Accurate forecasts of the effectiveness of interventions against Ebola may require models that account for variations in symptoms during infection

AUTHORS

W.S. Hart[1], L.F.R. Hochfilzer[1], N.J. Cunniffe[2], H. Lee[3], H. Nishiura[3], R.N. Thompson[1,4,5,*]

*Correspondence to: robin.thompson@chch.ox.ac.uk

AFFILIATIONS

[1]Mathematical Institute, University of Oxford, Andrew Wiles Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, UK

[2]Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, UK

[3]Graduate School of Medicine, Hokkaido University, Hokkaido, Japan

[4]Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

[5]Christ Church, University of Oxford, St Aldates, Oxford OX1 1DP, UK

ABSTRACT

Epidemiological models are routinely used to predict the effects of interventions aimed at reducing the impacts of Ebola epidemics. Most models of interventions targeting symptomatic hosts, such as isolation or treatment, assume that all symptomatic hosts are equally likely to be detected. In other words, following an incubation period, the level of symptoms displayed by an individual host is assumed to remain constant throughout an infection. In reality, however, symptoms vary between different stages of infection. During an Ebola infection, individuals progress from initial non-specific symptoms through to more severe phases of infection. Here we compare predictions of a model in which a constant symptoms level is assumed to those generated by a more epidemiologically realistic model that accounts for varying symptoms during infection. Both models can reproduce observed epidemic data, as we show by fitting the models to data from the ongoing epidemic in the Democratic Republic of Congo and the 2014-16 epidemic in

Liberia. However, for both of these epidemics, when interventions are altered identically in the models with and without levels of symptoms that depend on the time since first infection, predictions from the models differ. Our work highlights the need to consider whether or not varying symptoms should be accounted for in models used by decision makers to assess the likely efficacy of Ebola interventions.

KEYWORDS

Ebola virus disease; Mathematical modelling; Epidemic forecasting; Infectious disease management; Disease control; Interventions

1. INTRODUCTION

Ebola epidemics have devastating consequences. The current epidemic in the Democratic Republic of Congo is the second largest in history, with 663 cases (614 confirmed and 49 probable) having been recorded as of 15th January 2019 [1]. Mathematical models are increasingly used for exploring the effects of different possible control interventions during Ebola epidemics [2–4]. The values of model parameters are chosen so that the model output matches observed epidemic data (model fitting; Fig 1A), and then interventions are introduced in the fitted model to predict how the course of the epidemic is altered by different possible control strategies (intervention testing; Fig 1B).

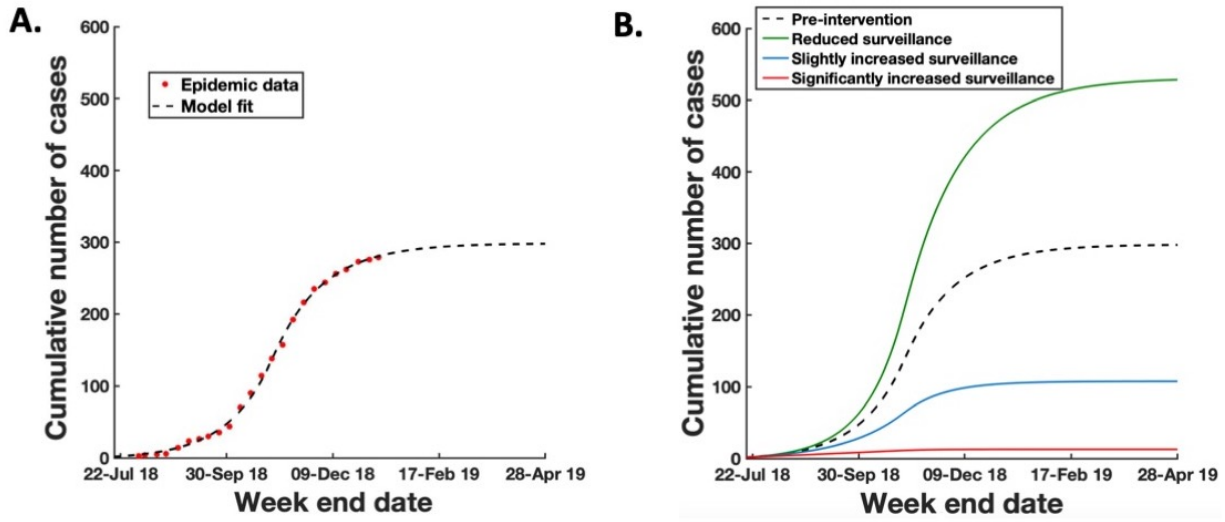

Figure 1. Schematic showing how a model can be used to predict the effect of changing control interventions on epidemic dynamics. Here, an intensification of surveillance is assumed to lead to improved detection and control of infectious hosts, thereby reducing the total number of cases. A. Model fitting. Model parameters are chosen so that the model output (black dotted) approximates epidemic data (red stars) B. Intervention testing. A range of alternative control interventions are introduced into the fitted model, and predicted dynamics under these new control interventions can be observed – predictions of the effects of reduced surveillance (green), slightly intensified surveillance (blue) and significantly intensified surveillance (red).

A commonly used model for characterising epidemics of diseases including Ebola is the Susceptible-Exposed-Infectious-Recovered (SEIR) model [5–7], and extensions to this basic model include explicit incorporation of transmission from Ebola deceased hosts [8–10] or accounting for mismatches between symptoms and infectiousness [11,12]. Possible interventions include isolation of symptomatic hosts, which can be included in the SEIR model by removing individuals from the infectious class. All individuals in the infectious class are usually assumed to be symptomatic, with the level of symptoms being assumed constant and therefore independent of the stage of infection. As an example, Chowell *et al.* [13] assume that symptomatic individuals are isolated at a constant rate, and Meakin *et al*. [4] assume that symptomatic individuals are hospitalised at a constant rate.

However, in reality, it is not the case that all symptomatic hosts are equally symptomatic. During an Ebola infection, an infected host progresses through different stages [14] – from initial non-specific symptoms (fever, headache and myalgia) to a gastrointestinal phase (diarrhoea, vomiting, abdominal symptoms and dehydration), and then either to a deterioration phase (collapse, neurological manifestations and bleeding) or recovery. Individuals with non-specific symptoms are less likely to be observed and treated/isolated than individuals who have progressed further through infection and have developed more specific and more serious symptoms.

Here, we investigate whether explicitly accounting for variations in symptom expression during the course of an Ebola infection leads to different epidemiological model dynamics compared to assuming a constant level of symptoms. To do this, we compare predictions derived from a model in which infectious hosts have a constant level of symptoms (the constant symptoms model – see Methods) with those from a model in which variable symptoms during infection are accounted for (the variable symptoms model). We parameterise our models using data from the ongoing Ebola epidemic in the Democratic Republic of Congo. We find that both models can be fitted closely to data from the epidemic. However, when control interventions in the models are altered, for example to explore the effects of intensifying surveillance and control, forecasts generated by the models are very different. We find the same qualitative result when we instead parameterise our models using data from the largest Ebola epidemic in history: the 2014-16 epidemic in west Africa.

