

# Potential for large outbreaks of Ebola virus disease

## Electronic supplementary material

A. Camacho<sup>a,\*</sup>, A. J. Kucharski<sup>a,\*</sup>, S. Funk<sup>a</sup>, J. Breman<sup>b</sup>, P. Piot<sup>c</sup>, W. J. Edmunds<sup>a</sup>

<sup>a</sup>*Centre for the Mathematical Modelling of Infectious Diseases, Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, United Kingdom*

<sup>b</sup>*Fogarty International Center, National Institutes of Health, United States*

<sup>c</sup>*London School of Hygiene and Tropical Medicine, London, United Kingdom*

---

---

### Contents

<b>S1 MCMC inference</b>	<b>2</b>
<b>S2 Simulation study</b>	<b>2</b>
<b>S3 Basic reproduction number</b>	<b>3</b>
<b>S4 Model comparison</b>	<b>4</b>

### List of Figures

S1	Posterior distribution of parameters . . . . .	6
S2	Simulated data . . . . .	7
S3	Model fit to simulated data . . . . .	8
S4	Posterior distribution of parameters (simulation study) . . . . .	9
S5	Course of $R(t)$ (simulation study) . . . . .	10
S6	Joint posterior of $R_{0i}$ and $R_{0d}$ . . . . .	11

### List of Tables

S1	Summary of case data . . . . .	12
S2	Estimates from simulated data . . . . .	13
S3	DIC of tested models . . . . .	14

---

\*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)

## S1. MCMC inference

Model simulation and fitting were performed using the **SSM** library (Dureau et al., 2013), freely available at <https://github.com/JDureau/ssm>. This library implements a Metropolis-Hastings adaptive MCMC algorithm with a multivariate normal proposal distribution. The adaptive procedure of the proposal kernel operates in two steps. First, the size of the covariance matrix is adapted at each iteration to achieve an optimal acceptance rate around 23% (Roberts et al., 1997). Second, its shape is adapted by using the empirical covariance matrix, computed from the accepted samples and updated at each iteration, thus leading to an optimal proposal distribution (Roberts and Rosenthal, 2009).

The **SSM** library (Dureau et al., 2013) also implements a simplex algorithm, which was used to maximise the (non-normalized) posterior distribution and thus initialise the MCMC close to the mode of the target. Since the simplex algorithm only guaranty convergence to a local maximum, we ran 1000 independent simplex initialised from parameter sets sampled from the prior distribution. We selected the simplex that converged to the highest posterior density value, and used the outputed parameter set to initialise 5 independent MCMC chains of 200,000 iterations. We visually checked that the 5 chains converged to the same stationary distribution and combined them after appropriate burning and thinning. The posterior distribution of the model parameters (Figure S1) and the model fitting (Figure 3) were plotted using the **R** (R Core Team, 2014) package `fitR`, which is freely available at <https://github.com/sbfnk/fitR>. Median and 95% CI for all parameters can be found in Table 2.

## S2. Simulation study

We checked the inference procedure and assessed potential bias in our estimates by performing the same analysis as in the main paper on a simulated dataset. We set a “true” parameter vector  $\theta^*$  as the maximum a posteriori probability estimate (MAP) obtained from the main analysis. Since our Bayesian inference procedure integrates strong priors for several parameters, we fixed those parameters at the mode of their respective priors (we note however that the posterior and prior distributions almost match for those parameters). As such we would expect  $\theta^*$  to coincide with the MAP of the distribution obtained with the inference procedure.

We simulated an outbreak under  $\theta^*$  using the deterministic model described by Equation (3). Then, we generated the four random daily incidence time-series using the same Poisson observation processes as described in the method section of the main paper. We selected a random dataset which was as representative as possible of the expected dataset under  $\theta^*$ ; that is, the dataset with every observation corresponding to the expected observation. Because Poisson expectations don't need to be integers, we achieved this by picking a random dataset whose cumulated incidence was close to that of the expected dataset (we used an overall tolerance of 8 incidence counts). Both used and expected datasets are shown in Figure S2.

