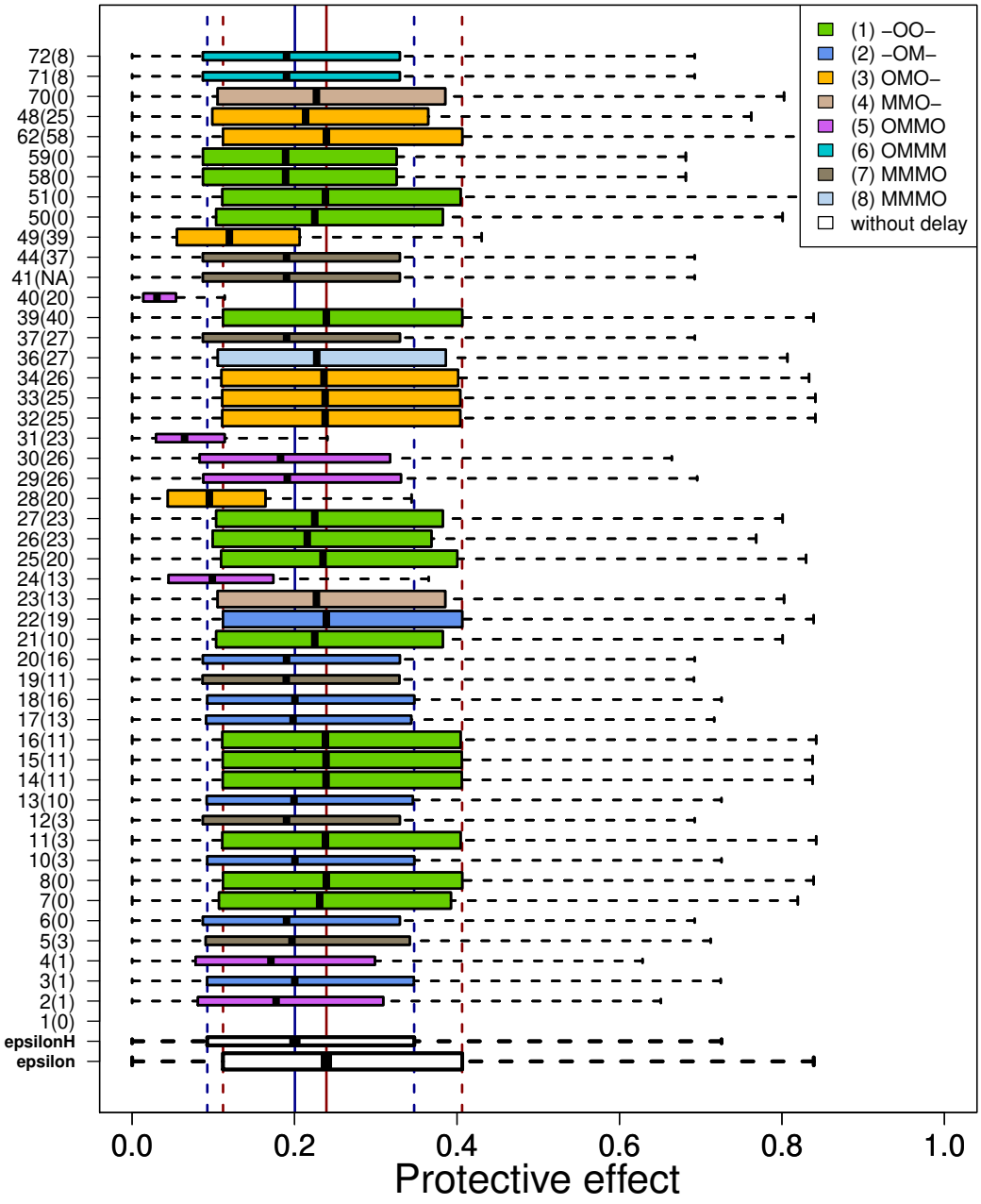


Figure 11: **The posterior distribution of the individual protective effect of case isolation.** The color boxes show the distributions of the individual protective effect except for the index case who was not isolated. The colors represent categories based on the patterns of observed time events. O represents observed, M indicates missing and $-$ indicates either observed or missing, and the order of events follows the dates of exposure, symptom onset, hospitalization and death or discharge. The white boxes show the protective effect without time delay of known (epsilon) and unknown hospitalization (epsilonH). The solid and dashed lines represent the mean and interquartile range while the red and blue lines represent the ones of known and unknown hospitalization. The thicker and thinner boxes represent the ones of known and unknown hospitalization, respectively.

