Potential for large outbreaks of Ebola virus disease

A. Camacho[a,*], A. J. Kucharski[a,*], S. Funk[a], J. Breman[b], P. Piot[c], W. J. Edmunds[a]

[a]*Centre for the Mathematical Modelling of Infectious Diseases, Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, United Kingdom*
[b]*Fogarty International Center, National Institutes of Health, United States*
[c]*London School of Hygiene and Tropical Medicine, London, United Kingdom*

Abstract

Outbreaks of Ebola virus can cause substantial morbidity and mortality in affected regions. The largest outbreak of Ebola to date is currently underway in West Africa, with 3,944 cases reported as of 5th September 2014. To develop a better understanding of Ebola transmission dynamics, we revisited data from the first known Ebola outbreak, which occurred in 1976 in Zaire (now Democratic Republic of Congo). By fitting a mathematical model to time series stratified by disease onset, outcome and source of infection, we were able to estimate several epidemiological quantities that have previously proved challenging to measure, including the contribution of hospital and community infection to transmission. We found evidence that transmission decreased considerably before the closure of the hospital, suggesting that the decline of the outbreak was most likely the result of changes in host behaviour. Our analysis suggests that the person-to-person reproduction number was 1.34 (95% CI: 0.92–2.11) in the early part of the outbreak. Using stochastic simulations we demonstrate that the same epidemiological conditions that were present in 1976 could have generated a large outbreak purely by

*Corresponding authors. These authors contributed equally to the work.
Email addresses: anton.camacho@lshtm.ac.uk (A. Camacho), adam.kucharski@lshtm.ac.uk (A. J. Kucharski)

chance. At the same time, the relatively high person-to-person basic reproduction number suggests that Ebola would have been difficult to control through hospital-based infection control measures alone.

Keywords: Ebola, 1976 Zaire outbreak, mathematical model, basic reproduction number

Introduction

There have been more than twenty-five known outbreaks of Ebola virus disease in Africa since the disease was first identified in Zaire (now Democratic Republic of Congo) in 1976 (Centers for Disease Control and Prevention, 2014). Five *ebolavirus* strains have been identified in total, the most virulent of which appears to be the Ebola Zaire variant (EBOV); it was responsible for over a dozen outbreaks between 1976–2008, with overall case fatality rate of 79% (95% CI: 0.76–0.81) (Centers for Disease Control and Prevention, 2014; Breman et al., 1978; Formenty et al., 2003; Georges et al., 1999; Heymann et al., 1980; Khan et al., 1999; Leroy et al., 2004; Nkoghe et al., 2011; Pattyn, 1978; Report of an International Commission, 1978). Transmission occurs as a result of direct contact with the body fluids of infected individuals, and is unlikely to occur during the incubation period (Breman et al., 1978; Dowell et al., 1999). In March 2014, a new outbreak of EBOV was identified in West Africa. Cases were reported first in Guinea (Baize et al., 2014), then in Liberia, Sierra Leone, Nigeria and Senegal. The outbreak is the largest to date: as of 5th September 2014, 3,944 cases have been reported by the World Health Organisation, and 1,759 deaths (World Health Organisation, 2014). Unlike previous outbreaks, which were centred on rural communities, infections have also been detected in large urban areas in 2014. It is therefore crucial to develop a better understanding of the transmission dynam-

ics of EBOV, and the implications it could have for control measures.

There have been a number of modelling studies of Ebola, which have focused on two historical outbreaks (Table 1). For the 1995 outbreak in Democratic Republic of Congo, estimates of the basic reproduction number have ranged from 1.4 to 3.7 (Chowell et al., 2004; Ferrari et al., 2005; Legrand et al., 2007; Lekone and Finkenstädt, 2006; Ndanguza et al., 2013; White and Pagano, 2008); for the 2000/1 outbreak in Uganda, estimates span 1.3 to 2.7 (Chowell et al., 2004; Ferrari et al., 2005; Legrand et al., 2007; McKinley et al., 2009). These studies fitted models of varying complexity to time series with date of disease onset and/or death. However, in both outbreaks, hospital-based infection played a substantial role in transmission (Borchert et al., 2011; Khan et al., 1999; Francesconi et al., 2003). As the data were not stratified by likely source of infection, it was not possible to identify the relative contribution of different transmission routes to the reproduction number. It therefore remains unclear to what extent person-to-person transmission contributed to past Ebola outbreaks, and how community and hospital-specific control measures influenced the reproduction number in each setting.

To gain further insights into the dynamics of Ebola, we revisited case data from the first known EBOV outbreak in 1976. These data included information on the likely source of infection, as well as date of onset and outcome. We used a transmission model to infer the basic reproduction number in different settings, and assessed the contribution of hospital and community infection to disease transmission. Having characterised the dynamics of EBOV, we used stochastic simulations to investigate alternative outcomes that could have been generated with the same epidemiological conditions present in 1976, and assessed the potential for a large outbreak of the disease. Finally, we discuss the implications of our results for other Ebola outbreaks.

Methods

Data

Between August and November 1976, there were 318 reported cases of Ebola in the Yandongi collectivity of Zaire, with 280 deaths. The outbreak was centred around the Yambuku Mission Hospital. With only five syringes issued each day, exposure to contaminated syringes and needles during routine outpatient visits was a common route of transmission; infected hosts then returned to their villages, and in some cases infected others in the community (Breman et al., 1978).

In our analysis, we used a line list of 262 cases, taken from the original epidemiological investigations (Breman et al., 1978; Report of an International Commission, 1978). The data (Supplementary File S1) reported: date of disease onset; outcome (death/recovery); date of outcome; and likely source of transmission (syringe during outpatient visit/person-to-person transmission/both/other). The progression of the outbreak is shown in Figure 1. Of the reported 262 cases, 250 had a likely source of infection recorded and 8 dates of onset and outcome were missing (Table S1). We used the line list to compile four daily time series: onset of disease following hospital infection via syringe (87 cases in total); onset of disease following person-to-person infection (140 cases in total); reported deaths (248 cases in total); and reported recoveries (11 cases in total).

Model

We used a compartmental model of infection to analyse the temporal dynamics of Ebola (Legrand et al., 2007). The model structure is outlined in Figure 2. We assumed that individuals start off susceptible to infection (S). Upon infection they enter an incubation period (E), then become symptomatic and infectious in the community (I). We therefore assume that the latent and incubation periods are

equivalent. After this point, they either: enter a recovered state (R); remain infectious and go into hospital (H); or die and remain infectious (D) until buried (B). Following hospitalisation, infectious hosts also move either into the recovered or dead compartment.

We assumed susceptible hosts in the community could become infected in three different ways: person-to-person transmission from an infectious host in the community, at rate $\beta_i(t)$, or from a dead but not buried patient during a traditional funeral ceremony, at rate $\beta_d(t)$; or hospital transmission via syringe during outpatient visits, at rate $\beta_h(t)$.

