Epidemiological Dynamics of Ebola Outbreaks

Thomas House*

Abstract

Ebola is a deadly virus that causes frequent disease outbreaks in the human population. Here, we analyse its rate of new introductions, case fatality ratio, and potential to spread from person to person. The analysis is performed for all completed outbreaks, and for a scenario where these are augmented by a more severe outbreak of several thousand cases. The results show a fast rate of new outbreaks, a high case fatality ratio, and an effective reproductive ratio of just less than 1.

Introduction

Ebola virus disease is an often-fatal disease of humans that is not vaccine-preventable and has no specific treatment. A total of 25 outbreaks, believed to have arisen due to zoonotic transmission from wild mammals, have occurred since the first observed cases in humans in 1976 [1]. The current epidemic is the largest to date [2]. This gives particular urgency to quantitative estimation of epidemiological quantities relevant to Ebola such as case fatality ratio, timing of new outbreaks, and the strength of human-to-human transmission.

The most important epidemiological quantity to estimate for an infectious disease is typically the basic reproductive ratio, R_0, defined as the expected number of secondary cases produced per primary case early in the epidemic [3]. When R_0 is greater than 1, the expectation is that a new epidemic will eventually infect a significant percentage of the population if it is not stopped by interventions or chance extinction; conversely, when R_0 is less than 1, chance events may lead to a large number of cases, but these are always expected to be much less numerous than the total population size.

Previous attempts to estimate R_0 for Ebola have found values between 1.34 and 3.65 by fitting compartmental epidemic models to the incidence over time of the large outbreaks in the Democratic Republic of Congo in 1995 and Uganda in 2000 [4, 5, 6], with similar results obtained for the ongoing outbreak [7]. This leads to the question of why all completed outbreaks numbered at most several hundreds, with the typical answer being that the medical and social response to an outbreak reduces transmission, leading to an effective reproductive ratio $R_t < R_0$ [4, 6], although it is also important to note that heterogeneity in transmission can lead to extremely high probabilities of an outbreak becoming extinct even if R_t is slightly greater than 1 [8].

*Warwick Infectious Disease Epidemiology Research Centre (WIDER) and Warwick Mathematics Institute, University of Warwick, Coventry, CV4 7AL, UK. T.A.House@warwick.ac.uk

Methods

Description

Here, a different approach is taken based on using the time between outbreaks, number of deaths, and final number of cases, for all 24 completed Ebola outbreaks reported by the World Health Organisation [1]. Full mathematical details of the approach are given below.

First, we model the start of new outbreaks as a 'memoryless' Poisson process with a rate λ. Secondly, we assume that each new outbreak has a case fatality ratio (CFR – the probability that a case will die) picked from a Beta distribution. Thirdly, the final size model involves two components: (i) A geometrically distributed number of cases, A, which includes cases arising from animal-to-human and pre-control transmission; (ii) A branching process model of human-to-human transmission [9, 10], whose offspring distribution has mean R_t, generating Z cases. The final size is then $K = A + Z|A$. This quantity should be interpreted as arising from a combination of R_t, R_0 and timing of interventions.

Bayesian MCMC with uninformative priors was used to fit all models [11]. Since doubts have been raised in the literature about the use of final size data for emerging diseases [12], a simulation study was also performed to test identifiability, although a recent study by Blumberg and Lloyd-Smith [13] of joint identifiability of two parameters in a related model is also highly relevant in this context.

Finally, the final size data were augmented by an outbreak of unknown size in the range 1000-5000 (with mathematical details given by Eq. (5) below) and the model was refitted. Due to the significant uncertainty in the severity of the current outbreak, this is not intended to be a real-time analysis, but rather to show how the modelling approach responds to such a scenario in general.