These analyses demonstrate that models with or without variable symptoms can reproduce observed disease incidence time series, but that predictions from the models are different when interventions are altered, even when the change in interventions is identical in both models. Our results highlight the need to consider whether variations in symptom expression during infection should be included in models of Ebola epidemics. Without accounting for variable symptoms, predictions of the possible effects of interventions may be incorrect.

2. METHODS

Datasets

To show that our results are not conditioned on particular properties of data from a single epidemic, we conducted two separate analyses in which we considered data from two different Ebola epidemics.

In the first analysis, we used data on the numbers of cases in approximately weekly time intervals from the ongoing Ebola epidemic in the Democratic Republic of Congo. It has recently been suggested by Dr Peter Salama, Deputy Director-General of Emergency Preparedness and Response at the World Health Organization, that this epidemic comprises several distinct outbreaks in different affected areas. Indeed, disease incidence time series display distinct phases (large numbers of cases at the end of July/beginning of August 2018, followed by low numbers of cases in September, and then larger numbers of cases again thereafter), probably due to spatial effects of spread of the virus which are not captured by standard non-spatial compartmental models [15]. For this reason, we focussed on data from the health zone of Beni, a city in the north-east of the Democratic Republic of Congo, and the neighbouring health zone of Kalunguta. This region has been severely impacted by the current epidemic. These data were obtained from World Health Organization disease outbreak news reports from 4th August 2018 to 10th January 2019 (Data S1, see also [16]).

In the second analysis, we considered data comprising of the numbers of cases in approximately weekly time intervals in Liberia during the 2014-16 Ebola epidemic, which were obtained from the World Health Organization (Data S2, see also [17]).

Mathematical model

In the commonly used SEIR model, individuals are classified according to whether they are (*S*)usceptible to infection, (*E*)xposed, (*I*)nfected by the pathogen or (*R*)emoved and

no longer infectious. We extended this model to account explicitly for case finding followed by isolation of infectious individuals. We also assumed that there were three distinct phases of infection, corresponding to different stages of an Ebola infection. This delivered the additional benefit that the infectious period (in the absence of control) was gamma distributed, rather than exponentially distributed – and gamma distributions have been found to characterise epidemiological periods accurately in a range of systems [18,19]. This gave rise to the $SEI_1I_2I_3RC$ model,

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\beta(t)S(I_1 + I_2 + I_3),$$

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \beta(t)S(I_1 + I_2 + I_3) - \gamma E,$$

$$\frac{\mathrm{d}I_1}{\mathrm{d}t} = \gamma E - \mu I_1 - \delta_1 I_1,$$

$$\frac{\mathrm{d}I_2}{\mathrm{d}t} = \mu I_1 - \mu I_2 - \delta_2 I_2,$$

$$\frac{\mathrm{d}I_3}{\mathrm{d}t} = \mu I_2 - \mu I_3 - \delta_3 I_3,$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \mu I_3,$$

$$\frac{\mathrm{d}C}{\mathrm{d}t} = \delta_1 I_1 + \delta_2 I_2 + \delta_3 I_3.$$

In this model, the *C* compartment represents the number of individuals that have ever been controlled (detected and isolated) until the current time. Since we wish to isolate the impacts on prediction of variable symptoms alone, in the baseline version of the model we assume that all infectious hosts are equally infectious – although we consider the effect of relaxing this assumption later.

In our analyses, we made the assumption widely used in Ebola models that the infection rate parameter in the model is temporally-varying [4,5], to reflect changes in transmissibility during the epidemic. This could, for example, indicate changes in behavioural responses or alterations to interventions (aside from detection and isolation, since we model that explicitly). In particular, we assumed that

166

167 $$\beta(t) = \begin{cases} \beta_0 \text{ for } t \leq T \text{ days}, \\ \beta_1 \text{ for } t > T \text{ days}. \end{cases}$$

168

169 We then considered two alternative versions of the model. In the first (the constant
170 symptoms model – Fig 2A), we assumed that all infectious individuals are successfully
171 detected and isolated at the same average rate per day, so that $\delta_1 = \delta_2 = \delta_3 = \delta$, say.
172 This assumption is common to any epidemiological model that includes interventions
173 aimed at symptomatic hosts, unless differences in symptom expression are accounted
174 for explicitly. The constant symptoms model is therefore similar to most epidemiological
175 models that have been used to represent Ebola epidemics previously (e.g. [4,13]).

176

177 We also considered the more realistic case in which symptoms become more severe as
178 infection progresses, so that $\delta_1 < \delta_2 < \delta_3$. We refer to the resulting model as the
179 variable symptoms model (Fig 2B). This model reflects the fact that, in reality,
180 individuals with initial mild symptoms are less likely to be detected and isolated to
181 prevent further transmission than individuals with more developed symptoms who are in
182 the gastrointestinal or deterioration phases.