There was evidence that hospital and person-to-person transmission declined over the course of the 1976 outbreak. Epidemiological reports note that the community stopped coming to the outpatient department as they associated the epidemic with the Yambuku Mission Hospital, which eventually was closed on 30th September. Also, as time went on the population became very suspicious and did not touch the corpses anymore, not even to bury them (Breman et al., 1978). We therefore used time-dependent smooth decreasing functions for $\beta_i(t)$, $\beta_d(t)$ and $\beta_h(t)$ (Chowell et al., 2004; Lekone and Finkenstädt, 2006; Ndanguza et al., 2013):

$$\begin{aligned}
\beta_i(t) &= \beta_i\big(1 - \delta_{pp}\sigma(t, \alpha_{pp}, \tau_{pp})\big) \\
\beta_d(t) &= \beta_d\big(1 - \delta_{pp}\sigma(t, \alpha_{pp}, \tau_{pp})\big) \\
\beta_h(t) &= \beta_h\big(1 - \sigma(t, \alpha_h, \tau_h)\big)\mathbb{1}_{t<T_h}
\end{aligned} \tag{1}$$

where $\mathbb{1}_{t<T_h}$ is the indicator function and σ is the following sigmoid function:

$$\sigma(t, \alpha, \tau) = \frac{1}{1 + \exp\big(-\alpha(t - \tau)\big)}. \tag{2}$$

As t grew large, we assumed that $\beta_i(t)$ and $\beta_d(t)$ approached values equal to a proportion δ_{pp} of their initial values. We also assumed that no further hospital transmission occurred after the hospital closed on 30 September (i.e. $\beta_d(t) = 0$),

and that no new cases entered the H compartment after this point.

Using our estimates of $\beta_i(0)$, $\beta_d(0)$ and $\beta_h(0)$, we were able to calculate the basic reproduction number, R_0, defined as the average number of secondary cases produced by a typical infectious host at the onset of the outbreak (i.e., in a completely susceptible population), see details in Text S3. At the start of the outbreak, the reproduction number, $R(t)$, defined as the average number of secondary cases produced by a typical infectious host at time t, was equal to R_0; as the outbreak progressed, $R(t)$ could vary depending on the values of $\beta_i(t)$, $\beta_d(t)$ and $\beta_h(t)$, as well as through depletion of susceptibles.

The full model was as follows (for brevity, the time dependencies of state variables are omitted):

$$\begin{aligned}
\frac{\mathrm{d}S}{\mathrm{d}t} &= -\big(\beta_i(t)I + \beta_h(t)H + \beta_d(t)D\big)\frac{S}{N} \\
\frac{\mathrm{d}E_{pp}}{\mathrm{d}t} &= \big(\beta_i(t)I + \beta_d(t)D\big)\frac{S}{N} - \epsilon E_{pp} \\
\frac{\mathrm{d}E_h}{\mathrm{d}t} &= \beta_h(t)H\frac{S}{N} - \epsilon E_h \\
\frac{\mathrm{d}I}{\mathrm{d}t} &= \epsilon(E_{pp} + E_h) - \Gamma_i(t)I \\
\frac{\mathrm{d}H}{\mathrm{d}t} &= \gamma_h\kappa_i(t)I - \big(\phi_h\nu_d + (1-\phi_h)\nu_r\big)H \\
\frac{\mathrm{d}D}{\mathrm{d}t} &= \gamma_d\big(1-\kappa_i(t)\big)\phi_i I + \nu_d\phi_h H - \mu_b D \\
\frac{\mathrm{d}R}{\mathrm{d}t} &= \gamma_r\big(1-\kappa_i(t)\big)(1-\phi_i)I + \nu_r(1-\phi_h)H \\
\frac{\mathrm{d}B}{\mathrm{d}t} &= \mu_b D
\end{aligned} \tag{3}$$

Parameters are summarised in Table 2. We used flat priors for case-fatality ratio and transmission-related parameters. We also used additional epidemiological information not in our time series (Supplementary File S2; (Breman et al., 1978; Report of an International Commission, 1978; Khan et al., 1999; Nkoghe et al., 2011; Okware et al., 2002)) to inform strong prior distributions for: proportion

of cases reported; proportion of cases hospitalised; incubation period; time from onset to hospitalization, death and recovery. A strong prior centred around 24 hours was used for the time from death to burial of individuals (Isaacson et al., 1978; Sureau et al., 1978). We fixed the initially susceptible population size at 60,000, as this was the number of people for which the Yambuku Mission Hospital served as principal point of care (Report of an International Commission, 1978). We assumed that the index case was introduced in the H compartment at a time T_0, which was also estimated. Indeed, the first reported case (25th August) was infected via a syringe and there is evidence that an unknown man came to the hospital with Ebola-like symptoms shortly before that date (Breman et al., 1978).

As the model included multiple transitions between compartments, we needed to define certain parameters carefully. To ensure that the overall case-fatality ratio was equal to ϕ, we defined ϕ_i and ϕ_h as follows:

$$\begin{aligned} \phi_i &= \frac{\phi\gamma_r}{\phi\gamma_r + (1-\phi)\gamma_d} \\ \phi_h &= \frac{\phi\nu_r}{\phi\nu_r + (1-\phi)\nu_d}. \end{aligned} \tag{4}$$

Similarly, $\kappa_i(t)$ was computed to ensure that the overall hospitalisation rate was equal to κ until hospital closure:

$$\kappa_i(t) = \frac{\kappa\big(\gamma_r(1-\phi_i) + \phi_i\gamma_d\big)}{\kappa\big(\gamma_r(1-\phi_i) + \phi_i\gamma_d\big) + (1-\kappa)\gamma_h}\mathbb{1}_{t<T_h}. \tag{5}$$

Finally, $\Gamma_i(t)$ denotes the total rate of exit from the I compartment:

$$\Gamma_i(t) = \gamma_h\kappa_i(t) + \gamma_d\big(1-\kappa_i(t)\big)\phi_i + \gamma_r\big(1-\kappa_i(t)\big)(1-\phi_i), \tag{6}$$

and ν_d and ν_r are the inverse of the mean time from hospitalisation to death and recovery respectively:

$$\begin{aligned} \nu_d &= \frac{\gamma_d\gamma_h}{\gamma_h - \gamma_d} \\ \nu_r &= \frac{\gamma_r\gamma_h}{\gamma_h - \gamma_r}. \end{aligned} \tag{7}$$