As shown in Figure S3, the model was able to fit the simulated data correctly. Comparison between  $\theta^*$  and the MAP is reported in Table S2 whereas the posterior estimates of the parameters are shown in Figure S4. Although we would not expect all true parameter values to coincide with the MAP (as a result of stochasticity during data generation) nor with the mode of the univariate posterior distributions (due to correlation between parameters), most parameter estimates are close to the “true” value, even some those with flat priors. In addition, we note that all true parameter values lied within the 95% CI of our posterior estimates. That said, the posterior distributions also reveal that it is difficult to separate the community ( $\beta_i$ ) and funeral ( $\beta_d$ ) transmission rates due to correlations between those two parameters (the data don't separate these two routes of transmission). We also note the flat posterior distributions of the shapes of the change in hospital-seeking ( $\alpha_h$ ) and contact ( $\alpha_{pp}$ ) behaviours, which suggests that the simulated data were not very informative for these parameters (note however that the shape of the sigmoid doesn't change significantly for values above 1). Despite this, Figure S5 shows that our inference procedure is able to accurately reconstruct the change of hospital-seeking and contact behaviours. Although this simulation study reveals some limitations of the data, it indicates that our inference procedure does not generate any substantial bias in estimates.

### **S3. Basic reproduction number**

It can be shown (Legrand et al., 2007) that the basic reproduction number ( $R_0$ ) for the model of Equation (3) can be split into a hospital ( $R_{0h}$ ) and person-to-person ( $R_{0pp}$ ) components, the latter can be further split into a community ( $R_{0i}$ ) and funeral ( $R_{0d}$ ) com-

ponents:

$$R_0 = \underbrace{\frac{\beta_h(0)\gamma_h\kappa_i(0)}{\Gamma_i(0)(\nu_d\phi_h + \nu_r(1 - \phi_h))}}_{R_{0h}} + \underbrace{\frac{\beta_i(0)}{\Gamma_i(0)} + \frac{\beta_d(0)\phi}{\mu_b}}_{R_{0pp}} \quad (\text{S1})$$

Posterior estimates for  $R_{0h}$  and  $R_{0pp}$  can be found on Table 3 of the main paper. Our choice of grouping both community ( $R_{0i}$ ) and funeral ( $R_{0d}$ ) into a single person-to-person component  $R_{0pp}$  is justified by correlation between these two quantities, as revealed by the simulation study of Section S2 and shown in Figure S6. Indeed, although our posterior density estimate suggests  $R_{0i} > R_{0d}$ , the 95% CI reveals that alternative scenarios cannot be excluded (e.g. larger reproduction number for dead cases than for alive cases).

The course of the effective reproduction number during the outbreak, as shown in Figure 4 of the main paper, can be obtained similarly by accounting for the time-dependence of the parameters as well as the proportion of susceptible individuals in the population (note however that the latter was negligible in the 1976 Yambuku outbreak):

$$R(t) = \left( \frac{\beta_h(t)\gamma_h\kappa_i(t)}{\Gamma_i(t)(\nu_d\phi_h + \nu_r(1 - \phi_h))} + \frac{\beta_i(t)}{\Gamma_i(t)} + \frac{\beta_d(t)\phi}{\mu_b} \right) \frac{S(t)}{N} \quad (\text{S2})$$

#### S4. Model comparison

In order to test the role of person-to-person and hospital transmissions, we constructed, fitted and compared four different models:

1. The first model is the one presented in the main paper and includes hospital closure, change in hospital seeking behaviour and change in person-to-person contact behaviour. We refer to the main paper for a more detailed description of this complete model.
2. The second model assume that person-to-person contact behaviour did not change over time.
3. In the third model we only included hospital closure (i.e. no change of behaviour over time).
4. The fourth model is similar to the complete model but assume that hospital transmission was density-independent. The force of infection for this model is  $\lambda_h(t) = \beta_h(t)\mathbb{1}_{H \geq 1}$  instead of  $\lambda_h(t) = \beta_h(t)H$  in the complete model.

Models were compared using the Deviance Information Criterion (DIC), which measures how well a model fits the data, adjusting for model complexity (Spiegelhalter et al., 2002). The model with the smallest DIC is the model that would best predict a replicate dataset which has the same structure as that currently observed. Given the differences  $\Delta\text{DIC} \gg 10$ , we can confidently rule out models 2, 3 and 4 in favour of the complete model 1 (Spiegelhalter et al., 2002).

