

# Challenges and opportunities in outbreak analytics

## Diphtheria outbreak among Rohingya refugees, Bangladesh 2017-18

Jonathan Polonsky

Epidemiologist

World Health Organisation Emergencies Programme

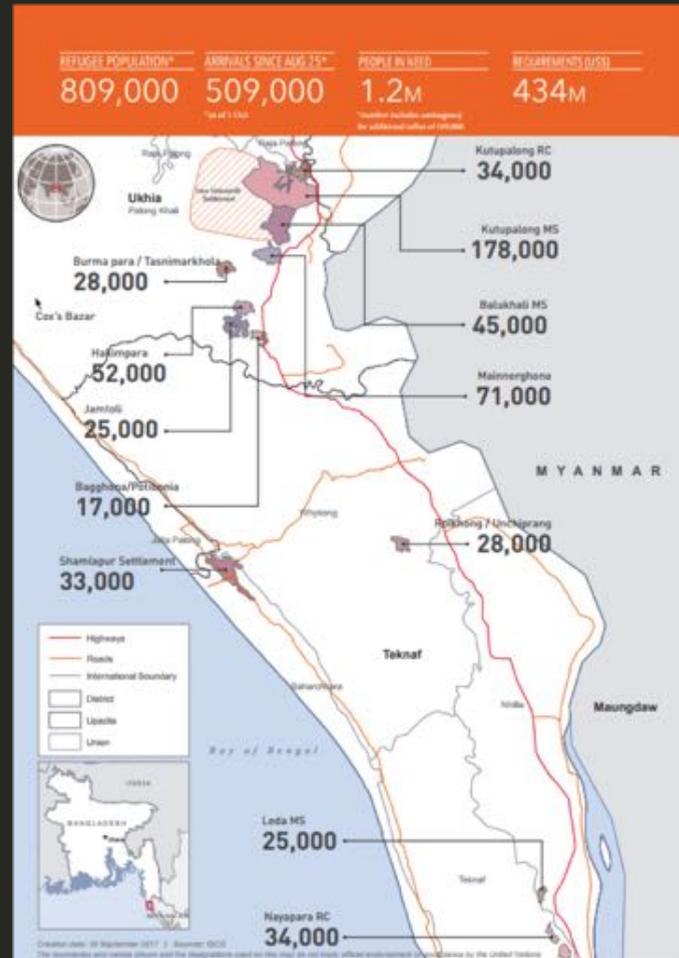
22 March 2018

# Background

# Context

- Renewal of inter-communal violence in Rakhine state, Myanmar, in August 2017
- Influx of > 650,000 new refugees into Bangladesh
- High risk of epidemic-prone diseases:
  - Very poor living conditions
    - Cramped, spontaneous settlements
    - High levels of faecally-contaminated drinking water (>80%)
  - Chronically underserved population
    - Poor access to health services
      - Low vaccination coverage
      - Emergency services, e.g. EmOC
    - Chronic and acute malnutrition

# Context



# Context



# Initial Terms of Reference

- Disease surveillance
  - Strengthen Early Warning Alert and Response System (EWARS)
- Preparedness
  - Assist preparedness efforts for AWD and other likely epidemic-prone diseases
- *Ad hoc* outbreak response (measles, cholera, AJS, ...)

# Diphtheria outbreak

- Rare outbreaks (recently several within short timeframe)
  - Yemen
  - Indonesia
  - Venezuela
- Ongoing for ~ 4 weeks, but became priority early Dec
- Immediate response activities:
  - Generate automated report using linelist data
  - Establish contact tracing
  - Forecasting requirements
    - Staffing & bed capacity
    - Chemoprophylaxis & DAT
  - Vaccination campaign

# Diphtheria outbreak

## Brief epidemiological description

### Case definition:

- Confirmed: +ive toxigenic *C. diphtheriae* strain by multiplex assay
- Probable: upper respiratory tract illness w/ laryngitis/nasopharyngitis/tonsillitis AND sore throat/difficulty swallowing and an adherent membrane/pseudomembrane OR gross cervical lymphadenopathy
- Suspected: any case with a clinical suspicion of diphtheria. Includes case-patients that are *unclassified due to missing values*

