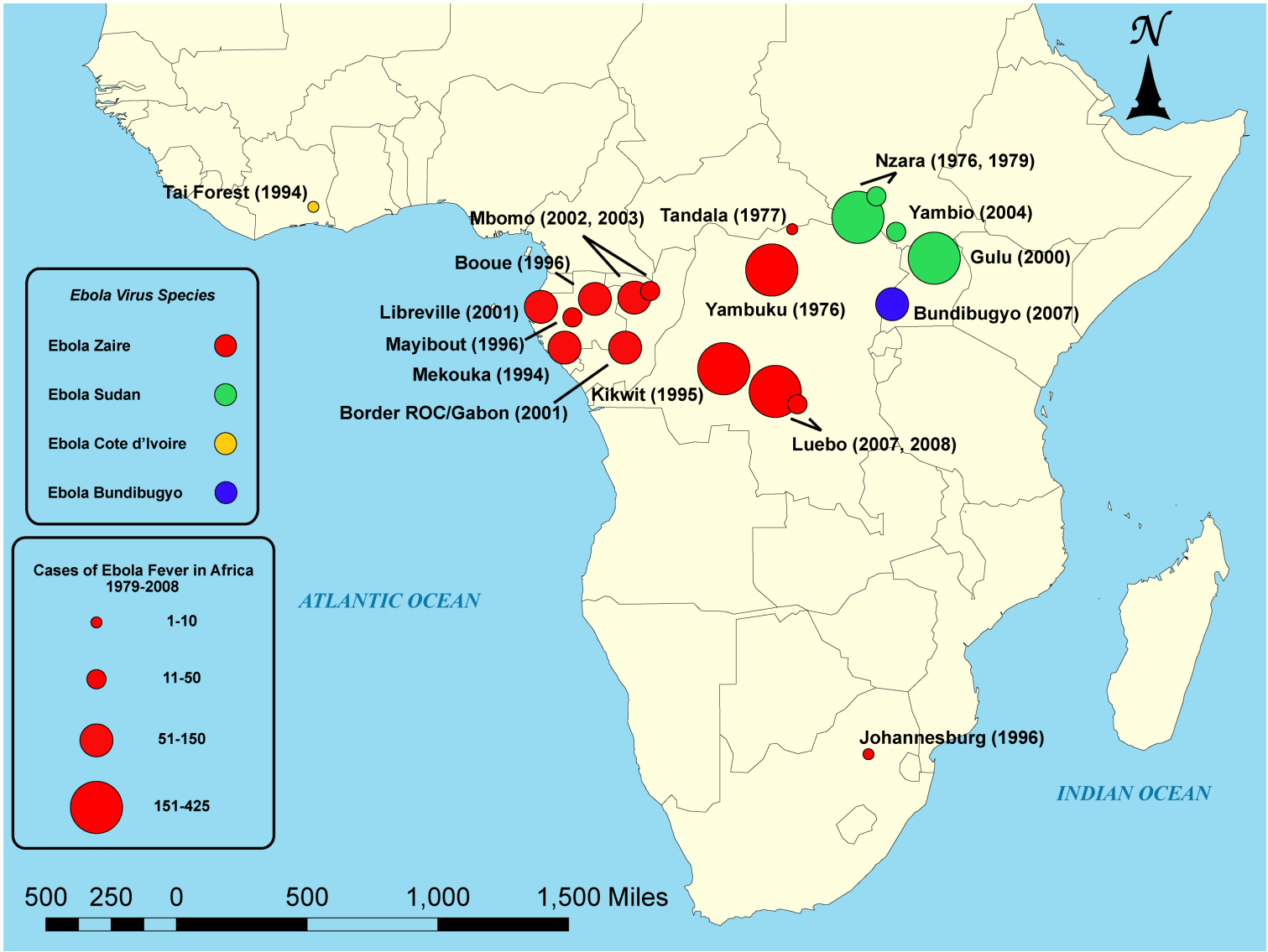
Statistical Inference in a Stochastic Epidemic SEIR Model with Control Intervention: Ebola as a Case Study

Authors: Lekone and Finkenstädt

Jon Azose

May 23, 2013

Ebola



(Image from Zach Orecchio via Wikipedia.)

Ebola

The disease can be horrific.