## References

- Dureau, J., Ballesteros, S., Bogich, T., 2013. SSM: Inference for time series analysis with State Space Models. arXiv 1307.5626v4.
- 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.
- R Core Team, 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria.
- Roberts, G.O., Gelman, A., Gilks, W.R., 1997. Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability* 7, 110–120.
- Roberts, G.O., Rosenthal, J.S., 2009. Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics* 18, 349–367.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64, 583–639.

## Figures

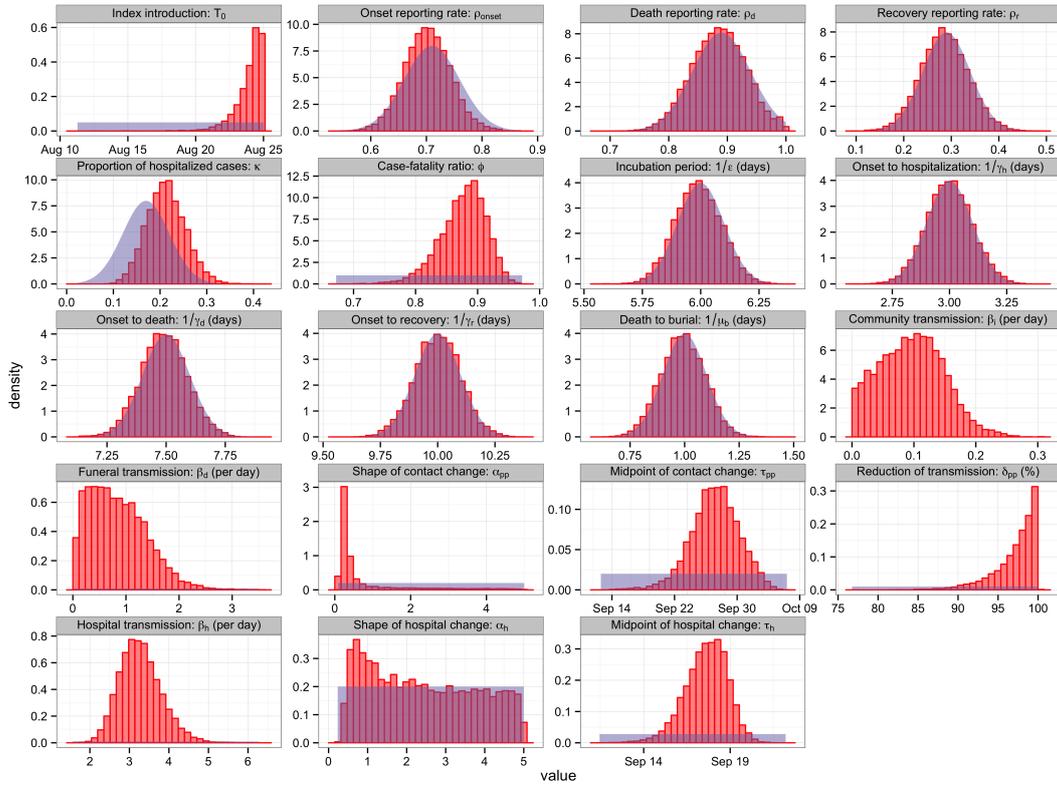


Figure S1: Posterior density of model parameters (red histogram) together with their prior density distribution (blue area), as specified in Table 2. As discussed in Section S3, there were correlations between  $R_{0i}$  and  $R_{0d}$ , which explain the somewhat wide marginal posterior distributions of  $\beta_i$  and  $\beta_d$ .