Technical details

Transmission model

Each outbreak has an initial number of cases A and a secondary number of cases Z. The total outbreak size is $K = A + Z|A$. We model the number of initial cases as a shifted geometric distribution,

$$\Pr[A = a|p] = (1-p)^{a-1}p \; . \tag{1}$$

We then model the number of secondary cases as the total progeny of a Galton-Watson branching process with A initial individuals and offspring distribution given by a geometrically distributed random variable ξ with mean $R_t =: (1-q)/q$. We adapt the results from Ball and Donnelly [10] to our model, giving

$$\Pr[Z = z|A = a, q] = \binom{2z + a - 1}{z} q^{z+a}(1-q)^z \; . \tag{2}$$

This gives a formula for the total size of the outbreak of

$$\Pr[K = k|p, q] = \sum_{a=1}^{k} \Pr[A = a|p]\Pr[Z = (k-a)|A = a, q] \; . \tag{3}$$

If the data D consists of a set of k_i (which is the size of outbreak i, with N the total number of outbreaks) then the likelihood function for the transmission model is

$$L(D|p,q) = \prod_i \Pr[K = k_i|p,q] \; . \tag{4}$$

When the data D' consists of the set of k_i augmented by an outbreak of size between κ_1 and κ_2 we use likelihood function

$$L(D'|p,q) = L(D|p,q) \sum_{k=\kappa_1}^{\kappa_2} \Pr[K = k|p,q] \; . \tag{5}$$

New outbreak model

We model the start of new outbreaks in the human population as a Poisson process of rate λ. If the time period over which N outbreaks is observed is T years, then the likelihood is

$$L(D|\lambda) = \frac{(\lambda T)^N \mathrm{e}^{-\lambda T}}{N!} \; . \tag{6}$$

We estimate $\lambda = 0.67[0.45, 0.98]$, with posterior distribution given in Figure 1C. The probability density function for t being the next outbreak time is

$$f(t) = \lambda \mathrm{e}^{-\lambda t} \; , \tag{7}$$

which is shown in Figure 1A.

Case fatality model

We let C_i be a random variable for the probability of fatality for a given case in outbreak i. We assume a parametric model in which this is drawn from a beta distribution, meaning that the probability density function is

$$\mathrm{Beta}(c|\alpha,\beta) = \frac{c^{\alpha-1}(1-c)^{\beta-1}}{B(\alpha,\beta)} \; , \quad B(\alpha,\beta) := \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\mathrm{d}x \; . \tag{8}$$

Then if $d_i \leq k_i$ is the number of fatalities in outbreak i, treating each fatality as independent, conditioned on infection, gives

$$\Pr[d_i|k_i,\alpha,\beta] = \binom{k_i}{d_i} \frac{B(\alpha + d_i, \beta + k_i - d_i)}{B(\alpha,\beta)} \; . \tag{9}$$

Then the likelihood is

$$L(D|\alpha,\beta) = \prod_i \Pr[d_i|k_i,\alpha,\beta] \; . \tag{10}$$

We estimate $\alpha = 6.1[2.8, 11]$ and $\beta = 3.1[1.5, 5.9]$, with posterior distributions given in Figure 1D,E.

Statistical methodology

The MCMC methodology used was Random-walk Metropolis Hastings with thinning to produce 10^3 uncorrelated samples, with each posterior ultimately derived from one long chain. The parameter spaces involved are low-dimensional enough that large-scale sweeps can be performed to check for multimodality, which was not observed, and convergence of the chains was observed to be fast and independent of initial conditions.

For the simulation study, the real-time incidence curves of Figure 3A are produced by modelling the geometric distributions as arising from Poissonian transmission with exponentially distributed rates. The times between new introductions are not explicitly modelled or shown.

Results

Figure 1 shows the results of fitting to times between outbreaks, with 1A showing the empirical distribution of times between outbreaks together with the fitted model distribution that has mean $1.49[1.02, 2.24]$ years between outbreaks, and 1C showing the posterior for the rate parameter. Figure 1 also shows the results of fitting CFR to number of deaths and final size, with 1B showing empirical CFRs for different outbreaks together with the fitted model distribution. Other plots in Figure 1 (D,E) show the posteriors for the Beta distribution parameters.

Figure 2 shows the results of fitting to completed outbreaks, with 2A,B giving the fitted distribution against data, 2C showing the posterior for the reproductive ratio, which is estimated to be $R_t = 0.88[0.64, 0.96]$. 2D shows the posterior for the geometric parameter, which is estimated to be $p = 0.089[0.029, 0.19]$.