Ebola

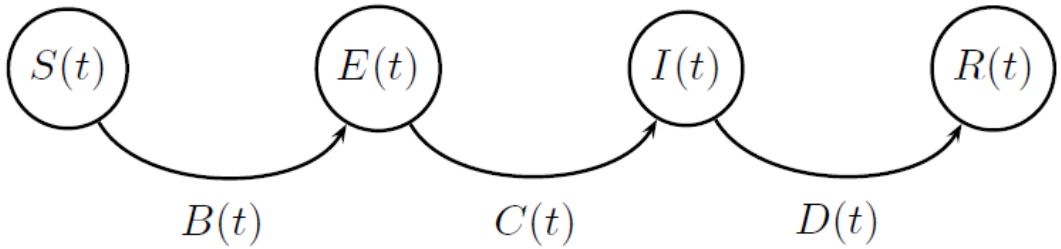
The disease can be horrific.

(Plushie from www.giantmicrobes.com)

Scientific question

How does preventative intervention influence the course of the epidemic?

Modeling choice: Use a susceptible-exposed-infectious-recovered (SEIR) model.



A model for the complete data

Assume a population of size N with one initial infection. At the individual level:

- Time until $S \to E$ transitions is exponentially distributed with rate $\beta(t)I(t)/N$, where

$$\beta(t) = \begin{cases} \beta & \text{if } t < t_* \\ \beta e^{-q(t-t_*)} & \text{if } t \geq t_*. \end{cases}$$

- Time until $E \to I$ transitions is exponentially distributed with rate ϱ.
- Time until $I \to R$ transitions is exponentially distributed with rate γ.

Parameters

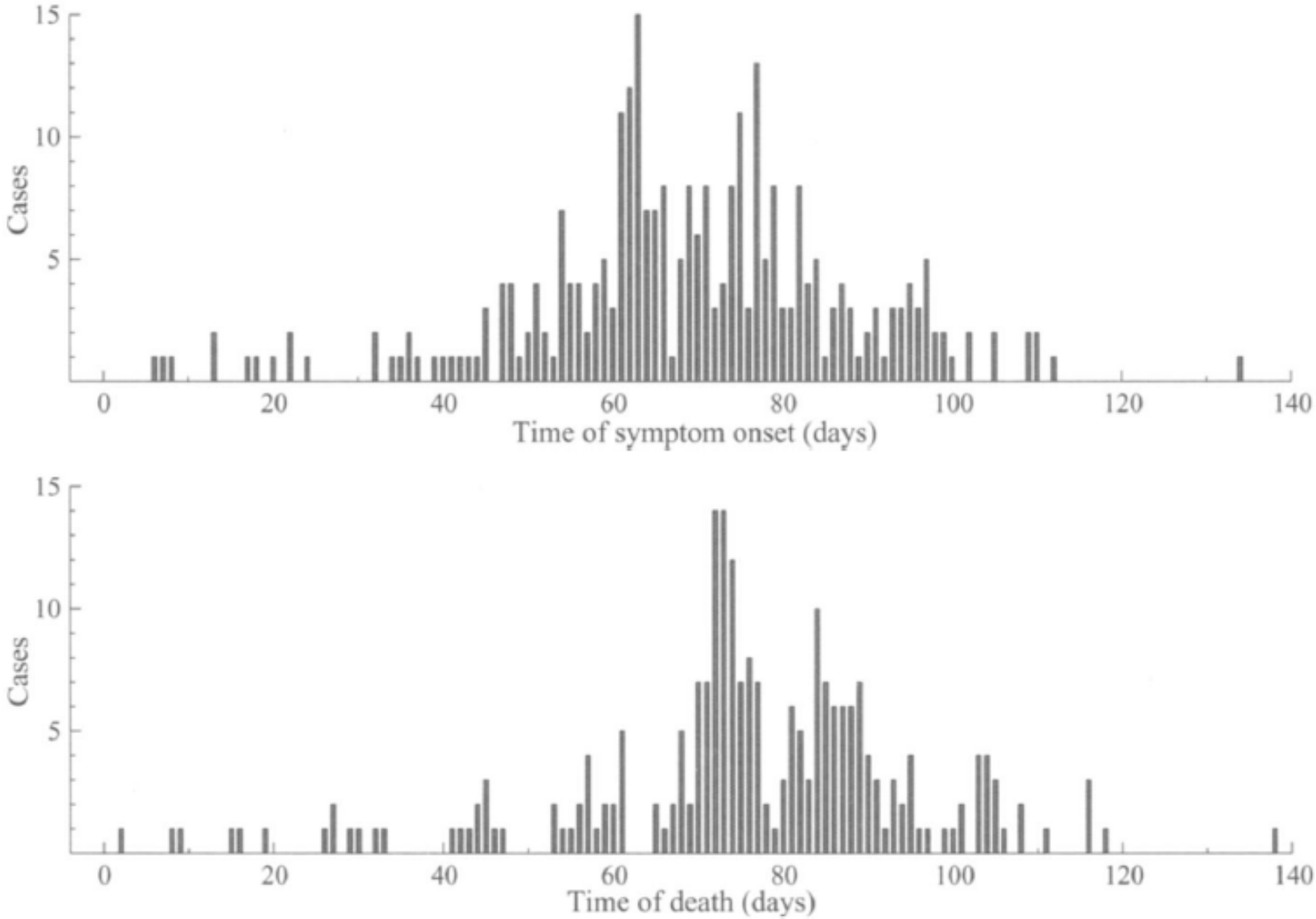
β: base transmission rate per infected individual (before intervention)

