



Published in final edited form as:

Epidemics. 2019 September ; 28: 100354. doi:10.1016/j.epidem.2019.100354.

Real-time predictions of the 2018–2019 Ebola virus disease outbreak in the Democratic Republic of the Congo using Hawkes point process models

J. Daniel Kelly^{a,h,*}, Junhyung Park^{b,1}, Ryan J. Harrigan^c, Nicole A. Hoff^d, Sarita D. Lee^b, Rae Wannier^a, Bernice Selo^e, Mathias Mossoko^e, Bathe Njoloko^e, Emile Okitolonda-Wemakoy^f, Placide Mbala-Kingebeni^g, George W. Rutherford^a, Thomas B. Smith^c, Steve Ahuka-Mundeke^g, Jean Jacques Muyembe-Tamfum^g, Anne W. Rimoin^d, Frederic Paik Schoenberg^b

^aDepartment of Epidemiology and Biostatistics, University of California, San Francisco, CA, USA

^bDepartment of Statistics, University of California, Los Angeles, CA, USA

^cCenter for Tropical Research, Institute of the Environment and Sustainability, University of California, Los Angeles, CA, USA

^dDepartment of Epidemiology, University of California, Los Angeles, CA, USA

^eMinistry of Health, Kinshasa, Congo

^fSchool of Public Health, University of Kinshasa, Kinshasa, Congo

^gInstitut National de Recherche Biomedicale, Kinshasa, Congo

^hF.I. Proctor Foundation, University of California, San Francisco, CA USA

Abstract

As of June 16, 2019, an Ebola virus disease (EVD) outbreak has led to 2136 reported cases in the northeastern region of the Democratic Republic of the Congo (DRC). As this outbreak continues to threaten the lives and livelihoods of people already suffering from civil strife and armed conflict, relatively simple mathematical models and their short-term predictions have the potential to inform Ebola response efforts in real time. We applied recently developed non-parametrically estimated Hawkes point processes to model the expected cumulative case count using daily case counts from May 3, 2018, to June 16, 2019, initially reported by the Ministry of Health of DRC and later confirmed in World Health Organization situation reports. We generated probabilistic estimates of the ongoing EVD outbreak in DRC extending both before and after June 16, 2019, and evaluated their accuracy by comparing forecasted vs. actual outbreak sizes, out-of-sample log-likelihood scores and the error per day in the median forecast. The median estimated outbreak

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

*Corresponding author at: Box 0886 3rd Floor, 550 16th Street, San Francisco, CA 94143, USA. Dan.Kelly@ucsf.edu (J.D. Kelly).

¹Co-first authors.

Declaration of Competing Interest
None to declare.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.epidem.2019.100354>.

sizes for the prospective three-, six-, and nine-week projections made using data up to June 16, 2019, were, respectively, 2317 (95% PI: 2222, 2464); 2440 (95% PI: 2250, 2790); and 2544 (95% PI: 2273, 3205). The nine-week projection experienced some degradation with a daily error in the median forecast of 6.73 cases, while the six- and three-week projections were more reliable, with corresponding errors of 4.96 and 4.85 cases per day, respectively. Our findings suggest the Hawkes point process may serve as an easily-applied statistical model to predict EVD outbreak trajectories in near real-time to better inform decision-making and resource allocation during Ebola response efforts.

Keywords

Ebola virus disease; Hawkes point process; Mathematical modelling; Democratic Republic of Congo; Compartmental models

1. Introduction

As of June 16, 2019, 2136 confirmed and probable cases of Ebola virus disease (EVD) were reported in North Kivu and Ituri Provinces of the Democratic Republic of the Congo (DRC) (WHO, 2019). Security issues resulting from activities of over 100 rebel and other insurgent groups, including attacks on Ebola treatment centers in Butembo and Katwa, have likely contributed to the ongoing nature of this EVD outbreak (Damon et al., 2018). Of the > 34 prior EVD outbreaks (CDC, 2019), none have occurred in a geographic region with a similar set of conflict issues. Moreover, the use of case counts from previous EVD outbreaks reported in the literature have proven unreliable in their ability to forecast an outbreak's size (Worden et al., 2018; Asher, 2018). It is likely that additional time and effort will be required before all of the contributing factors to this outbreak can be properly assessed, parameterized, and modeled.