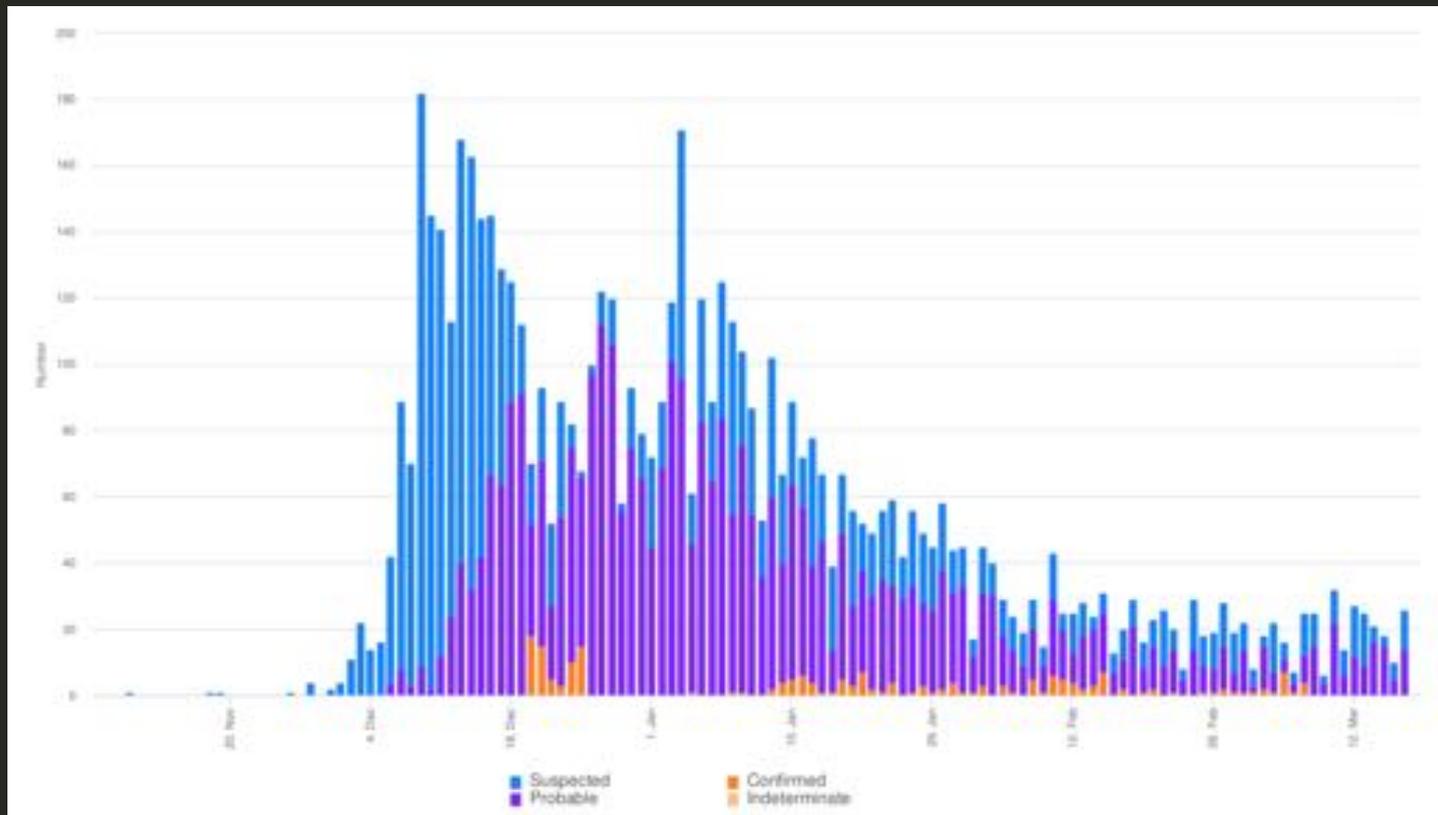
# Diphtheria outbreak

## Brief epidemiological description

Indicator	Cumulative <sup>1</sup>
<b>Epidemiology</b>	
<b>1.</b> Total number of cases <sup>2</sup>	6,250
Confirmed	3.1% (n= 195 )
Probable	52.9% (n= 3,306 )
Suspected	44.0% (n= 2,749 )
Indeterminate	0.0% (n= 0 )
<b>2.</b> Laboratory testing	
Positive	32.7% (n= 195 )
Indeterminate	0.0% (n= 0 )
Negative (excluded)	67.3% (n= 401 )
<b>3.</b> Total number of deaths	39
<b>4.</b> Attack rate <sup>3</sup> / 10,000 population	90.93