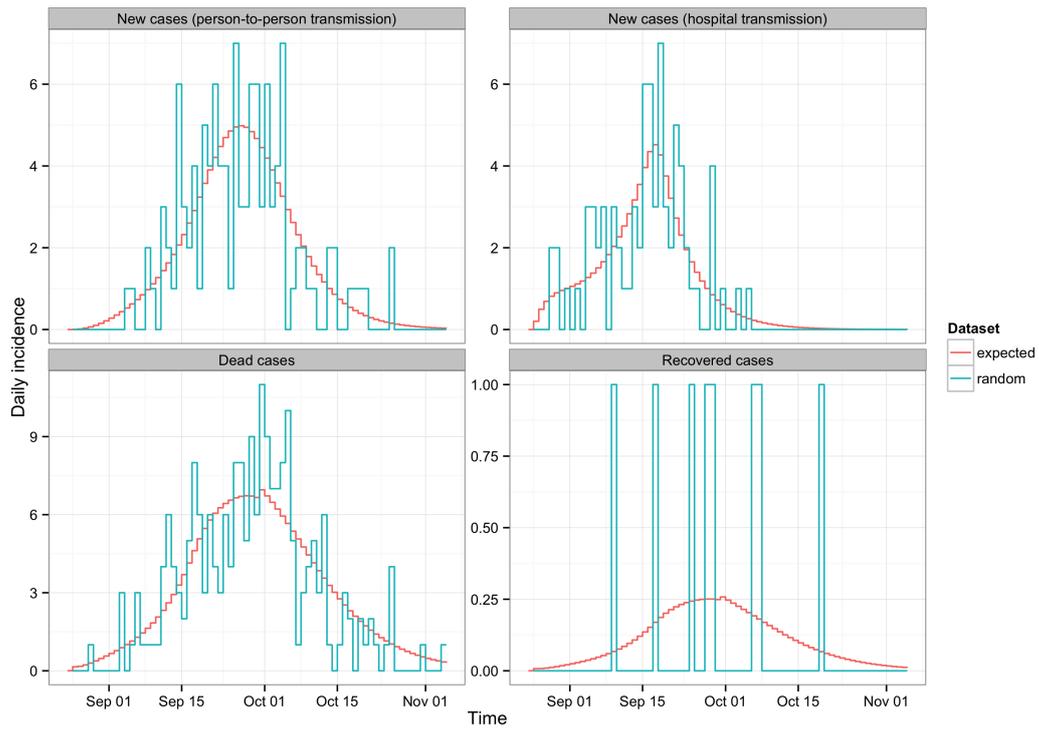


Figure S2: Simulated data (blue line) were chosen to be close to the expected observed time-series (red line) under the "true" parameter values  $\theta^*$ .

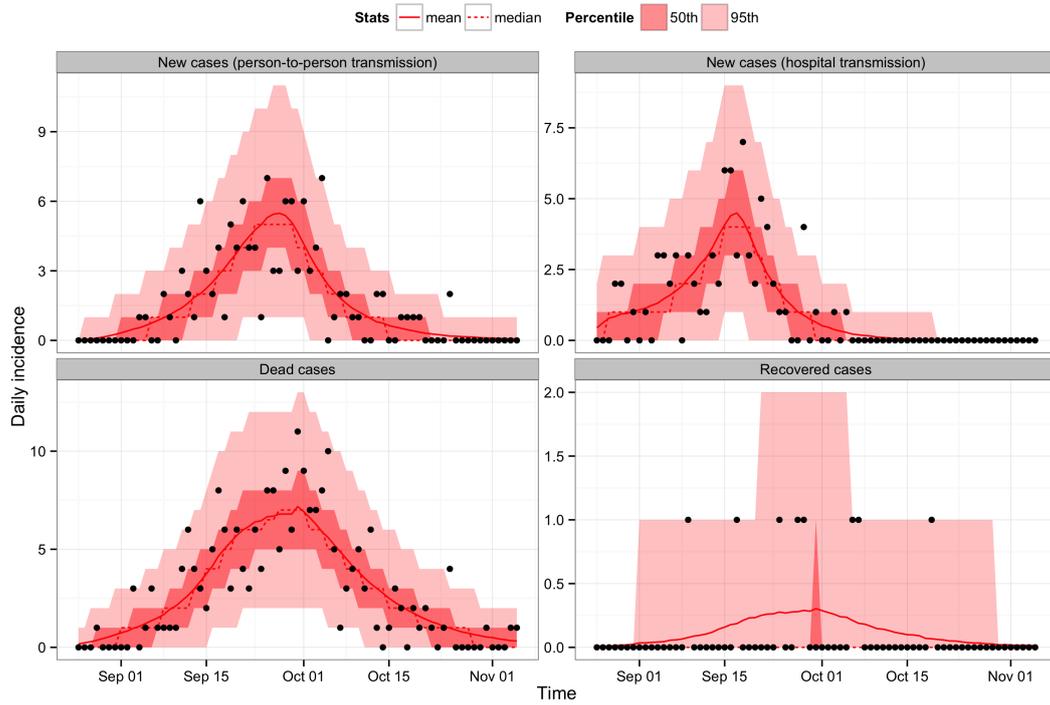


Figure S3: Comparison of our fitted model and simulated dataset (black dots). The mean and median fits are represented by solid and dashed red lines respectively. The dark and light red shaded areas correspond to the 50% and 95% credible intervals.