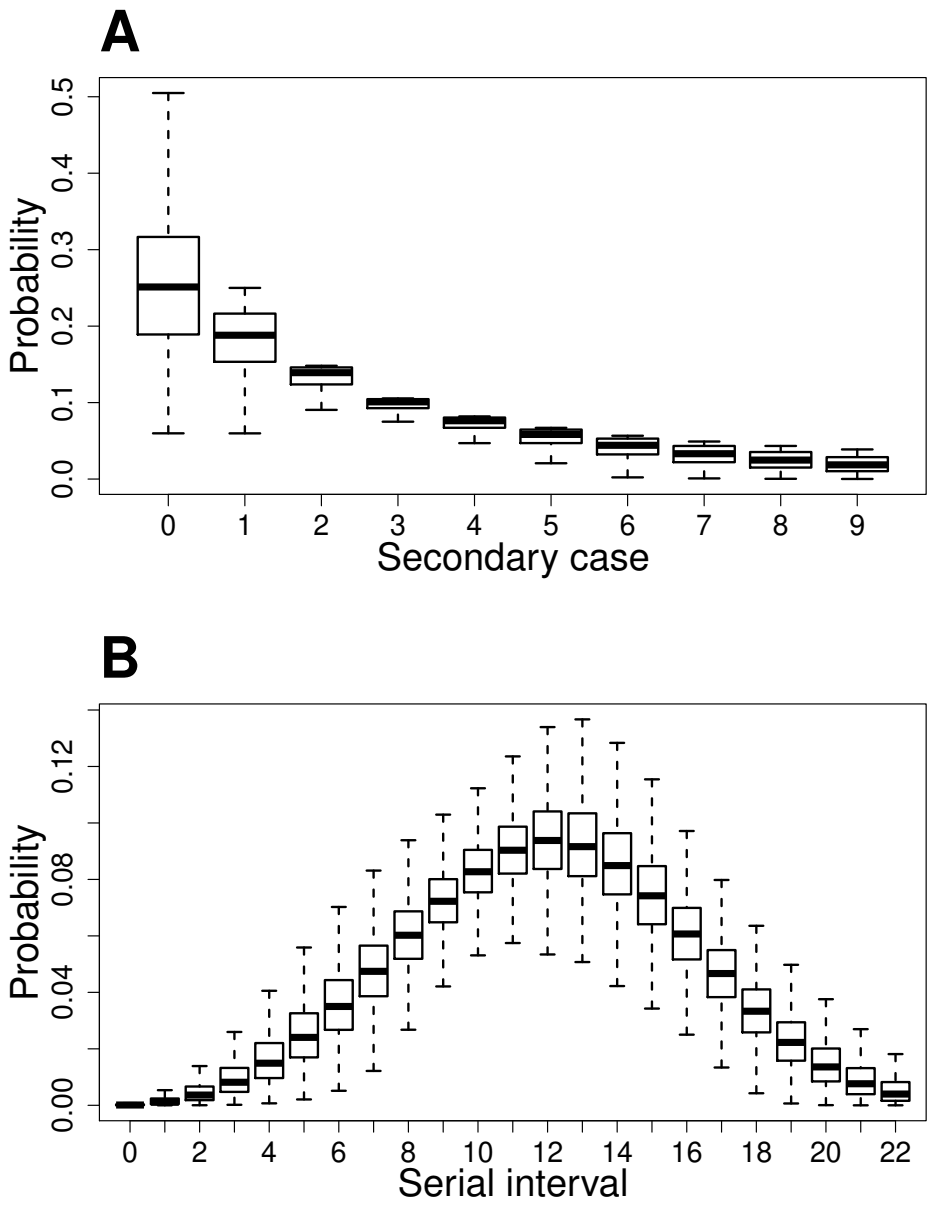


Figure 12: **The posterior distribution of secondary case number and serial interval.** A: The posterior distribution of the secondary case number per primary case $p(r)$ of which the mean as the reproduction number at time zero $R0$ was estimated at 3.0 (95%CI: 1.2-9.1). B: The posterior distribution of the unbiased serial interval $s(\tau)$ with mean at 12.6 (95%CI: 10.8-14.3) and SD at 4.0 (95%CI: 3.1-5.4) days. Both distributions are unbiased representing the situation in the absence of case isolation and were derived from the best model that employs geometric and Weibull distributions, respectively. The posterior distributions are shown using black boxes, ranging from the lower to the upper quartiles along with the median.

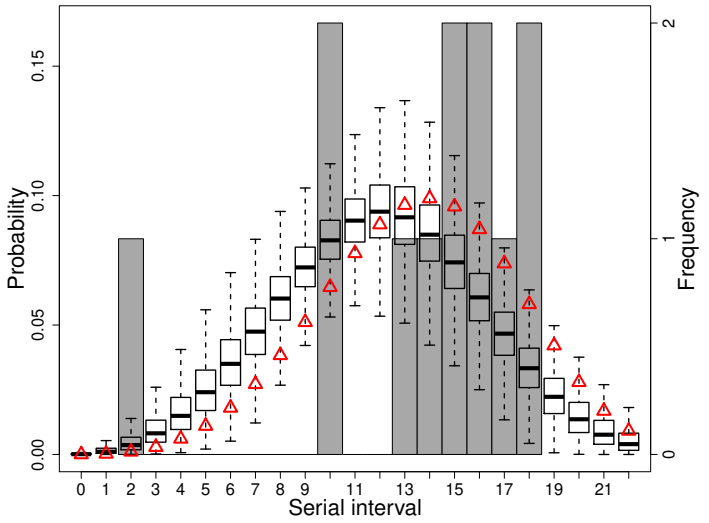


Figure 13: **The serial interval distribution.** The unbiased distribution $s(\tau)$ (i.e. in the absence of case isolation) is represented by using boxes as shown in Fig 12B. The mean and SD were 12.6 (95%CI: 10.8-14.3) and 4.0 (95%CI: 3.1-5.4) days. The corresponding biased distribution $\hat{s}(\tau)$ (i.e. in the presence of case isolation) is not shown due to various individual time from symptom onset to hospitalization. As an alternative, the frequency of observed serial intervals (i.e. biased / in the presence of case isolation) is shown as a histogram using gray bars. The sample means and SD were 13.7 and 4.5 days. The red circles represent the MLE fitted values by employing Weibull distribution, similar to Fig 8. The mean and SD were 14.1 and 3.9 days.

the primary case of Case 41 were similarly distributed in four scenarios. The reproduction number at time zero and both protective effect of case isolation were sensitive to the assumptions. The lowest panel of Fig 15 shows the mean of reproduction numbers over time. The reproduction numbers at time zero were much lower and the protective effect of known hospitalization became slightly higher while switching the dynamics to be constant. The protective effect of unknown hospitalization was the most sensitive to its dependence. The mean value increased from 20.0% (95%CI: 0.9%-62.2%) to 59.2% (95%CI: 9.2%-87.8%) while changing sampling to be an independent parameter. The individual protective effect was also sensitive as the ones of unknown hospitalization (Fig 16). However, the protective effect of known hospitalization remained unchanged. For the alternative scenarios on the proposal distribution, we found that the distribution of the primary case of Case 41 was sensitive to the assumption whereas the ranking orders remained the same. Fig 17 shows the top three cases estimated with much lower probabilities 10.7%, 9.7% and 9.3% given flat proposal.

All eight alternative scenarios were also compared using the selection criteria (Table 12). All three criteria preferred exponential to constant dynamics. Given the assumption of exponential reproduction number, the model posterior preferred the alternative as the independent protective effect of unknown hospitalization whereas AIC and BIC preferred the dependence due to a lower number of estimated parameters. On the contrary, the model posterior preferred the proposal using the baseline as using the transmission probability, whereas AIC and BIC (and