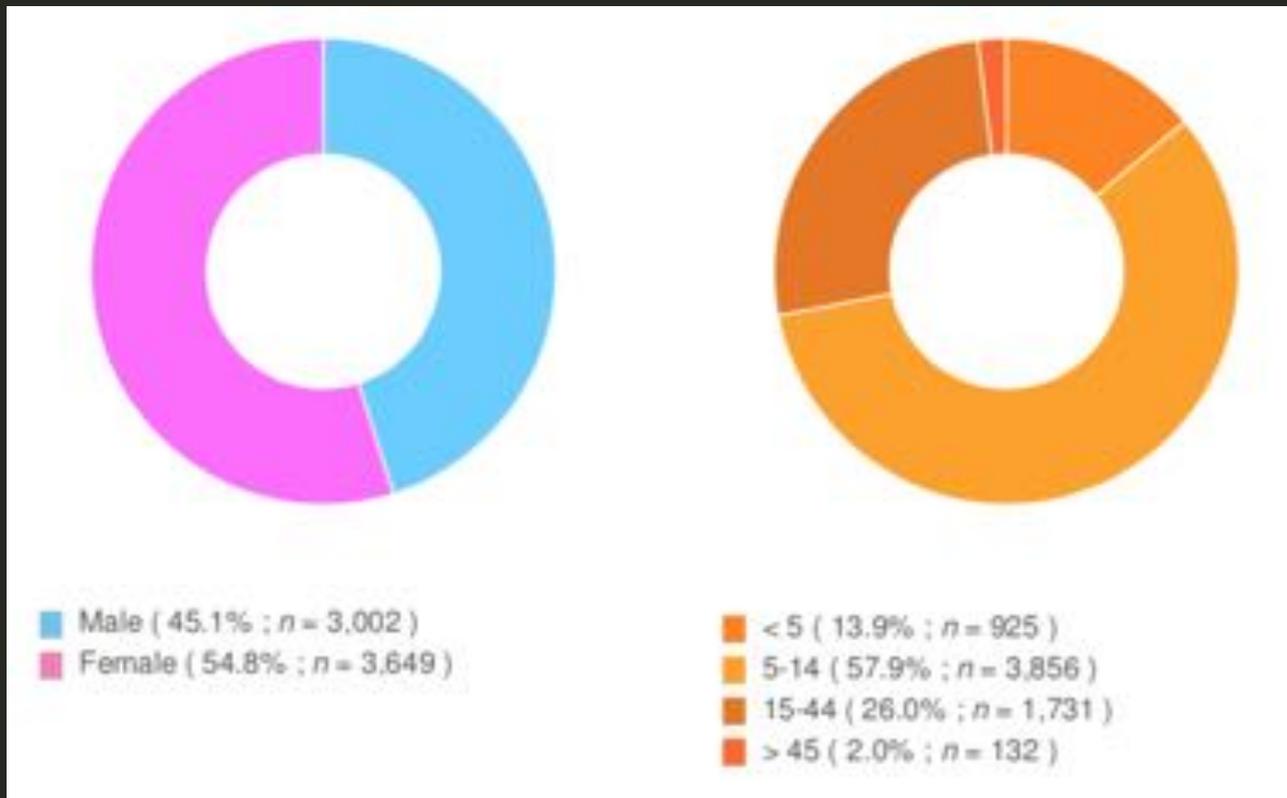
# Diphtheria outbreak

## Brief epidemiological description



# Diphtheria outbreak

## Brief epidemiological description



# Challenges

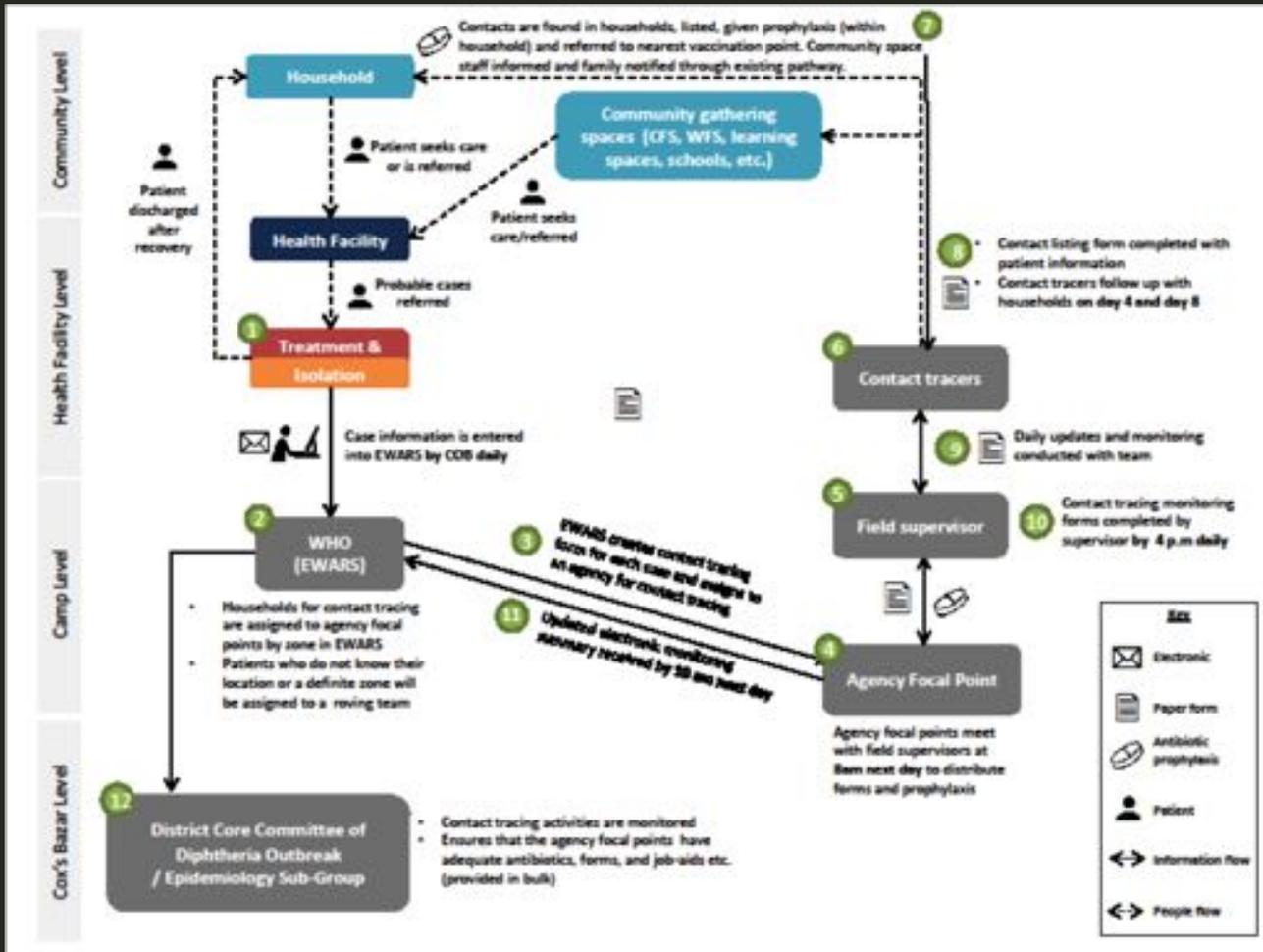
# Data management

- Information flow:
  - Paper forms filled at HF (initially 1, expanded to 7)
  - Collected daily by epidemiologists and entered into Excel 🤖
  - Nightly reporting to centralised email
  - Processed in R
- Issues:
  - Delays (poor internet, heavy file)
  - Incomplete data (staff capacity overwhelmed, difficulty updating records)
  - Management of Excel workbooks
    - Multiple Tx facilities
    - Multiple formats
  - Linelist subject to change
    - Evolving CD
    - # variables to include (>100!) 🤖
    - Updated forms - logistics & training

# Contact tracing monitoring

- Contacts: all people living under same roof
- Traced, given chemoprophylaxis, and followed up for 7 days
- No linelisting of contacts
  - Aggregated at HH level (total # contacts identified for follow up)
- Complicated system of data flow b/n facilities, WHO (system coordination), and CT partners