The Hawkes point process model, however, offers the Ebola modeling community a novel, rapid option to forecast outbreak size and spread (Meyer et al., 2012). Using modern methods, one can rapidly and nonparametrically estimate short-term outbreak size and rely on minimal modeling assumptions to do so (Schoenberg et al., 2017; Hawkes, 1971; Park et al., 2018). Decomposing peak history effects into the contribution of previous events and an average background rate, this point process model has long been used in the context of seismology to describe earthquakes and their aftershocks as well as other environmental science and biological phenomena (Hawkes, 1971; Gerhard et al., 2017; Schoenberg, 2004; Marsan and Lengliné, 2008). In some cases, Hawkes point process models have also been used to forecast the spatial and temporal spread of infectious disease outbreaks (Schoenberg et al., 2018; Meyer and Leonard, 2014; Meyer et al., 2012), including the 2013–2016 EVD outbreak in West Africa (Park et al., 2018).

There is an increasing body of evidence suggesting that short-term forecasts with few parameters are more reliable than long-term forecasts (particularly early in an outbreak) that determine the final outbreak size (Worden et al., 2018; Funk et al., 2018; Viboud et al., 2017; Chowell et al., 2017). In the context of an ongoing outbreak, many published statistical models have focused on long-term or final outbreak size (Meltzer et al., 2014;

Kelly et al., 2018; Valdez et al., 2015; Chretien et al., 2015; Siettos et al., 2015). Given the advantages of the Hawkes model and the limitations of other statistical models in the ongoing EVD outbreak setting (Chowell et al., 2017), we fit the Hawkes point process model to daily EVD case counts to forecast case counts over subsequent weeks. It is our hope that this application of the Hawkes point process model may further engage outbreak responders on the value of short-term forecasts when making important public health decisions related to resource allocations.

2. Methods

Data were collected from the Ministry of Health and World Health Organization (WHO) situation reports on EVD case counts occurring in the northeastern region of DRC. The Ministry of Health initially released daily case counts while WHO situation reports confirmed these case counts with weekly reports (WHO, 2019). Our dataset included probable and confirmed EVD cases that occurred from the start of the outbreak on May 3, 2018, until June 16, 2019 (Supplement 1). (We only included in our models case counts from the EVD outbreak in the northeastern region of DRC. In 2018, there was another EVD outbreak that occurred in the western region of DRC, and WHO declared the end of this outbreak on July 24. Although there was a temporal overlap of the EVD outbreaks in DRC, they occurred approximately 1500 miles apart and there has been no evidence of an epidemiological or viral genetic link between them).

We fit the Hawkes point process model to daily EVD case counts reported in the northeastern region of DRC. Details of this estimation method can be found elsewhere (Park et al., 2018). Briefly, for point processes, the expected rate at which points (or cases) accumulate at time t is characterized by the conditional intensity $\lambda(t)$. Although versions of Hawkes models have parameters that describe these types of data in space and time, to be comparable with the SEIR compartmental model here we consider a purely temporal Hawkes process (Hawkes, 1971) here, where $\lambda(t)$ is written as:

$$\lambda(t) = \mu + K \sum_{i: t_i < t} g(t - t_i).$$

The Hawkes model is estimated essentially by fitting a step function to the triggering density g , where the step heights and background rate μ are estimated by maximum likelihood, according to the method of Marsan and Lengliné (2008), and the step function is subsequently smoothed using a Gaussian kernel. The triggering density g indicates the rate at which infection is spread, and the fitted triggering density shows most secondary infections occurring within a week (Fig. 1).

