Real-time projections of Ebola outbreak size and duration with and without vaccine use in Équateur, Democratic Republic of Congo, as of May 27, 2018

Authors
J. Daniel Kelly (a, b)*
Lee Worden (a, b)
Rae Wannier (a)
Nicole A. Hoff (c)
Patrick Mukadi (d)
Cyrus Sinai (c)
Sarah Ackley (a)
Xianyun Chen (e)
Daozhou Gao (e)
Bernice Selo (f)
Mathais Mossoko (f)
Emile Okitolonda-Wemakoy (g)
Eugene T. Richardson (h, i)
George W. Rutherford (a)
Thomas M. Lietman (a, b)
Jean Jacques Muyembe-Tamfum (d)
Anne W. Rimoin (c)
Travis C. Porco (a, b)

Institutions
a. School of Medicine, University of California, San Francisco (UCSF), San Francisco, CA, USA
b. F.I. Proctor Foundation, UCSF, San Francisco, CA, USA
c. School of Public Health, University of California, Los Angeles, Los Angeles, CA, USA
d. National Institute of Biomedical Research, Kinshasa, Democratic Republic of Congo
e. Mathematics and Science College, Shanghai Normal University, Shanghai, China
f. Ministry of Health, Kinshasa, Democratic Republic of Congo
g. School of Public Health, University of Kinshasa, Kinshasa, Democratic Republic of Congo
h. Harvard Medical School, Boston, MA, USA
i. Brigham and Women's Hospital, Boston, MA, USA

Corresponding author:
J. Daniel Kelly, MD, MPH
Department of Medicine – Infectious Diseases
University of California, San Francisco
513 Parnassus Ave., S-384
San Francisco, CA 94143
Phone: (415) 476-6393
Abstract

**Background:** As of May 27, 2018, 54 cases of Ebola virus disease (EVD) were reported in Équateur Province, Democratic Republic of Congo. We used reported case counts and time series from prior outbreaks to estimate the current outbreak size and duration with and without vaccine use.

**Methods:** We modeled Ebola virus transmission using a stochastic branching process model with a negative binomial distribution, using both estimates of reproduction number $R$ declining from supercritical to subcritical derived from past Ebola outbreaks, as well as a particle filtering method to generate a probabilistic projection of the future course of the outbreak conditioned on its reported trajectory to date; modeled using 0%, 44%, and 62% estimates of vaccination coverage. Additionally, we used the time series for 18 prior Ebola outbreaks from 1976 to 2016 to parameterize a regression model predicting the outbreak size from the number of observed cases from April 4 to May 27.

**Results:** With the stochastic transmission model, we projected a median outbreak size of 78 EVD cases (95% credible interval: 52, 125.4), 86 cases (95% credible interval: 53, 174.3), and 91 cases (95% credible interval: 52, 843.5), using 62%, 44%, and 0% estimates of vaccination coverage. With the regression
model, we estimated a median size of 85.0 cases (95% prediction interval: 53.5, 216.6).

Conclusions: This outbreak has the potential to be the largest outbreak in DRC since 2007. Vaccines are projected to limit outbreak size and duration but are only part of prevention, control, and care strategies.

Introduction

On May 8, 2018, the World Health Organization (WHO) announced the occurrence of an outbreak of Ebola virus disease (EVD) in the Democratic Republic of Congo (DRC).\(^1\) From April 4 through May 7, 21 suspected EVD cases were reported in Iboko and Bikoro, Équateur Province. On May 7, blood samples from five hospitalized patients had been sent to Kinshasa for Ebola-PCR testing, and two were confirmed PCR-positive.\(^2\) On May 21, vaccination of healthcare workers started.\(^3\) By May 27, the ring vaccination campaign was being rolled out, and there were 54 suspected, probable and confirmed EVD cases, including four confirmed cases in Mbandaka, the provincial capital of Équateur.\(^4\)