# Contact tracing monitoring

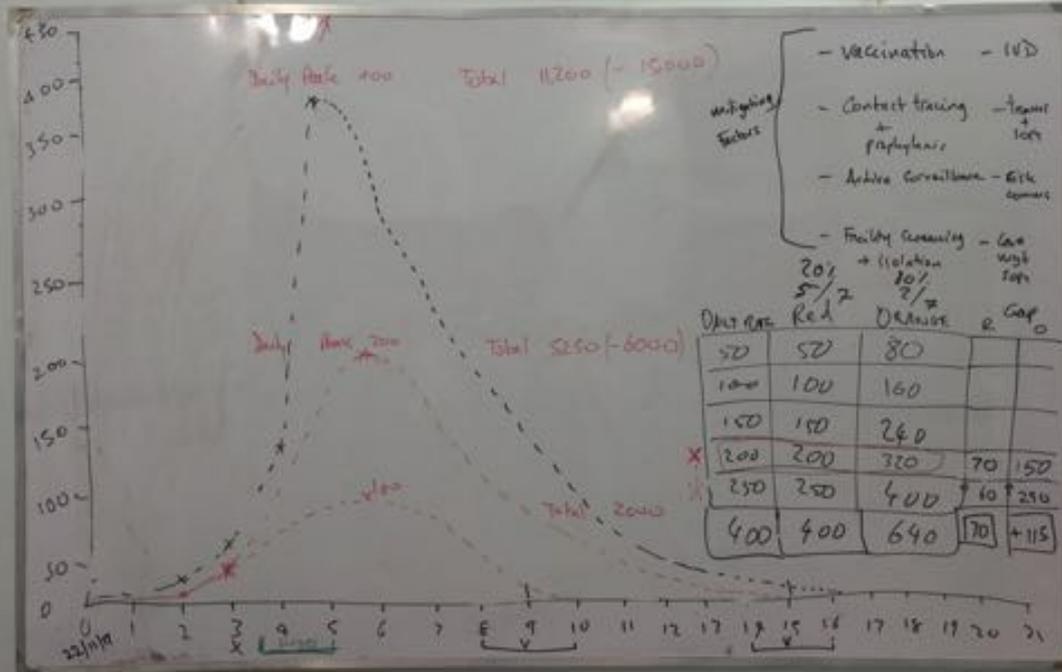


# Forecasting for operational needs purposes

- Bed capacity (engagement with GOARN partners and EMTs)
- Diphtheria Antitoxin (limited global supply)
- Antibiotics (treatment + contact tracing)



# Forecasting for operational needs purposes



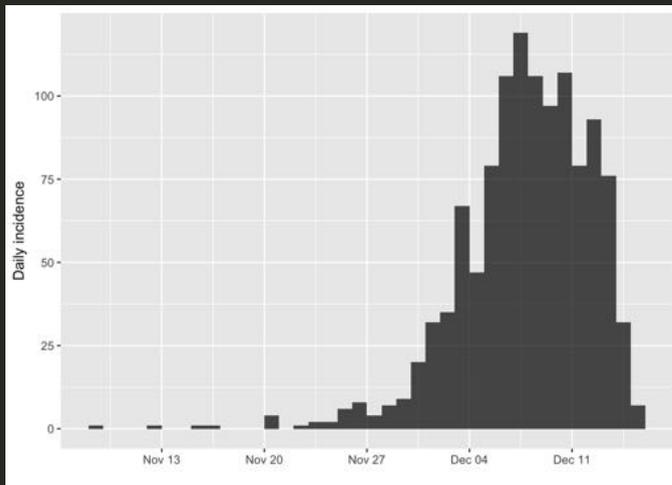
# Opportunities

(from the reconverse™)

