An IDEA for Short Term Ebola Outbreak Projection: Nearcasting Using the Basic Reproduction Number

Eleyine Zarour

November 26, 2014

Overview

The need for short-term outbreak projection tools

- Early outbreaks are described by limited preliminary data.
  - Short time-scale
  - Noisy and small data (e.g. unreported, imported)
  - Microbiological or serological diagnosis unavailable

  $\rightarrow$ inaccurate SIR parameter estimates, most notably R_0
- But early epidimiological characterization of outbreaks are crucial to inform control policies.

- The effective reproduction number R changes in response to many factors including
  - the depletion of the pool of susceptible individuals due to acquired immunity
  - transmission-reducing behavioral change (as a result of intervention measures and media coverage)
  - antiviral use and vaccination
  - spread to subpopulation types with different transmission rates e.g. influenza initiated among children
  - imported infective cases, most notably via air transport

Example: Imported infective cases in the H1N1 influenza early outbreak

- Germany: 47% of first 198 cases
- Ireland: 84% of first 156 cases
- Spain: 78% of first 98 cases
- Turkey: 77% of first 111 cases
- Western Australia: 54% of first 100 cases being imported from Victoria, Australia

Example: Rapid early decline in estimated R_0 in the 2009 H1N1 outbreak

- Japan: 2 to 1.3
- Mexico from 1.6 to below 1
- Chile from 2.0 to 1.6
- New Zealand from 2.1 to 1.7

The Incidence Decay and Exponential Adjustment (IDEA) Model

- Simple descriptive model
- Input: daily incidence count (I), average serial interval

Serial Interval

Time between symptoms developing in an index case and symptoms developing in a secondary case

- Fit only two-parameters: R_0 and time-dependent dampening factor d

The effective reproduction number R

Number of infectives in a completely susceptible population

$$I = \mathbf{R_0}^t$$

Effective Reproduction number

$$\mathbf{R} = \mathbf{R_0}\frac{S}{N}$$

The Dampening Factor d

Control is modelled empirically in a time-varying manner.

The IDEA Model

$$I = \left[\frac{R_0}{(1+d)^t}\right]^t \tag{1}$$

I: Incident Case Counts, d: Discount Factor

To assess the model's performance, we simulate epidemics using a simple difference equation model under different assumptions about infectiousness and varying orders of control. We then fit the IDEA model by minimizing the root-mean-squared differences (RMSD) between generation-specific case counts by adjusting the R_0 and d parameters.

The classical SIR model

$t =$ single disease generation

$$S_{t+1} = S_t - \mathbf{R} I_t$$

$$I_{t+1} = \mathbf{R} I_t$$

$$R_{t+1} = R_t + I_t$$

$$N = S + I + R$$

Note: **R** is the "Reproductive" number (the average number of successful transmissions per infected person) and R is the number "Removed" individuals (I didn't come up with the notation.)

Expanding the SIR Model: the Difference Equation Model

$\mathbf{R}e_t$ accounts for control activities and dynamic changes in population behavior that may reduce transmissibility of infection.
$t =$ single disease generation

$$S_{t+1} = S_t - \mathbf{R}e_t I_t$$

$$I_{t+1} = \mathbf{R}e_t I_t$$

$$R_{t+1} = R_t + I_t$$

The Control Factor in the Difference Equation Model

$$e_t = RR^{t^n}$$

RR = relative risk of transmission and n = "order" of control

For $n = 0$, $e_t = RR$ and Re is simply reduced by a constant fraction throughout the epidemic.

For $n = 1$, $e_t = RR^t$ and Re is reduced in a manner that accelerates with time.

For $n = 2$, $e_t = RR^{t^2}$ represents accelerated acceleration of control etc.

