Epidemics xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Epidemics

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

Using phenomenological models for forecasting the 2015 Ebola challenge

Bruce Pell [a,d,*], Yang Kuang [a], Cecile Viboud [c], Gerardo Chowell [b,c]

[a] School of Mathematical and Statistical Sciences, Arizona State University, AZ, USA
[b] School of Public Health, Georgia State University, Atlanta, GA, USA
[c] Division of International Epidemiology and Population Studies, Fogarty International Center, National Institutes of Health, Bethesda, MD, USA
[d] Department of Mathematics, Statistics, and Computer Science, St. Olaf College, MN, USA

ARTICLE INFO

Article history:
Received 30 June 2016
Received in revised form 1 November 2016
Accepted 15 November 2016
Available online xxx

Keywords:
Logistic growth model
Richards model
Generalized Richards model
Ebola challenge

ABSTRACT

Background: The rising number of novel pathogens threatening the human population has motivated the application of mathematical modeling for forecasting the trajectory and size of epidemics.
Materials and methods: We summarize the real-time forecasting results of the logistic equation during the 2015 Ebola challenge focused on predicting synthetic data derived from a detailed individual-based model of Ebola transmission dynamics and control. We also carry out a post-challenge comparison of two simple phenomenological models. In particular, we systematically compare the logistic growth model and a recently introduced generalized Richards model (GRM) that captures a range of early epidemic growth profiles ranging from sub-exponential to exponential growth. Specifically, we assess the performance of each model for estimating the reproduction number, generate short-term forecasts of the epidemic trajectory, and predict the final epidemic size.
Results: During the challenge the logistic equation consistently underestimated the final epidemic size, peak timing and the number of cases at peak timing with an average mean absolute percentage error (MAPE) of 0.49, 0.36 and 0.40, respectively. Post-challenge, the GRM which has the flexibility to reproduce a range of epidemic growth profiles ranging from early sub-exponential to exponential growth dynamics outperformed the logistic growth model in ascertaining the final epidemic size as more incidence data was made available, while the logistic model underestimated the final epidemic even with an increasing amount of data of the evolving epidemic. Incidence forecasts provided by the generalized Richards model performed better across all scenarios and time points than the logistic growth model with mean RMS decreasing from 78.00 (logistic) to 60.80 (GRM). Both models provided reasonable predictions of the effective reproduction number, but the GRM slightly outperformed the logistic growth model with a MAPE of 0.08 compared to 0.10, averaged across all scenarios and time points.
Conclusions: Our findings further support the consideration of transmission models that incorporate flexible early epidemic growth profiles in the forecasting toolkit. Such models are particularly useful for quickly evaluating a developing infectious disease outbreak using only case incidence time series of the early phase of an infectious disease outbreak.

1. Introduction

The rising number of novel pathogens with transmission potential threatening the human population has motivated the development of mathematical and computational modeling approaches for forecasting epidemic impact (Colizza et al., 2006; Balcan et al., 2009; Merler et al., 2015; Chretien et al., 2015). While epidemic models of disease spread have been used for decades primarily with the goal of gaining insight into the transmission dynamics and potential effect of different control strategies, researchers have only recently started to harness available computational power to simulate, calibrate, and generate forecasts of epidemic spread using a variety of epidemic models ranging from classic compartmental models to detailed agent-based models. Yet, besides significant increases in computational power, detailed epidemic data about the transmission characteristics and theoretical advances are needed in order to more realistically account for transmission and control mechanisms for different disease and social contexts.

* Corresponding author at: Arizona State University, Tempe, AZ, USA.
E-mail addresses: Bepell@asu.edu, pell1@stolaf.edu (B. Pell).

http://dx.doi.org/10.1016/j.epidem.2016.11.002

Because epidemics associated with infectious diseases of rapid dissemination typically comprise only a few disease generations of transmission, epidemic assessment using forecasting models is crucial during the early epidemic growth phase in order assess the potential disease burden posed by the infectious agent and approximate the scale of interventions needed to achieve epidemic containment. Unfortunately, the availability of detailed epidemiological data particularly during the early epidemic stages of an evolving epidemic outbreak is hindered by delays in detecting the first transmission events or releasing data to the public, or the particular characteristics of the surveillance system. For instance, during the 2014–15 Ebola epidemic in West Africa, publicly available epidemiological data from the World Health Organization (WHO) was not available during the first weeks during which the virus was to gain a solid foothold in populations of Guinea, Liberia and Sierra Leone. Moreover, data was largely limited to aggregated weekly Ebola case counts at the country level, which was the primary publicly available dataset documenting the Ebola epidemic in West Africa. Case count data at the subnational level (e.g. county/district levels) that later become available revealed substantial spatial heterogeneity in transmission patterns across the affected areas in West Africa, which could have influenced epidemic forecasts and assessments of the transmission potential (Chowell et al., 2015).