q: decay in rate of transmission after intervention

$1/\varrho$: mean incubation period

$1/\gamma$: mean infectious period

The real data won't work with that model.



Ideal versus actual data

Complete data would be:

- linked
- continuous
- complete (i.e. have $B(t)$, $C(t)$, and $D(t)$ with no missing records.)

Actual data are:

- unlinked
- discrete
- incomplete

The plan for making inferences

- Define a model that gives a likelihood function for unlinked, discrete-time data given a set of parameters.
- Choose prior distributions for parameters.
- Alternate sampling from distributions for (missing data|parameters) and (parameters|data).
- Make inferences based on posterior distribution of parameters.

We don't need linked, continuous, or complete data.

We don't need linked data:

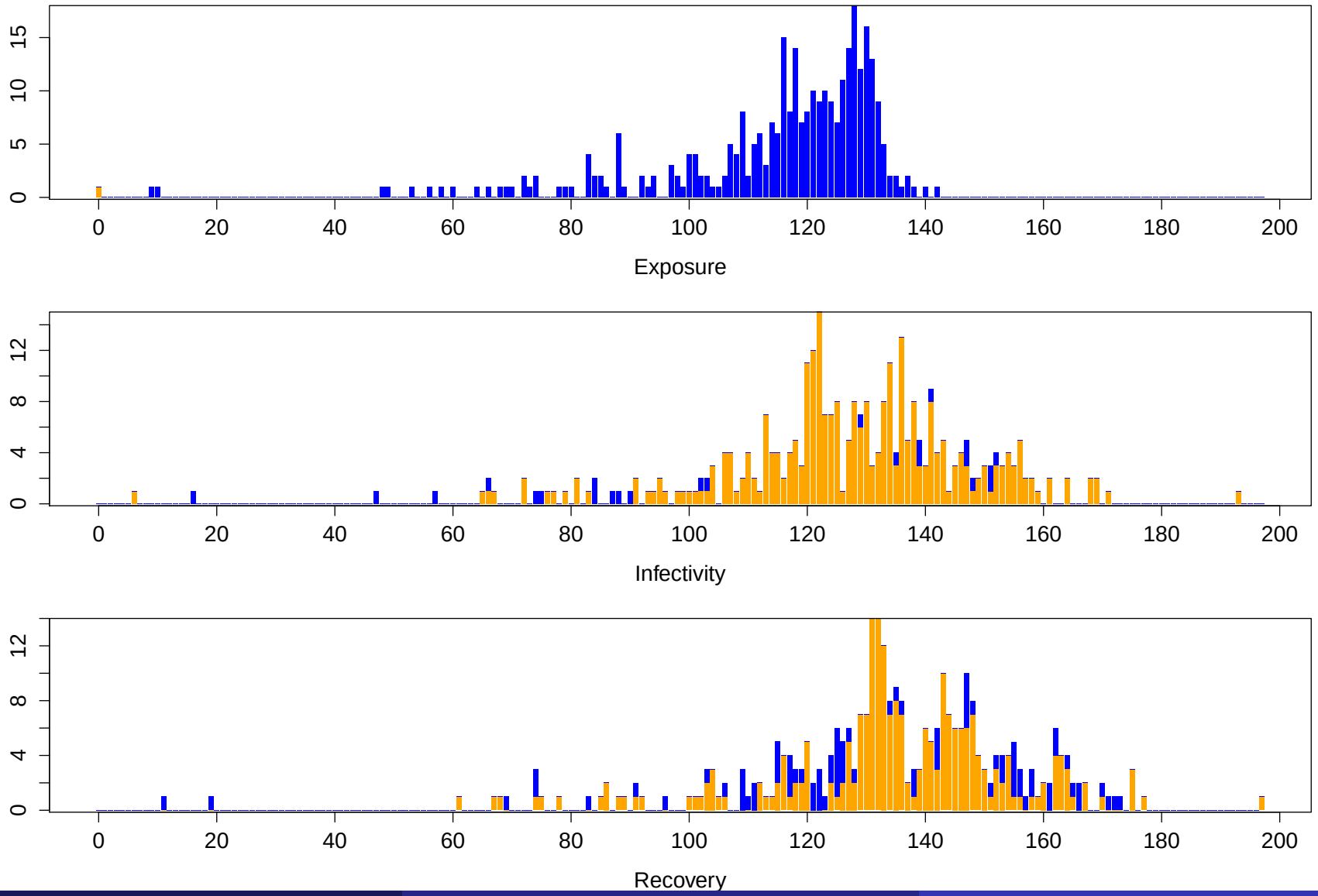
- Exponential distribution is memoryless.
- Sufficient statistics look like dwell times summed over all individuals.