Other Handy IDEA Metrics

By manipulating equation (1), we can obtain the generation where the number of new cases < 1.

$$t_{max} \geq \frac{\ln R_0}{\ln(1+d)}$$

Other Handy IDEA Metrics

In addition, integrating (1) over t provides an expression to estimate the total outbreak size:

$$I_{total} = \frac{exp\left(\frac{\ln(R_0)^2}{4\ln(1+d)}\sqrt{\frac{\pi}{\ln(1+d)}}\right)}{2}$$
$$\cdot[erf(x-\mu)\sqrt{\ln(1+d)} - erf(-\mu)\sqrt{\ln(1+d)}]$$

Where

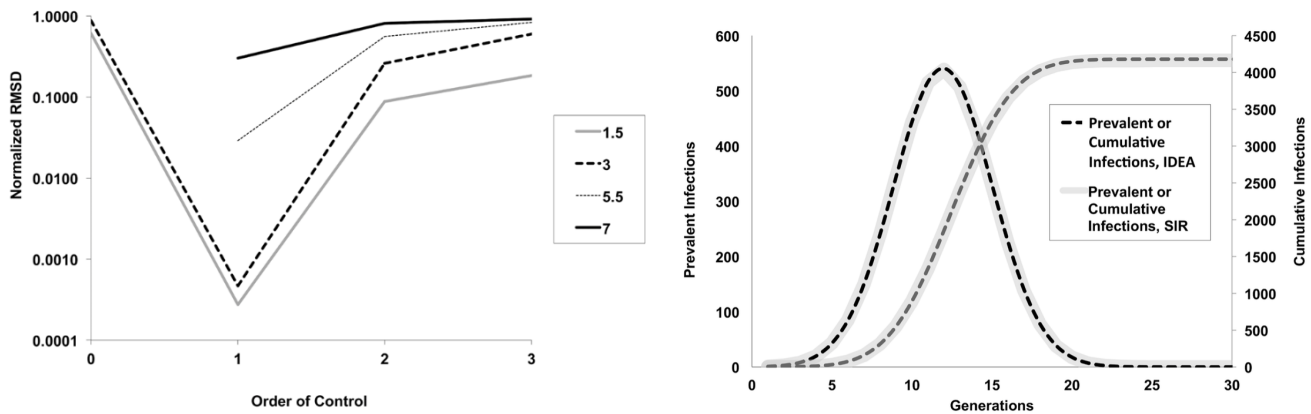
$$\mu = \frac{\ln R_0}{2\ln(1+d)}$$

Caveats

- Recent studies identified new mutations in Ebola generating genetically distinct sequence clades with substantial growth difference between clades.
- Model does not differentiate between locally transmitted cases and imported cases.
- Model is purely descriptive and consequently it is not possible to attribute mechanisms to the decay parameter d.

Best-performig Parameters

Figure : Best fits are achieved with first order control and low R_0



How convenient, Ebola has a low R_0.

Under-reporting

The authors evaluated fits to the SIR model outputs where increasing fractions of cases are unobserved.

Figure : IDEA model fits are stable as long as case reporting fractions exceed 5%

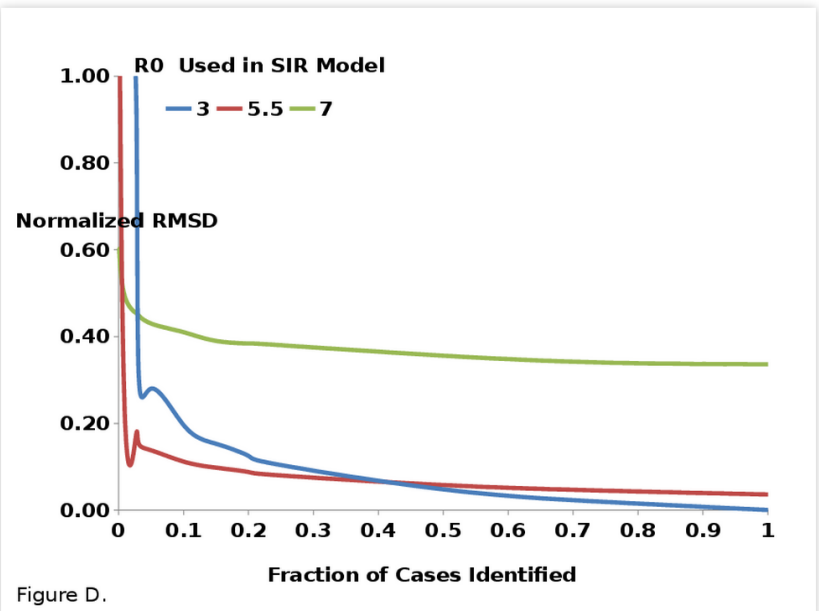


Figure D.