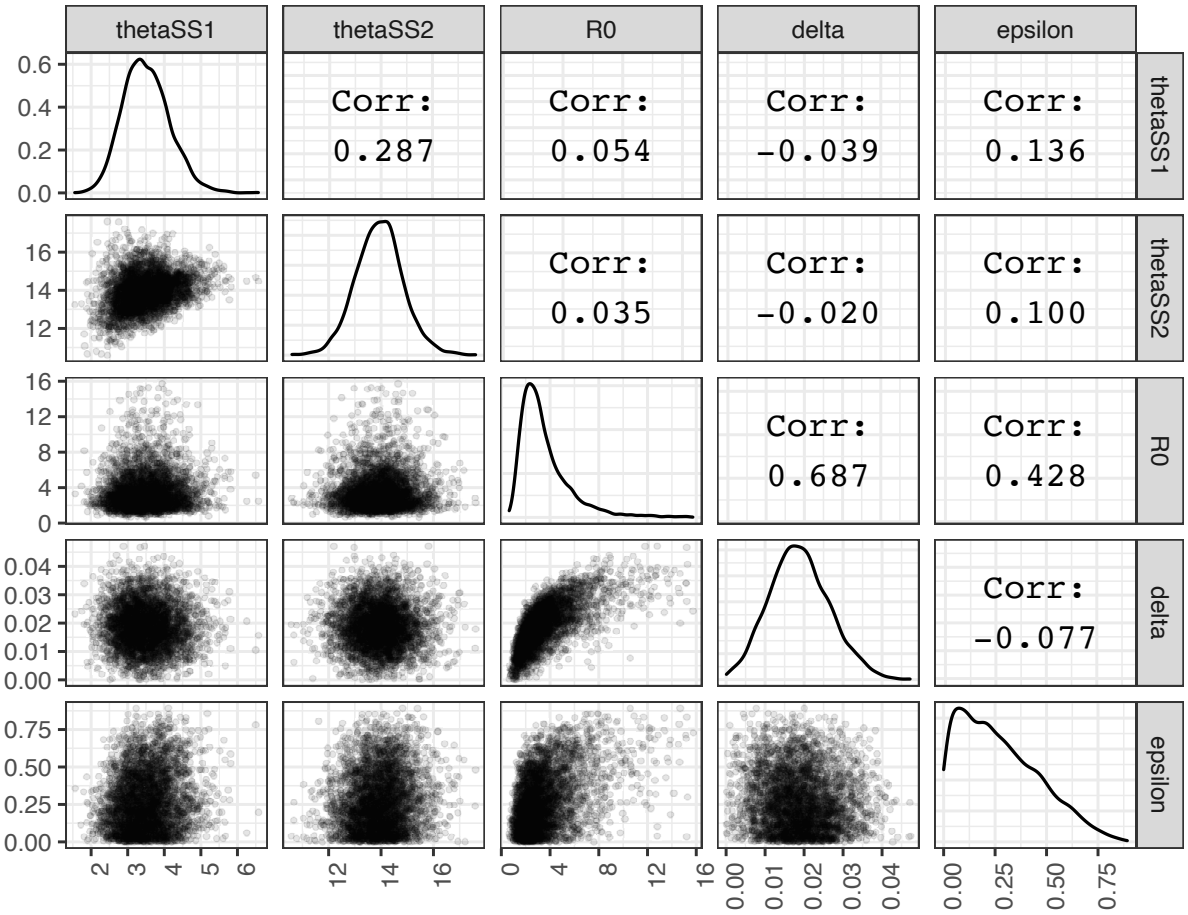


Figure 14: **The marginal and pairwise posterior distribution of parameters.** The posterior distribution of parameters derived from the best model that employs geometric and Weibull distributions to describe the observed secondary case number and unbiased serial interval, respectively. The estimated parameters include the shape (thetaSS1) and scale (thetaSS2) of the serial interval, the reproduction number at time zero (R0), the exponential rate of secondary transmission (delta) and the protective effect of case isolation without time delay (epsilon). The diagonal, lower diagonal and upper diagonal plots show the marginal distributions, dependence between parameters by scatter plots and correlation between parameters, respectively.

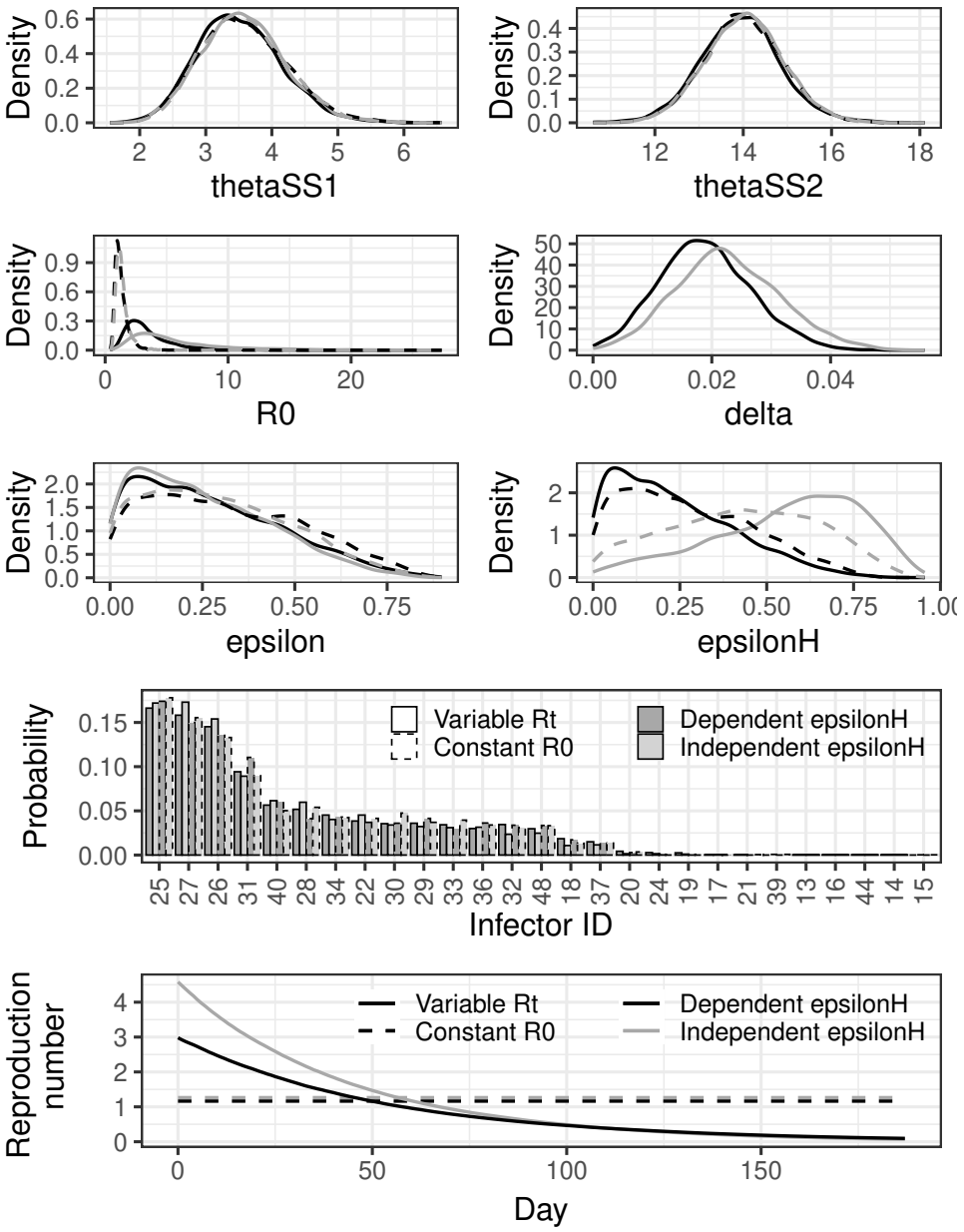
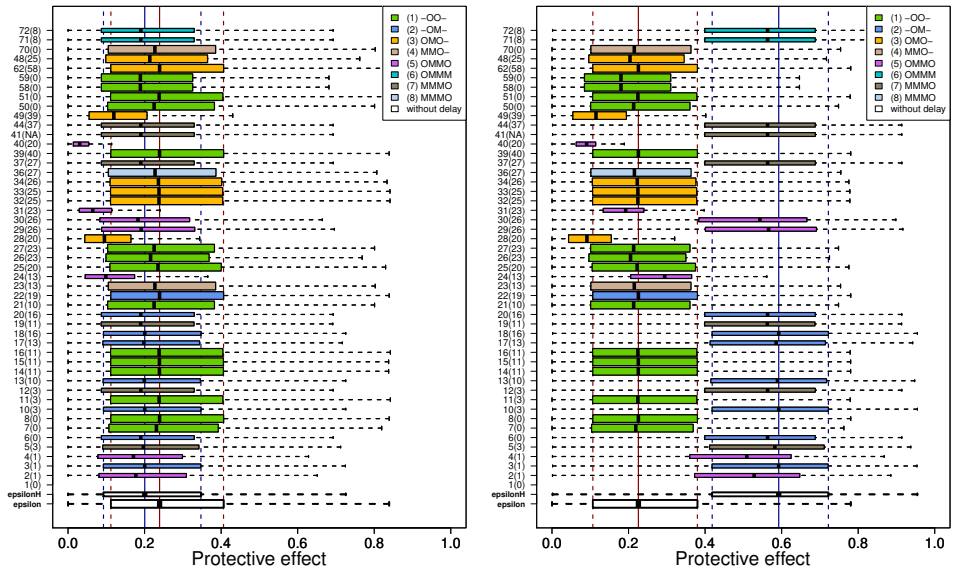