We don't need continuous data. On a discrete time scale, we can use

$$B(t) \sim \text{Bin}\left(S(t), \left(1 - \exp\left[-\frac{\beta(t)}{N} I(t)\right]\right)\right)$$
$$C(t) \sim \text{Bin}\left(E(t), \left(1 - e^{-\varrho}\right)\right)$$
$$D(t) \sim \text{Bin}\left(I(t), \left(1 - e^{-\gamma}\right)\right).$$

We don't need complete data. Impute missing data with Metropolis-Hastings. (If you want to see the algorithm, ask me later.)

Example imputation



Results

Inferred posterior mean and standard deviation for parameters:

	Paper results	My results
β	0.243 (0.020)	0.243 (0.028)
q	0.161 (0.009)	0.366 (0.161)
$1/\varrho$	9.431 (0.620)	11.222 (1.198)
$1/\gamma$	5.712 (0.548)	7.671 (0.889)

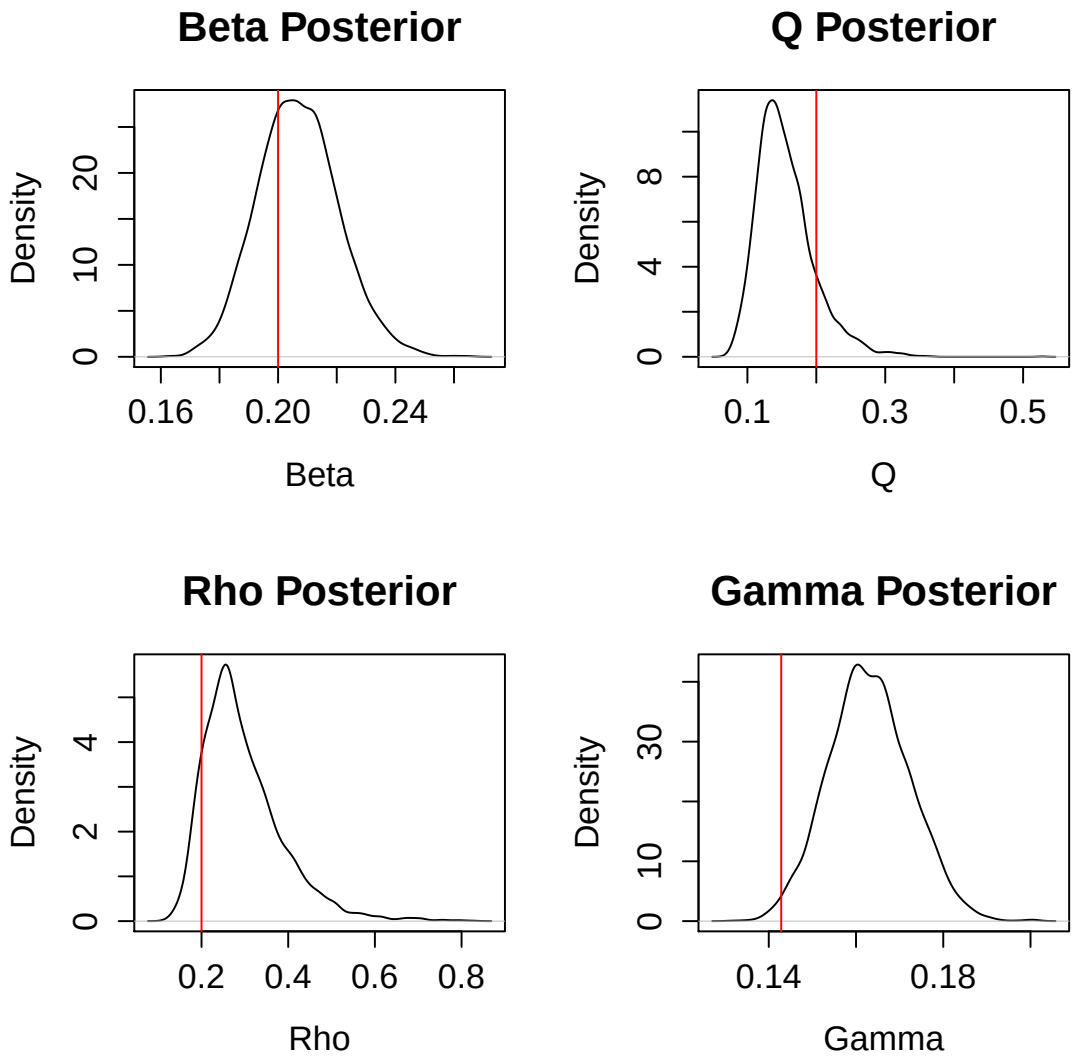
I trust my results

I believe that we're getting a representative sample from the posterior distribution. This is confirmed by:

- Trace plots and autocorrelation plots
- Replication across runs
- Geweke diagnostic
- Extremely long chains to check that we've explored the parameter space fully