Multi-Wave Epidemics

The structure of the IDEA model made it difficult to fit to multi-wave epidemics. However, an important indicator of the emergence of a new wave of infection was an increasing Δd where:

$$\Delta d = d_i - d_{i-1}$$

Multi-Wave Epidemics

Figure : Epidemic wave onset is characterized by increasing Δd

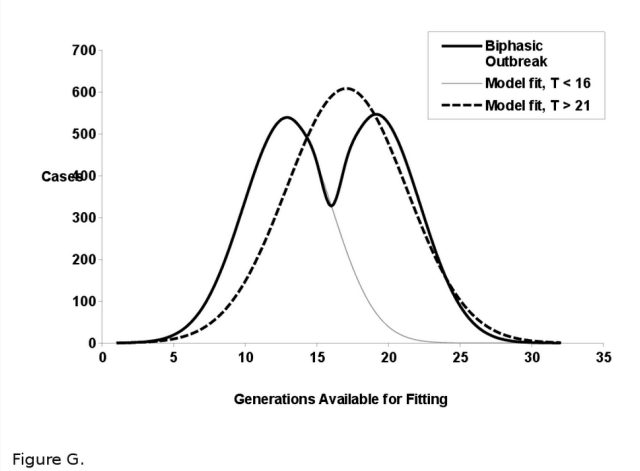


Figure G.

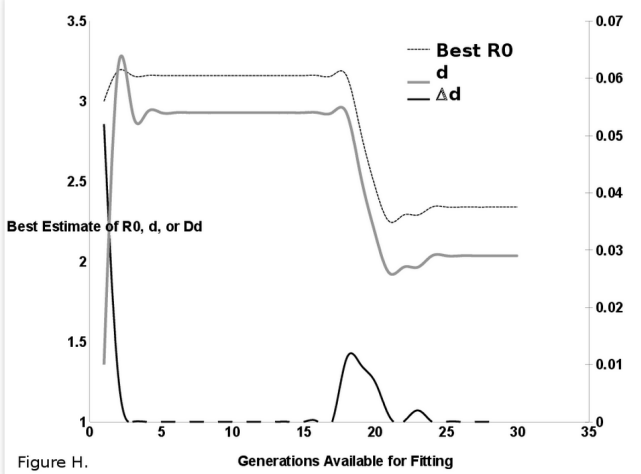


Figure H.

Partial Summary

IDEA model has several attractive properties.

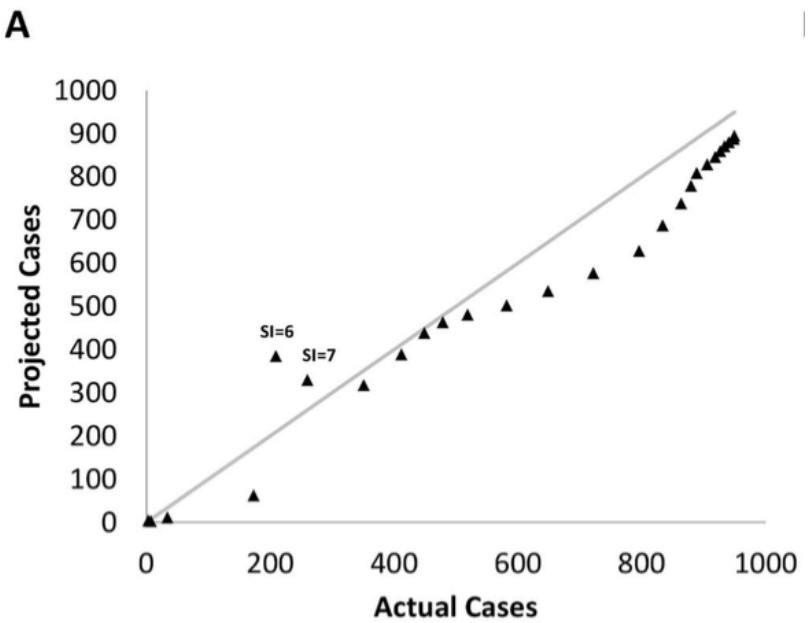
- can be parameterized by fitting to either incidence or cumulative incidence data
- requires no assumptions regarding immune status in the population
- provides reasonably accurate projections about epidemic size and duration (in the absence of change in control efforts) based on pre-peak epidemic data when R is low or moderate
- comparison with simulations suggests model can identify multi-wave epidemics and abrupt changes in control

