Epidemics xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Epidemics

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

A simple approach to measure transmissibility and forecast incidence

Pierre Nouvellet[a,b], Anne Cori[a], Tini Garske[a], Isobel M. Blake[a], Ilaria Dorigatti[a], Wes Hinsley[a], Thibaut Jombart[a,b], Harriet L. Mills[a], Gemma Nedjati-Gilani[a], Maria D. Van Kerkhove[a,c], Christophe Fraser[a,b], Christl A. Donnelly[a,b], Neil M. Ferguson[a,b], Steven Riley[a,b,*]

[a] MRC Centre for Outbreak Analysis and Modelling, Imperial College London, Faculty of Medicine, London, UK
[b] National Institute for Health Research Health Protection Research Unit in Modelling Methodology, Imperial College London, Faculty of Medicine, London, UK
[c] Center for Global Health, Institute Pasteur, Paris, France

ARTICLE INFO

Article history:
Received 14 June 2016
Received in revised form 1 February 2017
Accepted 17 February 2017
Available online xxx

Keywords:
Forecasting
Rapid response
Branching process
Renewal equation
MCMC

ABSTRACT

Outbreaks of novel pathogens such as SARS, pandemic influenza and Ebola require substantial investments in reactive interventions, with consequent implementation plans sometimes revised on a weekly basis. Therefore, short-term forecasts of incidence are often of high priority. In light of the recent Ebola epidemic in West Africa, a forecasting exercise was convened by a network of infectious disease modellers. The challenge was to forecast unseen "future" simulated data for four different scenarios at five different time points. In a similar method to that used during the recent Ebola epidemic, we estimated current levels of transmissibility, over variable time-windows chosen in an ad hoc way. Current estimated transmissibility was then used to forecast near-future incidence. We performed well within the challenge and often produced accurate forecasts. A retrospective analysis showed that our subjective method for deciding on the window of time with which to estimate transmissibility often resulted in the optimal choice. However, when near-future trends deviated substantially from exponential patterns, the accuracy of our forecasts was reduced. This exercise highlights the urgent need for infectious disease modellers to develop more robust descriptions of processes – other than the widespread depletion of susceptible individuals – that produce non-exponential patterns of incidence.

1. Introduction

In epidemiology, and particularly in the context of outbreaks, mathematical modelling is now frequently used to forecast future incidence (Chretien et al., 2015; Nsoesie et al., 2014). Such forecasts were initially performed to improve the situational awareness of key stakeholders. Increasingly, forecasting incidence is used in the context of advocacy planning, to monitor the situation, and to help implement, prioritise and evaluate control strategies. During the recent Ebola epidemic in West Africa, such forecasts were almost continuously performed: many were shared with policy makers with some published in peer-reviewed literature (WHO Ebola Response Team, 2015a; Meltzer et al., 2014; WHO Ebola Response Team, 2014; Gomes et al., 2014; Merler et al., 2015).

While all methods for forecasting future incidence seek to characterise the central predicted trend and the dispersion around it based on covariates, they vary according to the nature of the underlying model, with some methods relying on a purely statistical approaches (Goldstein et al., 2011) and some relying on a mechanistic models of disease transmission (Meltzer et al., 2014). Recent forecasting exercises in the context of influenza (Influenza Forecasting, 2017), Dengue (Dengue Forecasting, 2017) or Chikungunya (Chikungunya Forecasting, 2017) highlight the diversity of possible models with some clearly belonging to one of the aforementioned categories while others take a more nuanced approach perhaps best described as semi-mechanistic. In all models, a careful balance must be reached between obtaining accurate forecasts while accounting for all uncertainties, both in the data themselves and in the dynamics of transmission.

During the recent Ebola epidemic, our team helped support the World Health Organization (WHO) and the Ministries of Health of the three most affected countries (Guinea, Liberia and Sierra Leone). In a wide collaborative effort, we were able to gain valu-

* Corresponding author at: MRC Centre for Outbreak Analysis and Modelling, Imperial College London, St Mary's Campus, Norfolk Place, London W2 1PG, UK.
E-mail address: s.riley@imperial.ac.uk (S. Riley).

http://dx.doi.org/10.1016/j.epidem.2017.02.012

Table 1
Estimated instantaneous reproduction numbers (R_t) and serial intervals (in days) for the 5 time-points and 4 scenarios.