Figure 15: **The marginal posterior distributions of parameters in alternative scenarios.** The solid and dashed lines show the distributions under the assumption of reproduction number decreased exponentially and remained constant, respectively. The black and gray colors represent the ones under the assumption of epsilon-dependent and independent epsilonH, respectively. The parameters include the shape (thetaSS1) and scale (thetaSS2) parameters of the unbiased serial interval, the reproduction number at time zero (R0), the exponential rate of secondary transmission (delta), the protective effect of case isolation without time delay of known (epsilon) and unknown hospitalization (epsilonH) and the primary case of Case 41. The lowest panel shows the mean value of the estimated reproduction number during the outbreak.



(a) Variable Rt; Dependent epsilonH.

(b) Variable Rt; Independent epsilonH.

(c) Constant R0; Dependent epsilonH.

(d) Constant R0; Independent epsilonH.

Figure 16: **The posterior distribution of the individual protective effect of case isolation in alternative scenarios.** Given the p_{ij} proposal, there are four alternative scenarios: (A) The baseline scenario: Variable Rt; Dependent epsilonH (i.e. Fig 11). (B) Variable Rt; Independent epsilonH. (C) Constant R0; Dependent epsilonH. (D) Constant R0; Independent epsilonH. The color boxes show distributions of the individual protective effect except for the index case who was not isolated. The colors represent categories based on patterns of observed time events. O represents observed, M indicates missing and $-$ indicates either observed or missing, and the order of events follows the dates of exposure, symptom onset, hospitalization and death or discharge. The white boxes show the protective effect without time delay of known (epsilon) and unknown hospitalization (epsilonH). The solid and dashed lines represent the mean and interquartile range while the red and blue lines represent the ones of known and unknown hospitalization. The thicker and thinner boxes represent the ones of known and unknown hospitalization, respectively.

Table 11: **The parameters estimated in alternative scenarios.** Four alternative scenarios represent different assumptions on the reproduction number and the protective effect of unknown hospitalization. The parameters include the protective effect of case isolation without time delay of known (epsilon) and unknown hospitalization (epsilonH), the reproduction number at time zero (R0), the exponentially decreasing rate of secondary transmission (delta) and the shape (thetaSS1) and scale (thetaSS2) of the unbiased serial interval. The corresponding mean and SD of the unbiased serial interval $s(\tau)$ are included.

-	Scenario A	Scenario B	Scenario C	Scenario D
R	Variable Rt	Variable Rt	Constant R0	Constant R0
epsilonH	Dependent	Independent	Dependent	Independent
Parameters	Mean (95%CI)	Mean (95%CI)	Mean (95%CI)	Mean (95%CI)
epsilon (%)	23.9 (1.1-69.5)	22.6 (1.1-65.6)	29.5 (1.8-74.2)	27.2 (1.3-69.8)
epsilonH (%)	20.0 (0.9-62.2)	59.2 (9.2-87.8)	24.8 (1.5-67.0)	43.2 (3.7-80.2)
R0	3.0 (1.2-9.1)	4.6 (1.5-16.5)	1.2 (0.7-2.7)	1.3 (0.7-2.9)
delta	0.019 (0.005-0.035)	0.022 (0.007-0.041)	0	0
thetaSS1	3.5 (2.4-4.9)	3.5 (2.4-4.9)	3.5 (2.4-5.0)	3.5 (2.5-4.9)
thetaSS2	13.9 (12.2-15.8)	14.0 (12.3-15.8)	14.0 (12.2-15.7)	14.0 (12.3-15.8)
$s(\tau)$ mean	12.6 (10.8-14.3)	12.7 (11.0-14.2)	12.6 (10.9-14.2)	12.6 (11.0-14.2)
$s(\tau)$ SD	4.0 (3.1-5.4)	3.9 (3.0-5.5)	3.9 (3.0-5.5)	3.9 (3.0-5.4)