Principle

- The IDEA model is able to rapidly determine whether the outbreak is growing or stabilizing, based on the change in t_{max} and the change in Δd.
- The IDEA model is able to compare actual versus projected cases as a means of judging whether the outbreak is under control.

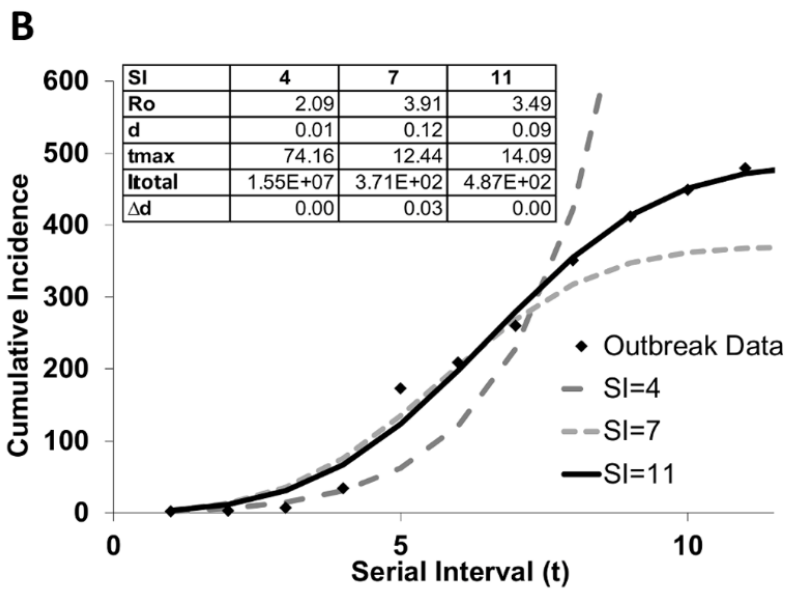
IDEA applied to the Nunavut H1N1 Influenza

Figure : When $y > x$, the model is projecting excess cases, implying that at this snapshot in the outbreak, the current generation had slowed its growth.



IDEA applied to the Nunavut H1N1 Influenza

Figure : Δd can be used to predict the onset of a new epidemic wave.



Ebola Dataset

Open-source WHO data available at `https://github.com/cmrivers/ebola/`.

Figure : Aggregated Case Count Time Series by Country

Date	Day	Cases_Guinea	Cases_Liberia	Cases_SierraLeone	Cases_Nigeria	Cases_Senegal	Cases_UnitedStates	Cases_Spain	Cases_Mali
11/18/2014	241	2047	7082	6190	20	1	4	1	6
11/16/2014	239	1971		6073	20	1	4	1	5
11/15/2014	238		7069						
11/11/2014	234	1919		5586	20	1	4	1	4
11/10/2014	233		6878						
11/9/2014	232	1878		5368	20	1	4	1	1
11/8/2014	231		6822						
11/4/2014	227		6619	4862	20	1	4	1	1
11/3/2014	226	1760							
11/2/2014	225	1731		4759	20	1	4	1	1
10/31/2014	222		6525						
10/29/2014	220	1667		5338	20	1	4	1	1
10/27/2014	218	1906		5235	20	1	4	1	1

Ebola Parameter Ranges

In addition, we can use prior reports on key Ebola descriptors to derive a range for the several parameters:

Serial Interval Heuristic

$$t = \text{incubation} + \frac{1}{2}\text{infective period}$$

$$\begin{cases} \text{incubation period} \approx 13\text{days} \\ \text{infectivity} = [3, 5]\text{days} \end{cases} \rightarrow t \in [12, 18]\text{days}$$