Inference

To compare the model output with observed data, we calculated the incidences corresponding to the four time series on each day:

$$\begin{aligned}
\Delta I_{pp}(t) &= \int_t^{t+1} \epsilon E_{pp} \mathrm{d}t \\
\Delta I_h(t) &= \int_t^{t+1} \epsilon E_h \mathrm{d}t \\
\Delta D(t) &= \int_t^{t+1} \Big(\gamma_d\big(1-\kappa_i(t)\big)\phi_i I + \nu_d \phi_h H\Big)\mathrm{d}t \\
\Delta R(t) &= \int_t^{t+1} \Big(\gamma_r\big(1-\kappa_i(t)\big)(1-\phi_i) I + \nu_r (1-\phi_h) H\Big)\mathrm{d}t
\end{aligned} \tag{8}$$

We assumed that onset (i.e. $\Delta I_{pp}(t), \Delta I_h(t)$), death ($\Delta D(t)$) and recovery ($\Delta R(t)$) data were reported according to Poisson processes with constant reporting rates ρ_{onset}, ρ_d and ρ_r respectively. We allowed for three potentially different reporting rates because i) not all cases had reported onsets; ii) 30 of the 262 were reported with both, other or unknown source of infection rather than person-to-person or syringe. These cases were therefore included in the outcome time series but not in the onset one; and iii) only 11 of 38 (29%) recovery cases versus 248 of 280 (89%) death cases were reported.

The probability of observing $y_{pp}(t)$ new onsets resulting from person-to-person transmission on day t, given parameter set θ (summarised in Table 2) was therefore as follows:

$$L_{pp}(y_{pp}(t) \mid \theta) = \frac{[\rho_{\text{onset}} \Delta I_{pp}(t)]^{y_{pp}(t)} e^{-\rho_{\text{onset}} \Delta I_{pp}(t)}}{y_{pp}(t)!} \,. \tag{9}$$

Similar expressions (L_h, L_D and L_R) were derived for the other incidence data and combined into the likelihood function:

$$L(y_{pp}, y_h, y_D, y_R \mid \theta) = \prod_{t=1}^{72} \prod_{k \in \{pp,h,D,R\}} L_k(y_k(t) \mid \theta) \tag{10}$$

We used a Bayesian framework to fit the model to all four time series simultaneously and make inference on the parameter set θ. Given the likelihood function L and the chosen prior distribution of the parameters, the posterior distribution is known up to a normalising constant. Markov chain Monte Carlo methods construct Markov chains whose stationary distribution is the distribution of interest, when it cannot be directly simulated. We used the SSM library (Dureau et al., 2013), which implements an adaptive Metropolis-Hastings algorithm (Roberts and Rosenthal, 2009) to generate sequences of draws from the posterior distribution of the parameters. We refer to Text S1 for more details. To test the accuracy of our inference framework, we fitted the model to observed data, generated a set of simulated time series from our fitted model, then estimated the parameters from the simulated data; our inference framework was able to recover the parameters in question (Text S2, Table S2 and Figures S2–S5).

Results

Our model was able to capture the dynamics of Ebola virus disease, including infections resulting from exposure to contaminated syringes and person-to-person transmission, and the timing of outcomes (Figure 3). By fitting to multiple time series, we were able to jointly estimate a number of key epidemiological parameters (Table 2 and Figure S1). Using these estimates, we calculated the contribution of person-to-person transmission (via infection from living and dead hosts in the community) and hospital-based transmission (via contaminated syringe) to the overall basic reproduction number, R_0 (see Text S3). We found that the overall R_0 was 4.71 (95% CI: 3.92–5.66) at the onset of the epidemic. Most of this number was the result of hospital-based transmission, although we found evidence that the person-to-person basic reproduction number was potentially above 1 (Table 3).

Person-to-person transmission was separated into two components in the model:

infections occurring while the case was alive and those occurring after death (i.e. before the patient had been buried). This meant fitting both transmission rates β_i and β_d. However, due to limited data on these specific transmission routes, the relative contribution to infection from living and dead patients in the community was not fully identifiable (see Text S1). When we fitted both transmission rates independently, the contributions from community and funeral cases to R_0 were highly correlated (Figure S6). As it was not possible to identify the contribution from community and funeral infection to person-to-person transmission, we therefore gathered the two measurements together into a single person-to-person basic reproduction number, denoted R_{0pp}, which could be estimated from the available data.

As the epidemic progressed, we found that the overall reproduction number decreased due to changes in the contact rate within the community and within the hospital. Splitting the overall reproduction number into its person-to-person and hospital components, we found that although hospital transmission was dominant during the early stages of the epidemic, it had dropped significantly by mid September (Figure 4). Our results suggest the hospital reproduction number R_h was below 1 well before the hospital closed on the 30 of September. Moreover, we found that hospital closure alone could not explain the observed data; when changes in person-to-person and hospital-based transmission were excluded, the model performed significantly worse (Table S3). We estimated that the drop in person-to-person transmission occurred later and less sharply than the reduction in exposure to contaminated syringes. However, the reduction in person-to-person transmission was still enough to drive the overall reproduction number below 1 by the end of September. Overall, these results are consistent with the observations reported by the epidemiological investigation team (Breman et al., 1978).

To examine the possible range of dynamics for an outbreak with the same

characteristics as the one observed in the 1976 Yambuku outbreak, we ran 10,000 stochastic simulations of our model under the maximum a posteriori probability estimates of the parameters (Figure 5). We found that although most simulated epidemics were of similar size to the one in 1976, major outbreaks could also occur. Although only 2.6% of simulations resulted in a major outbreak (i.e. more than 1000 cases), the cumulated number of cases could reach up to several thousands in the worst-case scenario (Figure 6A). In the context of the 1976 epidemic, such a major outbreak could have arisen if – by chance – a sufficiently high number of infections had occurred before the change of community contact and hospital seeking behaviours.

To understand how different control measures could affect outbreak size, we also considered several alternative scenarios in our simulation study. First, we set the hospital closure date in the model to be 7 days after the onset date of the first case in the line list i.e. on 1st September rather than 30th. Although early closure resulted in fewer cases, with no outbreaks generating more than 1000 cases, there were still occasionally outbreaks consisting of several hundred cases (Figure 6C). Next, we examined the effect of a smaller reduction in person-to-person transmission, assuming that the hospital reduction remained the same. We found that if the person-to-person transmission rate was reduced by 50% – rather than 98% as in our median estimates – transmission could persist longer in the community, and hence 28% of simulations resulted in an outbreak of at least 1000 cases (Figure 6D). The number of reported cases in historical Ebola outbreaks has varied greatly, from a few infections to more than 3500 (Figure 6B); our results suggest that such variability might be expected given the transmission dynamics of Ebola.

To explore the possibility of large outbreak occurring without the large contribution from hospital transmission, we also considered a model with only person-

to-person transmission (i.e. $R_{0h} = 0$). We assumed that the index case started in the community (I compartment). In the absence of control measures, we found that 35% of outbreaks resulted in more than 1000 cases.