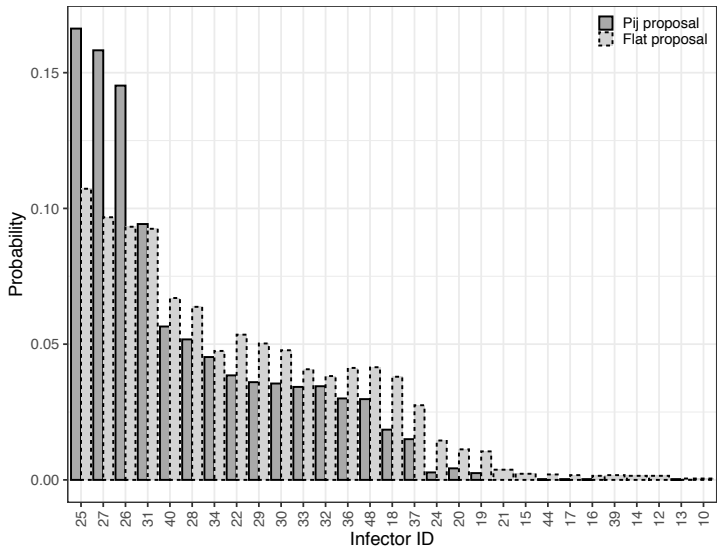


Figure 17: **The posterior distribution of the primary case of Case 41 in alternative scenarios on proposal distribution.** The solid line with darker gray bars and the dotted line with lighter gray bars represent the scenario of transmission probability (p_{ij}) proposal and flat proposal, respectively. Both scenarios indicate that the top three were Case 25, 27 and 26. The corresponding probabilities were 16.6%, 15.8% and 14.5% given p_{ij} proposal (i.e. the baseline scenario shown in Fig 10) whereas the probabilities were 10.7%, 9.7% and 9.3% given flat proposal, respectively.

maximum likelihood) slightly preferred the flat proposal.

Table 12: **Selection criteria of alternative scenarios.** Given the best model that employs geometric and Weibull distributions to describe the observed secondary case number and unbiased serial interval, eight alternative scenarios were compared using three different criterion values. AIC and BIC were calculated using the maximum values of log-likelihood $\ln(\hat{L})$ among MCMC samples. Model posteriors were calculated using marginal likelihoods.

Proposal	epsilonH	Reproduction number	$\ln(\hat{L})$	AIC	BIC	Posterior (%)
p_{ij}	Dependent	Variable Rt	-180.8	373.6	384.9	19.4
Flat	Dependent	Variable Rt	-180.8	373.6	385.0	10.5
p_{ij}	Independent	Variable Rt	-180.1	374.2	387.4	32.2
Flat	Independent	Variable Rt	-180.0	374.0	387.2	22.9
p_{ij}	Dependent	Constant R0	-183.0	376.0	385.4	8.6
Flat	Dependent	Constant R0	-183.0	376.0	385.5	1.9
p_{ij}	Independent	Constant R0	-182.8	377.5	388.9	0.7
Flat	Independent	Constant R0	-182.8	377.7	389.0	3.7

Discussion

This study quantified the impact of intervention measures during the 2014 EVD outbreak in Pujehun District, Sierra Leone. The assessment went through quantifying the protective effect of case isolation as a relative reduction of transmissibility and applied on a partially observed data for 49 patients with known and unknown hospitalization. We developed a statistical model to demonstrate that the secondary transmission cases were reduced and the serial interval was shortened due to isolation of hospitalized cases, while the transmission dynamic was partially reconstructed and missing data were imputed. In the baseline scenario, we showed that the number of secondary cases generated by a primary case followed geometric distribution and the serial interval followed Weibull distribution in the absence of case isolation. The reproduction number at time zero was estimated yielding the mean at 3.0 (95%CI: 1.2-9.1) with the exponential rate of the mean at 0.019 (95%CI: 0.005-0.035). The mean and SD of the unbiased serial interval was estimated as 12.6 (95%CI: 10.8-14.3) and 4.0 (95%CI: 3.1-5.4) days. More importantly, the protective effect of case isolation given known hospitalization was 23.9% (95%CI: 1.1%-69.5%) indicating a successful contribution of intervention measures due to case isolation during the local outbreak. The unknown primary case of infection was identified that Cases 25, 27 and 26 were the most likely patients infected Case 41 with probabilities of 16.6%, 15.8% and 14.5%, respectively. At the time of the outbreak occurring, there was no vaccine available and while the effect of vaccination has been broadly discussed, the present study focuses on non-pharmaceutical intervention strategies with the modeling approach of reconstructing transmission network and time events. In the following, we discuss three important points from this study and they are the individual protective effect and data imputation, the protective effect of this local outbreak, and the estimates of reproduction number and serial interval.

The individual protective effect of case isolation ε_i of each hospitalized patient was explicitly estimated and highly varied among individuals due to various patterns observed in the data. In general, the 3rd, 5th and 6th categories using the incubation period in imputation would lead to a lower protective effect, as observed in Cases 24, 28, 31, 40 and 49. However, it was not all cases in the categories with a significantly lower protective effect. The second biasing reason was the unknown hospitalization, of which four patients were not identified. With the uncertainty of evidence, the estimates obeyed an average among cases and the patients were therefore with lower values compared with those with evidence. Additionally, we compared with the alternative scenario of using the independent protective effect of unknown hospitalization. It verified the estimates of the protective effect of known hospitalization as remaining unchanged. As independence, the protective effect of unknown hospitalization increased significantly. However, one of the main factors negatively influencing the impact of interventions was the traditional burials at the initial stage of the outbreak. Series of practices during the burial ceremonies provided a chance of exposure to the virus through contact with infectious bodily fluid, although the patients had been under isolation since hospitalization. Such infec-

tion was considered as a failure of protection in this modeling and was indicated by the less than one-third protective effect. There was an improvement due to the effort by the EVD burial teams conducting safe burials. However, our model did not capture the temporal change of protectiveness and significantly affected by 11.9% (5/42) of traditional burials in the community. According to our previous study on the outbreak in Nigeria, the protective effect was estimated to be 39.7%. Despite a comparison between the local level of this outbreak in Sierra Leone and the national level of the one in Nigeria, the gap would be related to vulnerability. The infectious disease vulnerability index (IDVI) was 0.22 and 0.27 across all domains and; 0.06 and 0.19 in the health care domain for Sierra Leone and Nigeria respectively (Moore et al. 2017). The index was introduced following a study on the assessment of public health risk based on factors potentially influencing vulnerability to the 2014 EVD outbreak (Gelfeld et al. 2015). Other factors such as the timing of initiating the outbreaks would also lead to a different result.