This outbreak has several features that are worrisome for widespread transmission. Cases have been reported over a 168-kilometer distance, including the densely populated Mbandaka, a city of over 700,000 inhabitants situated on the Congo River and bordering Congo-Brazzaville.\(^5\) Moreover, travel to Kinshasa is frequent from Mbandaka. Given these risk factors, early epidemic growth profiles,\(^6\) and evidence of undocumented infection from previous outbreaks,\(^7,8\) the risk of a substantially larger outbreak cannot be ignored.
The factors causing epidemic growth to peak are debated. Delayed detection and resulting widespread distributions of EVD have significantly contributed to epidemic growth. Containment strategies which prioritize isolation over care can be counterproductive by discouraging sick individuals from presenting to high-mortality Ebola units. Unreported cases occurred at higher rates during the early period of the 2013-2016 West Africa outbreak, and urban transmission may also occur at a higher rate than rural transmission. Change to subcritical transmission (reproduction number below 1) tends to occur when Ebola response organizations deploy control, prevention and care measures, communities adopt more protective behaviors, and/or transmission decreases in a social network. In the current Ebola outbreak in DRC, scientific advances with rapid diagnostics and vaccines from the West Africa outbreak has the potential to limit Ebola virus transmission as they are being deployed.

We used reported case counts from the current outbreak and/or time series from 18 prior outbreaks to estimate the total outbreak size and duration with and without the use of vaccines. These projections are intended to help organizations anticipate and allocate sufficient resources to the 2018 Ebola outbreak in DRC.

Methods

We used the following methods to generate projections: a stochastic branching process model, statistical regression based on prior outbreaks, and Gott's Law. On May 27, 54 EVD cases were reported in three locations (Iboko,
Bikoro, and Mbandaka) (Figure 1). Based on EVD situation reports from DRC, we assumed the ring vaccination campaign would start on May 26.

**Stochastic branching process model**

We modeled Ebola virus transmission using a stochastic branching process model, parameterized by transmission rates estimated from the dynamics of prior Ebola outbreaks, and conditioned on agreement with reported case counts from the current outbreak to date. We incorporated high and low estimates of vaccination coverage into this model. Then we generated a set of probabilistic projections of the size and duration of simulated outbreaks in the current setting.

To estimate the reproduction number $R$ as a function of the number of days from the beginning of the outbreak, we included reported cases by date from thirteen prior outbreaks and excluded the first historical outbreak reported in those countries (e.g., 1976 outbreak in Yambuko, DRC) ([Supplement 1](#)). As there is a difference in the Ebola response system as well as community sensitization to EVD following a country's first outbreak, we employed this inclusion criterion to reflect the Ebola response system in DRC during what is now its ninth outbreak. We used the Wallinga-Teunis technique to estimate $R$ for each case and therefore for each reporting date in the outbreak. The serial interval distribution used for this estimation was a gamma distribution with a mean of 14.5 days and a standard deviation of 5 days, with intervals rounded to the nearest whole number of days, consistent with the understanding that the
serial interval of EVD cases ranges from 3 to 36 days with mean 14 to 15 days.

We fit a curve $R = R_0 e^{-\tau d}$ to each outbreak's estimates of $R$ by day $d$, initial reproduction number $R_0$ and quenching rate $\tau$ (Supplement 2).

We modeled transmission using a stochastic branching process model in which the number of secondary cases caused by any given primary case is drawn from a negative binomial distribution whose mean is the reproduction number $R$, and variance is controlled by a dispersion parameter $k$.\textsuperscript{35,36} All transmission events were assumed to be independent. The interval between date of detection of each primary case and that of each of its secondary cases is assumed gamma distributed with mean 14.5 days and standard deviation 5 days, rounded to the nearest whole number of days, as above. The pair of parameters $R_0$ and $\tau$ estimated for the different past outbreaks used, and dispersion parameter $k$, were used in all possible combinations (with $R_0$ and $\tau$ taken as a unit) to simulate outbreaks.