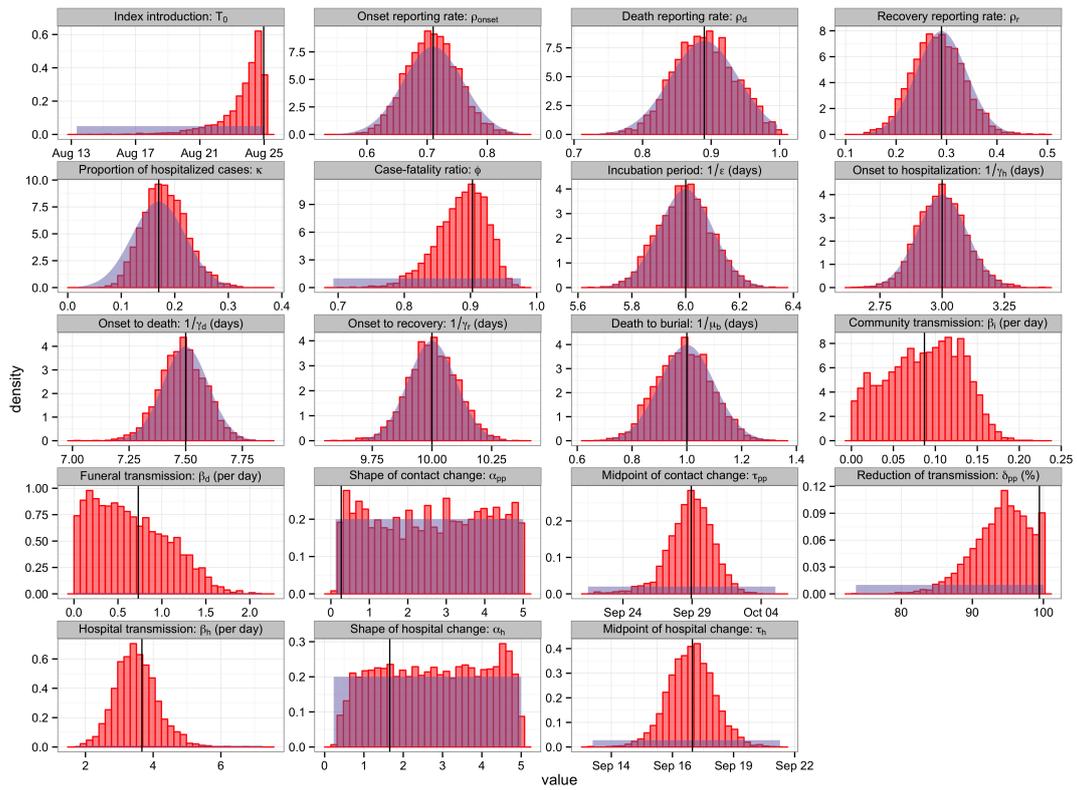


Figure S4: Posterior density of model parameters for the simulation study, where true values are mapped as vertical black lines. See legend of Figure S1 for more details.