Scenario	Line-list	Case-count	Field-report	Time-point	R0 (median)	R0 (IQR)	SI (median)	SI (IQR)
1	√	√	√	1	1.03	[0.86; 1.25]	15.4	[11.3; 18.7]
				2	1.33	[1.27; 1.40]	13.3	[10.1; 16.0]
				3	0.87	[0.85; 0.90]	12.5	[9.8; 14.8]
				4	0.87	[0.85; 0.90]	12.5	[9.8; 14.8]
				5	0.79	[0.75; 0.82]	12.7	[10.3; 14.7]
2		√	√	1	1.62	[1.49; 1.75]	14.2[a]	
				2	0.89	[0.86; 0.92]		
				3	1.00	[0.96; 1.05]		
				4	0.91	[0.89; 0.94]		
				5	0.72	[0.70; 0.74]		
3		√	√	1	1.69	[1.55; 1.83]	14.2[a]	
				2	1.28	[1.20; 1.37]		
				3	1.32	[1.28; 1.37]		
				4	1.05	[1.02; 1.08]		
				5	0.69	[0.67; 0.71]		
4		√	√	1	1.43	[1.29; 1.58]	14.2[a]	
				2	1.39	[1.31; 1.46]		
				3	1.12	[1.09; 1.15]		
				4	0.88	[0.85; 0.91]		
				5	0.98	[0.96; 0.99]		

IQR: interquartile range.
[a] Indicated that in the absence a line-list, the distribution of the serial interval was taken from WHO Ebola Response Team (2015a). Unknown at the time of challenge, accuracy of data and reports progressively decreased from scenario 1 to scenario 4.

able insights into the transmissibility, epidemiology and impact of intervention strategies (WHO Ebola Response Team, 2015a, 2014, 2015b; Nouvellet et al., 2015; WHO Ebola Response Team, 2016; Garske et al., 2017). We were also involved in producing regular forecast of future incidence (e.g. WHO Ebola Response Team, 2015a, 2014), using a semi-mechanistic model based on a renewal equation (Fraser, 2007).

As the Ebola epidemic was declining, the Research and Policy for Infectious Disease Dynamics (RAPIDD) program, from the US National Institute of Health's Fogarty International Center, gave eight teams (including us) the opportunity to assess their models against simulated data. Simulated data (based on Gomes et al., 2014; Merler et al., 2015) for 4 outbreak scenarios, differing in the assumptions underlying transmissibility and degree/quality of data reporting, at 5 different time-points during the outbreak were provided together with 'field reports' outlining the epidemiological situation (see SI.1). For each scenario and time-point, we were tasked with providing short-term forecasts (4 weeks into the future) and an estimate of the current level of transmissibility. Here we present the method used by the 'Imperial College Team' and how it performed.

2. Methods

At each of the five time points, and for each scenario, we were provided with a case-count dataset that consisted of weekly counts of newly confirmed cases (Table 1). A field report was also provided, containing information on interventions, e.g.: timing of a recently implemented intervention or increased bed capacity (see SI.1).

Our approach was to estimate the current reproduction number (the average number of secondary cases generated by a typical infected individual, R_t) and to use that to forecast future incidence (Figs. 1–2). The current reproduction number was estimated using the case-count dataset, assuming constant transmissibility during a chosen time-window (see the Estimation and Forecast sections below).

For scenario 1, we were also provided with a line-list. The line-list contained detailed data for each individual and was used exclusively to infer a serial interval distribution and gain epidemiological insights into the current situation (see preliminary analyses and Fig. 3). The line-list focused on confirmed cases, and was affected by both under-reporting and delays in reporting (see 'Preliminary analyses' below).

2.1. Estimation of the reproduction number

The reproduction number used to forecast future incidence was estimated from the case-count data.

Several methods to estimate the reproduction number exist, e.g. see Van Kerkhove et al. (2015) for various methods linked to the estimation of the basic and effective reproduction of Ebola virus. Here we relied on a well-established and simple method that assumed the daily incidence, I_t, could be approximated with a Poisson process following the renewal equation (Fraser, 2007):

$$I_t \sim Pois\left(R_t \sum_{s=0}^{t} I_{t-s}\omega_s\right),$$

where R_t is the instantaneous reproduction number and ω the serial interval distribution. From this a likelihood of the data given a set of model parameters can be calculated, as well the posterior distribution of R_t given previous observations of incidence and knowledge of the serial interval (Cori et al., 2013). The serial interval was assumed to be gamma distributed with parameters taken either from the literature (WHO Ebola Response Team, 2015a) (i.e. for scenario 2–4), or estimated from the line-list (i.e. scenario 1, see preliminary analyses below).

We used this approach to estimate R_t over three alternative time-windows defined by assuming a constant R_t for either the 2, 3 or 4 weeks prior to the most recent data-point. We made no assumptions regarding the epidemiological situation and transmissibility prior to each time-window. Therefore, no data prior to the time-window were used to estimate R_t and instead we jointly estimated R_t as well as back-calculated the incidence before the time-window. Specifically, we jointly estimated the R_t and the incidence level 100 days before the time-widow. Past incidence was then calculated using the known relationship between the serial interval, growth rate and reproduction number (Wallinga and Lipsitch, 2007). The joint posterior distribution of R_t and the

Please cite this article in press as: Nouvellet, P., et al., A simple approach to measure transmissibility and forecast incidence. Epidemics (2017), http://dx.doi.org/10.1016/j.epidem.2017.02.012