We compared the reproduction number under the assumption on time independence and the estimate of mean in the alternative scenario reduced to 1.2 (95%CI: 0.7-2.7), which complied with another study (Ajelli et al. 2015). Ajelli et al. (Ajelli et al. 2015) estimated that the reproduction number follows negative binomial distribution yielding the estimate of mean at 1.63 (considering the first eight cases representing the early stage of the outbreak) and the basic reproduction number in the absence of interventions was 2.24 (95%CI: 1.52-4.51). Furthermore, the reproduction number in the national level was estimated to be 2.02 (95%CI: 1.79-2.26) (WHO Ebola Response Team 2014). The values were relatively low due to an absence of a super-spreader, which was found for example in the 2003 SARS outbreak (Wallinga & Teunis 2004) and the 2014 EVD outbreak in Nigeria (see Chapter 1). The serial interval distributed similarly between scenarios and found to be shortened compared to the sample mean of 13.7 days or a direct MLE fitting using Weibull distribution of the mean at 14.1 days in the presence of case isolation (Fig 13). The estimates were slightly different than the ones following gamma distribution at both national and global levels of which 11.6 and 15.3 days, respectively (WHO Ebola Response Team 2014). Regardless, the estimates of a two-week interval resembled the other published studies (WHO Ebola Response Team 2014, Chowell & Nishiura 2014, Nishiura & Chowell 2015).

Nonetheless, our modeling study has several limitations. First, we imputed time events according to the observational patterns shown in Table 8 while there are various ways of reconstructing the missing events using other alternative combinations of time lags or directions with observed events. The choices of time lag distributions included the incubation period estimated from a larger scale dataset (WHO Ebola Response Team 2014) have shown that the individual protective effect was highly sensitive to the structure of data imputation. We also assumed that all individuals share the same time lag distributions estimated from observed data even though each case has its own probability depending on the normalizing constant calculated using the range of possible days. We note that the ranges shown in Fig 7 illustrate possible dates, where

some cases especially those with wide ranges had low probabilities (Fig 9). Second, we assumed an average impact among all cases without evidence of hospitalization in the baseline scenario. The corresponding individual protective effect was reduced in relation to the probability of hospitalization $p^h = 82\%$. We tested the alternative as formulating using the independent parameter, of which estimates had a three-fold increase. One can instead capture the chance of hospitalization using a random draw from Bernoulli distribution, however, requiring a higher computational cost. Third, we assumed that the protective effect of case isolation remained constant over time and could not capture temporal change during the outbreak. The protectiveness should be dynamical and expected to be lower at the early stage with traditional burials. It then became higher with better strategic planning and interventions such as safe burials in the later stage. Fourth, we assumed that patients became infectious since the onset of symptoms and thus the model focused on dates of symptom onset instead of exposure. While the same approach applies to other diseases such as HIV/AIDS, generation time defined as the time interval between exposure time of cases should replace the usage of serial interval. Fifth, we assumed that the dataset was reliable and captured the transmission dynamics of individuals properly. This is a data-driven modeling approach that heavily relies on the observed data. Some outliers might influence the resulting values and create bias due to the limited sample size. For example, it was observed that Case 18 died on the same day of developing symptoms and Case 22 recovered and discharged on the next day of symptom onset. We, however, did not discard those events due to limited available information.

Overall, the sampling process of this present study required a heavy computational cost due to the large part of missing information. We split the analysis into two parts as point estimation using MLE and Bayesian inference with MCMC sampling. One should have included imputation at once using Bayesian approach draw missing information in each iteration and each individual would have own dynamics of more freedom. Similar approaches of data imputation and modeling estimation was employed in other studies of the 1967 smallpox outbreak in Nigeria (Eichner & Dietz 2003, Stockdale et al. 2017). Additionally, the present study focuses on quantifying the impact of case isolation while the influence of contact tracing was covered in the parameter of protective effect. One interesting modeling approach should include quantifying the impact of contact tracing such as the speed of tracing and coverage of contacts.

In summary, this study demonstrated that case isolation of the 2014 EVD outbreak in Pujehun District, Sierra Leone reduced and shortened the secondary transmission cases and serial interval respectively. The local outbreak was controlled successfully by using the effective intervention measures suspending further spreading via human-to-human transmission. In the presence of successful control, the final size was restricted to 49 while the duration was limited to about 6 months with the declaration of Ebola-free. While vaccination is available as a post-epidemic stage, non-pharmaceutical interventions are extremely important in controlling disease spreading at the initial stage of an outbreak. The pharmaceutical interventions includ-

ing vaccine and drugs are not immediately available, while a new unknown pathogen invades into the human population, for instance, the new coronavirus discovered in Wuhan, China creating an epidemic and declared as public health emergency of international concern (PHEIC) in 2020 (Wu et al. 2020). Mathematical modeling of analyzing epidemiological data in nowcasting and forecasting the potential spreading domestically and internationally remains an important task to provide non-pharmaceutical control strategies and prevent new pandemic occurring in the future.

Conclusion

In this present work, the statistical model was devised and applied to the 2014 EVD epidemics. The data from two local outbreaks in Nigeria and Sierra Leone were analyzed in Chapters 1 and 2 respectively. The impact of case isolation was quantified by the protective effect as the relative reduction of the secondary transmission rate whereas other model parameters represent transmission dynamics in the absence of case isolation. While the data were partially observed, the missing pieces of information including individual time events and sources of infection were reconstructed. Moreover, the sensitivity analysis was performed by considering different scenarios on model assumptions. The followings list the significant findings precisely.

- The protective effect of case isolation was estimated to be 39.7% (95%CI: 2.4%-82.1%) and 23.9% (95%CI: 1.1%-69.5%) for the outbreaks in Nigeria and Sierra Leone, respectively. For most of the cases, the individual protective effect was similar to the ones above.

- The secondary case numbers were fitted by geometric distributions in both outbreaks. Under the assumption of exponential dynamics, the estimates of reproduction number at beginning of two outbreaks yield the mean at 10.0 (95%CI: 3.0-18.5) and 3.0 (95%CI: 1.2-9.1) with the declining rates at 0.14 (95%CI: 0.07-0.23) and 0.019 (95%CI: 0.005-0.035), respectively.

- The unbiased serial interval in the absence of case isolation follows gamma and Weibull distribution for the data from Nigeria and Sierra Leone. The mean values were 15.3 (95%CI: 14.2-16.6) and 12.6 (95%CI: 10.8-14.3) days and SD were 2.3 (95%CI: 1.6-3.5) and 4.0 (95%CI: 3.1-5.4) days, respectively.

- The key individual time events of all cases were reconstructed using various time lag distributions and illustrated probabilistically in timelines during the entire outbreaks. The dates of symptom onset were the most important timing to reveal transmission dynamics, whereas the dates of hospitalization were crucial to determine the effectiveness of public health interventions. The time from symptom onset to hospitalization follows Weibull distributions in both outbreaks. The mean values were 4.0 and 5.0 days and SD were 2.3 and 2.5 days for the data from Nigeria and Sierra Leone, respectively.