This model generated randomly varying simulated outbreaks with a range of case counts per day. The outbreak was assumed to begin on April 4 with a single case. The simulation process occurs as follows: proposed epidemic trajectories are generated in an initial step based on the above branching process, and these are then subsequently filtered by discarding all but those whose cumulative case counts match the known counts of the current DRC outbreak on known dates (Table 1). The filtration accepts epidemics within a range of 3 cases more or less than each recorded value. This one-step particle filtering technique produced an ensemble of model outbreaks, filtered on
agreement with the recorded trajectory of the outbreak to date. This filtered ensemble is then used to generate projections of the eventual outcome of the outbreak.\textsuperscript{14}

To model vaccination coverage with respect to total transmission (unreported and reported), we multiplied the estimate of vaccine effectiveness by low and high estimates of reported cases. In a ring vaccination study at the end of the West Africa outbreak, the overall estimated rVSV-vectored vaccine efficacy was 100\% and vaccine effectiveness was 64.6\% in protecting all contacts and contacts of contacts from EVD in the randomized clusters, including unvaccinated cluster members.\textsuperscript{22} We designed the stochastic model using vaccine effectiveness. Then, we used past estimates of the proportion of unreported cases to estimate the proportion of exposed individuals not covered by the vaccination process. Based on a Sierra Leonean study from the 2013-2016 outbreak,\textsuperscript{13} we estimated that the proportion of reported cases in DRC would rise over time from a low of 68\% to a high of 96\%. Given these low and high estimates of reported cases and the estimate of vaccine effectiveness, a low estimate of vaccination program coverage was 44\% (68\% * 64.6\%) and a high estimate of vaccination program coverage was 62\% (96\% * 64.6\%). We modeled the course of the outbreak with and without the vaccination program, with the program beginning on May 26, based on dates available from situation reports.\textsuperscript{4}

In the 2013-2016 West Africa outbreak, the mean time from symptom onset to reporting was found to be approximately 4 days.\textsuperscript{37} Based on that estimate, we assumed a four-day period of transmission, with mean time of a
transmission event falling two days before reporting. Based on this, we modeled
transmission from a primary case to a secondary case using the value
of $R$ corresponding to two days before the reporting date of the primary case.

From 122,683,392 simulated outbreaks, 904 were retained after filtering
on approximate agreement with DRC case counts. The simulated outbreaks that
were retained after filtering were continued until they generated no further cases.
We used this ensemble to derive a distribution of outbreak sizes and durations.
We calculated mean and median values and 95% credible intervals using the 2.5
and 97.5 percentiles of simulated outbreak size and duration. We conducted the
analyses using R 3.4.2 (R Foundation for Statistical Computing, Vienna, Austria).

Regression model

For contrast with the stochastic model above, we conducted a simple
regression forecast based solely on outbreaks of size 10 or greater. Time series
for all 18 such prior outbreaks were obtained, including seven prior ones from
DRC, dating back to 1976 (Supplement 1 & 3).\textsuperscript{4,26-33,38-44} No exclusions were
made; no attempt to model specific features of this outbreak was conducted. The
regression model predicted the outbreak size based on values of the outbreak
size at a specific earlier time. The beginning of each outbreak was not reliably
characterized; therefore, all time series were aligned on the day they reached 10
cases. In the current outbreak, we observed cases over the period from April 4 to
May 27 (day 0 to day 53). May 27 corresponded to day 34 since reaching 10
reported cases. For the prior 18 outbreaks, we used linear interpolation to obtain
the number of cases on day 34 (after reaching 10 cases). To reduce the influence of outliers and high leverage points, and to improve linearity, we calculated the pseudologarithm transform \( f(x) = \text{arcsinh}(x/2) \), asymptotically logarithmic but well-behaved at 0. We used non-parametric Theil-Sen regression (R-package \textit{mblm}) followed by calculation of the resulting prediction interval for a new observation.\textsuperscript{45,46} Finally, we reported the median and 95% central coverage intervals for the prediction distribution conditional on the value being no smaller than the observed value on day 34. Sensitivity analysis was conducted using ordinary least squares regression. We conducted these analyses using R 3.4.2 (R Foundation for Statistical Computing, Vienna, Austria).