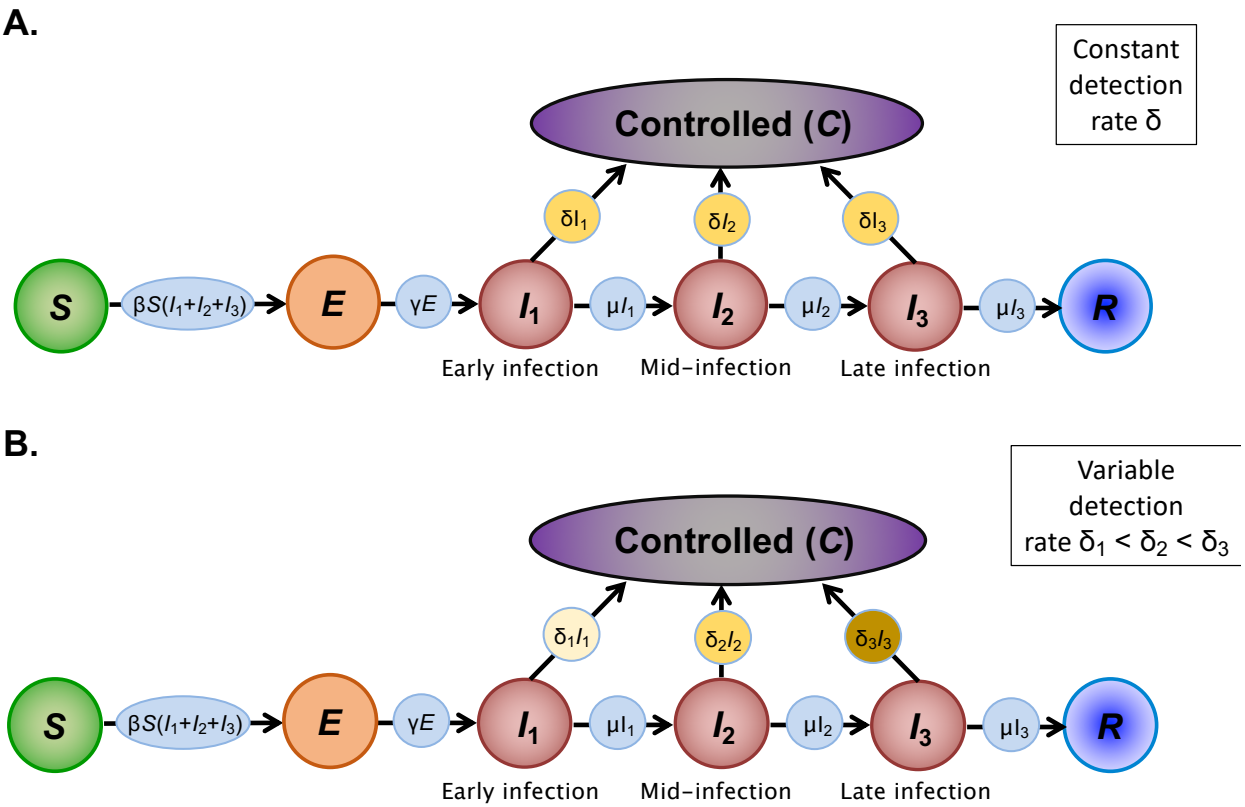

Figure 2. Schematic of the different models that we considered. A. Constant symptoms model, in which individuals in each stage of infection are equally likely to be detected and isolated (so that the detection rate, δ, is equal for all three infectious classes); B. Variable symptoms model, in which symptoms are assumed to intensify during an Ebola infection (so that the detection rate is smaller for individuals in earlier infection compared to later infection, i.e. $\delta_1 < \delta_2 < \delta_3$). We also show how additional epidemiological complexity can be included in these models (see Supplementary Material).

Model fitting and parameters

We considered the numerical solutions of the models described above in a host population of size of $S + E + I_1 + I_2 + I_3 + R + C = N$ individuals and a basic reproduction number at the beginning of the epidemic and in the absence of surveillance given by $R_0 = \frac{3\beta_0 N}{\mu}$.

The default parameter values used in our analyses are given in Table 1 (for the Democratic Republic of Congo in 2018-19) and Table 2 (for Liberia in 2014-16).

However, as described in the Results, we also checked the robustness of our results to these particular parameter values. The values of the infection rates β_0 and β_1, as well as the date on which the infection rate changes, T, were obtained by fitting the outputs of the models to the epidemic data. The start date of the epidemic, T_0, was also estimated during the fitting procedure. Model fitting was performed using least squares estimation – i.e. choosing parameter values to minimise the sum of squares distance between the cumulative numbers of detected or removed hosts in the model ($C + R$) and the cumulative numbers of cases in the data. Numerical solutions were generated starting with a single host in the E compartment at the start time of the epidemic, T_0, with all other individuals susceptible.

The values of the parameters characterising the rate of Ebola detection and isolation, i.e. δ_1, δ_2 and δ_3, depend on the level of surveillance, which includes various passive and active case finding strategies. We did not model explicitly the wide range of different surveillance activities that take place during an Ebola response (see Discussion). However, to provide a concrete setting in which to illustrate the principle that forecasts are different under the constant symptoms and variable symptoms models, we instead considered a simplified scenario in which each host is checked for infection on average every Δ days. Each time monitoring occurs, there is a detection probability of p_i for individuals in class I_i (for $i = 1,2, \text{or } 3$). As a result,

$$\delta_i = \frac{1}{\Delta\left(\frac{1}{p_i} - \frac{1}{2}\right)}.$$

This expression is derived in the Supplementary Material (for a similar approach, see also [20]).

In our main analyses, we considered two different surveillance regimes. When the models were fitted, under weak surveillance, we assumed that the default surveillance period was $\Delta = 21$ days. When the fitted models were then used to predict the impacts of intensified surveillance, the surveillance period was changed to $\Delta = 14$ days. We

assumed that the detection probability in the constant symptoms model was $p_1 = p_2 = p_3 = 0.6$. When we accounted for the possibility that symptoms change as hosts progress through infection, we instead used default values of $p_1 = 0.1$, $p_2 = 0.8$ and $p_3 = 0.9$ so that the mean value of p_1, p_2 and p_3 was equal to the value of these parameters in the constant symptoms model. In other words, conditional on not being detected previously, a host chosen at a random time in the infectious period was equally likely to be detected in both models.

3. RESULTS

As described in the Introduction, fitted models are often used to test potential control interventions (Fig 1). We therefore considered fitting models to two different datasets – one from the current Ebola epidemic in the Democratic Republic of Congo, and another from the historical Ebola epidemic in west Africa in 2014-16.

First, we considered data from the ongoing Ebola epidemic in the Democratic Republic of Congo (Fig 3A). We fitted the constant symptoms model and variable symptoms model to these data in turn, and found that both of these models could replicate the observed dynamics of the epidemic (Figs 3B). We then used these fitted models to predict how the epidemic dynamics would have been altered under a different control intervention. In particular, we increased the rate of detection in the fitted models, to represent predictions under an intensification of surveillance and control efforts (see Methods).

When surveillance was intensified, the prediction of the constant symptoms model differed substantially from that of the variable symptoms model (Fig 3C). In particular, for the parameter values displayed here, the constant symptoms model predicted 24% fewer cases than the more epidemiologically realistic variable symptoms model (108 cases in the constant symptoms model as opposed to 142 cases in the variable symptoms model). Consequently, even though the observed dynamics of the models

appear identical when fitted to data, they produce different predictions when control interventions are changed.

We then repeated our analysis, instead using data from the 2014-16 Ebola epidemic in west Africa (Figs 3D-F). In this case, the constant symptoms model predicted 35% fewer cases than the variable symptoms model (Fig 3F). Since the total number of cases in this epidemic was so large, this corresponded to 641 cases difference between the forecasts of the two models.

We also considered the robustness of our results to the values of the model parameters, including the level of surveillance assumed when fitting to data and the extent to which surveillance was intensified in the models (Figs S1-S7). In addition, we considered different values of the detection probability whenever surveillance occurs (Figs S8 and S9). In each case, we found qualitatively identical results – both the constant symptoms and variable symptoms models could reproduce epidemic time series data, but when interventions were changed in the models the predicted epidemic dynamics then differed. As well as considering an intensification of surveillance, we also examined cases in which surveillance was relaxed (e.g. Figs S4B and S4C). Whenever surveillance was intensified, the constant symptoms model underestimated the total number of cases compared to the more realistic variable symptoms model. However, when surveillance was instead reduced, the constant symptoms model led to overestimation of the total number of cases.

We also considered the effect of enhancing surveillance at different times during the epidemic. In particular, we considered increasing the surveillance level once the epidemics had already been in progress until a certain date, for different possible dates of surveillance intensification. The earlier that surveillance was intensified, the larger the error when using the constant symptoms model rather than the more biologically realistic variable symptoms model, since early surveillance intensification then allows more time for model predictions to differ (Figs 4 and S10).

Until this point, so that we could isolate the effect of variable symptoms alone on the predicted outcomes of interventions, we assumed that at any time during an epidemic all infected and uncontrolled hosts generated new infections at a constant rate. However, we also conducted an analysis in which the infection rate also varied throughout the course of an Ebola infection, by considering cases in which infectiousness was either correlated with or correlated against the level of symptom expression (Fig 5 and Supplementary Material). In cases in which higher levels of symptoms were associated with reduced infectiousness – for example due to a lower level of mixing in the population compared to hosts with less serious symptoms – our result that predictions are different between the constant symptoms and variable symptoms models was enhanced (Figs 5B and 5D).