The log-likelihood of an observed sequence of infections according to an estimated Hawkes model is:

$$l(\hat{\theta}) = \sum_i \log[\lambda(t_i; \hat{\theta})] - \int_0^T \lambda(t; \hat{\theta}) dt.$$

Here, $\hat{\theta} = (\cdot)$ is the vector of parameter estimates. The log-likelihood can be computed on the data used to estimate the parameters or can be computed on data outside of the training sample. The log-likelihood is a measure of fit and is closely related to the entropy or information gain of the estimated model relative to a stationary Poisson model (Harte and Vere-Jones, 2005).

One application of the Hawkes model is to enable real-time forecasting of an EVD outbreak. Using the median of 1000 simulations of the fitted Hawkes model, we predicted the number of cases expected to occur over a nine-, six-, and three-week period, starting on April 14, 2019, May 5, 2019 and May 26, 2019, all ending on June 16, 2019, where each subsequent forecast uses model parameters re-estimated with updated data. Then using data up to June 16, we generated probabilistic projections of three-, six-, and nine- weeks based on prior research showing the degradation of epidemic forecasting accuracy over the long term (Worden et al., 2018; Chowell et al., 2017). We evaluated the accuracy of our probabilistic projections by comparing projected vs. actual outbreak sizes, the log-likelihood (information) score (Brocker and Smith, 2005) and the error per day in the median forecast. On April 14, 2019, May 5, 2019 and May 26, 2019 there were 1312, 1667 and 1956 reported EVD cases, respectively. We conducted all analyses using R 3.4.2 (R Foundation for Statistical Computing, Vienna, Austria).

3. Results

As of June 16, 2019, there were 2136 reported EVD cases across 22 health zones in the provinces of North Kivu and Ituri, DRC. Of these EVD cases, about 95.7% were confirmed and 4.3% were probable. We used the Hawkes model to generate nine-, six- and three-week probabilistic forecasts (all ending June 16, 2019) (Fig. 2). The median simulated outbreak size on June 16 was 1892 (95% prediction interval [PI]: 1525, 2641), 2236 (95% PI: 1881, 2773) and 2206 (95% PI: 2079, 2401) respectively. The errors in the median forecasts for the nine-, six-, and three-week forecasts were respectively 6.73 cases, 4.96 cases, and 4.85 cases per day. The log-likelihood (per case) evaluated on the data after the forecasts were made for the nine-, six- and three-week forecasts were 1.60, 1.43 and 1.24, respectively. The higher log-likelihood per day for the 9-week forecasts appears to be attributable to the increased number of observed cases during the first few weeks of the forecasting period causing a sharp increase in the sum of $\log(\lambda)$ term in the log-likelihood.

In our forecast of the unobserved period using data up to June 16, the three-, six-, and nine-week probabilistic projections of median outbreak size were respectively as follows: 2317 (95% PI: 2222, 2464); 2440 (95% PI: 2250, 2790); and 2544 (95% PI: 2273, 3205) (Fig. 3). The log-likelihood score (per case) of the estimated Hawkes models in Figs. 2a–c and 3 are 0.69, 0.93, 1.04 and 1.06, respectively. Projected and actual outbreak sizes followed a near linear increase (Fig. 2).

4. Discussion

We employed a non-parametrically estimated Hawkes point process model to generate multiple probabilistic projections of the ongoing 2018–2019 EVD outbreak size in DRC. As

seen in Fig. 2, the median nine-week projection experienced some degradation with forecast errors of 6.73 cases per day, whereas the six- and three-week projections were more reliable, with errors in the median forecasts of 4.96 and 4.85 cases per day, respectively, and with the observed number of cases falling well within the estimated 95% prediction window obtained using simulations of the fitted Hawkes model for the three-week period. These findings were consistent with other modeling studies that have shown how even short-term forecasts can degrade over longer periods of time (Funk et al., 2018; Worden et al., 2018). Our results support earlier work performed using Hawkes point process models to predict the size of infectious disease outbreaks; our models of the 2013–2016 EVD outbreak in West Africa reduced root mean squared error (RMSE) by as much as 38% when compared to traditional compartmental models (Park et al., 2018). Growing evidence, including the work presented here, suggests that point process models can provide accurate estimates of caseloads for a wide variety of epidemics, including both ongoing and previous Ebola outbreaks (Schoenberg et al., 2017; Park et al., 2018).