While the model is designed not to depend explicitly on the temporal dynamics of Ebola virus disease, Figure 3A shows a set of 24 outbreaks simulated from a continuous-time Markov chain with the same probability distribution for final size as the estimated model. These show behaviour that is typical of near-critical branching processes, which often becoming extinct early, but also often grow to significant size before extinction. Figure 3B plots the likelihood surface for these simulated data showing parameter identifiability.

Figure 4 shows the results of fitting to completed outbreak final sizes augmented by an outbreak of uncertain size in the range 1000-5000. Here 4A gives the fitted distribution against data, and 4B shows the posterior for the probability of the additional outbreak, which is estimated to be $0.023[0.0015, 0.088]$. 4C shows the posterior for the reproductive ratio, which is estimated to be $R_t = 0.94[0.87, 0.99]$, and 4D shows the posterior for the geometric parameter, which is estimated to be $p = 0.11[0.054, 0.21]$.

Discussion

The results obtained point to the following conclusions about Ebola transmission dynamics. (i) The rate of new epidemics and CFR are both high, but with significant variability from outbreak to outbreak. (ii) The effective reproductive ratio R_t for person-to-person transmission is just below 1. (iii) There is extremely large variability in the final size of outbreaks.

It is also important to consider the sensitivity of these conclusions. A larger final size for the current outbreak (but still significantly less than the population size of a country) as suggested by

the analysis above will tend to lead to a narrower posterior about a value of R_t closer to 1; this can be understood from general properties of branching processes [9]. Such a finely tuned constant value of R_t would, however, become increasingly difficult to interpret as a fundamental property of the outbreak and a modelling approach in which R_t was allowed to vary in time – along with the public health and behavioural responses – would be preferred.

Also, it is possible that a number of small outbreaks were not recorded by the WHO. This could be addressed through incorporation of additional variability into the model through introduction of explicit overdispersal parameters as in [8, 13], although for the data currently available there was no strong evidence for overdispersal beyond that implied by the geometric distributions.

All of these conclusions suggest no reason for complacency and give support to appeals for greater resources to respond to the ongoing epidemic [14].

Acknowledgements

Work supported by the UK Engineering and Physical Sciences Research Council. I would like to thank Deirdre Hollingsworth, Matt Keeling and Graham Medley for helpful discussions, and the Editors and Reviewers for helpful comments and suggestions.

References

[1] World Health Organisation. Ebola virus disease, Fact sheet Number 103. http://www.who.int/mediacentre/factsheets/fs103/en/ (Accessed 21 August 2014).

[2] World Health Organisation. Global Alert and Response, Ebola virus disease. http://www.who.int/csr/don/archive/disease/ebola/en/ (Accessed 21 August 2014).

[3] O. Diekmann, J. A. P. Heesterbeek, and J. A. J. Metz. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, 28(4):365–382, 1990.

[4] G. Chowell, N. W. Hengartner, C. Castillo-Chavez, P. W. Fenimore, and J. M. Hyman. The basic reproductive number of Ebola and the effects of public health measures: the cases of Congo and Uganda. *Journal of Theoretical Biology*, 229(1):119–126, 2004.

[5] M. J. Ferrari, O. N. Bøjrnstad, and A. P. Dobson. Estimation and inference of R_0 of an infectious pathogen by a removal method. *Mathematical Biosciences*, 198(1):14–26, 2005.

[6] J. Legrand, R. Grais, P. Boëlle, A. Valleron, and A. Flahault. Understanding the dynamics of ebola epidemics. *Epidemiology and Infection*, 135(04):610–621, 2007.

[7] C. L. Althaus. Estimating the reproduction number of Zaire ebolavirus (EBOV) during the 2014 outbreak in West Africa. Pre-print submitted to PLOS Currents Outbreaks available at arXiv:1408.3505v1.

[8] J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, and W. M. Getz. Superspreading and the effect of individual variation on disease emergence. *Nature*, 438(7066):355–359, 2005.

[9] K. B. Athreya and P. E. Ney. *Branching Processes*. Oxford University Press, 1992.

[10] F. Ball and P. Donnelly. Strong approximations for epidemic models. *Stochastic Processes and their Applications*, 55(1):1–21, 1995.

[11] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman and Hall/CRC, Boca Raton, 1995.