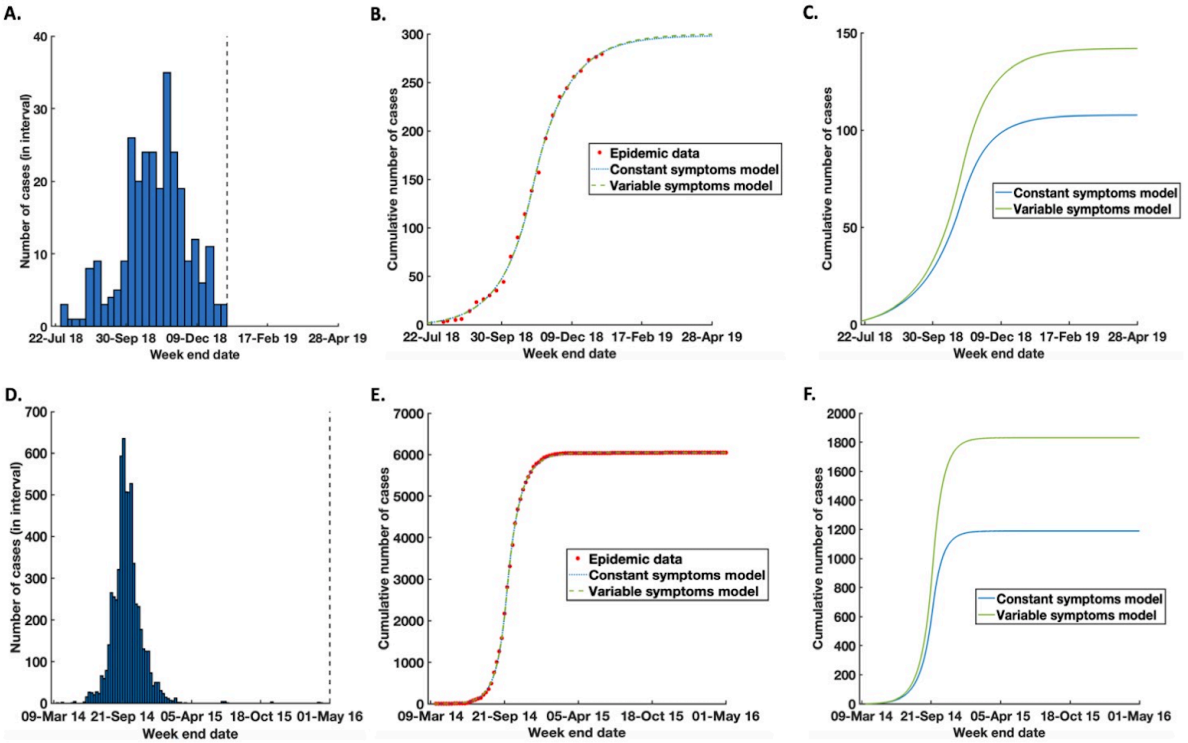

Figure 3. Using the constant symptoms model and variable symptoms model to predict alternative interventions. A. Number of cases each week in Beni and Kalunguta health zones in the Democratic Republic of Congo. The dotted black vertical line represents the time before which data were available. B. Model fits to the data (red stars), using the constant symptoms model (blue dotted) and variable symptoms model (green dash) C. Predictions of the effect of intensified surveillance, using the constant symptoms

model (blue) and variable symptoms model (green). D-F. Equivalent figures to A-C, using the data from the 2014-16 Ebola epidemic in Liberia. For model parameters, see Methods and Tables 1 and 2.

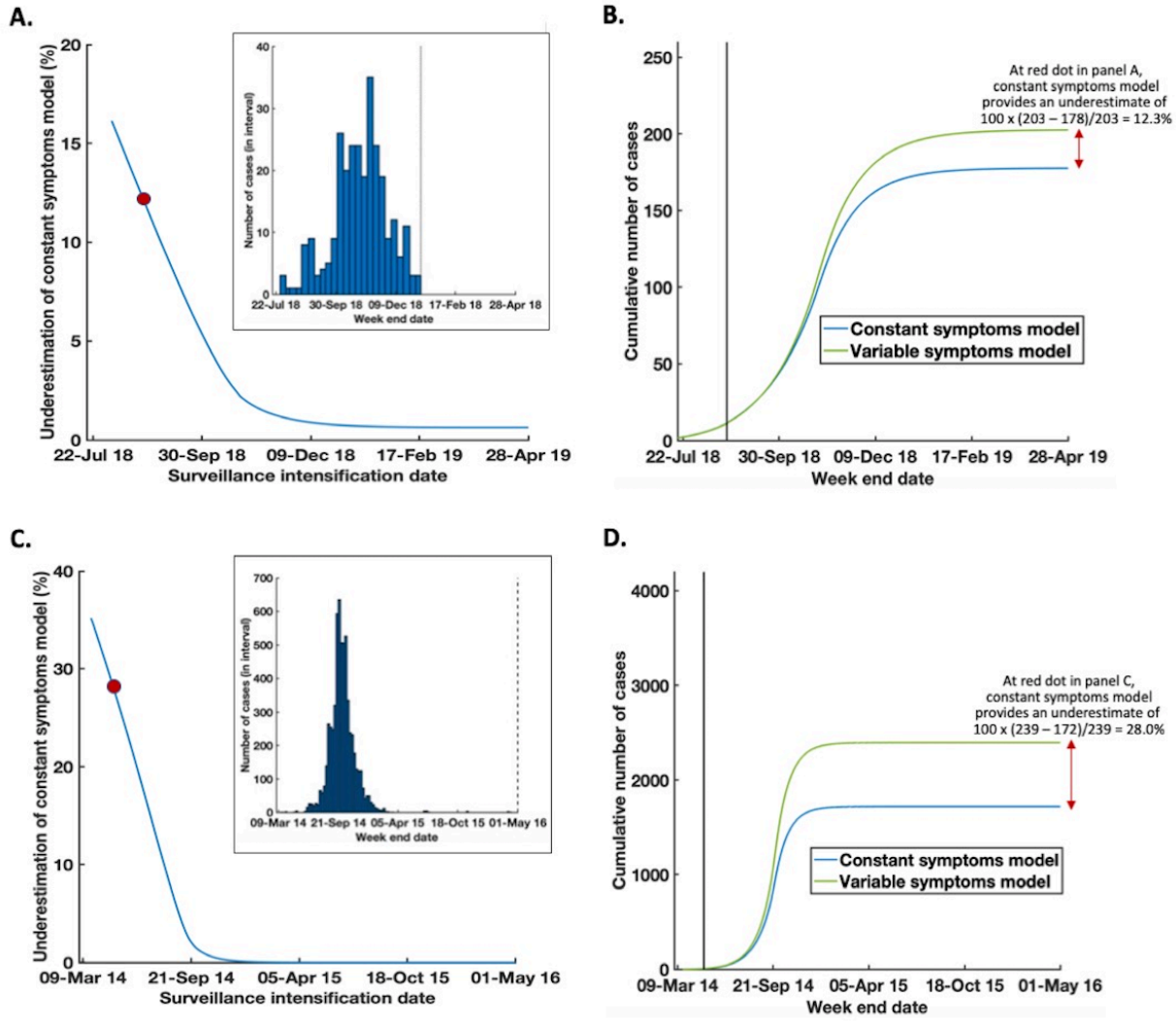

Figure 4. Using the constant symptoms model and variable symptoms model to predict alternative interventions, for different times of surveillance intensification. A. Reduction in the total number of cases predicted by the constant symptoms model compared to the variable symptoms model, expressed as a percentage of the number of cases predicted by the variable symptoms model, for different times of surveillance intensification. The models were fitted to data from the ongoing epidemic in the Democratic Republic of Congo. B. Illustration of how the values in panel A were calculated. In the graph shown, intensified surveillance began on 23rd August 2018 in the models fitted to the ongoing epidemic in the Democratic Republic of Congo. This corresponds to the time denoted by the red dot in panel A. C-D. Equivalent panels to A-B, for the 2014-16 Ebola epidemic in Liberia. In panel D, intensified surveillance began on 2nd May 2014. In panels A and C, we only consider intensifying surveillance at or after the time