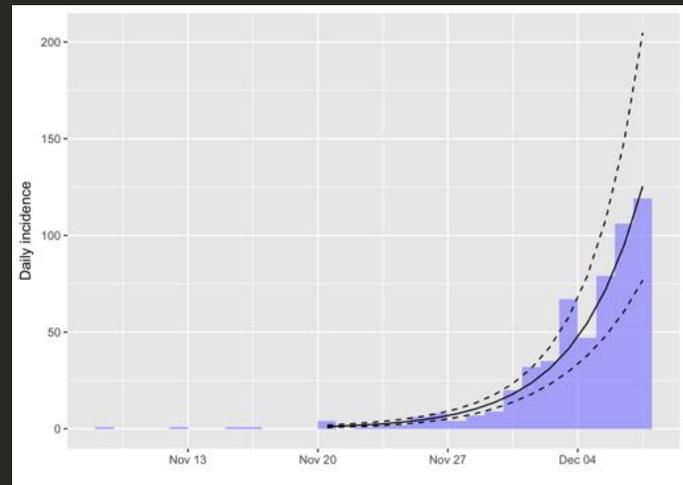


## 2. reconhub::incidence

```
i <- incidence(df$date_onset)
plot(i)
```



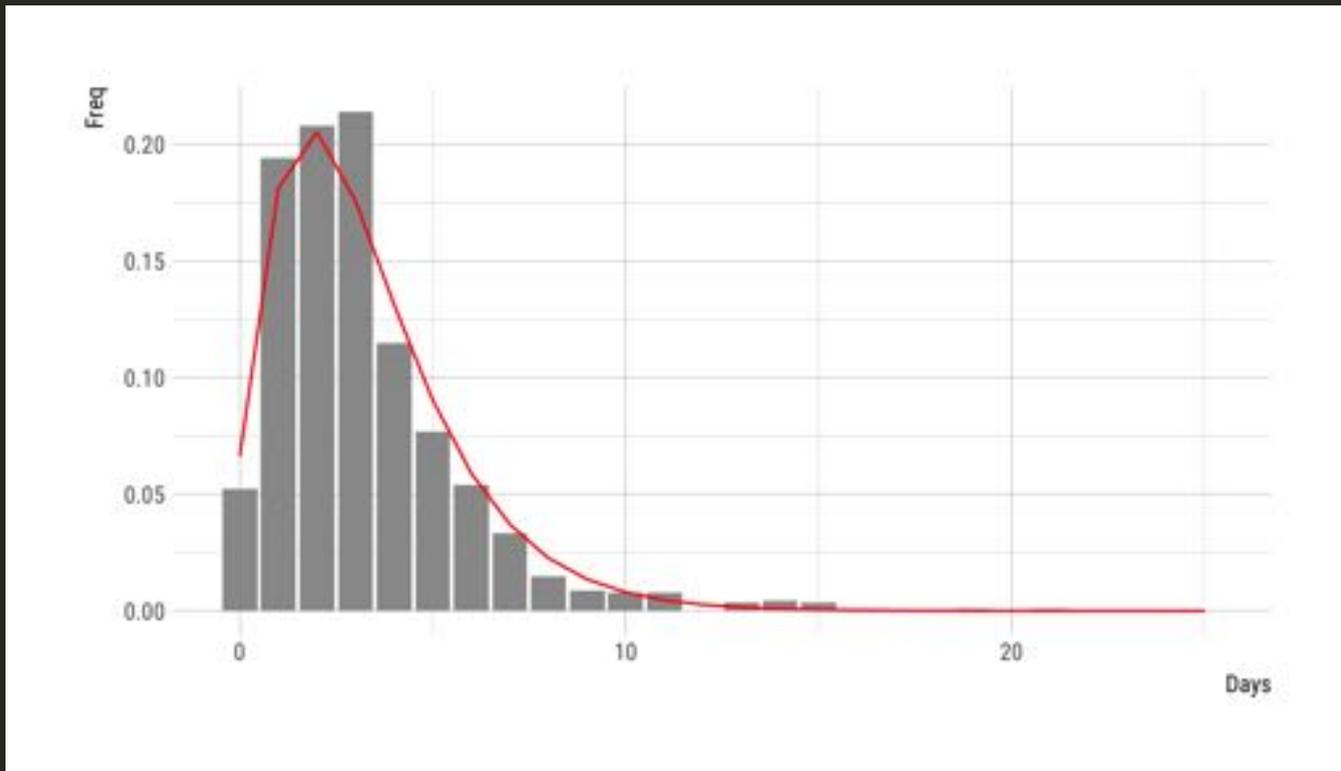
```
f_up <- fit(i[12:30])
plot(i[1:30], fit = f_up, color =
```



- Log-linear model of incidence on the growing phase of the outbreak for incidence prediction and estimating doubling time
- The doubling time over this early period was 2.5 days [95% CI 2.1: 3.1 days].