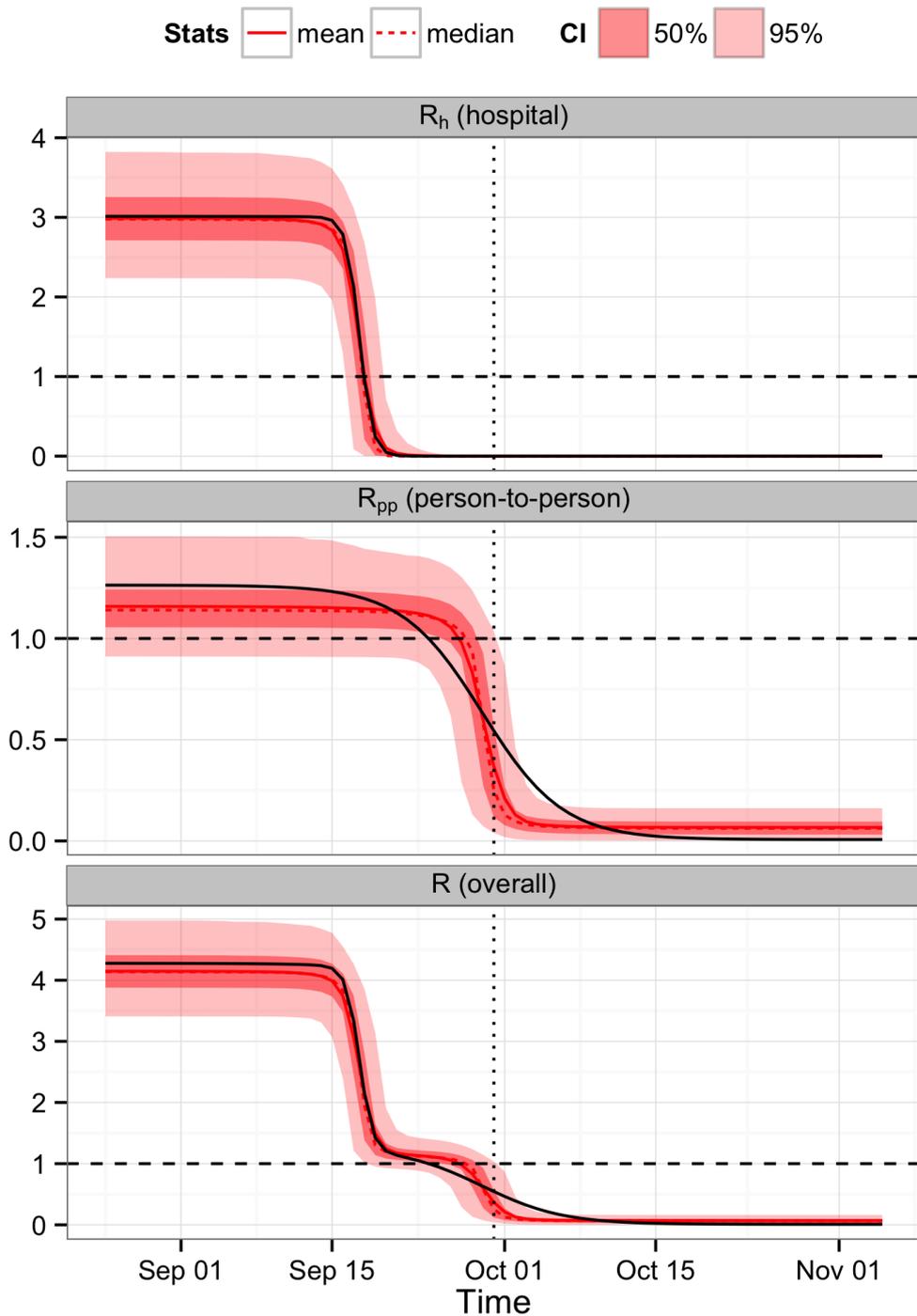


Figure S5: Drop in the effective reproduction number ( $R(t)$ ) owing to change of behaviour in community contacts and visit of outpatients to the hospital and comparison with the true drop (black solid line). The overall  $R$  (lower panel) can be split into an hospital (upper panel) and person-to-person (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 S3.