As well as allowing us to model setting-specific transmission, the line list made it possible to directly calculate the case-fatality ratio (CFR) in different settings. We found that the probability of survival varied depending on route of transmission. The overall CFR – defined as the proportion of cases that died – across the 262 cases in our line list was $251/262 = 0.96$ (binomial 95% CI: 0.93–0.98). The CFR for cases that resulted from person-to-person transmission was 0.92 (0.87–0.96); in contrast, the CFR for cases that were exposed via a contaminated syringe was 1.00 (0.96–1.00). Note that these empirical CFR estimates are based on a subset of 262 of the 318 reported cases in the Yambuku outbreak, for whom individual data were available. The CFR based on all 318 reported cases was $280/318 = 0.88$ (95% CI: 0.86–0.92) (Breman et al., 1978).

Discussion

Using a model of Ebola virus transmission, we examined the role of different transmission routes during the 1976 outbreak in DRC. We found that the basic reproduction number (R_0) associated with hospital transmission was significantly above one. Our analysis also suggests that the person-to-person reproduction number R_{pp} could have been above 1 for the early part of the outbreak. This has profound implications: it suggests that a large outbreak (involving thousands of cases) could have happened even without changing any epidemiological conditions. We estimated the probability of such a large outbreak (>1000 cases) to be around 3%. This means that given the same initial conditions, Ebola outbreaks would have been occasionally been large, just by chance. Moreover, a relatively high person-to-person transmission component ($R_{0pp} \approx 1$) implied that the 1976

epidemic would have been difficult to control via hospital-based infection control measures alone. If the reduction in community transmission had been smaller, or infection had been seeded into a number of different communities, the outbreak could have continued for some time.

Our results also suggest that changes in behaviour caused a significant reduction in both hospital-to-community and within-community transmission. Although Yambuku Mission hospital closed on the 30th September, we found that the reduction in transmission occurred well before this point, most likely from susceptible hosts having less contact with infected patients, and making fewer routine outpatient visits to the hospital (Breman et al., 1978). As well as contributing to transmission, infections from syringes also appeared to have a higher case fatality ratio (CFR) than person-to-person infections. This could have been the result of a larger viral inoculum during contact with a contaminated syringe. With more data on transmission events – including chains of person-to-person infection – it would be possible to further investigate the role of exposure in the natural history of Ebola infection.

Even with four time series, it was not possible to robustly distinguish between person-to-person transmission resulting from contact with community cases and funeral attendance. Additional case data, such as dates on which patients took care of an infected case or attended a funeral ceremony could allow us to disentangle the relative role of these two routes of community transmission. However, it is plausible that individuals had similar contact rates with infected and dead patients. Epidemiological investigations in 1976 found that 86% of hosts infected from person-to-person transmission reported prior contact with alive Ebola patients; the same proportion reported attending the funeral of an infected case (Breman et al., 1978). Assuming similar transmissibility for both types of contact, this would be equivalent to setting $\beta_i = \beta_d$ in our model.

There are some additional limitations to the model. First, we assumed that hosts mixed randomly both in the community and hospital. This was a reasonable assumption given that we stratified the data by route of transmission and outcome. However, there was evidence that certain groups, such as women aged 15–29, were more likely to attend clinics at Yambuku Mission Hospital in 1976 and hence be exposed to syringes (Report of an International Commission, 1978). To model the dynamics of the infection at a finer resolution, for instance by comparing model outputs to age-stratified case data, it would be necessary to account for such heterogeneity. We also assumed that occurrence of reported cases was Poisson distributed, and the proportion reported did not vary over time or by location. This might be plausible when cases occur in a relatively short outbreak in a small geographic region, but during outbreaks that span a much larger geographic area and persist for several months, reporting could change with time and vary between different settings. Moreover, if the dynamics of Ebola were to be modelled in real-time, it would be important to account for potential delays in reporting of cases and outcomes.

In our stochastic scenario analysis we also assumed that timing and magnitude of changes in transmission rate were independent of epidemic size. Our simulations that used parameters from the fitted model (Figure 6A) therefore assumed that identification and control of the infection would not have occurred quicker if more individuals had been infected earlier. However, we tested the sensitivity of our results to timing of hospital closure by assuming that the hospital closed one week after the first case (Figure 6C); we also explored the effects of a smaller change in magnitude in person-to-person risk (Figure 6D). Ideally, it would be possible to define a functional relationship between incidence and changes in transmission rate (Funk et al., 2009). However, this relationship is likely to be complex and setting-specific: in 1976, behavioural changes reduced

transmission (Breman et al., 1978); in other Ebola outbreaks, large amounts of infection have increased fear and mistrust in the community, which might also have increased transmission (Borchert et al., 2011; World Health Organisation, 2014).

The modelling tools illustrated here could easily be adapted for other Ebola outbreaks, and highlight the benefits of having data on likely source of infection and time of onset, hospitalisation and outcome for each patient. Previous Ebola modelling studies have examined the 1995 outbreak in Kikwit, DRC (Chowell et al., 2004; Ferrari et al., 2005; Legrand et al., 2007; Lekone and Finkenstädt, 2006; White and Pagano, 2008; McKinley et al., 2009; Ndanguza et al., 2013), and the 2000/1 outbreak in Uganda (Chowell et al., 2004; Ferrari et al., 2005; Legrand et al., 2007). As in Yambuku in 1976, hospital-based transmission played a substantial role in both outbreaks (Khan et al., 1999; Francesconi et al., 2003). However, modelling studies so far have incorporated time series for onset and/or death only, which meant that it was not possible to robustly infer the role of different routes of infection, such as the contribution of hospital and community transmission. In contrast, by fitting a transmission model to time series stratified by transmission route, we were able to estimate the contribution of different sources of infection to the dynamics of the epidemic.

We estimated that the overall R_0 was 4.71 (95% CI: 3.92–5.66) for the 1976 Yambuku outbreak. This is high compared to estimates of R_0 in the 1995 and 2000/1 outbreaks, which ranged from 1.34–3.65 (Table 1). However, our analysis suggests that most of the R_0 in 1976 consisted of transmission via syringe; the person-to-person basic reproduction number was 1.34 (0.92–2.11). Given data on likely source of infection in 1995 and 2000/1, it would be possible to establish whether person-to-person transmission contributed a similar amount to overall transmission during these outbreaks.

Our estimate of a person-to-person basic reproduction number $R_{0pp} \approx 1$ in

1976 suggests that Ebola would have been capable of generating a wide range of outbreak sizes in the absence of any extrinsic variation in epidemiological conditions. This implies that effective reduction in person-to-person transmission was crucial in reducing the potential size of the outbreak; stochastic simulations suggest Ebola could still have generated a large number of cases if hospital transmission was absent in 1976. Measures to reduce person-to-person transmission – including isolation of patients, follow-up surveillance of their contacts, and education to curtail infection in the community – are therefore likely to form a crucial part of the response to Ebola outbreaks (Borchert et al., 2011; Khan et al., 1999; Okware et al., 2002).