\textit{Gott’s Law}

We applied Gott’s Law to estimate the outbreak size using data from early in the current outbreak period.\textsuperscript{25} With Gott’s Law, we assume we have no special knowledge of our position on the epidemic curve. If we assume a non-informative uniform prior for the portion of the epidemic that still remains, the resulting probability distribution for the remaining number of cases \( y \) is \( P(Y=y) = \frac{Y_0}{y^2} \). We estimated the median outbreak size, along with the two-sided 95% confidence interval.

\textbf{Results}

As of May 27, 2018, there were 54 EVD cases (\textbf{Table 1, Figure 1}). Bikoko had ten confirmed cases, 11 probable cases, and one suspected case. Iboko had
21 confirmed cases, two probable cases, and one suspected case. Mbandaka had four confirmed cases and one suspected cases. In addition to cases, 906 contacts have been recorded and are actively being monitored. Twenty-five (46%) of 54 EVD cases have died.  

In the absence of any vaccination program, the projected median outbreak size based on the stochastic model was 91 cases (mean 161.0; 95% credible interval: 52, 843.5). Median duration of projected outbreaks was 112 days (mean 126.9; 95% credible interval: 54, 256.5). Using a lower estimate of 44% vaccination coverage, the median outbreak size was 86 cases (mean 91.4; 95% credible interval: 53, 174.3) and median duration was 96 days (mean 100.8; 95% credible interval: 54, 174.1). Using a higher estimate of 62% vaccination coverage, the median size was 78 EVD cases (mean 82.3; 95% credible interval: 52, 125.4), and the median duration was 89 days (mean 91.5; 95% credible interval: 54, 136.4) (Figure 2).

With the regression based on past outbreaks, the median outbreak size was estimated to be 85.0 cases (95% prediction interval: 53.5, 215.6). Removal of data from the 2013-2016 West Africa outbreak from the regression yielded a median size of 71.4 cases (95% central coverage interval for outbreak size distribution: 53.0, 127.1), while use of ordinary least squares produced a median size of 114.4 cases (95% central coverage interval: 54, 433). One (5.6%) of the 18 prior outbreaks in our sample (the 2013-2016 West Africa outbreak) achieved a large size (over 28,000 cases).
Gott's Law suggests that given 54 EVD cases, the median estimate of outbreak size was 108 EVD cases (95% CI: 55, 2160).

Discussion

We estimated that the epidemic curve of this Ebola outbreak has peaked and new EVD cases in Équateur Province are declining at a rate consistent with subcritical transmission. This outbreak still has the potential to become the largest outbreak in DRC since 2007, regardless of vaccine use. Introduction of Ebola virus into Mbandaka with its over 700,000-person population resulted in multiple transmission chains and redoubles the importance of an adequate influx of support for response team efforts to implement interventions that control and prevent the spread of EVD. While real-time modeling projections historically overestimate outbreak size and duration, this study exemplifies how simple mathematical models can be useful for advising real-time decision making because they provide rapid projections and similar estimates of $R$ as compared to complex models.

Vaccine use, regardless of 62% or 44% coverage levels, is projected to provide a limited preventive benefit, which may be partially due to timing of implementation (after predicted peak of the epidemic curve). As such, the massive mobilization of resources to roll it out should not detract from the prioritization of patient and survivor care, lest we recapitulate the control-over-care paradigm that marred the response to the 2013-2016 outbreak in West Africa. Thus far, there has been a strong local and international response, and
deployment of vaccines and rapid diagnostic tests (RDT) occurred early in response efforts. RDTs are being used to screen Ebola suspects while the vaccines are being administered to high-risk groups for EVD, including healthcare workers, contacts, and contacts of contacts. To limit epidemic growth from unreported cases, particularly those who have non-specific symptoms but are screened negative by the WHO case definition, more decentralized use of RDTs should be considered. We found that the impact of vaccines on curbing transmission would have been greater if vaccines were implemented prior to the peak of the epidemic curve (data not shown). Our assumptions about vaccine coverage in the current outbreak may be obviated if evaluations of vaccine use and coverage on transmission reduction are conducted.