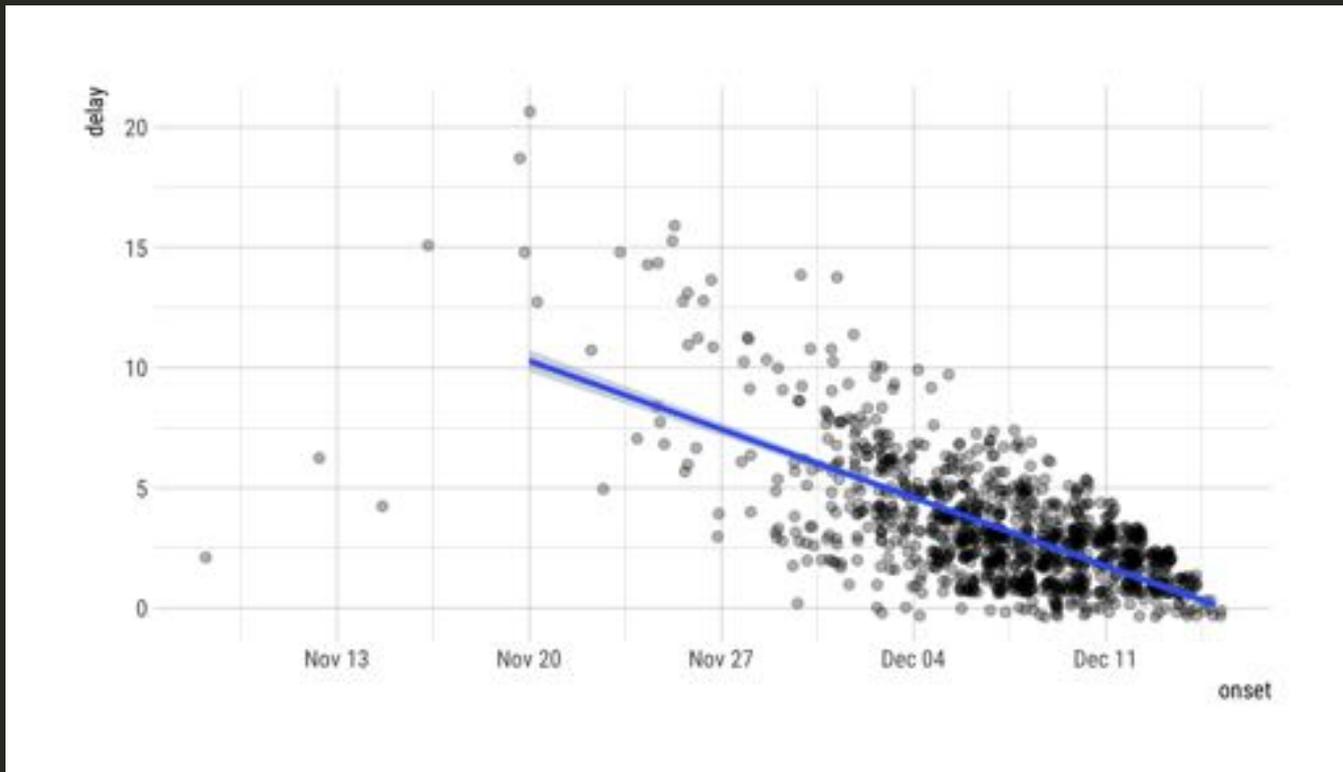
### 3. reconhub::epitrix

```
delay <- as.integer(df$date_seen - df$date_onset)
delay_dist <- epitrix::fit_disc_gamma(delay)
delay_freq <- sapply(0:25, function(i) sum(i == delay))
delay_freq <- delay_freq / sum(delay_freq)
```



### 3. reconhub::epitrix

```
df %>%  
  ggplot(aes(x = onset, y = delay)) +  
  geom_jitter(alpha = .3) +  
  geom_smooth(method = "lm")
```



# 4. reconhub::EpiEstim

## EpiEstim App

This web application generates an estimate of infectious disease transmissibility throughout an outbreak. The time-dependent reproduction number ( $R_t$ ) is inferred from disease incidence time series and patient data or estimates of the serial interval. For more information about how to use this application, please see the [wiki](#). For the most part, we assume that users will want to upload their own data, however there are also some [example datasets](#) built in, to illustrate how the application might be used.

For more information on uploading your own data and on the required format of the data, please [click here](#).

The authors request users to cite the original publication when referring to this tool, any results generated from it, or the R software application on which this tool is based (EpiEstim 2):

Thompson RN, Stockwin JE, van Gaalen RD, Polinsky JA, et al. Improved inference of time-varying reproduction numbers during infectious disease outbreaks. Submitted (2018).

Ready

Step 7 of at least 9. View the [interactive documentation](#) for this state.

### Serial Interval (SI) Input

n1

Choose the number of serial interval distributions (n1) to be sampled (see parameters below)

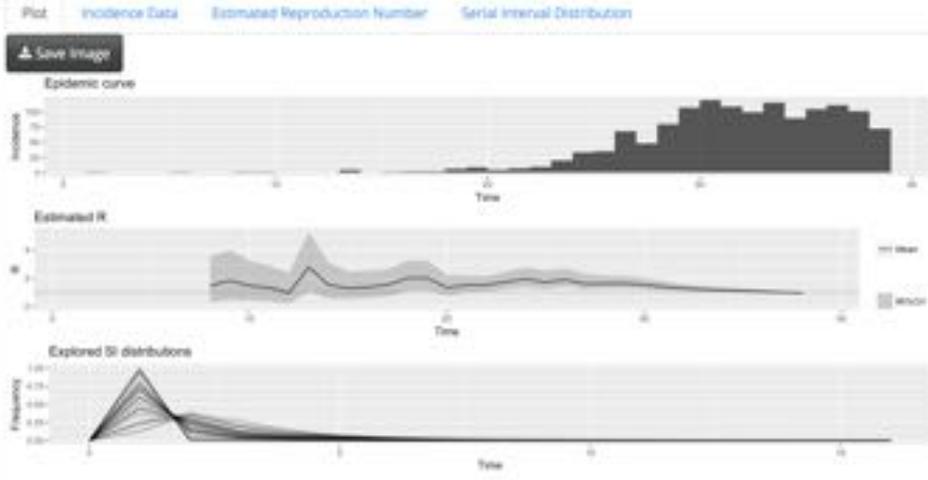
n2

Choose the number of serial interval values (n2) to be sampled from each of the n1 serial interval distributions in order to estimate  $R_t$ .