As well as variation in social and cultural factors between different regions, the stochastic nature of Ebola outbreaks means that inference to other settings must be done with caution. Our analysis concentrates on a single outbreak of 318 cases, rather than a set of past Ebola outbreaks, which have ranged from a small number of cases to several thousand (Figure 6B). Analyses of data from a large number of historical outbreaks simultaneously would help reduce this stochastic uncertainty and allow comparative studies to be performed. By making the line listing of the 1976 outbreak available (Supplementary File S1), we hope to stimulate such work. Comparative studies could potentially shed further light on which underlying factors contribute to the differences in outcome of Ebola outbreaks, and which control measures are likely to be most effective.

Acknowledgements

This work was supported by the Medical Research Council (AC, fellowship MR/J01432X/1; AJK, fellowship MR/K021524/1; SF, fellowship MR/K021680/1). We thank Prof. Bernard Cazelles for access to the computational resources used in this work, which were funded by an investment grant from the Région Île-de-

France through the scientific program DIM MALINF.

References

Baize, S., Pannetier, D., Oestereich, L., Rieger, T., Koivogui, L., Magassouba, N., Soropogui, B., Sow, M.S., Keïta, S., De Clerck, H., Tiffany, A., Dominguez, G., Loua, M., Traoré, A., Kolié, M., Malano, E.R., Heleze, E., Bocquin, A., Mély, S., Raoul, H., Caro, V., Cadar, D., Gabriel, M., Pahlmann, M., Tappe, D., Schmidt-Chanasit, J., Impouma, B., Diallo, A.K., Formenty, P., Van Herp, M., Günther, S., 2014. Emergence of Zaire Ebola Virus Disease in Guinea - Preliminary Report. N Engl J Med 0.

Borchert, M., Mutyaba, I., Van Kerkhove, M.D., Lutwama, J., Luwaga, H., Bisoborwa, G., Turyagaruka, J., Pirard, P., Ndayimirije, N., Roddy, P., et al., 2011. Ebola haemorrhagic fever outbreak in Masindi District, Uganda: outbreak description and lessons learned. BMC infectious diseases 11, 357.

Breman, J., Piot, P., Johnson, K., White, M., Mbuyi, M., Sureau, P., Heymann, D., Van Nieuwenhove, S., McCormick, J., Ruppol, J., et al., 1978. The epidemiology of Ebola hemorrhagic fever in Zaire, 1976. Ebola virus haemorrhagic fever (Pattyn, SR ed.) , 85–97.

Centers for Disease Control and Prevention, 2014. Outbreaks Chronology: Ebola Hemorrhagic Fever. http://www.cdc.gov/vhf/ebola/resources/outbreak-table.html .

Chowell, G., Hengartner, N.W., Castillo-Chavez, C., Fenimore, P.W., Hyman, J.M., 2004. The basic reproductive number of Ebola and the effects of public health measures: the cases of Congo and Uganda. J Theor Biol 229, 119–26.

Dowell, S.F., Mukunu, R., Ksiazek, T.G., Khan, A.S., Rollin, P.E., Peters, C., 1999. Transmission of Ebola hemorrhagic fever: a study of risk factors in family members, Kikwit, Democratic Republic of the Congo, 1995. Journal of Infectious Diseases 179, S87–S91.

Dureau, J., Ballesteros, S., Bogich, T., 2013. SSM: Inference for time series analysis with State Space Models. arXiv 1307.5626v4.

Ferrari, M.J., Bjørnstad, O.N., Dobson, A.P., 2005. Estimation and inference of R_0 of an infectious pathogen by a removal method. Math Biosci 198, 14–26.

Formenty, P., Libama, F., Epelboin, A., Allarangar, Y., Leroy, E., Moudzeo, H., Tarangonia, P., Molamou, A., Lenzi, M., Ait-Ikhlef, K., et al., 2003. L'épidémie de fievre hémorragique avirus Ebola en Republique du Congo, 2003: une nouvelle stratégie. Med Trop (Mars) 63, 291–5.

Francesconi, P., Yoti, Z., Declich, S., Onek, P.A., Fabiani, M., Olango, J., Andraghetti, R., Rollin, P.E., Opira, C., Greco, D., et al., 2003. Ebola hemorrhagic fever transmission and risk factors of contacts, Uganda. Emerging infectious diseases 9, 1430.

Funk, S., Gilad, E., Watkins, C., Jansen, V.A.A., 2009. The spread of awareness and its impact on epidemic outbreaks. Proc Natl Acad Sci U S A 106, 6872–6877.

Georges, A.J., Leroy, E.M., Renaut, A.A., Benissan, C.T., Nabias, R.J., Ngoc, M.T., Obiang, P.I., Lepage, J., Bertherat, E.J., Bénoni, D.D., et al., 1999. Ebola hemorrhagic fever outbreaks in Gabon, 1994–1997: epidemiologic and health control issues. Journal of Infectious Diseases 179, S65–S75.

Heymann, D., Weisfeld, J., Webb, P., Johnson, K., Cairns, T., Berquist, H., 1980. Ebola hemorrhagic fever: Tandala, Zaire, 1977–1978. Journal of Infectious Diseases 142, 372–376.

Isaacson, M., Ruppol, J., Collas, R., Matundu, N., Tshibamba, Omombo, K., 1978. Containment and surveillance of a hospital outbreak of Ebola virus disease in Kinshasa, Zaire, 1976, in: Ebola virus haemorrhagic fever (Pattyn, SR ed.), Elsevier/North-Holland Biomedical Press Amsterdam, The Netherlands. pp. 122–127.

Khan, A.S., Tshioko, F.K., Heymann, D.L., Le Guenno, B., Nabeth, P., Kerstiëns, B., Fleerackers, Y., Kilmarx, P.H., Rodier, G.R., Nkuku, O., et al., 1999. The reemergence of Ebola hemorrhagic fever, Democratic Republic of the Congo, 1995. Journal of Infectious Diseases 179, S76–S86.

Legrand, J., Grais, R.F., Boelle, P.Y., Valleron, A.J., Flahault, A., 2007. Understanding the dynamics of Ebola epidemics. Epidemiol Infect 135, 610–21.

Lekone, P.E., Finkenstädt, B.F., 2006. Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study. Biometrics 62, 1170–7.

Leroy, E.M., Rouquet, P., Formenty, P., Souquière, S., Kilbourne, A., Froment, J.M., Bermejo, M., Smit, S., Karesh, W., Swanepoel, R., Zaki, S.R., Rollin, P.E., 2004. Multiple Ebola virus transmission events and rapid decline of central African wildlife. Science 303, 387–90.