[12] J. M. Drake. Limits to forecasting precision for outbreaks of directly transmitted diseases. *PLoS Medicine*, 3(1):e3, 11 2005.

[13] S. Blumberg and J. O. Lloyd-Smith. Inference of R_0 and transmission heterogeneity from the size distribution of stuttering chains. *PLoS Computational Biology*, 9(5):e1002993, 2013.

[14] Médecins Sans Frontières. Ebola in West Africa: "The epidemic is out of control". http://www.msf.org.uk/node/25511 (Accessed 21 August 2014).

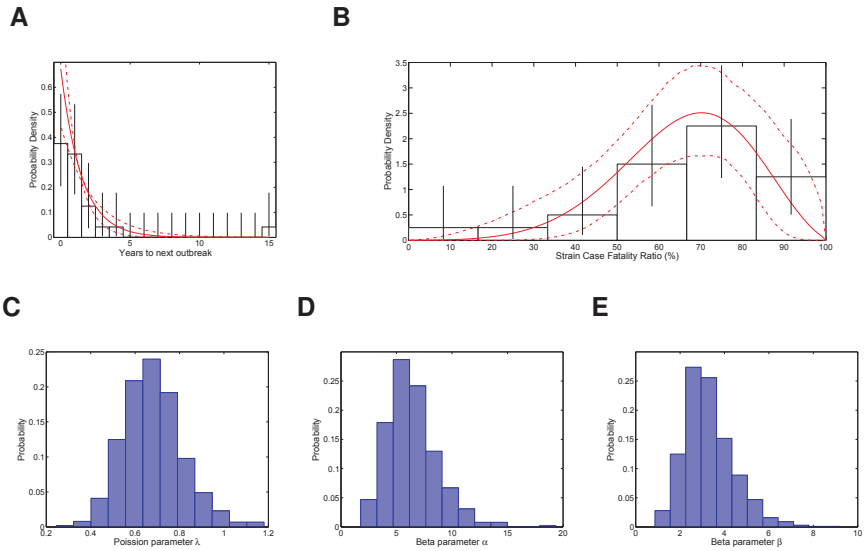

Figure 1: Analysis of rate of new outbreaks and case fatality ratio. (A) shows empirical data and 95% CI (black lines) together with fitted distribution and 95% CI (red lines) for rate of new outbreaks. (B) shows empirical data and 95% CI (black lines) together with fitted distribution and 95% CI (red lines) for case fatality ratio. (C) shows the posterior density for rate of new outbreaks λ, while (D,E) show the posterior density for the Beta distribution parameters α and β respectively.

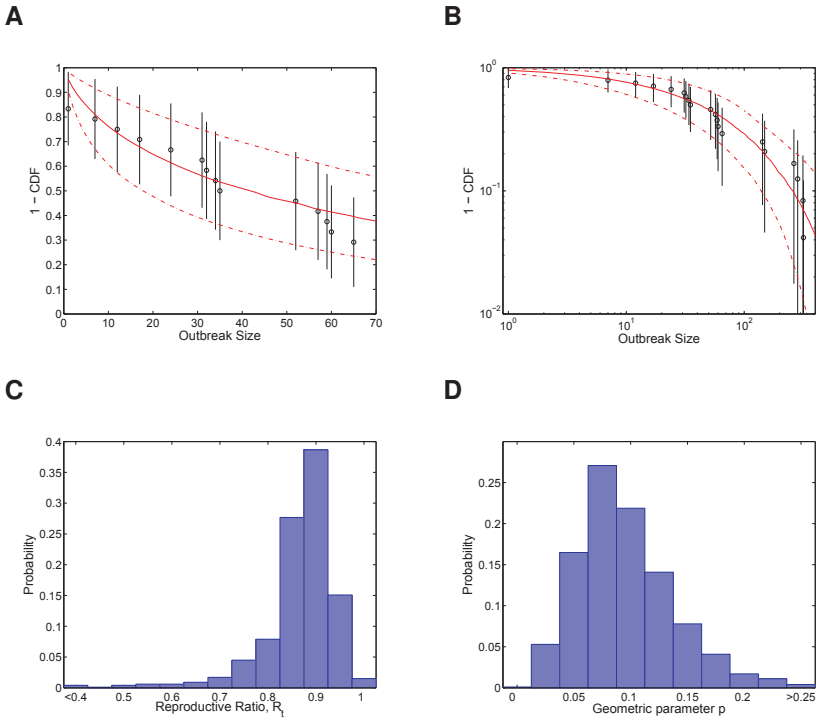

Figure 2: Analysis of transmission dynamics for completed outbreaks. (A, B) Model (solid red line) and 95% CI (dash-dot red line) versus data (black circles) and 95% CI (solid black lines) for different axis scales. (C) Posterior for values of the reproductive ratio R_t. (D) Posterior for the geometric parameter p.