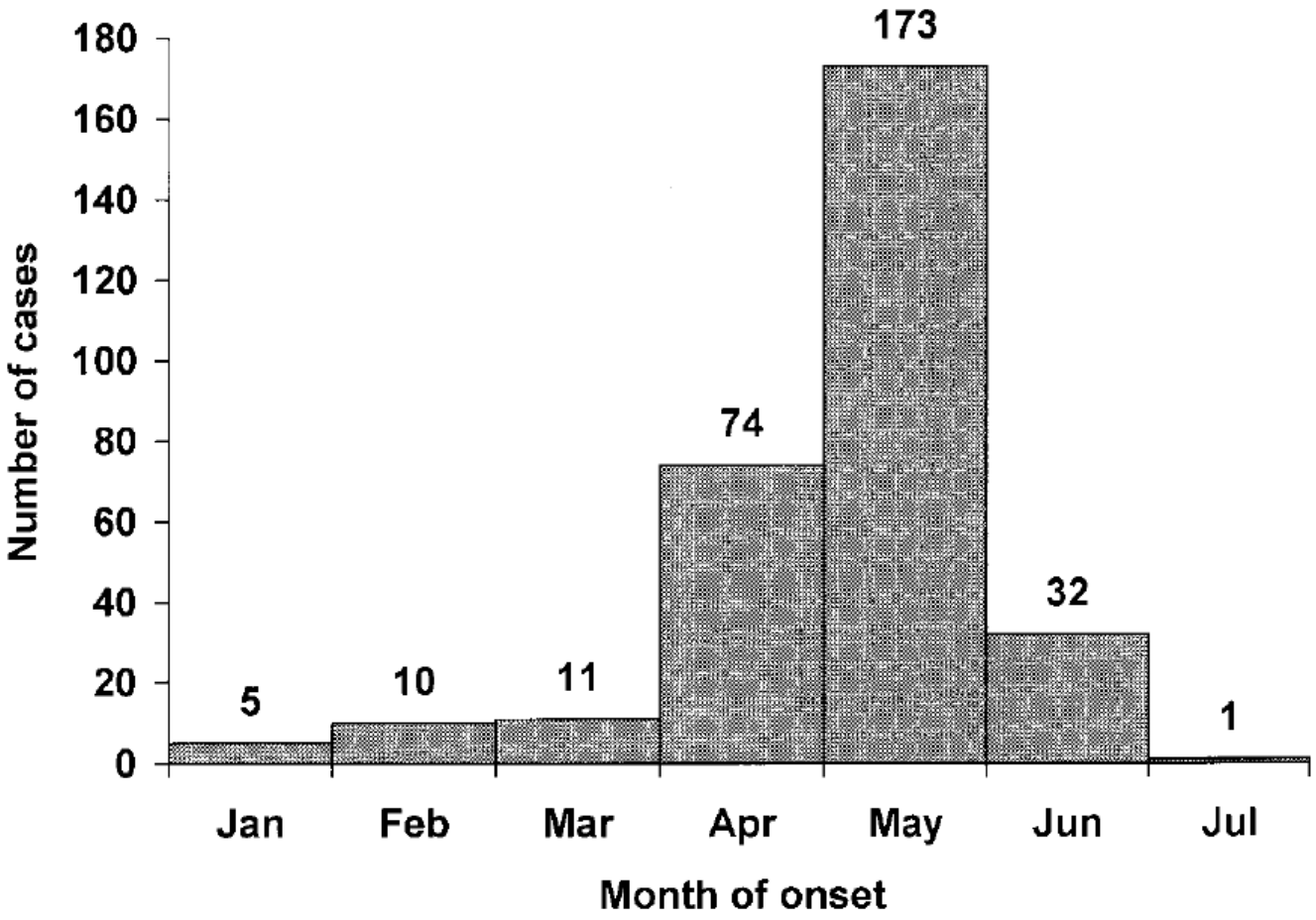
I trust my results

A simulation study confirms that my code makes reasonable inferences.



Why the discrepancy?

I *think* they're using additional data.



Critique

There's very little model validation.

The transmission rate per infected individual is:

$$\beta(t) = \begin{cases} \beta & \text{if } t < t_* \\ \beta e^{-q(t-t_*)} & \text{if } t \geq t_*. \end{cases}$$

How do we know this is a good model for transmission rates? The "right" model?

Did we restrict $\beta(t)$ too much?

After intervention, exposure rates decay to zero. Is that realistic?

Chowell et al. (2004) propose to use

$$\beta(t) = \begin{cases} \beta_0 & \text{if } t < t_* \\ \beta_1 + (\beta_0 - \beta_1)e^{-q(t-t_*)} & \text{if } t \geq t_*, \end{cases}$$

which allows transmission rates to decay to the non-zero quantity β_1.

What if there were no "intervention"?

Even without formal intervention, people tend to reduce contact with infected.

Summary

Scientific contribution: How does intervention impact the spread of Ebola?