McKinley, T., Cook, A.R., Deardon, R., 2009. Inference in epidemic models without likelihoods. Int J Biostat 5, 24.

Ndanguza, D., Tchuenche, J., Haario, H., 2013. Statistical data analysis of the 1995 Ebola outbreak in the Democratic Republic of Congo. Afrika Matematika 24, 55–68.

Nkoghe, D., Kone, M.L., Yada, A., Leroy, E., 2011. A limited outbreak of Ebola haemorrhagic fever in Etoumbi, Republic of Congo, 2005. Transactions of the Royal Society of Tropical Medicine and Hygiene 105, 466–472.

Okware, S., Omaswa, F., Zaramba, S., Opio, A., Lutwama, J., Kamugisha, J., Rwaguma, E., Kagwa, P., Lamunu, M., 2002. An outbreak of Ebola in Uganda. Tropical Medicine & International Health 7, 1068–1075.

Pattyn, S. (Ed.), 1978. Ebola virus haemorrhagic fever. Elsevier.

Report of an International Commission, 1978. Ebola haemorrhagic fever in Zaire, 1976. Bull World Health Organ 56, 271–293.

Roberts, G.O., Rosenthal, J.S., 2009. Examples of adaptive MCMC. Journal of Computational and Graphical Statistics 18, 349–367.

Sureau, P., Piot, P., Breman, G., Ruppol, F., Masamba, M., Berquist, H., Heymann, D., Kintoki, V., Koth, M., Mandiangu, M., et al., 1978. Containment and surveillance of an epidemic of Ebola virus infection in Yambuku area, Zaire, 1976, in: Ebola virus haemorrhagic fever (Pattyn, SR ed.), Elsevier/North-Holland Biomedical Press Amsterdam, The Netherlands. pp. 116–121.

White, L.F., Pagano, M., 2008. A likelihood-based method for real-time estimation of the serial interval and reproductive number of an epidemic. Stat Med 27, 2999–3016.

World Health Organisation, 2014. Disease outbreak news. WHO online .

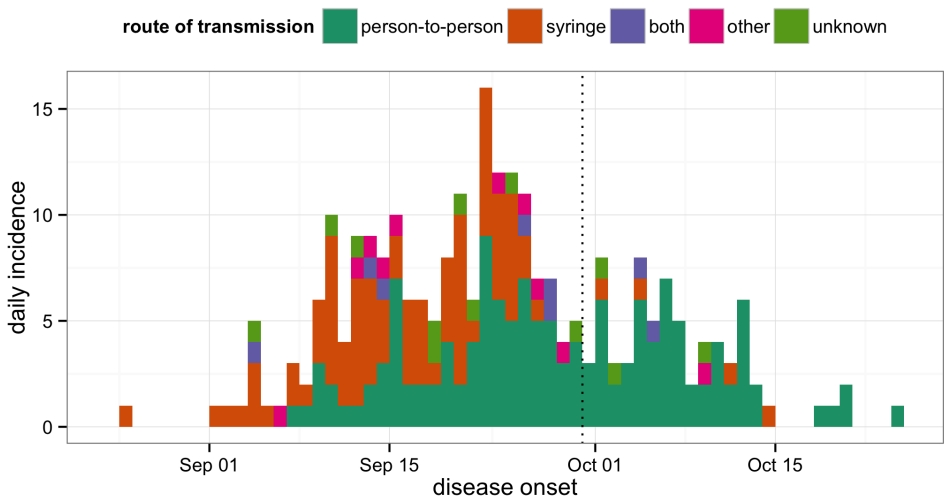

Figure 1: Daily incidence time series of Ebola virus disease onsets in 1976. Cases are coloured by route of transmission, as reported by the epidemiological investigation team (Breman et al., 1978). 'Both' indicates infections that could have come from syringe or person-to-person transmission; 'other' denotes alternative infection routes (mainly congenital). The dotted line corresponds to the hospital closure date (30th September).

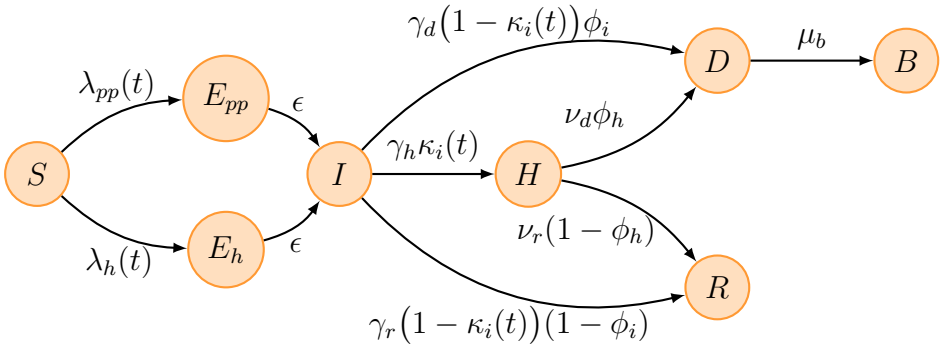

Figure 2: Schematic of model structure. Individuals start off susceptible to infection (S). Upon infection they enter an incubation period (E), then at symptom onset they become infectious in the community (I). After this point, they either: enter a recovered state (R); remain infectious and go into hospital (H); or die and remain infectious (D) until buried (B). Hospitalised infectives also move either into the recovered or dead compartment. Finally, the E compartment is split according to the route of transmission in order to keep track whether a case was infected via contaminated syringes at the hospital (E_h) or by person-to-person contact (E_{pp}) with either an infective in the community or a dead but not buried case. The forces of infection for the two transmission processes are $\lambda_h(t) = \beta_h(t)H/N$ and $\lambda_{pp}(t) = \big(\beta_i(t)I + \beta_d(t)D\big)/N$, where $\beta_h(t)$, $\beta_i(t)$ and $\beta_d(t)$ are the time-varying transmission rates given by Equation (1). Other parameters are as follows: ϵ, inverse of the mean incubation period; γ_h, γ_d and γ_r, inverse of the mean duration from symptom onset to hospitalization, death and recovery respectively; ν_d and ν_r, inverse of the mean duration from hospitalization to death and recovery respectively (see Equation (7)); μ_b, inverse of the mean duration from death to burial; $\kappa_i(t)$ is computed to ensure that the overall hospitalisation rate is equal to κ until hospital closure (see Equation (5)); ϕ_i and ϕ_h are computed to ensure that the overall case-fatality ratio is equal to ϕ (see Equation (4)). Parameter values and prior assumptions can be found in Table 2. The model was simulated by integrating the set (3) of ordinary differential equations using the SSM library (Dureau et al., 2013).