In this article we summarize the forecasting results from using the logistic equation to forecast the 2015 Ebola challenge. After summarizing these results, we present the results of a post-challenge systematic comparative analysis of the logistic growth model, which assumes an early exponential growth phase (chowell and Viboud, 2016), and the generalized Richards model (GRM) (Chowell et al., 2016a), which incorporates a flexible range of early epidemic growth profiles including early sub-exponential and exponential growth epidemics. We compare the performance of these models in the context of the 2015 Ebola challenge based on synthetic data derived from a detailed individual-based model of Ebola transmission. Specifically, we analyze the reproduction number, forecasts of the epidemic trajectory and the final epidemic size. In addition to model comparison, we compare two uncertainty methods of the best fit solutions to the synthetic data.

2. Materials and methods

2.1. Model description

The well-known logistic growth model was previously employed for epidemic forecasting the 2015 Ebola epidemic (Chowell et al., 2014), and was the model originally employed by the Arizona State Team (BP & YK) during the 2015 Ebola Challenge. This simple model is given by the following differential equation:

$$C' = rC\left(1 - \frac{C}{K}\right) \quad (1)$$

where $C'(t)$ models the rate of change in the number of new cases at week t. The logistic model relies on two parameters, the intrinsic infection rate, r, and the final epidemic size K.

For comparative purposes, we also analyzed the performance of the recently introduced generalized Richards model (GRM) (Chowell et al., 2016a), which has been recently devised in order to capture the possibility of early sub-exponential growth epidemics and is given by:

$$C' = rC^p\left(1 - \left(\frac{C}{K}\right)^a\right) \quad (2)$$

The GRM is an enhanced version of the Richards model (Wang et al., 2012) by integrating the generalized-growth model (GGM; $C' = rC^p(t)$) (Viboud et al., 2016). Specifically, the GRM incorporates a deceleration of growth parameter p to model a range of early epidemic growth profiles ranging from constant incidence ($p = 0$), polynomial ($0 < p < 1$) and exponential growth dynamics ($p = 1$). The GRM model was recently employed to generate forecasts of the Zika epidemic in Antioquia, Colombia (Chowell et al., 2016a). All parameter values are positive: r is the growth rate, K is the final epidemic size, and a is a parameter that modulates the peak timing.

2.2. Data

The Research and Policy for Infectious Disease Dynamics (RAPIDD) Ebola Challenge was designed to test the forecasting ability of mathematical models during an epidemic in real-time (Ebola Challenge website, 2016). The challenge was motivated by the need to develop and test an ensemble of mathematical models for use in forecasting developing infectious disease epidemics and to foster collaborations across different scientific domains. Goals of the contest included:

1. Improving predictive capabilities for future emergencies
2. Guiding the implementation of control measures
3. Illustrating how data quality and availability affect prediction accuracy

In this spirit, synthetic epidemic data was generated by a modified version of the model published by Merler et al. that was calibrated for an EVD outbreak in Liberia (Merler et al., 2015). Synthetic epidemic data was released at five different time points with a test release on Sept. 18, 2015. Model predictions were due two weeks later after each time point. For model calibration, we only used the country level incidence time series data for predictions.

Contained in each of the five batches of released data, four scenarios representing different epidemiological conditions, behavioral changes, intervention measures and data availability were prepared for use in forecasting the epidemic (chowell and Viboud, 2016). In addition, each scenario dataset contained outbreak situation reports, transmission tree data and weekly reported new EVD cases at the county and country level. New EVD cases were forecasted at one, two, three and four weeks past each time point, see Fig. S1.

2.3. The generation time

The generation time is defined as the time elapsed between infection in an index case patient and infection in a patient infected by that index case (Chowell et al., 2006). We used transmission tree data (Ebola Challenge website, 2016) that was made available as part of the challenge for scenarios 1, 3 and 4 to derive their generation time distributions, respectively. For scenario 2 we used estimations from scenario 1.

2.4. The effective reproduction number

The effective reproduction number, $R_e(t)$, is defined as the average number of new infections generated by one infectious individual in the population at time t (Nishiura and Chowell, 2009). $R_e(t)$ was numerically evaluated by training each model on an increasing amount of data (Chowell et al., 2016a,b) using the discretized renewal equation (Nishiura and Chowell, 2009; Chowell et al., 2016b; Fraser, 2007):

$$R_e(t_i) = \frac{I_i}{\sum_{j=0}^{i} I_{i-j}\rho_j} \quad (3)$$

where I_i denotes incidence at time t_i, ρ_j denotes the discretised probability distribution of the generation interval, which we

Please cite this article in press as: Pell, B., et al., Using phenomenological models for forecasting the 2015 Ebola challenge. Epidemics (2016), http://dx.doi.org/10.1016/j.epidem.2016.11.002