There are limitations to our projections. Projection distributions are right-skewed, with long tails (and we therefore report the median instead of the mean). While there have been 22 observed EVD outbreaks with a case count greater than ten cases, we were unable to include all prior outbreaks in our estimates due to data availability. Note that the simple regression projection is based entirely on past outbreaks of Ebola virus disease (measured and reported in different ways), and cannot account for the improved control measures and vaccination in the way that a mechanistic model does. We included, however, as much real-time information into our estimates as possible, but situations such as the introduction of EVD into a large urban population and implementation of RDTs and vaccines are new to DRC. We did not include vaccination of healthcare workers in the stochastic model. Our estimates of vaccination...
effectiveness, reported cases, and mean time from symptom onset to reporting were obtained from West Africa not DRC studies. A strength of our approach was the use of multiple methods to estimate the outbreak size, even though Gott’s Law has not been validated for outbreak projections.

As of May 27, the epidemic curve of the EVD outbreak is thought to be declining, even though EVD transmission entered an urban setting. We do not believe this outbreak will be a repeat of the 2013-2016 West African one. Introduction of vaccines will limit transmission, even with lower levels of vaccination coverage, and DRC has the significant experience of eight prior outbreaks. As vaccine coverage is scaled up, an influx of support is warranted to support and bolster the evolving rapid response and be continued during the post-outbreak period. Stronger surveillance, preparedness, and health systems are needed to detect EVD cases before transmission threatens an even larger populace during future outbreaks.

Acknowledgments

We thank the Ebola responders for their efforts in the current outbreak. We appreciate the Faucett Catalyst Fund for funding Ebola response efforts of the UCLA-DRC research program. TCP and LW are supported by the National Institutes of General Medical Sciences (NIGMS) under Award Number U01-GM087728.

Authors' contributions
J.D.K. contributed to concept and design, analysis, interpretation, drafting, revising, and final approval; L.W. contributed to concept and design, analysis, interpretation, drafting, revising, and final approval; R.W. contributed to acquisition, analysis, interpretation, revising, and final approval; N.A.H. contributed to interpretation, revising, and final approval; P.M. contributed to interpretation, revising, and final approval; C.S. contributed to acquisition, revising, and final approval; S.A. contributed to interpretation, revising, and final approval; X.C. contributed to acquisition, revising, and final approval; D.G. contributed to acquisition, revising, and final approval; E.T.R. contributed to concept and design, interpretation, revising, and final approval; G.W.R. contributed to concept and design, interpretation, revising, and final approval; T.L. contributed to concept and design, analysis, interpretation, revising, and final approval; J.J.M. contributed to interpretation, revising, and final approval; A.W.R. contributed to concept and design, interpretation, revising, and final approval; and T.C.P. contributed to concept and design, analysis, interpretation, revising, and final approval.

**Competing interests:** None

**References**


---

**Table 1.**

Cumulative case counts used in filtering stochastic model.

<table>
<thead>
<tr>
<th>Date</th>
<th>Case Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>May 7 (day 33)</td>
<td>21</td>
</tr>
<tr>
<td>May 11 (day 37)</td>
<td>34</td>
</tr>
<tr>
<td>May 14 (day 40)</td>
<td>41</td>
</tr>
<tr>
<td>May 18 (day 44)</td>
<td>45</td>
</tr>
<tr>
<td>May 20 (day 46)</td>
<td>51</td>
</tr>
<tr>
<td>May 23 (day 49)</td>
<td>52</td>
</tr>
<tr>
<td>May 27 (day 53)</td>
<td>54</td>
</tr>
</tbody>
</table>
Figure 1. Map of cases of Ebola virus disease per health zone in Equateur Province, Democratic Republic of Congo
Figure 2. Distribution of outbreak projections from stochastic branching process model. a. Mean and credible interval of cumulative case count by day in model projections, by proportion of vaccine coverage. Target case numbers and dates are marked with + signs. b. Mean and credible interval of weekly average of per-day case count in model projections. c. Distribution of outbreak size over all model projections. d. Distribution of outbreak duration over all model projections.
d.