The Hawkes model performed well during this outbreak with minimal modeling assumptions, and could be a valuable tool for real-time decision making amidst ongoing outbreak of EVD or other diseases. In its non-parametric form, a disadvantage of the Hawkes model may be its inability to parameterize contexts that may help explain the current epidemic trajectory. While these factors (e.g., contact tracing and clinical care) may be considered in future iterations of the Hawkes model (Funk et al., 2017), developing these parameters can also delay model development and application. Even with such parameters estimated, some factors in real biological epidemics, such as political unrest or armed conflict that affect disease transmission rates, can be challenging to parameterize in statistical models.

While the Hawkes model's simplicity has advantages, it can also be viewed as a limitation when, for instance, inhomogeneity of the background rate or changes in productivity lead to overestimation of finescale clustering, leading to a triggering function estimate that may be less biologically plausible. Unanticipated shocks (e.g., introduction of EVD into a large metropolitan area) that occur after predictions may decrease the model's accuracy beyond our uncertainty estimates. In addition, these models estimate future cases via the triggering function, which requires scrutiny due to its tendency to underestimate secondary transmission rates. Should dynamics of the disease rapidly change at a time period for which data was not included (e.g. driving productivity to a value greater than 1), a relatively simple point process model may not account for such rapid shifts in triggering and may not be able to anticipate them in forecasts. While the Hawkes model might be able to adjust to decreasing numbers of cases as data become available and parameter estimates change, it may well be that the Hawkes model fails to perform well as the disease cases wane near the end of an outbreak, and this behavior should be a major subject of future research in assessing the forecasts made here.

As such, we do not suggest here that these models replace traditional compartmental models (SIR and their relatives). Rather, we see these models as complementary, and in the particular case of requiring rapid response and prediction of caseloads, a valuable addition for efforts that attempt to limit the impacts of an outbreak. In an effort to continue to

evaluate the efficacy of these models in predicting outbreak rates and cumulative cases in real-time (or near real-time, given the time it takes for corrected caseload data to be released), we have constructed a free, publicly-accessible website that can track this and other outbreaks, with purely prospective forecasts and results updated weekly as new data become available (for full details, see <http://www.stat.ucla.edu/~frederic/ebola>).

In conclusion, we are encouraged by the ability of non-parametric Hawkes point process models to describe epidemic events over the short term and in real time that are consistent with the 2018–2019 EVD outbreak in DRC. The Hawkes point process is a relatively simple statistical model, and results suggest that statistical modelers in the public health community should consider the Hawkes model in their ensemble when engaging in decision-making and resource allocation of EVD and other emerging infectious disease outbreaks.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank the Ebola responders for their efforts in the 2018–2019 EVD outbreak in North Kivu and Ituri Provinces, Democratic Republic of Congo.

Funding sources

This work was supported by the U.S. National Science Foundation NSF (grant number PD-08-1269), awarded to F.P.S and R.J.H.