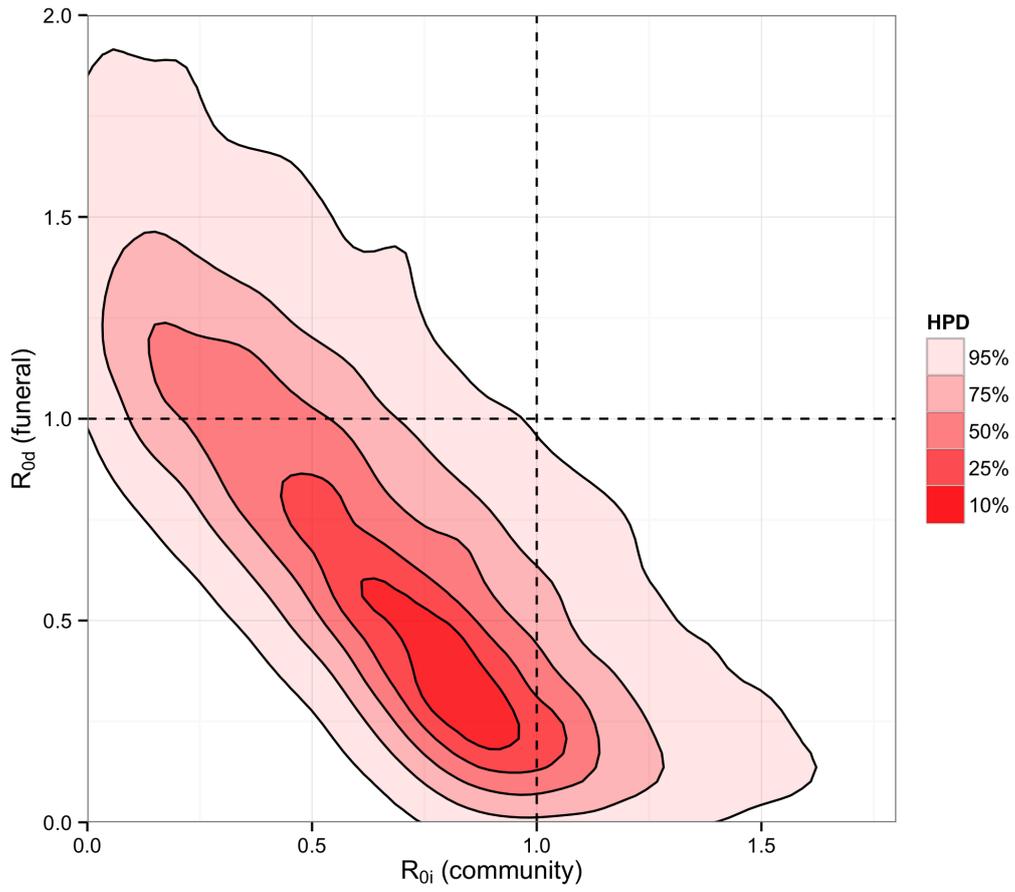


Figure S6: Joint estimate of the contribution of alive and dead (but not yet buried) cases to person-to-person reproduction number. Several high posterior density (HPD) regions are indicated.

## Tables

Source of infection	Cases	With date of onset	With date of outcome
Person to person	142	140	141
Syringe	90	87	88
Both	8	8	8
Other	10	10	10
Unknown	12	12	12
All	262	257	259

Table S1: Summary of case data.

Parameter	Description	$\theta^*$	MAP
$T_0$	Date of introduction of index case to H compartment	Aug 24	Aug 24
$\rho_{\text{onset}}$	Proportion of onsets reported	0.71	0.72
$\rho_d$	Proportion of death reported	0.89	0.89
$\rho_r$	Proportion of recovery reported	0.29	0.26
$\kappa$	Proportion of cases hospitalised until hospital closure	0.17	0.19
$\phi$	Case-fatality ratio	0.90	0.88
$1/\epsilon$	Incubation period (days)	6.00	6.00
$1/\gamma_h$	Mean time from onset to hospitalisation (days)	3.00	3.05
$1/\gamma_d$	Mean time from onset to death (days)	7.50	7.53
$1/\gamma_r$	Mean time from onset to recovery (days)	10.00	9.95
$1/\mu_b$	Mean time from death to burial (days)	1.00	1.07
$\beta_i$	Transmission rate in the community at the onset of the epidemic	0.09	0.07
$\beta_d$	Transmission rate during traditional burial at the onset of the epidemic	0.73	0.66
$\alpha_{pp}$	Shape of the change of person-to-person contact behaviour in community and during traditional burial	0.27	3.45
$\tau_{pp}$	Midpoint date for the change of person-to-person contact behaviour	Sep 28	Sep 29
$\delta_{pp}$	Reduction of the person-to-person transmission rate following change of contact behaviour (%)	99	95
$\beta_h$	Transmission rate in hospital at the onset of the epidemic	3.67	3.49
$\alpha_h$	Shape of the change of hospital seeking behaviour from outpatients	1.66	1.45
$\tau_h$	Midpoint date for the change of hospital seeking behaviour	Sep 17	Sep 17

Table S2: Comparison between the true parameter values ( $\theta^*$ ) and the MAP obtained in the simulation study.

Model	Description	DIC	$\Delta$ DIC
1	Complete model	711	0
2	No change of person-to-person contact behaviour	827	117
4	Density-independent hospital transmission	970	259
3	No change of hospital seeking or person-to-person contact behaviours	1191	480

Table S3: Deviance Information Criterion for the four models tested.