Forecasting the Progress of the West African Ebola Epidemic, 2014-2015: A Retrospective Analysis

Keith Ord, Georgetown University
Arthur Getis, San Diego State University
Herman Stekler

Herman was one of the founding fathers of our subject. His forecasting activity in print started in 1959 following on his PhD at MIT. His interests were wide-ranging, from sports to macroeconomics and he never hesitated to evaluate both forecasting successes and failures in these fields. His commitment to research, both his own and through students and colleagues was outstanding. We will all miss him, both for his contributions, but also for his gentle but firm critiques and his warm personality.

[Adapted from an appreciation by Robert Fildes that appeared in Oracle, the newsletter of the IIF]
Aim of the Paper and Outline

• The aim of the paper is to develop a forecasting procedure using such data as was available in real time, with a view to predicting the total size of the epidemic and hence the rate of new arrival of new cases). This objective is important as such predictions drive the deployment of resources.

OUTLINE

• Local geography and the nature of Ebola
• The spatial spread of the epidemic
• Spatial clustering
• Model-building to describe the epidemic
• Data sources
• Data analysis
• Discussion
Map of West Africa

Liberia

Sierra Leone

Guinea
Nature of Ebola

• The probable source was a fruit bat, as they carried the type of Ebola virus that was identified in West Africa during the epidemic. A bat carrying the virus or an animal infected by a bat bite was eaten by a person who then became infected.

• The first case, in December 2013, was diagnosed after the death of a young boy in the rural community of Meliandou in Guinea, which is about 8 kilometers from the regional center of Guéckédou which, in turn, is 6 kilometers from the Liberia border and 16 kilometers from the Sierra Leone border.

• An infected person transmits the virus to those that come in contact with one of the infected’s bodily fluids. The disease is highly contagious and it is likely that all of those in direct contact with a diseased person also became infected unless immune.

• A newly infected individual displays Ebola disease characteristics from two to 21 days after contracting the disease. The initial death rate was over 50 percent of cases before treatment became available.

• In the early stages of the epidemic, as is true for most epidemics, the number of cases was small, but because the entire population is susceptible to the disease, the number of cases increased rapidly in the Spring of 2014, eventually crossing the borders into the nearby countries.
Cumulative number of cases in Weeks 20, 40 and 60 (since onset)
Number of new cases in Weeks 20, 40 and 60 (since onset)
Measuring spatial clustering

• We can assess spatial clustering by means of local averages for each location. For location $i$, with neighboring areas specified by $N(i)$, we calculate

$$G_i^* = \sum_{j \in N(i)} (x_j - \bar{x})$$

• Large positive values of $G$ indicate areas of high activity. For details see Getis and Ord, 1992; Ord and Getis, 1995).
G* statistics for weeks 10, 20, ..., 80 (from onset)
How to forecast the ultimate size of the epidemic?

1. The virus is highly contagious (Lau et al., 2017, suggest most infections were local and spatial spread was due to a small number of index cases or super spreaders.)

2. Spread is governed by the number of active cases in the population and the number of their contacts.

3. Initially, contact with infected cases was considerable. As the transmission process became better understood, cases were isolated, people at risk were quarantined and medical personnel wore protective clothing, so the number of contacts per infective was greatly reduced.

THUS

• Classical notions of uniform mixing of susceptibles and infectives are not appropriate (Kermack and McKendrick, 1927)

• We must allow for the reduction in contacts per infective over time.
Caitlin Rivers, et al. model

\[ S \xrightarrow{\lambda} E \xrightarrow{\alpha} I \xrightarrow{\gamma_H \ell} H \xrightarrow{\gamma_F \delta_2} F \xrightarrow{\gamma_H (1 - \delta_2)} R \]

S = Susceptibles  
E = Exposed  
I = Infecteds  
H = Hospitalized  
F = Funeral  
R = Recovered
Basis for the Model

• Let $E_i$ denote the number of exposures related to the i-th case. As new cases come in they are (conceptually) numbered in order of arrival (confirmation); the exact order is not relevant to this part of the argument.