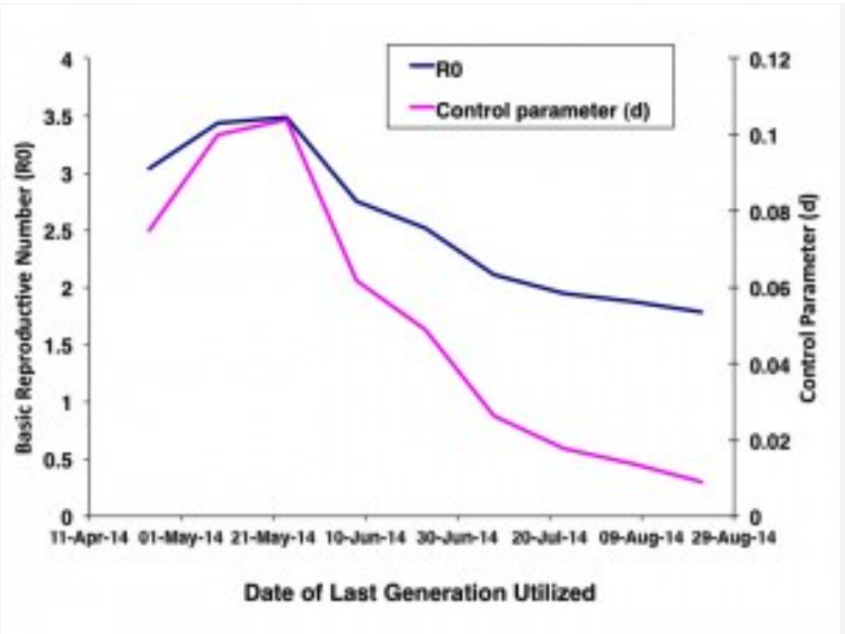
Assumption: Incubation is equivalent to latency for this virus.