that the first cases were observed (i.e. 3rd August 2018 for the ongoing epidemic in the Democratic Republic of Congo, and 23rd March 2014 for the 2014-16 Liberia epidemic). The original epidemic datasets are shown as insets to panels A and C. For model parameters, see Supplementary Material and Tables 1 and 2.

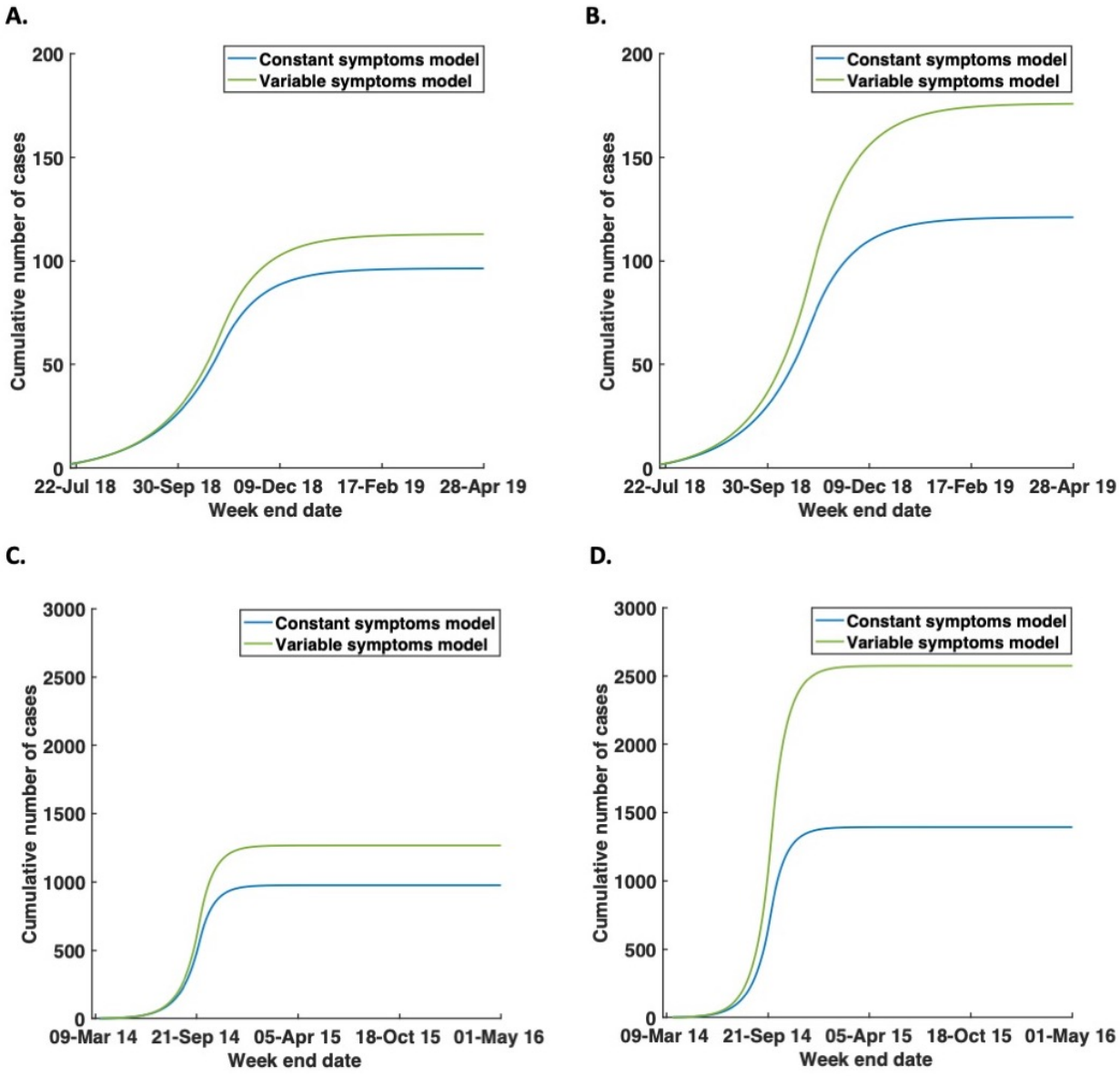

Figure 5. Using the constant symptoms model and variable symptoms model to predict alternative interventions, if infectiousness is also assumed to depend on the stage of infection. A. Predictions of the effect of intensified surveillance using the constant symptoms model (blue) and variable symptoms model (green), fitted to data from the ongoing epidemic in the Democratic Republic of Congo and assuming that infectiousness increases during an Ebola infection as described in the Supplementary Material. B. Predictions of the effect of intensified surveillance using the constant symptoms model (blue) and variable

symptoms model (green), fitted to data from the ongoing epidemic in the Democratic Republic of Congo and assuming that infectiousness decreases during an Ebola infection as described in the Supplementary Material. C-D. Equivalent figures to A-B but fitted to data from the 2014-16 Ebola epidemic in Liberia. The values of parameters other than the fitted parameters (i.e. the infection rates, the start times of the epidemics and the times at which the infection rates change), are identical to those described in Tables 1 and 2. Surveillance intensification is assumed to occur at the beginning of the epidemic. Model fits are not shown, since they appear identical by eye to those in Figs 3B and 3E.

4. DISCUSSION

Epidemiological models that have been fitted to data are often used to predict the dynamics of an epidemic under different control interventions (e.g. [21–25]). However, most models considering interventions targeted at symptomatic hosts assume that those hosts display a constant level of symptoms that does not change during the course of infection. For a number of infectious diseases, however – including Ebola virus disease – there are different stages of an infection, and in each of these stages the level of symptoms is likely to be different. Individuals in early infection tend to have milder symptoms than those in later infection. As a result, symptomatic hosts in early infection are less likely to have appeared in the surveillance data that are routinely collected during an epidemic than those in later infection, who not only have had a longer period during which to be detected but are also likely to have developed more severe and recognisable symptoms.