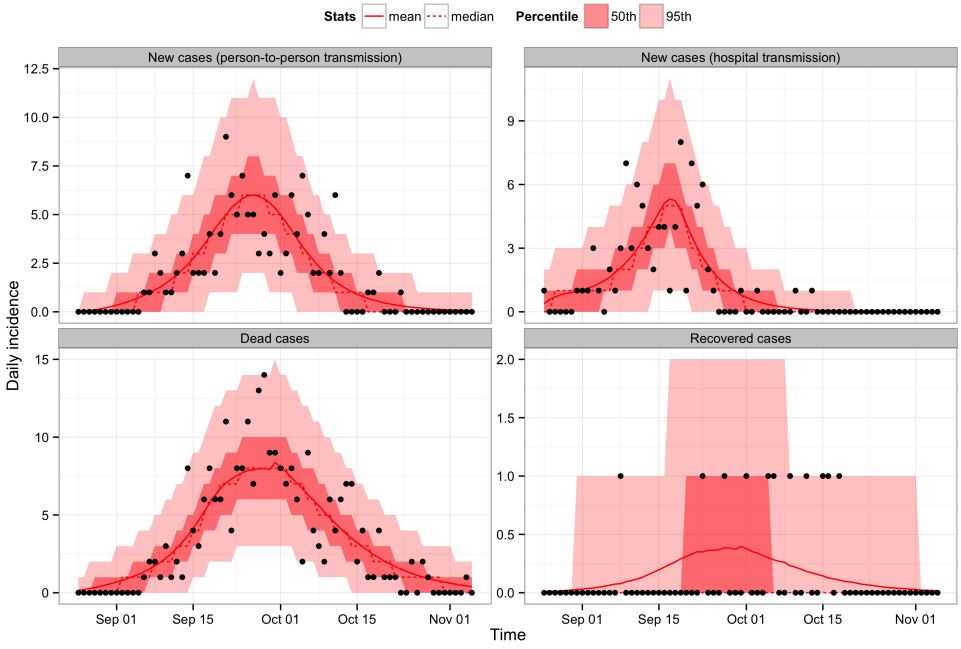

Figure 3: Comparison of our fitted model and observed daily incidence time series (black dots) reconstructed from the line list of Ebola cases in Zaire in 1976. The mean and median fits are represented by solid and dashed red lines respectively. The dark and light red shaded areas correspond to the 50% and 95% credible intervals.

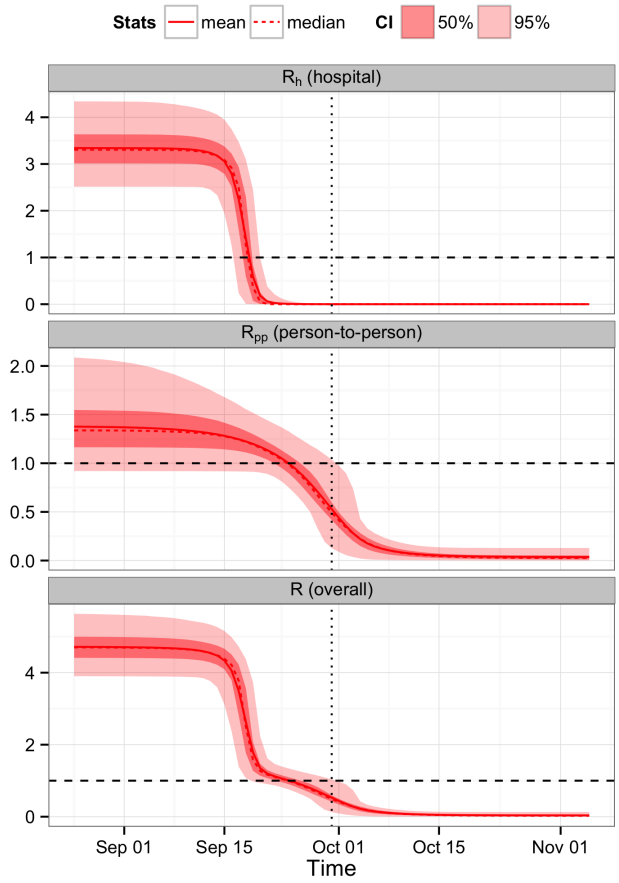

Figure 4: Drop in the reproduction number ($R(t)$) owing to change of behaviour in community contacts and visit of outpatients to the hospital. The overall R (lower panel) can be split into an hospital (upper panel) and person-to-person (middle panel) component. The dashed line indicates the epidemic threshold ($R = 1$) and the dotted line corresponds to the hospital closure (30th September). Solid, dashed and shaded red lines/area as in Figure 3.

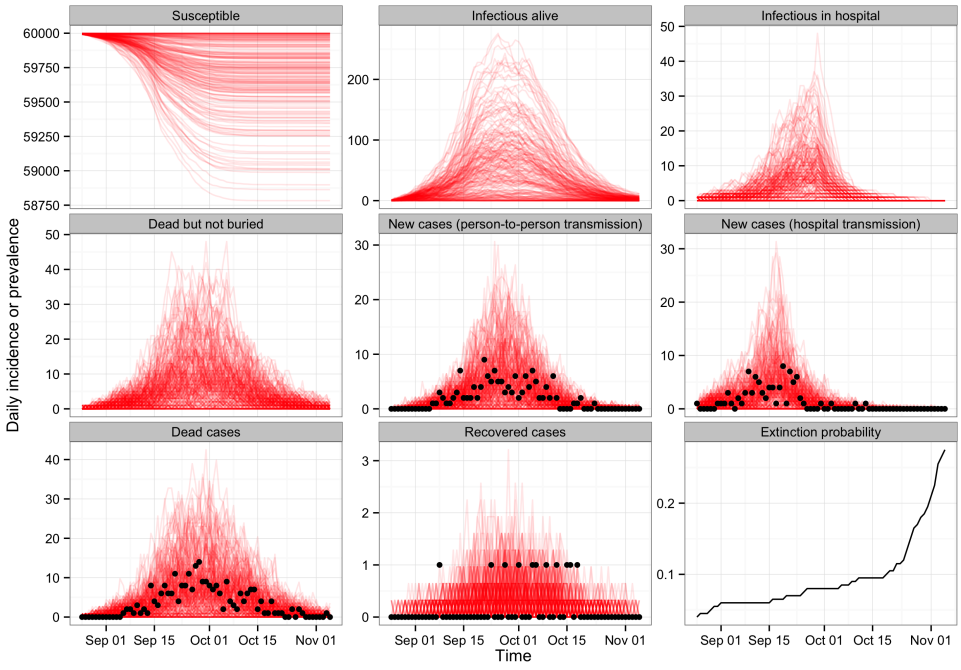

Figure 5: Potential alternative trajectories of an Ebola outbreak in Yambuku. Ten thousand stochastic simulations were run with parameter values taken from the maximum a posteriori probability estimate (for readability only the first 200 are plotted). For comparison, data are plotted as black dotted points.