Reproductive Number

$$R_0 \in [1.5, 2.7]$$

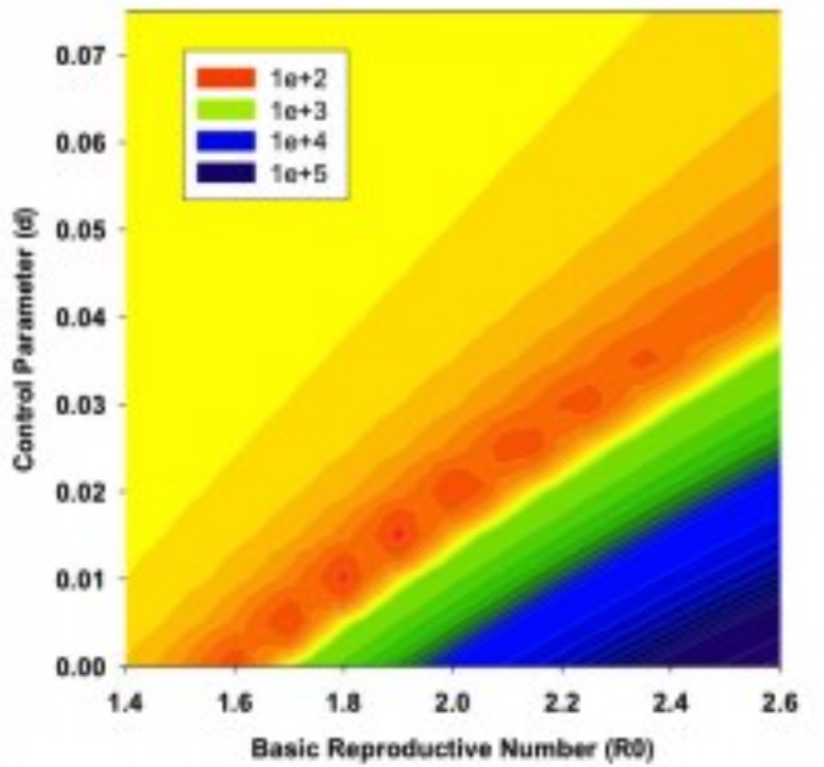
Ebola Nearcasting Results

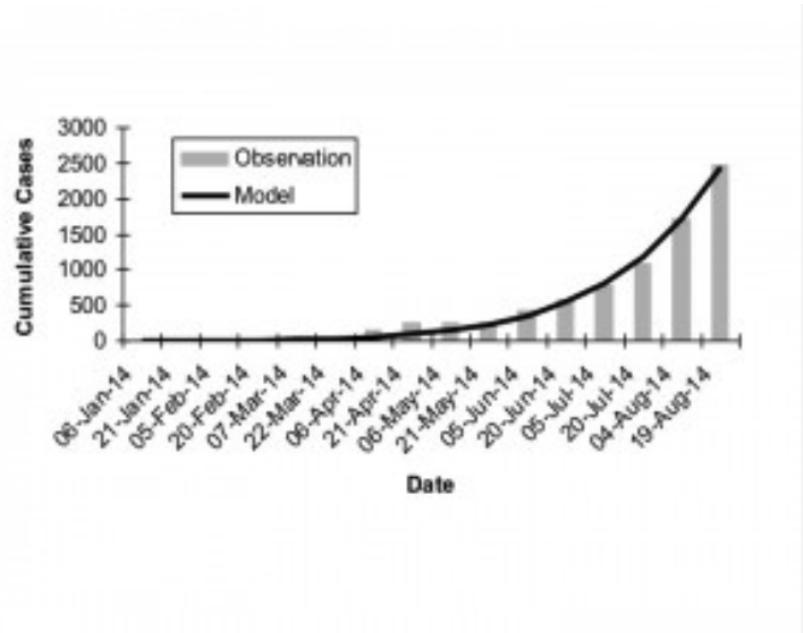
Figure : No abrupt changes in d mean there doesn't seem to be multi-wave outbreaks.



Ebola Nearcasting Results

Figure : RMSD is lowest, by an order of magnitude, for R values close to 1.8, and d values close to 0.01





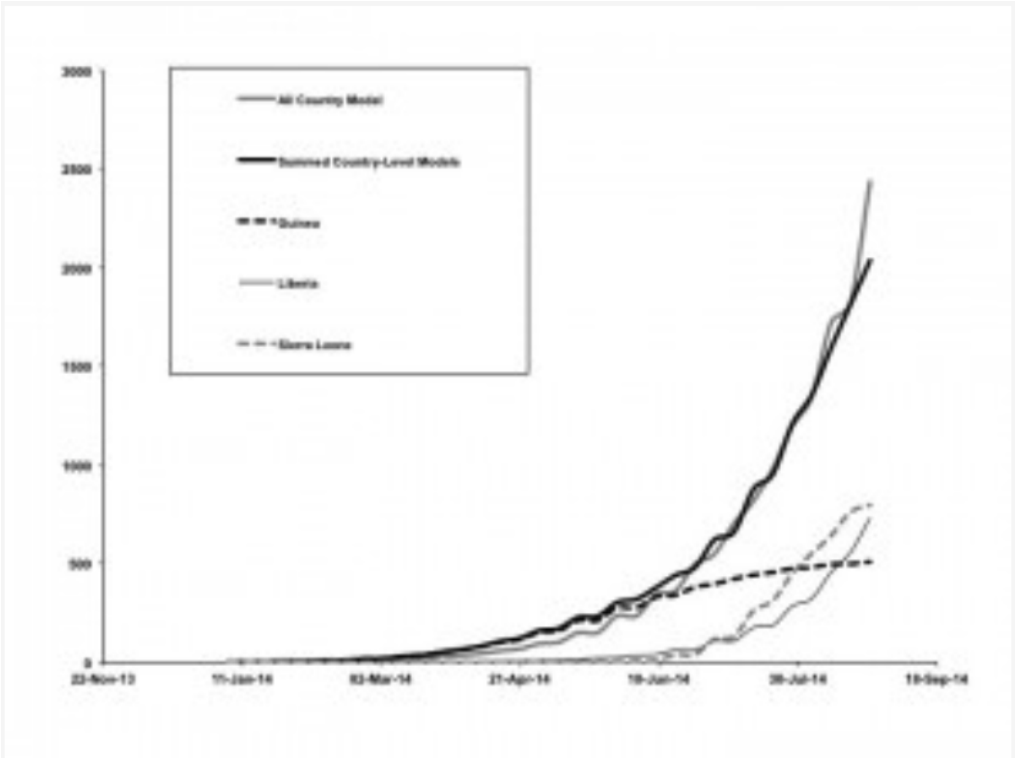
Ebola Nearcasting Results

Figure : None of the different case analyses provide estimates of R_0 and d that differ markedly from those dervied in the base case.