• We assume that after $m$ time periods no further cases are observed, so the total size of the epidemic ($M$) is

$$M = \sum_{i=1}^{m} E_i$$

• By time $t$, the number of recorded cases will be

$$M(t) = \sum_{i=1}^{m(t)} E_{it}$$
The Logistic Model

• We assume homogeneous mixing among **infectives** and those **exposed** to the virus (and are ultimately infected), leading to the model:

\[
dM(t)/dt = \phi M(t)[M - M(t)]
\]

• In turn, this leads us to a Logistic curve of the form:

\[
M(t) = M/[1 + \exp(-\phi(t - t_0))]
\]

• The parameters can be estimated by nonlinear least squares; time records may not be regular but that is not a problem **(within reason)** for an estimation procedure based upon cumulative numbers.
Data sources

- The study that motivated our work was that of Caitlin Rivers et al. (2014). Dr. Rivers was very generous in allowing us access to the extensive database she had created.
- We focused, as she did, upon Sierra Leone because of data quality. Liberia and Guinea had serious breaks in their series that made analysis very difficult.
- The Sierra Leone government maintained good, almost daily, records throughout the epidemic and these were publicly available. We used data at the regional level.
- We focused on the number of confirmed cases as these records were the most reliable, but it does mean that there could be a considerable time lag between onset and reporting.
- The WHO developed accurate records of onset for each case but these data were only available after the fact. Thus, they were not used for forecasting but serve a valuable role in validation.
Data Analysis

• We considered the Logistic and also a Gompertz model. In the interests of time, only the (better fitting) Logistic is discussed in detail.

• We examined the cumulative curves of numbers of confirmed cases at both regional and national levels at three different times:
  • September 14 – the cutoff considered by Rivers et al. [34 days recorded]
  • December 31 – when the epidemic was well underway but in retrospect was about at the inflection point [142 days]
  • March 31 - when the epidemic had largely run its course [232 days]

• The models used only the data that would have been available at that point in time.

• The plots show the results for the national figures. Grouped regional results were somewhat more patchy but generally OK provided the regions within a group had similar start times.
Cumulative numbers of cases to (A) Sept 14, 2014, (B) Dec 31, 2014, (C) Mar 31, 2015
Actual values and fitted Logistic curves
Cumulative numbers of cases to (A) Sept 14, 2014, (B) Dec 31, 2014, (C) Mar 31, 2015
Actual values and fitted Gompertz curves
Rural 1: Kailahun and Kenema (red)
Urban: Freetown area (blue)
### Numerical results for Logistic

The region **Rural, 1** refers to the combination of two regions close to the Liberia/Guinea borders that were the first to report cases.

The **Urban** wave refers to cases in and around the capital city of Freetown.

**National** includes all cases.

<table>
<thead>
<tr>
<th>Region</th>
<th>End date</th>
<th>Size Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural, 1</td>
<td>09/14/14</td>
<td>1255</td>
</tr>
<tr>
<td></td>
<td>12/31/14</td>
<td>1056</td>
</tr>
<tr>
<td></td>
<td>03/31/15</td>
<td>1059</td>
</tr>
<tr>
<td>Actual</td>
<td></td>
<td>1063</td>
</tr>
<tr>
<td>Urban</td>
<td>09/14/14</td>
<td>598</td>
</tr>
<tr>
<td></td>
<td>12/31/14</td>
<td>3482</td>
</tr>
<tr>
<td></td>
<td>03/31/15</td>
<td>4519</td>
</tr>
<tr>
<td>Actual</td>
<td></td>
<td>5172</td>
</tr>
<tr>
<td>National</td>
<td>09/14/14</td>
<td>9877</td>
</tr>
<tr>
<td></td>
<td>12/31/14</td>
<td>7750</td>
</tr>
<tr>
<td></td>
<td>03/31/15</td>
<td>7706</td>
</tr>
<tr>
<td>Actual</td>
<td></td>
<td>7516</td>
</tr>
</tbody>
</table>
Discussion

• The estimates are generally in the right ballpark, which is what is needed for planning purposes. Original estimates by Rivers et al. and others put the epidemic size at a much higher figure. Authorities built field hospitals that were largely under-utilized.

• The World Health Organization (WHO) subsequently tracked each case and produced a “gold standard” data set of the progression of cases. Analysis of this data set produced similar results to our analysis of the field data, suggesting that reasonable field data can be a sound basis for decision-making.

• By contrast, the records for Liberia and Guinea were too spotty for accurate analysis.

• The growth patterns for different regions tended to be similar, meaning that results and estimates from the ‘lead’ regions could provide useful prior information for later developments in other regions.

• Successful modeling requires some understanding of the transmission process for the particular disease. Nevertheless, contact between infectives and susceptibles is the prime cause, indicating that isolation and quarantine are often the most effective tools. Vaccination to reduce the number of susceptibles will be effective in the longer term.
References [ordk@georgetown.edu]


Postscript: Congo Outbreak

- Initial outbreak in remote rural areas confirmed on May 8, 2018 then some cases in major city of Mbandaka
- Congolese medical response system quick to identify outbreak
- Concerns about major outbreak in urban areas
- Quick response from international community: WHO, CDC, Doctors Without Borders
- Use of vaccine to form buffer rings of contacts and high risk individuals
- Outbreak was declared to be over by July 24, 2018. Total deaths were 33.
- In the last few weeks, reports of another outbreak.
THANK YOU