Here, we have considered Ebola virus disease as a case study, and used two models to predict the possible effects on the dynamics of two epidemics under different surveillance levels. We assumed that increased surveillance leads to improved detection and control of symptomatic hosts. We compared the output of a model in which, if an individual is surveyed, the probability of detection is constant at each time during the infectious period (the constant symptoms model) to the equivalent predictions from a model in which the probability of detection increases throughout infection (the variable symptoms model). We found that both these models can be fitted closely to data from Ebola epidemics (Figs 3B and 3E). However, when the level of surveillance in the models is increased, we found that the more epidemiologically realistic variable

symptoms model predicted a smaller number of cases (e.g. Figs 3C and 3F). Thus, it might be important to use the more realistic model to assess the quantitative effects of interventions aimed at reducing the impacts of Ebola epidemics.

Variations in symptoms between different stages of infection, as well as the signature of such variable symptoms in those types of data that are collected during an epidemic, have to date received little attention. Until now, the impact of variable symptoms on predictions of models used for testing Ebola interventions has never been rigorously assessed. However, our approach of splitting the infectious and symptomatic period into different compartments was inspired by the so-called “method of stages” [18,26], a technique most often used to model gamma distributed epidemiological periods [11,27,28]. Within that framework, varying infectiousness – rather than symptoms – over the course of infection has been considered previously. For example, Cunniffe *et al.* [19] consider a model of plant disease epidemics in which the rate of sporulation (production of viable spores by each infected host) is a function of the time since infection, and implement this in an SEIR model by splitting the *E* and *I* classes into compartments and assigning different infection rates to hosts in the different *I* classes. A similar modelling framework could be adopted in our work, using a large number of compartments so that the level of symptoms is represented by a continuous curve (rather than being at constant levels within the different stages of infection). However, we do not pursue this here, since discrete changes in symptom expression in each symptomatic host are sufficient to make our underlying point that accurate forecasts of the effects of Ebola interventions may require models that account for variations in symptoms.

To conduct our analyses, we sought to develop a simple model in which the level of symptoms increases during infection, and to compare the results from this model to those from the analogous model in which there is a constant level of symptoms during infection. Practical use of either model during an Ebola epidemic would require adjustment for the particular epidemic under consideration. For example, transmission in different settings could be included in a single model, such as spread in hospitals, community care centres, at funerals or in the wider community [3,4]. We modelled

detection and isolation of symptomatic hosts here, but other interventions such as vaccination could be modelled explicitly [29]. If an Ebola vaccine is not perfectly effective, as has been suggested for the vaccine used in the ongoing Ebola epidemic [30], the possibility that vaccination might mask symptoms while not completely stopping infectiousness could be included in our approach. A model that includes spatial spread of the pathogen or transmission through social contact networks might be required to replicate observed data [31,32], or different geographical areas could be considered separately [2,4]. To demonstrate the principle that variable symptoms can affect predictions of the effects of interventions, we assumed that all infected individuals pass through three stages of an Ebola infection (from non-specific symptoms, to a gastrointestinal phase and then to a deterioration phase), whereas in reality some hosts might recover rather than passing to the deterioration phase [14]. At the cost of an additional parameter to be estimated, it would be straightforward to include this in a compartmental epidemiological model (see preliminary analysis in Supplementary Material and Fig S11, in which some hosts recover rather than passing to the final stage of infection). We also modelled surveillance in a simple fashion, by assuming that hosts are surveyed on average at periodic intervals and that there is a particular probability of detection whenever a host is surveyed. For forecasting, it would be necessary to consider the wide range of different surveillance approaches used in practice including contact tracing from known cases [33] and rural village visitations to detect cases in locations where access to healthcare is limited [34], as well as disruptions to surveillance caused by factors including armed conflict [35].

We parameterised our models using the simplest possible approach – namely fitting the numbers of detected or removed individuals in the relevant classes of the models to data on the cumulative numbers of symptomatic cases using least squares estimation. We did not quantify the uncertainty in estimates of the values of model parameters, since the precise method of parameter inference was not central to our message. Instead, we sought to use the simplest possible fitting method. While this approach is used frequently during epidemics due to its ability to produce quick forecasts [5,36,37],

to properly quantify the uncertainty in forward projections it would be necessary to use non-cumulative incidence data and fit stochastic transmission models [38].

One advantage of the models that we used is that the surveillance level is assumed to impact on the epidemiological dynamics themselves, rather than simply the observed dynamics. This is not always the case in epidemiological models: a common method for accounting for under-reporting is simply to scale the incidence data up [39], thereby assuming a fixed percentage of infectious cases are detected with no impact on the numbers of cases generated by those individuals. Another approach is to assume that some individuals in the infectious class are unobserved [40]. In reality, detected hosts have a lower probability of transmitting the pathogen than undetected hosts due to the higher chance that those individuals are subject to interventions, and our models reflect this.

Here, we considered a control strategy of detection and isolation under different surveillance levels. The effects of including variable symptoms in models of other intervention strategies should be tested, to see whether it is always necessary to account for changing levels of symptoms throughout infection. We note that including additional epidemiological detail in forecasting models does not always improve predictions [12]. Simple models are easier to parameterise and interpret than more complex models, and so modellers should consider carefully, in each study, whether or not including variable symptoms will change model predictions. We also note that, in theory, it might be possible to deploy commonly used epidemiological models with altered parameter values as a proxy for explicit consideration of variable symptoms [41]. For example, if the chance of detection in early infection is low, then early non-specific symptoms could be considered as part of the incubation period. In that case, care should be taken when “lifting” the values of model parameters directly from the clinical literature, to ensure that the definitions of parameters in the model match those in the original studies.

In summary, including different levels of symptoms at different stages of infection in epidemiological models can alter predictions of the effects of intervention strategies compared to assuming a fixed level of symptoms. If variations in symptoms during infection – and their impacts on detectability – can be well characterised by epidemiologists and then included in predictive tools by modellers, decision makers will be able to make more informed choices as to which particular intervention, or combination of interventions, to pursue.

REFERENCES

1. World Health Organization. Ebola virus disease - Democratic Republic of the Congo [Internet]. 2019 p. https://www.who.int/csr/don/17-january-2019-ebola-.
2. WHO Ebola Response Team. Ebola virus disease in west Africa — The first 9 months of the epidemic and forward projections. N Engl J Med. 2014;371: 1481–1495. doi:10.15678/EBER.2017.050110
3. Camacho A, Kucharski A, Aki-Sawyerr Y, White MA, Flasche S, Baguelin M, et al. Temporal changes in Ebola transmission in Sierra Leone and implications for control requirements: A real-time modelling study. PLoS Curr. 2015;7: 1–12. doi:10.1371/currents.outbreaks.406ae55e83ec0b5193e3085
4. Meakin SR, Tildesley MJ, Davis E, Keeling MJ. A metapopulation model for the 2018 Ebola outbreak in Equateur province in the Democratic Republic of the Congo. bioRxiv. 2018; 1–30.
5. Chowell G, Hengartner NW, Castillo-Chavez C, Fenimore PW, Hyman JM. The basic reproductive number of Ebola and the effects of public health measures: The cases of Congo and Uganda. J Theor Biol. 2004;229: 119–126. doi:10.1016/j.jtbi.2004.03.006
6. Althaus CL. Estimating the reproduction number of Ebola virus (EBOV) during the 2014 outbreak in West Africa. PLoS Curr. 2014;6. doi:10.1371/currents.outbreaks.91afb5e0f279e7f29e7056095255b288
7. Nishiura H, Chowell G. Transmision dynamics and control of Ebola virus disease: a review. BMC Med. 2014;12: 196. Available:

http://www.biomedcentral.com/content/pdf/s12916-014-0196-0.pdf

8. Agusto FB, Teboh-Ewungkem MI, Gumel AB. Mathematical assessment of the effect of traditional beliefs and customs on the transmission dynamics of the 2014 Ebola outbreaks. BMC Med. 2015;13: 1–17. doi:10.1186/s12916-015-0318-3