Alternate Assumption	R0	d
Base case	1.78	0.009
12 day generation time	1.68	0.009
18 day generation time	1.94	0.013
Outbreak recognized generation 3	2.19	0.022
Outbreak recognized generation 7	1.70	0.011
Outbreak 50% under-reported	1.92	0.013
Outbreak 100% under-reported	2.02	0.015
Virologically confirmed cases only	1.74	0.011
Deaths only	1.66	0.008
Guinea cases only	2.46	0.050
Liberia cases only	1.72	0
Sierra Leone cases only	8.33	0.22

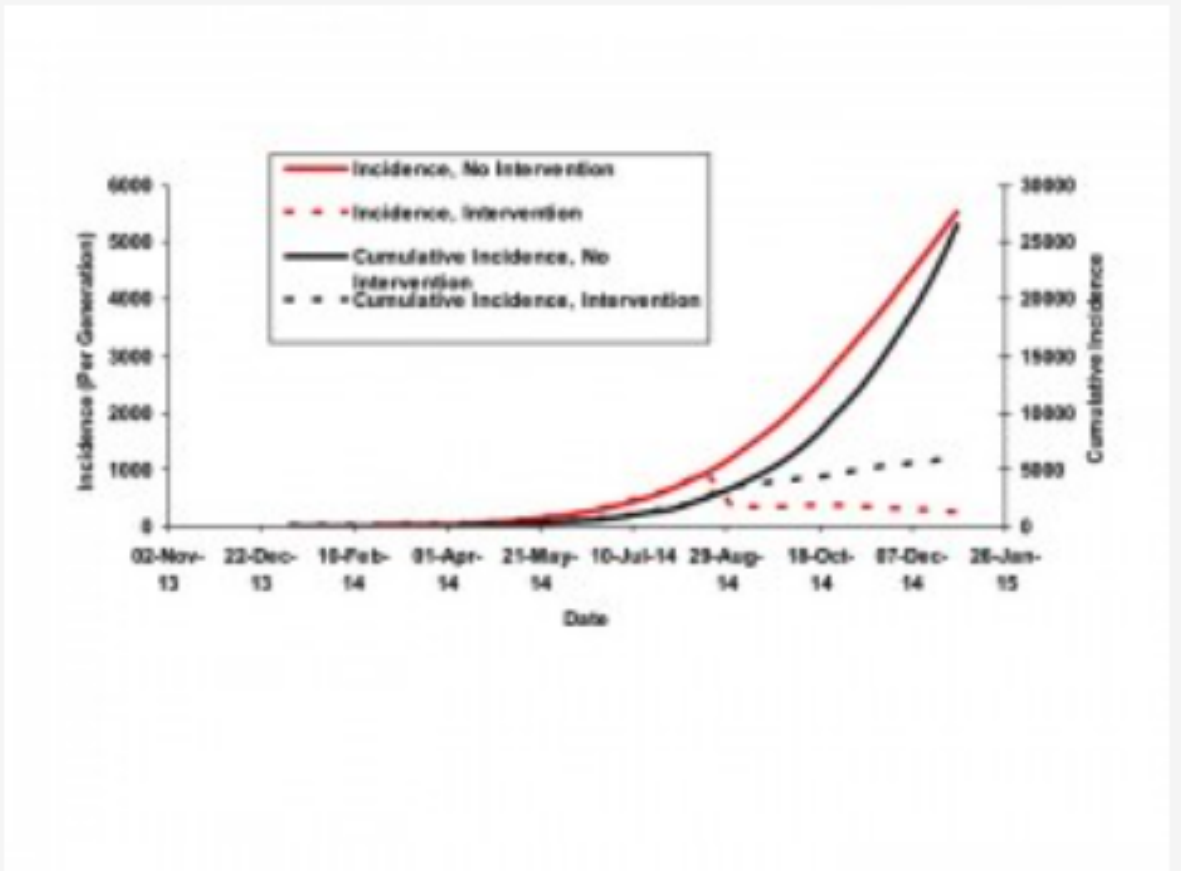
Ebola Nearcasting Results

Figure : Summing individual country curves reproduced the overall epidemic curve well.