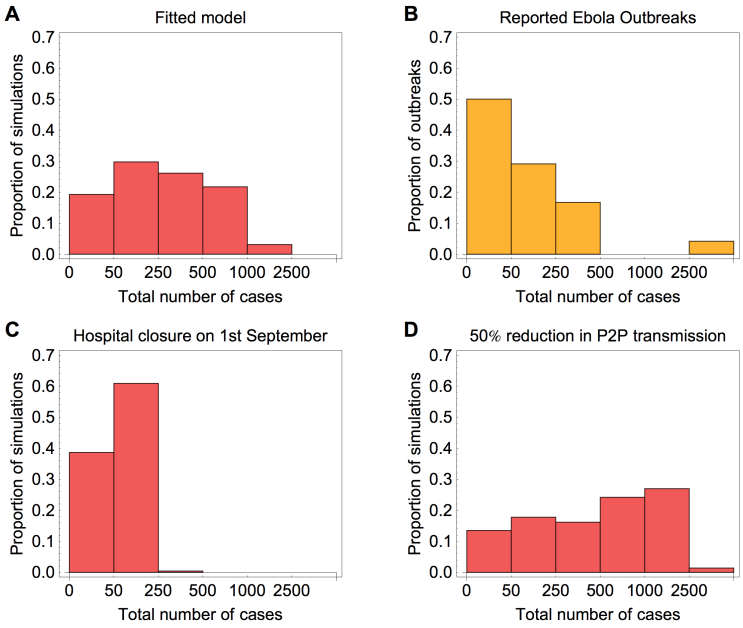

Figure 6: Distribution of Ebola outbreak sizes in different scenarios. (A) Outbreak size distribution from 10,000 stochastic simulations using the maximum a posteriori probability estimate. (B) Distribution of number of cases reported in Ebola outbreaks in Africa from 1976 to present. (C) Outbreak size distribution from 10,000 stochastic simulations when hospital is closed 7 days after the date of the first onset (i.e. 1st September). All other parameters remain the same. (D) Outbreak size distribution from 10,000 stochastic simulations when person-to-person transmission is reduced by 50% rather than 98%. The final category includes all outbreaks with more than 2500 cases.

Location	Date	R_0	95% CI (if given)	Reference
DRC	1995	1.83		Chowell et al. (2004)
		3.65	3.05–4.33	Ferrari et al. (2005)
		2.7	1.9–2.8	Legrand et al. (2007)
		1.38		Lekone and Finkenstädt (2006)
		2.22	1.9–2.73	Ndanguza et al. (2013)
		1.93	1.74–2.78	White and Pagano (2008)
Uganda	2000/1	1.34		Chowell et al. (2004)
		1.79	1.52–2.30	Ferrari et al. (2005)
		2.7	2.5–4.1	Legrand et al. (2007)

Table 1: Previously published estimates of basic reproduction number, R_0, for Ebola.

Parameter	Description	Prior	Estimates: median (95% CI)
N	Population size	Fixed	$60,000$
T_0	Date of introduction of index case to H compartment	$\mathcal{U}[\text{Aug } 05 - \text{Aug } 25]$	$\text{Aug } 24\,(\text{Aug } 21 - \text{Aug } 24)$
T_h	Date of hospital closure	Fixed	Sep 30
ρ_{onset}	Proportion of onsets reported	$\mathcal{N}(0.71, 0.05)$	$0.70\,(0.62 - 0.79)$
ρ_d	Proportion of death reported	$\mathcal{N}(0.89, 0.05)$	$0.89\,(0.80 - 0.97)$
ρ_r	Proportion of recovery reported	$\mathcal{N}(0.29, 0.05)$	$0.28\,(0.19 - 0.39)$
κ	Proportion of cases hospitalised until hospital closure	$\mathcal{N}(0.17, 0.05)$	$0.21\,(0.14 - 0.30)$
ϕ	Case-fatality ratio	$\mathcal{U}[0 - 1]$	$0.88\,(0.80 - 0.94)$
$1/\epsilon$	Incubation period (days)	$\mathcal{N}(6, 0.1)$	$5.99\,(5.80 - 6.18)$
$1/\gamma_h$	Mean time from onset to hospitalisation (days)	$\mathcal{N}(3, 0.1)$	$3.00\,(2.81 - 3.20)$
$1/\gamma_d$	Mean time from onset to death (days)	$\mathcal{N}(7.5, 0.1)$	$7.49\,(7.30 - 7.69)$
$1/\gamma_r$	Mean time from onset to recovery (days)	$\mathcal{N}(10, 0.1)$	$10.00\,(9.80 - 10.19)$
$1/\mu_b$	Mean time from death to burial (days)	$\mathcal{N}(1, 0.1)$	$0.99\,(0.80 - 1.18)$
β_i	Transmission rate in the community at the onset of the epidemic	$\mathcal{U}[0 - 100]$	$0.10\,(0.01 - 0.20)$
β_d	Transmission rate during traditional burial at the onset of the epidemic	$\mathcal{U}[0 - 100]$	$0.78\,(0.08 - 2.00)$
α_{pp}	Shape of the change of person-to-person contact behaviour in community and during traditional burial	$\mathcal{U}[0 - 5]$	$0.30\,(0.14 - 4.17)$
τ_{pp}	Midpoint date for the change of person-to-person contact behaviour	$\mathcal{U}[\text{Aug } 25 - \text{Oct } 14]$	$\text{Sep } 27\,(\text{Sep } 20 - \text{Oct } 03)$
δ_{pp}	Reduction of the person-to-person transmission rate following change of contact behaviour (%)	$\mathcal{U}[0 - 100]$	$98.00\,(90.00 - 100.00)$
β_h	Transmission rate in hospital at the onset of the epidemic	$\mathcal{U}[0 - 100]$	$3.24\,(2.36 - 4.43)$
α_h	Shape of the change of hospital seeking behaviour from outpatients	$\mathcal{U}[0 - 5]$	$2.29\,(0.49 - 4.85)$
τ_h	Midpoint date for the change of hospital seeking behaviour	$\mathcal{U}[\text{Aug } 25 - \text{Sep } 30]$	$\text{Sep } 17\,(\text{Sep } 14 - \text{Sep } 19)$

Table 2: Parameter definitions and corresponding estimates. Prior distributions used during model fitting are also shown.

Parameter	Route of transmission	Estimates: median (95% CI)
R_{0h}	Hospital via syringe	$3.32\,(2.53 - 4.34)$
R_{0pp}	Person-to-person (in community and during funeral)	$1.34\,(0.92 - 2.11)$
R_0	Overall	$4.71\,(3.92 - 5.66)$

Table 3: Estimates of the basic reproduction number, R_0, split into different component transmission routes.