References

- Asher J, 2018. Forecasting Ebola with a regression transmission model. *Epidemics* 22, 50–55. 10.1016/j.epidem.2017.02.009. [PubMed: 28342787]
- Brocker J, Smith LA, 2005. Scoring probabilistic forecasts: the importance of being proper. *Am. Meteorol. Soc.* 4 (November). 10.1175/WAF966.1. Available at: .
- Centers for Disease Control and Prevention Ebola (Ebola Virus Disease): History of Ebola Virus Disease: 2014–2016 Ebola Outbreak in West Africa: Case Counts. Available at: <https://www.cdc.gov/vhf/ebola/history/2014-2016-outbreak/case-counts.html>.
- Chowell G, Viboud C, Simonsen L, Merler S, Vespignani A, 2017. Perspectives on model forecasts of the 2014–2015 Ebola epidemic in West Africa: lessons and the way forward. *BMC Med.* 15 (1), 42. 10.1186/s12916-017-0811-y. [PubMed: 28245814]
- Chretien JP, Riley S, George DB, 2015. Mathematical modeling of the West Africa Ebola epidemic. *Elife* 4 10.7554/eLife.09186.
- Damon IK, Rollin PE, Choi MJ, Arthur RR, Redfield RR, 2018. New tools in the Ebola Arsenal. *N. Engl. J. Med.* 379 (21), 1981–1983. 10.1056/NEJMp1811751. [PubMed: 30281389]
- Funk S, Ciglenecki I, Tiffany A, Gignoux E, Camacho A, Eggo RM, et al. , 2017. The impact of control strategies and behavioural changes on the elimination of Ebola from Lofa County, Liberia. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* 372 (1721). 10.1098/rstb.2016.0302. Central.
- Funk S, Camacho A, Kucharski A, Lowe R, Eggo R, Edmunds J, 2018. Assessing the performance of real-time epidemic forecasts: a case study of the 2013–16 Ebola epidemic. *BioRxiv* 23 (November) Available at: <https://www.biorxiv.org/content/early/2018/11/23/177451>.
- Gerhard F, Deger M, Truccolo W, 2017. On the stability and dynamics of stochastic spiking neuron models: nonlinear Hawkes process and point process GLMs. *PLoS Comput. Biol.* 13 (2), e1005390. 10.1371/journal.pcbi.1005390. [PubMed: 28234899]

- Harte D, Vere-Jones D, 2005. The entropy score and its uses in earthquake forecasting. *Pure Appl. Geophys.* 162 (6–7), 1229–1253.
- Hawkes AG, 1971. Point spectra of some mutually exciting point processes. *J. R. Stat. Soc. B* 33, 438–443.
- Kelly JD, Worden L, Wannier R, Hoff NA, Mukadi P, Sinai C, Ackley S, Chen X, Gao D, Selo B, Mossoko M, Okitolonda-Wemakoy E, Richardson ET, Rutherford GW, Lietman TM, Muyembe-Tamfum JJ, Rimoin AW, Porco TC, 2018. Real-time projections of Ebola outbreak size and duration with and without vaccine use in Equateur, Democratic Republic of Congo. *BioRxiv*. 10.1101/331447.. As of May 27, 2018. Available at: <https://www.biorxiv.org/content/early/2018/06/04/331447.1>.
- Marsan D, Lengliné O, 2008. Extending earthquakes' reach through cascading. *Science* 319 (5866), 1076–1079. 10.1126/science.1148783.. [PubMed: 18292339]
- Meltzer MI, Atkins CY, Santibanez S, Knust B, Petersen BW, Ervin ED, et al. , 2014. Estimating the future number of cases in the Ebola epidemic—Liberia and Sierra Leone, 2014–2015. *MMWR Suppl.* 63 (3), 1–14
- Meyer S, Elias J, Höhle M, 2012. A space-time conditional intensity model for invasive meningococcal disease occurrence. *Biometrics* 68 (2), 607–616. 10.1111/j.1541-0420.2011.01684.x. [PubMed: 21981412]
- Meyer S, Leonard H, 2014. Power-law models for infectious disease spread. *Ann. Appl. Stat.* 8 (3), 1612–1639.
- Park J, Chafee AW, Harrigan RJ, Schoenberg FP, 2018. A Non-parametric Hawkes Model of the Spread of Ebola in West Africa. 2 December. Available at: University of California, Los Angeles. <http://www.stat.ucla.edu/~frederic/papers/chaffeepark107.pdf>.
- Schoenberg FP, 2004. Testing separability in spatial-temporal marked point processes. *Biometrics* 60 (2), 471–481. 10.1111/j.0006-341X.2004.00192.x. [PubMed: 15180673]
- Schoenberg FP, Hoffmann M, Harrigan RJ, 2017. A recursive point process model for infectious diseases. *Arxiv* 23 (March) Available at: <https://arxiv.org/pdf/1703.08202.pdf>.
- Schoenberg FP, Gordon JS, Harrigan R, 2018. Analytic computation of nonparametric Marsan-Lengline estimates for Hawkes point processes. *J Nonparametr Stat.* 30 (3), 742–775.
- Siettos C, Anastassopoulou C, Russo L, Grigoras C, Mylonakis E, 2015. Modeling the 2014 Ebola Virus Epidemic - agent-based simulations, temporal analysis and future predictions for Liberia and Sierra Leone. *PLoS Curr.* 7 10.1371/currents.outbreaks.8d5984114855fc425e699e1a18cdc6c9.
- Valdez LD, Aragão Rêgo HH, Stanley HE, Braunstein LA, 2015. Predicting the extinction of Ebola spreading in Liberia due to mitigation strategies. *Sci. Rep.* 5 (12172). 10.1038/srep12172. Epub 2015/07/20. Central.
- Viboud C, Simonsen L, Chowell G, Vespignani A, 2017. The RAPIDD Ebola forecasting challenge special issue: preface. *Epidemic.* 10.1016/j.epidem.2017.10.003. Epub 2017/10/24..
- World Health Organization Regional Office for Africa. Health topics: Ebola virus disease. Available at: <http://www.afro.who.int/health-topics/ebola-virus-disease>.
- Worden L, Wannier R Hoff NA, Musene K, Selo B, Mossoko M, Okitolonda-Wemakoy E, Muyembe-Tamfum JJ, Rutherford GW, Lietman TM, Rimoin AW, Porco TC, Kelly JD, 2018. Real-time projections of epidemic transmission and estimation of vaccination impact during an Ebola virus disease outbreak in the Eastern region of the Democratic Republic of Congo. *BioRxiv* 5 (November). 10.1101/461285. Available at: <https://www.biorxiv.org/content/biorxiv/early/2018/11/05/461285.full.pdf>.