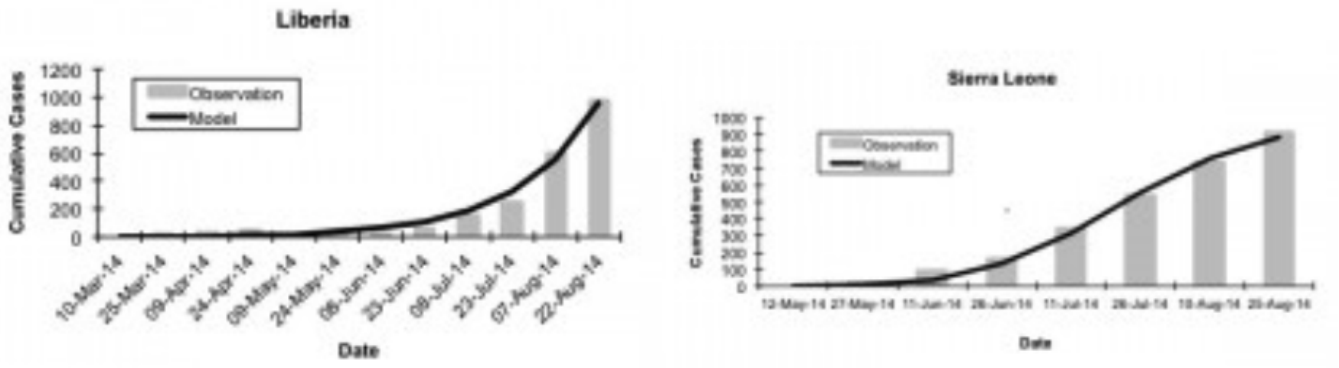
Ebola Nearcasting Results

Figure : The effect of intervention on incidence



Ebola Nearcasting Results

Figure : Country-specific Model fits: Liberia shows little evidence for slowing of transmission.



Project Directions

The authors have utilized epidemic time series up till August 22, 2014.
In addition to cross-checking their results by implementing their method independently, I will also carry out an analysis of the most current data.

References

1. Fisman DN, Hauck TS, Tuite AR, Greer AL. An IDEA for short term outbreak projection: nearcasting using the basic reproduction number. PLoS One. 2013;8(12):e83622. PubMed PMID:24391797.
2. Fisman, David, Edwin Khoo, and Ashleigh Tuite. Early epidemic dynamics of the West African 2014 Ebola outbreak: estimates derived with a simple two-parameter model. Plos Currents Outbreaks (2014).
3. Mercer, G. N., Glass, K., & Becker, N. G. (2011). Effective reproduction numbers are commonly overestimated early in a disease outbreak. Statistics in Medicine, 30(9), 98494. doi:10.1002/sim.4174
4. Ridenhour, B., Kowalik, J. M., & Shay, D. K. (2014). Unraveling R0: considerations for public health applications. American Journal of Public Health, 104(2), e3241. doi:10.2105/AJPH.2013.301704
5. Gire SK, Goba A, Andersen KG, Sealfon RS, Park DJ, et al. (2014) Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. Science 345: 13691372
6. Arino, J., & Driessche, P. Van Den. (n.d.). The Basic Reproduction Number in a Multi-city Compartmental Epidemic Model, 135142.
7. Luksza, M., Bedford, T., & Lassig, M. (2014). Epidemiological and evolutionary analysis of the 2014 Ebola virus outbreak, (June), 119. Populations and Evolution. Retrieved from http://arxiv.org/abs/1411.1722
8. Hufnagel, L., Brockmann, D., & Geisel, T. (2004). Forecast and control of epidemics in a globalized world, (Track II).

Thanks! Questions?