Epidemics xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Epidemics

journal homepage: www.elsevier.com/locate/epidemics

Real-time forecasting of infectious disease dynamics with a stochastic semi-mechanistic model

Sebastian Funk*, Anton Camacho, Adam J. Kucharski, Rosalind M. Eggo, W. John Edmunds

Centre for the Mathematical Modelling of Infectious Diseases, London School of Hygiene & Tropical Medicine, London, United Kingdom

ARTICLE INFO

Article history:
Received 8 July 2016
Received in revised form
17 November 2016
Accepted 17 November 2016
Available online xxx

Keywords:
Forecasting
Real-time modelling
Infectious disease dynamics
Outbreak

ABSTRACT

Real-time forecasts of infectious diseases can help public health planning, especially during outbreaks. If forecasts are generated from mechanistic models, they can be further used to target resources or to compare the impact of possible interventions. However, paremeterising such models is often difficult in real time, when information on behavioural changes, interventions and routes of transmission are not readily available. Here, we present a semi-mechanistic model of infectious disease dynamics that was used in real time during the 2013–2016 West African Ebola epidemic, and show fits to a Ebola Forecasting Challenge conducted in late 2015 with simulated data mimicking the true epidemic. We assess the performance of the model in different situations and identify strengths and shortcomings of our approach. Models such as the one presented here which combine the power of mechanistic models with the flexibility to include uncertainty about the precise outbreak dynamics may be an important tool in combating future outbreaks.

1. Introduction

Forecasting the incidence of infectious diseases is an important part of public health and intervention planning. This was especially true during the 2013–2016 West African Ebola epidemic, when the rapid expansion of the outbreak triggered an enormous national and international public health response in late summer 2014. From November 2014, the Centre for the Mathematical Modelling of Infectious Diseases (CMMID) at the London School of Hygiene & Tropical Medicine produced weekly situation reports presenting updates of publicly available epidemiological data, model fits and forecasts, and estimates of key epidemiological parameters. These reports were distributed to a wide range of public health planners, policy makers, field workers and academics in several countries by email, and were made publicly available on a dedicated web site (Center for the Mathematical Modelling of Infectious Diseases, 2015).

The forecasts in these situation reports were produced using a stochastic semi-mechanistic model of Ebola transmission (Camacho et al., 2015b). The model was mechanistic in the sense that it relied on a compartmental description of the epidemiological status of the population based on known aspects of Ebola infection such as the incubation rate or infectious period. To model transmission between individuals, however, we used a more phenomenological, stochastic approach. During an emergency such as the Ebola epidemic, it is difficult to determine the precise factors underlying disease transmission that are required to inform a fully mechanistic model. Information about the relative importance and intensity of transmission in the community, hospital or at funerals (Faye et al., 2015), about the exact extent of control measures and their impact (WHO Ebola Response Team, 2015), about behavioural changes in the community (Funk et al., 2014) as well as about the potential role of seasonality (Groseth et al., 2007) or genetic changes in the virus (Carroll et al., 2015) were not available in real-time. To capture the overall change in transmission arising from these different mechanisms, we modelled transmission between individuals using a time-varying stochastic rate.

Capturing the uncertainty in transmission in a stochastic term gives the model the flexibility to match the data in the presence of noise and uncertainty. In addition, the inferred trajectories of the transmission rate can directly be interpreted as change in the reproduction number and thus provide valuable information for decision makers, for example by indicating how far the outbreak is from being under control.

* Corresponding author.
E-mail address: sebastian.funk@lshtm.ac.uk (S. Funk).

http://dx.doi.org/10.1016/j.epidem.2016.11.003

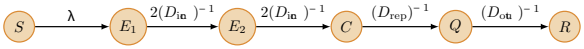

Fig. 1. Flow between compartments of the transmission model.

Here, we present model fits and forecasts generated as part of the Ebola Forecasting Challenge conducted in 2015 after the true epidemic had waned. The challenge was based on four scenarios of synthetic data inspired from the outbreak in Liberia outbreak, with increasing levels of noise and uncertainty (see Vespignani et al., in this issue). We used a similar model to the one used during the Ebola epidemic fitted to the simulated outbreak trajectories generated as part of the Challenge to produce forecasts of the number cases at upcoming time points. We particularly focus on methodological issues and forecasting performance. We report results on county-level data of scenario 1, and assess forecasts at time points 1, 2, 4 and 5 (time points 3 was not considered for logistical reasons). Results for the other scenarios are shown in the supplementary material.

2. Methods

Our semi-mechanistic model of Ebola dynamics is a modified Susceptible-Exposed-Infectious-Recovered (SEIR) model, accounting for delays in notification and time-varying transmission (Fig. 1). At time t, the force of infection experienced by susceptible individuals is $\lambda_t = \beta_t I_t/N$, where I_t is the overall number of infectious individuals, N the population size and β_t the time-varying transmission rate which follows a random walk (Dureau et al., 2013):

$$d\log\beta_t = \sigma dW_t, \tag{1}$$

Here, W_t denotes a Wiener process (Durrett, 1984), σ the volatility of the transmission rate and the log-transform ensures positivity of β_t. Modelling the time-varying transmission rate as a random walk means it is auto-correlated: the transmission rate on any day is most likely to be the same as on the previous one.