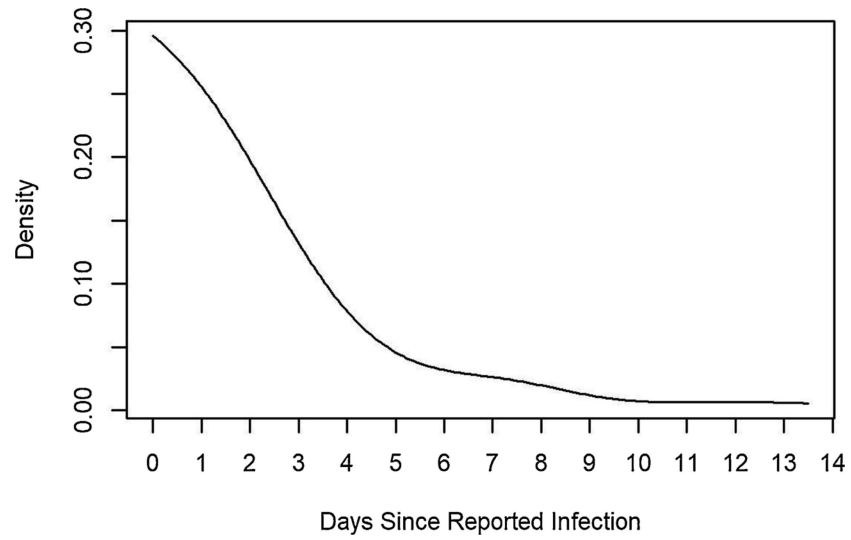


Fig. 1. Fitted triggering density (smoothed) for the 2018–2019 EVD outbreak from May 3, 2018 to June 16, 2019. The x-axis represents days since infection as reported by WHO, where infection day was in some cases estimated based on how long patients were symptomatic.