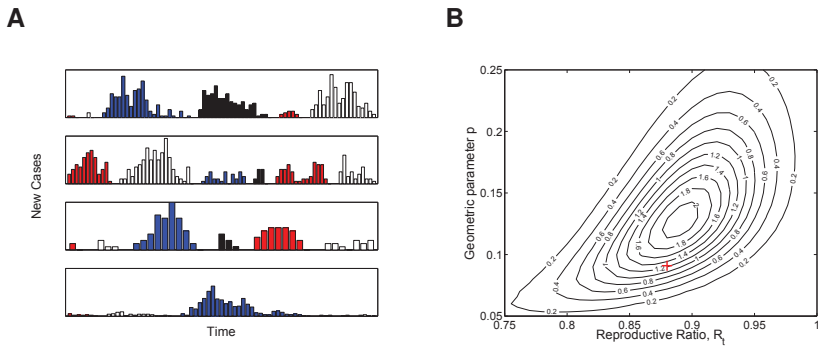

Figure 3: Simulation study. (A) Real-time model simulations, with change in colour denoting a new outbreak. (B) Likelihood contours (black lines and values multiplied by an unimportant constant) together with parameters used to simulate (red cross) showing that the parameters are identifiable from such data.

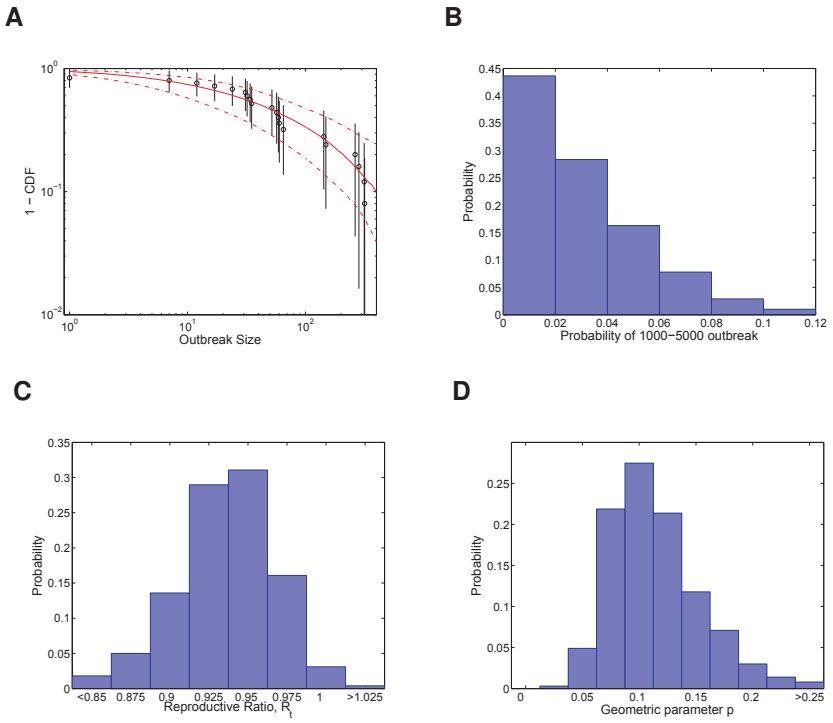

Figure 4: Analysis of transmission dynamics for completed outbreaks, plus one outbreak of size 1000-5000. (A) Model (solid red line) and 95% CI (dash-dot red line) versus data (black circles) and 95% CI (solid black lines). (B) Posterior for the probability of the large uncertain outbreak. (C) Posterior for values of the reproductive ratio R_t. (D) Posterior for the geometric parameter p.