- The sources of infection were identified given that there was one missing link of each transmission tree. The possible candidates were ranked by posterior, of which one and three cases were regarded as the most likely infectors to fill the gap in the transmission network in Nigeria and Sierra Leone, respectively.

- Given the alternative scenarios, the reproduction numbers were preferred to be exponentially declining rather than remaining constant over time in both local outbreaks due to their relatively small scales. In addition, this is partially contributed by the index case that

caused the super-spreading event in Nigeria.

- For the outbreak in Sierra Leone, the protective effect of unknown hospitalized cases could be independent of those with hospitalized evidence. Nevertheless, the protective effect of those with evidence remained nearly the same between scenarios, which verified the estimates of the protective effect.

Any epidemics or pandemics can potentially be avoided if efficient control strategies are implemented and carried out effectively. When pharmaceutical interventions are not immediately available especially at the initial stage of an outbreak, isolation of infectious cases and quarantine of potential cases are essential to reduce contacts and hence infections between individuals. As modeling to estimate the impact of public health intervention measures, our approach can be applied to other diseases and outbreaks. This current work provides a statistical framework for analyzing individual surveillance data collected through contact tracing to evaluate the effect of any implemented policies on social contact restrictions. However, transmission routes and dynamics could be different between infectious diseases. Therefore, it might require massive modification and extension of our model. As an example, the 2020 coronavirus pandemic must be the top priority as the declaration of a public health emergency of international concern (PHEIC). To capture the spread of the virus and create a complete transmission network, all infectious individuals including asymptomatic cases should be taken into account. In spite of evaluating the impact of control strategies, it is important to raise public awareness and change individual behavior to protect against the disease.

Acknowledgements

This work was supported by the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan; the Japan Agency for Medical Research and Development (AMED) [grant numbers: JP18fk0108050]; the Japan Society for the Promotion of Science (JSPS) KAKENHI [grant numbers, 17H04701, 17H05808, 18H04895 and 19H01074]; the Inamori Foundation; and the Japan Science and Technology Agency (JST) CREST program [grant number: JPMJCR1413].

Disclosure of Conflict of Interest

The author declares no conflict of interest.

References

Ajelli, M., Parlamento, S., Bome, D., Kebbi, A., Atzori, A., Frasson, C., Putoto, G., Carraro, D. & Merler, S. (2015), 'The 2014 Ebola virus disease outbreak in Pujehun, Sierra Leone: epidemiology and impact of interventions', *BMC medicine* **13**(1), 1.

Bowen, E., Platt, G., Lloyd, G., Baskerville, A., Harris, W., Vella, E. et al. (1977), 'Viral haemorrhagic fever in southern Sudan and northern Zaire. Preliminary studies on the aetiological agent.', *Lancet* pp. 571–573.

Camacho, A., Kucharski, A., Aki-Sawyerr, Y., White, M. A., Flasche, S., Baguelin, M., Pollington, T., Carney, J. R., Glover, R., Smout, E. et al. (2015), 'Temporal changes in Ebola transmission in Sierra Leone and implications for control requirements: a real-time modelling study', *PLoS currents* **7**.

Cauchemez, S., Nouvellet, P., Cori, A., Jombart, T., Garske, T., Clapham, H., Moore, S., Mills, H. L., Salje, H., Collins, C. et al. (2016), 'Unraveling the drivers of MERS-CoV transmission', *Proceedings of the National Academy of Sciences* p. 201519235.

CDC (2019), '2014-2016 ebola outbreak in west africa', (Accessed 23 December 2019).
URL: *https://www.cdc.gov/vhf/ebola/history/2014-2016-outbreak/index.html*

Chowell, D., Castillo-Chavez, C., Krishna, S., Qiu, X. & Anderson, K. S. (2015), 'Modelling the effect of early detection of Ebola', *The Lancet Infectious Diseases* **15**(2), 148–149.

Chowell, G. & Nishiura, H. (2008), 'Quantifying the transmission potential of pandemic influenza', *Physics of Life Reviews* **5**(1), 50–77.

Chowell, G. & Nishiura, H. (2014), 'Transmission dynamics and control of Ebola virus disease (EVD): a review', *BMC medicine* **12**(1), 196.

Drake, J. M., Kaul, R. B., Alexander, L. W., O'Regan, S. M., Kramer, A. M., Pulliam, J. T., Ferrari, M. J. & Park, A. W. (2015), 'Ebola cases and health system demand in Liberia', *PLoS biology* **13**(1), e1002056.

Eichner, M. & Dietz, K. (2003), 'Transmission potential of smallpox: estimates based on detailed data from an outbreak', *American Journal of epidemiology* **158**(2), 110–117.

Fang, L.-Q., Yang, Y., Jiang, J.-F., Yao, H.-W., Kargbo, D., Li, X.-L., Jiang, B.-G., Kargbo, B., Tong, Y.-G., Wang, Y.-W. et al. (2016), 'Transmission dynamics of Ebola virus disease and intervention effectiveness in Sierra Leone', *Proceedings of the National Academy of Sciences* **113**(16), 4488–4493.

Folarin, O. A., Ehichioya, D., Schaffner, S. F., Winnicki, S. M., Wohl, S., Eromon, P., West, K. L., Gladden-Young, A., Oyejide, N. E., Matranga, C. B. et al. (2016), 'Ebola Virus Epidemiology and Evolution in Nigeria', *Journal of Infectious Diseases* p. jiw190.

Fraser, C. (2007), 'Estimating individual and household reproduction numbers in an emerging epidemic', *PLoS One* **2**(8), e758.

Gelfeld, B., Efron, S., Moore, M. & Blank, J. (2015), 'Mitigating the Impact of Ebola in Potential Hot Zones: A Proof-of-Concept Approach to Help Decisionmakers Prepare for High-Risk Scenarios Outside Guinea, Liberia, and Sierra Leone', *RAND Corporation* .

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. & Rubin, D. B. (2013), *Bayesian data analysis*, CRC press.

Halloran, M. E., Longini, I. M., Struchiner, C. J. & Longini, I. M. (2010), *Design and analysis of vaccine studies*, Vol. 18, Springer.

Henao-Restrepo, A. M., Camacho, A., Longini, I. M., Watson, C. H., Edmunds, W. J., Egger, M., Carroll, M. W., Dean, N. E., Diatta, I., Doumbia, M. et al. (2017), 'Efficacy and effectiveness of an rVSV-vectored vaccine in preventing Ebola virus disease: final results from the Guinea ring vaccination, open-label, cluster-randomised trial (Ebola Ça Suffit!)', *The Lancet* **389**(10068), 505–518.