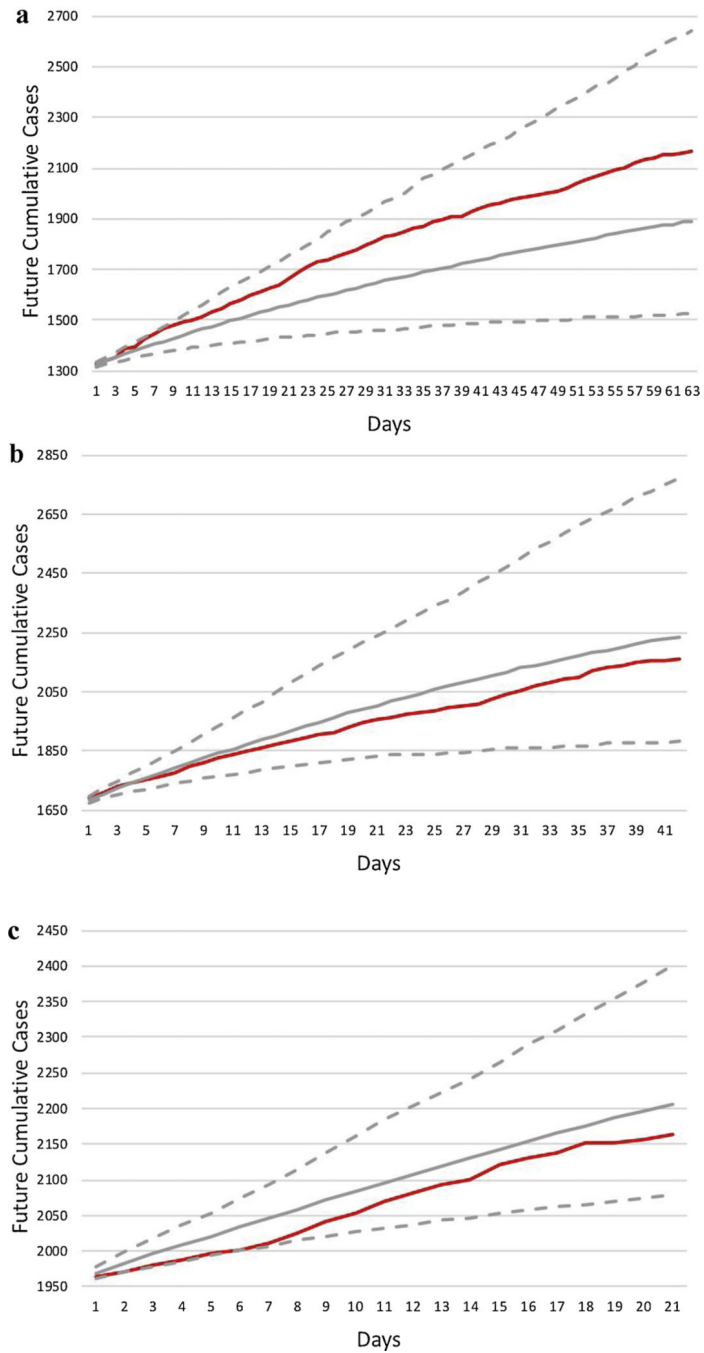


Fig. 2. Median estimate of projected cumulative case counts (grey line) from April 14, 2019 (Fig. 2a), May 5, 2019 (Fig. 2b) and May 26, 2019 (Fig. 2c), all ending June 16, 2019, and the 95% prediction interval (dotted). Actual cumulative case counts are plotted for comparison (red line) but were not known at the time projections were made. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