9. Barbarossa MV, Dénes A, Kiss G, Nakata Y, Röst G, Vizi Z. Transmission dynamics and final epidemic size of Ebola virus disease outbreaks with varying interventions. PLoS One. 2015;10: 1–21. doi:10.1371/journal.pone.0131398

10. Weitz JS, Dushoff J. Modeling post-death transmission of Ebola: Challenges for inference and opportunities for control. Sci Rep. 2015;5: 8751.

11. Thompson RN, Gilligan CA, Cunniffe NJ. Detecting Presymptomatic Infection Is Necessary to Forecast Major Epidemics in the Earliest Stages of Infectious Disease Outbreaks. PLoS Comput Biol. 2016;12. doi:10.1371/journal.pcbi.1004836

12. Thompson RN, Hart WS. Effect of confusing symptoms and infectiousness on forecasting and control of ebola outbreaks. Clin Infect Dis. 2018;67. doi:10.1093/cid/ciy248

13. Chowell D, Castillo-Chavez C, Krishna S, Qiu X, Anderson KS. Modelling the effect of early detection of Ebola. Lancet Infect Dis. 2015;15: 148–149. doi:10.1371/currents.outbreaks.9e4c4294ec8ce1adad2831

14. Beeching NJ, French M, Houlihan CF. Ebola virus disease. Brit Med J. 2014;349: 26–30. doi:10.1155/2014/527378

15. Keeling MJ, Rohani P. Modeling infectious diseases in humans and animals. Princeton University Press; 2011.

16. World Health Organization. Disease outbreak news (DONs) on the Ebola outbreak in Democratic Republic of the Congo [Internet]. 2019 p. https://www.who.int/ebola/situation-reports/drc-20.

17. World Health Organization. Ebola data and statistics [Internet]. 2016 p. http://apps.who.int/gho/data/view.ebola-sitrep.ebo.

18. Wearing HJ, Rohani P, Keeling MJ. Appropriate models for the management of infectious diseases. PLoS Med. 2005;2: 0621–0627. doi:10.1371/journal.pmed.0020174

19. Cunniffe NJ, Stutt ROJH, Bosch F van den, Gilligan CA. Time-dependent infectivity and flexible latent and infectious periods in compartmental models of plant disease. Phytopathology. 2012;102: 365–380.
20. Cunniffe NJ, Laranjeira FF, Neri FM, DeSimone RE, Gilligan CA. Cost-effective control of plant disease when epidemiological knowledge is incomplete: Modelling Bahia bark scaling of citrus. PLoS Comput Biol. 2014;10: e1003753. doi:10.1371/journal.pcbi.1003753
21. Ferguson NM, Donnelly CA, Anderson RM. The foot-and-mouth epidemic in Great Britain: Pattern of spread and impact of interventions. 2014;1155. doi:10.1126/science.1061020
22. Merler S, Marco A, Fumanelli L, Gomes MFC, Pastore y Piontti A, Luca Rossi L, et al. Spatio-temporal spread of the Ebola 2014 outbreak in Liberia and the effectiveness of nonpharmaceutical interventions: a computational modelling analysis. Lancet Infect Dis. 2014;15. doi:10.1016/S1473-3099(14)71074-6
23. Brooks-Pollock E, Roberts GO, Keeling MJ. A dynamic model of bovine tuberculosis spread and control in Great Britain. Nature. 2014;511: 228–231. doi:10.1038/nature13529
24. Kucharski AJ, Camacho A, Flasche S, Glover RE, Edmunds WJ, Funk S. Measuring the impact of Ebola control measures in Sierra Leone. Proc Natl Acad Sci U S A. 2015;112: 14366–14371. doi:10.1073/pnas.1508814112
25. Cunniffe NJ, Cobb RC, Meentemeyer RK, Rizzo DM, Gilligan CA. Modeling when, where, and how to manage a forest epidemic, motivated by sudden oak death in California. Proc Natl Acad Sci U S A. 2016;113: 5640–5645. doi:10.1073/pnas.1602153113
26. Lloyd AL. Destabilization of epidemic models with the inclusion of realistic distributions of infectious periods. R Soc London B Biol Sci. 2001;268: 985–993. doi:10.1098/rspb.2001.1599
27. Anderson D, Watson R. On the spread of a disease with gamma distributed latent and infectious periods. Biometrika. 1980;67: 191–198.
28. Ma J, Earn DJD. Generality of the Final Size Formula for an Epidemic of a Newly Invading Infectious Disease. Bull Math Biol. 2006;68: 679–702.

551 doi:10.1007/s11538-005-9047-7
552 29. Fisman D, Tuite A. Projected impact of vaccination timing and dose availability on
553 the course of the 2014 West African Ebola epidemic. PLoS Curr. 2014;6.
554 30. Metzger WG, Vivas-Martínez S. Questionable efficacy of the rVSV-ZEBOV Ebola
555 vaccine. Lancet. 2018;391: 1021.
556 31. Kramer AM, Pulliam JT, Laura W, Park AW, Rohani P, Drake JM, et al. Spatial
557 spread of the West Africa Ebola epidemic. R Soc Open Sci. 2016;3(8): 160294.
558 32. Kiskowski M, Chowell G. Modeling household and community transmission of
559 Ebola virus disease: Epidemic growth, spatial dynamics and insights for epidemic
560 control. Virulence. 2016;7: 163–173. doi:10.1080/21505594.2015.1076613
561 33. Pandey A, Atkins K E, Medlock J, Wenzel N, Townsend J P, Childs J E, et al.
562 Strategies for containing Ebola in West Africa. Science (80-). 2014;346: 991–995.
563 34. Namukose E, Bowah C, Cole I, Dahn G, Nyanzee P, Saye R, et al. Active Case
564 Finding for Improved Ebola Virus Disease Case Detection in Nimba County ,
565 Liberia , 2014 / 2015 : Lessons Learned. Adv Public Heal. 2018; 6753519.
566 35. Gostin LO, Kavanagh MM, Cameron E. Ebola and war in the Democratic Republic
567 of Congo. Jama. 2018;321: 2018–2019. doi:10.1001/jama.2018.19743
568 36. Rivers C, Lofgren E, Marathe M, Eubank S, Lewis B. Modeling the impact of
569 interventions on an epidemic of Ebola in Sierra Leone and Liberia. PLoS Curr.
570 2014; 1–10. doi:10.1371/currents.outbreaks.fd38dd85078565450b0be3fcd78f5ccf
571 37. Webb G, Browne C, Huo X, Seydi O, Seydi M, Magal P. A model of the 2014
572 ebola epidemic in West Africa with contact tracing. PLoS Curr. 2015;7: 1–7.
573 doi:10.1371/currents.outbreaks.846b2a31ef37018b7d1126a9c8adf22a
574 38. King AA, De Cellés MD, Magpantay FMG, Rohani P. Avoidable errors in the
575 modelling of outbreaks of emerging pathogens, with special reference to Ebola.
576 Proc R Soc B Biol Sci. 2015;282: 0–6. doi:10.1098/rspb.2015.0347
577 39. Meltzer MI, Santibanez S, Knust B, Petersen BW, Ervin ED, Nichol ST, et al.
578 Estimating the future number of cases in the Ebola epidemic - Liberia and Sierra
579 Leone, 2014–2015. MMWR Surveill Summ. 2015;63: 1–14. Available:
580 http://www.cdc.gov/mmwr/preview/mmwrhtml/su6303a1.htm?s_cid=su6303a1_w
581 40. Thompson RN, Morgan OW, Jalava K. Rigorous surveillance is necessary for