Henao-Restrepo, A. M., Longini, I. M., Egger, M., Dean, N. E., Edmunds, W. J., Camacho, A., Carroll, M. W., Doumbia, M., Draguez, B., Duraffour, S. et al. (2015), 'Efficacy and effectiveness of an rVSV-vectored vaccine expressing Ebola surface glycoprotein: interim results from the Guinea ring vaccination cluster-randomised trial', *The Lancet* **386**(9996), 857–866.

Hens, N., Calatayud, L., Kurkela, S., Tamme, T. & Wallinga, J. (2012), 'Robust reconstruction and analysis of outbreak data: influenza A(H1N1)v transmission in a school-based population', *American journal of epidemiology* p. kws006.

Johnson, K., Webb, P., Lange, J., Murphy, F. et al. (1977), 'Isolation and partial characterisation of a new virus causing acute haemorrhagic fever in Zaire.', *Lancet* pp. 569–571.

Kucharski, A. J., Camacho, A., Checchi, F., Waldman, R., Grais, R. F., Cabrol, J.-C., Briand, S., Baguelin, M., Flasche, S., Funk, S. et al. (2015a), 'Evaluation of the benefits and risks of introducing Ebola community care centers, Sierra Leone', *Emerging infectious diseases* **21**(3), 393.

Kucharski, A. J., Camacho, A., Flasche, S., Glover, R. E., Edmunds, W. J. & Funk, S. (2015b), 'Measuring the impact of Ebola control measures in Sierra Leone', *Proceedings of the National Academy of Sciences* **112**(46), 14366–14371.

Kucharski, A. J., Eggo, R. M., Watson, C. H., Camacho, A., Funk, S. & Edmunds, W. J. (2016), 'Effectiveness of ring vaccination as control strategy for Ebola virus disease', *Emerging infectious diseases* **22**(1), 105.

Lewnard, J. A., Mbah, M. L. N., Alfaro-Murillo, J. A., Altice, F. L., Bawo, L., Nyenswah, T. G.

& Galvani, A. P. (2014), 'Dynamics and control of Ebola virus transmission in Montserrado, Liberia: a mathematical modelling analysis', *The Lancet Infectious Diseases* **14**(12), 1189–1195.

Lokuge, K., Caleo, G., Greig, J., Duncombe, J., McWilliam, N., Squire, J., Lamin, M., Veltus, E., Wolz, A., Kobinger, G. et al. (2016), 'Successful control of Ebola virus disease: analysis of service based data from rural Sierra Leone', *PLoS neglected tropical diseases* **10**(3), e0004498.

Merler, S., Ajelli, M., Fumanelli, L., Gomes, M. F., y Piontti, A. P., Rossi, L., Chao, D. L., Longini Jr, I. M., Halloran, M. E. & Vespignani, A. (2015), 'Spatiotemporal spread of the 2014 outbreak of Ebola virus disease in Liberia and the effectiveness of non-pharmaceutical interventions: a computational modelling analysis', *The Lancet Infectious Diseases* **15**(2), 204–211.

Moore, M., Gelfeld, B. & Adeyemi Okunogbe, C. P. (2017), 'Identifying future disease hot spots: infectious disease vulnerability index', *Rand health quarterly* **6**(3).

Nishiura, H. & Chowell, G. (2015), 'Theoretical perspectives on the infectiousness of Ebola virus disease', *Theoretical Biology and Medical Modelling* **12**(1), 1.

Nishiura, H., Schwehm, M., Kakehashi, M. & Eichner, M. (2006), 'Transmission potential of primary pneumonic plague: time inhomogeneous evaluation based on historical documents of the transmission network', *Journal of epidemiology and community health* **60**(7), 640–645.

Pandey, A., Atkins, K. E., Medlock, J., Wenzel, N., Townsend, J. P., Childs, J. E., Nyenswah, T. G., Ndeffo-Mbah, M. L. & Galvani, A. P. (2014), 'Strategies for containing Ebola in west Africa', *Science* **346**(6212), 991–995.

Pattyn, S., Jacob, W., Van der Groen, G., Piot, P., Courteille, G. et al. (1977), 'Isolation of Marburg-like virus from a case of haemorrhagic fever in Zaire.', *Lancet* pp. 573–574.

Ponce, L., Kinoshita, R. & Nishiura, H. (2019), 'Exploring the human-animal interface of Ebola virus disease outbreaks.', *Mathematical biosciences and engineering: MBE* **16**(4), 3130–3143.

Rivers, C. M., Lofgren, E. T., Marathe, M., Eubank, S. & Lewis, B. L. (2014), 'Modeling the impact of interventions on an epidemic of Ebola in Sierra Leone and Liberia', *PLoS currents* **6**.

Shen, M., Xiao, Y. & Rong, L. (2015), 'Modeling the effect of comprehensive interventions on Ebola virus transmission', *Scientific reports* **5**, 15818.

Shuaib, F., Gunnala, R., Musa, E. O., Mahoney, F. J., Oguntimehin, O., Nguku, P. M., Nyanti,

S. B., Knight, N., Gwarzo, N. S., Idigbe, O. et al. (2014), 'Ebola virus disease outbreak—Nigeria, July–September 2014', *MMWR Morb Mortal Wkly Rep* **63**(39), 867–72.

Stockdale, J. E., Kypraios, T. & O'Neill, P. D. (2017), 'Modelling and Bayesian analysis of the Abakaliki smallpox data', *Epidemics* **19**, 13–23.

Vyshemirsky, V. & Girolami, M. A. (2007), 'Bayesian ranking of biochemical system models', *Bioinformatics* **24**(6), 833–839.

Wallinga, J. & Teunis, P. (2004), 'Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures', *American Journal of Epidemiology* **160**(6), 509–516.

Webb, G., Browne, C., Huo, X., Seydi, O., Seydi, M. & Magal, P. (2015), 'A model of the 2014 Ebola epidemic in West Africa with contact tracing', *PLoS currents* **7**.

WHO Ebola Response Team (2014), 'Ebola virus disease in West Africa—the first 9 months of the epidemic and forward projections', *New England Journal of Medicine* **371**(16), 1481–1495.

Wu, J. T., Leung, K. & Leung, G. M. (2020), 'Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study', *The Lancet* .

Yamin, D., Gertler, S., Ndeffo-Mbah, M. L., Skrip, L. A., Fallah, M., Nyenswah, T. G., Altice, F. L. & Galvani, A. P. (2015), 'Effect of Ebola progression on transmission and control in Liberia', *Annals of internal medicine* **162**(1), 11–17.