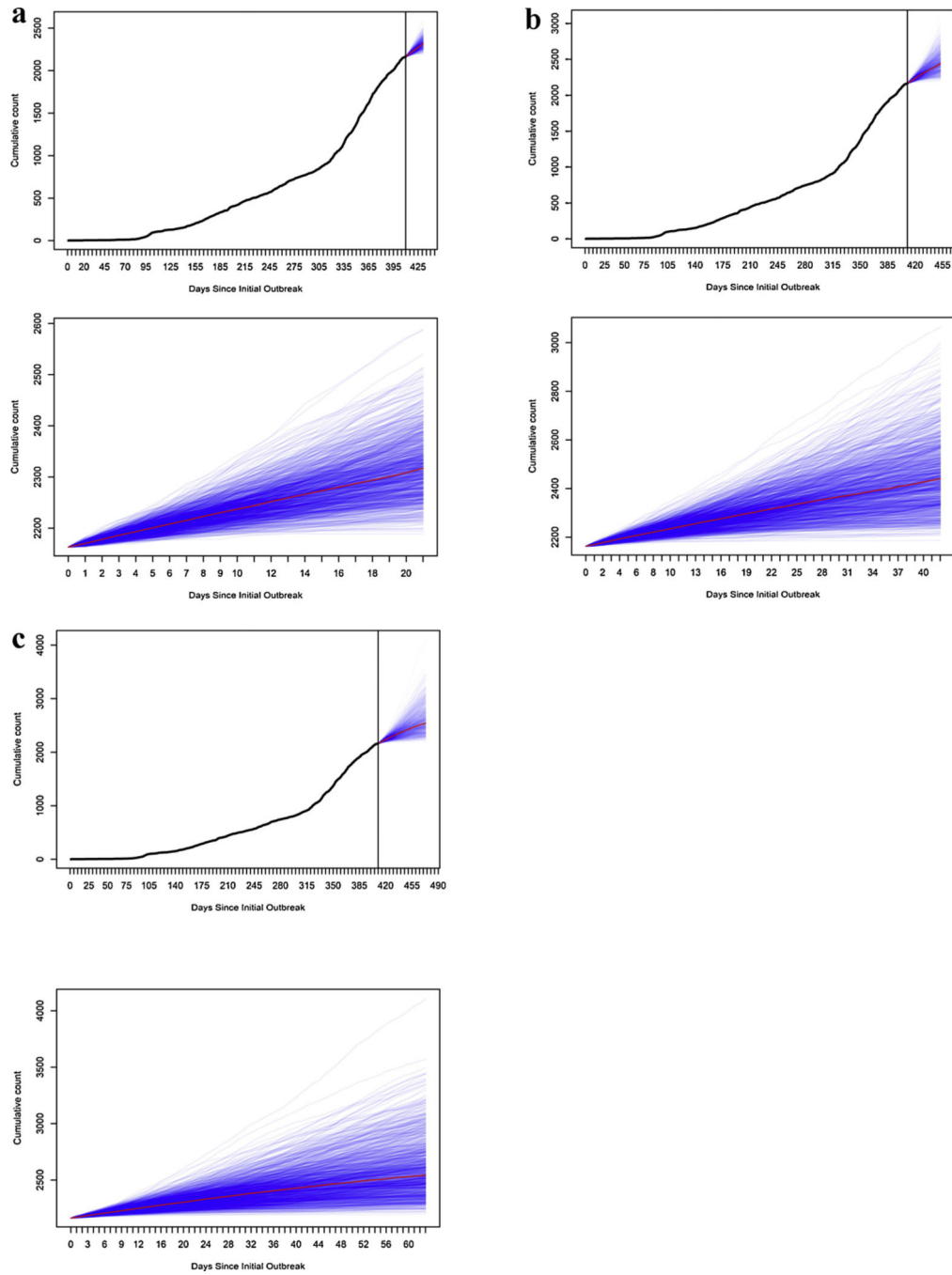


Fig. 3. Epidemic curve, as of June 16 (cutoff at vertical line), followed by three- (Fig. 3a), six- (Fig. 3b), and nine-week (Fig. 3c) probabilistic projections (blue lines) of case counts, using the Hawkes model (median, red line), both with outbreak history and zoomed-in. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)