high confidence in end-of-outbreak declarations for Ebola and other infectious diseases. Philos Trans R Soc B. 2019; doi:10.1098/rstb.2018.0431

41. Fraser C, Riley S, Anderson RM, Ferguson NM. Factors that make an infectious disease outbreak controllable. Proc Natl Acad Sci U S A. 2004;101: 6146–6151.

42. ReliefWeb. DRC (Nord Kivu) Ebola virus disease [Internet]. p. https://reliefweb.int/sites/reliefweb.int/files/re.

43. World Bank. World Bank Open Data - Liberia [Internet]. p. https://data.worldbank.org/country/liberia.

Parameter	Definition	Default Value	Justification
N	Population size	230,000	[42]
β_0	Infection rate early in epidemic	8.43×10^{-7} day^{-1} (constant symptoms model) or 8.1×10^{-7} day^{-1} (variable symptoms model)	Fitted to data
$tay\beta_1$	Infection rate late in epidemic	3.68×10^{-7} day^{-1} (constant symptoms model) or 3.59×10^{-7} day^{-1} (variable symptoms model)	Fitted to data
T	Infection rate switch time	25th October 2018 (constant symptoms model) or 24th October 2018 (variable symptoms model)	Fitted to data

$1/\gamma$	Latent/incubation period	7 days	[4]
$1/\mu$	Infectious period	9.8 days	[4]
p_1	Detection probability in early infection	0.6 (constant symptoms model) or 0.1 (variable symptoms model)	Assumption (for analyses with different values, see Supplementary Material)
p_2	Detection probability in mid infection	0.6 (constant symptoms model) or 0.8 (variable symptoms model)	Assumption (for analyses with different values, see Supplementary Material)
p_3	Detection probability in late infection	0.6 (constant symptoms model) or 0.9 (variable symptoms model)	Assumption (for analyses with different values, see Supplementary Material)
Δ	Sampling frequency	21 days (weak surveillance) or 14 days (intensified surveillance)	Assumption (for analyses with different values, see Supplementary Material)
δ_1	Detection/isolation rate in early infection	4.08×10^{-2} day^{-1} (constant symptoms model) or 5.01×10^{-3} day^{-1} (variable symptoms model)	Calculated using values of p_1 and Δ (see Methods)

δ_2	Detection/isolation rate in mid infection	4.08×10^{-2} day^{-1} (constant symptoms model) or 6.35×10^{-2} day^{-1} (variable symptoms model)	Calculated using values of p_2 and Δ (see Methods)
δ_3	Detection/isolation rate in late infection	4.08×10^{-2} day^{-1} (constant symptoms model) or 7.79×10^{-2} day^{-1} (variable symptoms model)	Calculated using values of p_3 and Δ (see Methods)
T_0	Start date of epidemic	25th June 2018 (constant symptoms model and variable symptoms model)	Fitted to data

593 Table 1. Default parameter values used in our analysis of data from the ongoing
594 Democratic Republic of Congo epidemic.
595
596

Parameter	Definition	Default Value	Justification
N	Population size	4,500,000	2015 estimate obtained from [43]
β_0	Infection rate early in epidemic	4.54×10^{-8} day^{-1} (constant symptoms model) or 4.33×10^{-8} day^{-1} (variable symptoms model)	Fitted to data

β_1	Infection rate late in epidemic	1.89×10^{-8} day^{-1} (constant symptoms model) or 1.81×10^{-8} day^{-1} (variable symptoms model)	Fitted to data
T	Infection rate switch time	21st September 2014 (constant symptoms model and variable symptoms model)	Fitted to data
$1/\gamma$	Latent/incubation period	7 days	[4]
$1/\mu$	Infectious period	9.8 days	[4]
p_1	Detection probability in early infection	0.6 (constant symptoms model) or 0.1 (variable symptoms model)	Assumption (for analyses with different values, see Supplementary Material)
p_2	Detection probability in mid infection	0.6 (constant symptoms model) or 0.8 (variable symptoms model)	Assumption (for analyses with different values, see Supplementary Material)
p_3	Detection probability in late infection	0.6 (constant symptoms model) or 0.9 (variable symptoms model)	Assumption (for analyses with different values, see Supplementary Material)
Δ	Sampling frequency	21 days (weak surveillance) or 14	Assumption

		days (intensified surveillance)	(for analyses with different values, see Supplementary Material)
δ_1	Detection/isolation rate in early infection	4.08×10^{-2} day^{-1} (constant symptoms model) or 5.01×10^{-3} day^{-1} (variable symptoms model)	Calculated using values of p_1 and Δ (see Methods)
δ_2	Detection/isolation rate in mid infection	4.08×10^{-2} day^{-1} (constant symptoms model) or 6.35×10^{-2} day^{-1} (variable symptoms model)	Calculated using values of p_2 and Δ (see Methods)
δ_3	Detection/isolation rate in late infection	4.08×10^{-2} day^{-1} (constant symptoms model) or 7.79×10^{-2} day^{-1} (variable symptoms model)	Calculated using values of p_3 and Δ (see Methods)
T_0	Start date of epidemic	21st March 2014 (constant symptoms model) or 19th March 2014 (variable symptoms model)	Fitted to data

Table 2. Default parameter values used in our analysis of data from the 2014-16 epidemic in Liberia.

DATA AVAILABILITY

The data used in our analyses are available in the supplementary files DataS1.csv and DataS2.csv. Analyses were performed in Matlab. Code is available for running the models, and is accessible at https://github.com/will-s-hart/EbolaVariableSymptoms.

COMPETING INTERESTS

We have no competing interests.

AUTHORS' CONTRIBUTIONS

RNT conceived the research; All authors designed the study; RNT, LFRH and WSH carried out the research; RNT and WSH drafted the manuscript; All authors revised the manuscript and gave final approval for publication.

FUNDING

RNT was funded by a Junior Research Fellowship from Christ Church, Oxford, and a JSPS Postdoctoral Research Fellowship (short-term, grant PE18029). WSH was funded by an EPSRC Excellence Award for his doctoral studies. LH was funded by an EPSRC undergraduate vacation bursary.