Choose values describing the mean serial interval distribution, and the uncertainty around these values.

mean\_si

std\_mean\_si

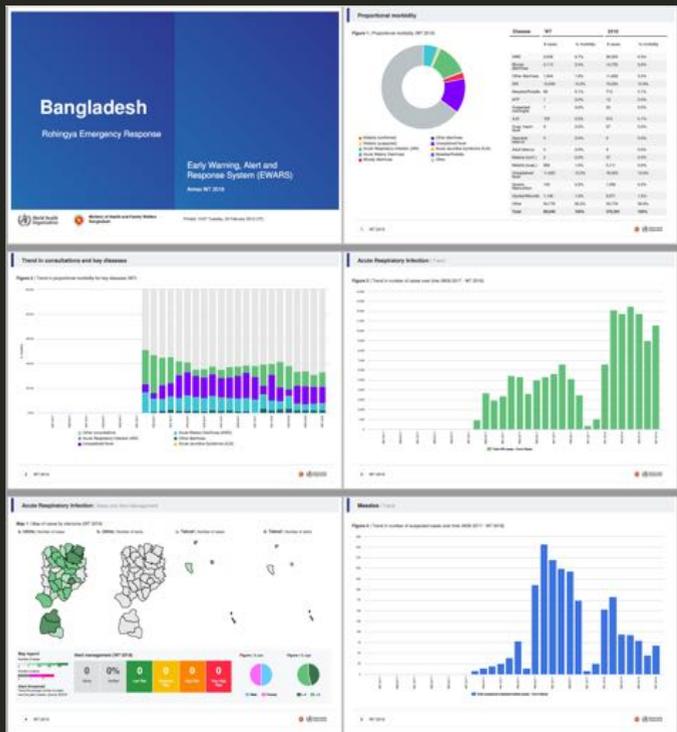


# Non-*R*-based opportunities

# EWARS-in-a-box

- WHO-led project
- System for:
  - Early warning
    - Weekly aggregated reporting from HFs
  - Alert management
    - Triggering & investigating alerts
    - Verification & risk assessment
    - Rapid response teams
  - Outbreak response
    - Linelisting of cases during active outbreaks
  - Automated analysis and reporting
- Managed by administrators at various levels in the country, e.g. MoH, GOARN deployees...

# EWARS-in-a-box



EWARS report



Diphtheria report

# GOData

- WHO-led project
- System for linelisting and linking cases and contacts for relevant epidemic-prone diseases
  - Diphtheria
  - Ebola
- Under development (missed opportunity?)
  - Possibly to integrate R & `htmlwidgets`

# Remaining challenges

# Cross-cutting issues

- Reproducibility:
  - Proliferation of tools & *ad hoc* solutions that do not fit field requirements
  - R pkgs:
    - Basic epi analyses trivially easy, *but* only if you know R
    - Tendency to break following handovers
    - *Shiny* w/ "remote control" is one option (but, again internet!)
- Continuing need for tools for simplifying/reducing redundancy in our work
  - Preferably open source w/ wide user base & support network
  - Designed with *front-line users* in mind
  - Introducing (yet!) more systems (resources- training, time, etc)
- Field constraints: Limited staff capacity & IT challenges

# Data management

- Still doing data management (and some analysis!) in Excel

Spreadsheets: a dystopian moonscape of unrecorded user actions

– *Gordon Shotwell*

- Editing data at peripheral levels
  - Adding outcome, updating signs & symptoms, ...
  - Mobile, offline functionality, w/ lightweight data synchronisation

# Data analysis

- Research vs. field epi perspective - don't always speak same language
  - field conditions (logistical and other constraints)
  - technical abilities (modelling)
- Do **more** with the data
  - Data quality issues
  - Accurate modelling of outbreak for predictive purposes:
    - Enormous potential for real-time operational planning *during* outbreaks
    - Confusion over methods, assumptions, parameters, uncertainty, etc
    - Difficult for non-specialists to grasp - interpretation & utility

# Acknowledgements

- Thibaut Jombart & Rich Fitzjohn (*incidence, epitrix*)
- Anne Cori et al (*EpiEstim*)
- Robin Thompson, Jake Stockwin et al (*EpiEstimApp*)
- WHO/GOARN field epi teams in CXB (past and present)
- MSF, IOM, & Samaritan's Purse for treating patients and sharing data
- Various NGOs involved in contact tracing