Statistical contribution: How can we make inference in an SEIR model when the $S \rightarrow E$ transitions are entirely unobserved?

Critiques: How do we know that exponential decay in exposure rate is a reasonable model? Are we sure that intervention *causes* the decrease in new cases?

We can impute $B(t)$, $C(t)$, and $D(t)$

How to sample from $B|C, D, \Theta$:

1. Start with some B that is not impossible. (e.g. B must be such that no new infections can occur if there are no infectious.)
2. Propose a configuration B' by randomly picking days t_+ and t_-. Take $B'(t_+) = B(t_+) + 1$ and $B'(t_-) = B(t_-) - 1$.
3. Accept the move to the proposed B' with probability

$$\min\left(\frac{\pi(B'|C, D, \Theta)}{\pi(B|C, D, \Theta)}\frac{p(B' \to B)}{p(B \to B')}, 1\right)$$

4. Repeat steps 2 and 3.

The same procedure will work on the missing observations from $C(t)$ and $D(t)$.

Results

Epidemic metrics with and without intervention

	With intervention	Without intervention
Epidemic length	$\sim$ 200 days	$\sim$ 950 days (or $\sim$ 1100)
Epidemic size	$\sim$ 300	$\sim$ 3.5 million (or $\sim$ 2.6 million)

(My results in red.)