Upon infection at rate λ_t, susceptible individuals (S) move from being exposed (E) to being infectious at a rate given by the reciprocal of the incubation period ($1/D_{\text{inc}}$). We used two exposed sub-compartments in sequence to obtain an Erlang-distribution of the incubation period with shape $k = 2$ (Lloyd, 2001), and split the estimated initial number of exposed individuals evenly between these two sub-compartments. To account for delays in reporting of new cases, the infectious compartment was split into two compartments representing infectious but not yet reported (C) and infectious and potentially reported (Q) cases. The transition between C and Q occurs at a randomly varying rate with a mean equal to the reciprocal of the average reporting delay ($1/D_{\text{rep}}$) and 10% overdispersion following a Gamma distribution. This stochastic variation is introduced to capture non-independence in the time until cases get reported, in order to capture situations where, for example, several members of the same family might be reported simultaneously (Camacho et al., 2015b). Lastly, infectious individuals are removed (R) when they recover or die from the Q compartment at a rate equal to the reciprocal of the difference between the infectious period and the reporting delay ($D_{\text{out}} = D_{\text{inf}} - D_{\text{rep}}$). The model can be formulated as a set of stochastic differential equations which was simulated with the noise term fixed for a time step of 1 day and a Runge–Kutta method solving the remaining ordinary differential components (Milstein and Tretyakov, 2004). The only stochastic components are the trajectory of the transmission rate and the reporting noise, in contrast to the model used for the situation reports during the true Ebola epidemic, which also included demographic noise.

The observation process was modelled to operate on the weekly incidence (Z_t), given by the number of infectious individuals entering the Q compartment. The observed incidence ($\tilde{Z}_t$) was assumed to follow a normal approximation (chosen for computational efficiency) to the negative binomial distribution with reporting probability p and overdispersion ϕ:

$$\tilde{Z}_t \sim N(pZ_t,\, p(1-p)Z_t + p^2 Z_t^2 \phi^2). \tag{2}$$

where standard deviations smaller than 1 were rounded up to 1 to avoid the singularity at $Z_t = 0$. Note that stochastic variation here captures variability in the probability that cases get reported, whereas the stochastic variation acting on the transition from C to Q captures variability in the delay until cases can get reported.

Table 1
Parameters used in the model and their values/ranges.

Parameter	Value or prior range	Description	Source
D_{inc}	variable	Mean delay from infection to symptoms	line list (where available)
	6 days		(Camacho et al., 2014)
D_{rep}	1 week	Mean delay from symptom onset to notification	assumption
D_{inf}	variable	Mean delay from symptom onset to outcome	line list (where available)
	7.8 days		(Camacho et al., 2014)
p	0.7	Proportion of cases reported	(Camacho et al., 2014)
σ	$\mathcal{U}(0, 0.5)$	Volatility of the transmission rate	Fitted
ϕ	$\mathcal{U}(0, 0.5)$	Overdispersion in reporting	Fitted
$E^\star$	$\mathcal{U}(0, 5)$	Initial number of exposed individuals	Fitted
$\mathcal{R}^\star$	$\mathcal{U}(0, 5)$	Initial reproduction number	Fitted

The model thus has 8 parameters, which we either estimated from the line list of cases, took from a study on a pre-2014 outbreak of Ebola (Camacho et al., 2014), or estimated from model fits process (Table 1). Prior ranges of the transmission rate volatility and reporting overdispersion were established in preliminary runs and chosen to be able to sufficiently capture sudden changes in cases without allowing a degree of variation that would render the algorithm unstable.

We used a Metropolis-Hastings particle Markov chain Monte-Carlo (pMCMC) algorithm to sample from the joint posterior distributions of the estimated parameters and states of the model (i.e. the trajectories). In brief, at each MCMC step, a particle filter is used to estimate the likelihood of the proposed parameter set, and to generate a sampled trajectory of the states of the model and the transmission rate β_t from their marginal posterior distribution (Andrieu et al., 2010).

Our forecasts were generated under a "no change" hypothesis: we assumed that the transmission rate would remain constant after the last observed data point. More precisely, we sampled 10,000 parameter sets from the posterior distribution in combination with the associated states and estimated values of β_t at the last observed data point, and simulated the model forward one year. The future number of reported cases was generated by applying the observation process to the forecast incidence. Predicted reported peak incidence, death counts and final sizes were calculated from the sampled forecast observation trajectories. County-level forecasts were obtained under the assumption that no transmission occurred between counties.

Model fits were generated using a fully automated algorithm applied to all the regional and national data sets as follows, implemented to facilitate convergence of the computationally intensive pMCMC sampler and to avoid long burn-in and low effective sample sizes: First, Metropolis-Hastings MCMC was run on the

Please cite this article in press as: Funk, S., et al., Real-time forecasting of infectious disease dynamics with a stochastic semi-mechanistic model. Epidemics (2016), http://dx.doi.org/10.1016/j.epidem.2016.11.003