The University of Nottingham

# Modelling and Bayesian inference for the Abakaliki smallpox data

Philip O'Neill

Mathematical Sciences
University of Nottingham

The University of Nottingham

Joint work with :

Theo Kypraios (Maths, Nottingham)

Jessica Stockdale (Maths, Nottingham)

EPSRC

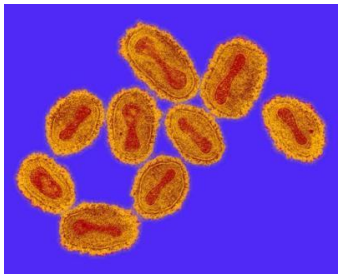
The University of Nottingham

1. The data and some history
2. Fitting epidemic models via DA-MCMC
3. Smallpox transmission model
4. Likelihood
5. Results
6. Concluding comments

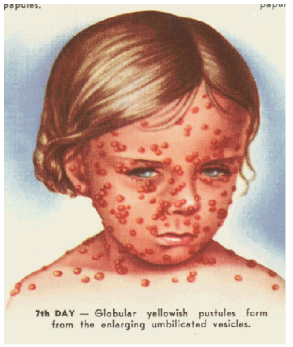
The University of Nottingham

1. The data and some history
2. Fitting epidemic models via DA-MCMC
3. Smallpox transmission model
4. Likelihood
5. Results
6. Concluding comments

# 1. The data & history



## Smallpox




- Infectious disease caused by *Variola* virus
- Transmission via inhalation of airborne virus
- Symptoms include fever and severe rash
- Overall case fatality around 30%
- Control via “ring-vaccination” (= isolation and local vaccination)
- Declared eradicated in 1980 by WHO
- Concerns over use as bioterrorist weapon

The University of Nottingham

# 1. The data & history

# 1. The data & history

## Abakaliki


- Town in South-Eastern Nigeria
- Mass smallpox and measles immunization (Feb 1967)
- Smallpox outbreak April – June 1967
- 32 cases, almost all members of FTC (Faith Tabernacle Church) who had refused vaccination
- Outbreak described in World Health Organization report (Thompson and Foege, 1968)

# 1. The data & history

## The recorded data

For each of the 32 cases:

- Date of onset of rash
- FTC member (yes/no)
- Vaccinated (yes + when/no)
- Compound number (dwelling)*
- Age
- Sex

* 4 individuals moved compound during outbreak

TABLE 1. LINE LISTING OF SMALLPOX CASES

| Case No. | Age | Sex | Onset of rash | Vaccination status Dates of vacc. | Vacc. scar | Member of FTC | Compound |
|---|---|---|---|---|---|---|---|
| 1 | 10 | F | 5 April | - | 0 | Yes | 1 |
| 2 | 25 | F | 18 April | - | 0 | " | 1 |
| 3 | 35 | M | 25 April | - | 0 | " | 1 |
| 4 | 4-1/2 | F | 27 April | - | 0 | " | 1 |
| 5 | 11 | M | 30 April | - | 0 | " | 1 |
| 6 | 1-1/2 | M | Last of April | - | 0 | " | 1 |
| 7 | 4 | F | Last of April | - | 0 | " | 1 |
| 8 | 8 | F | 1 May | 1966 | 0 | " | 2 |
| 9 | 12 | M | 5 May | 1963 | + | " | 2 |
| 10 | 2 | M | 10 May | - | 0 | " | 1 |
| 11 | 35 | M | 13 May | - | 0 | " | 4 |
| 12 | 28 | F | 15 May | - | 0 | " | 5 |
| 13 | 3-1/2 | M | 15 May | - | 0 | " | 1 |
| 14 | 1-1/2 | F | 17 May | - | 0 | " | 1 |
| 15 | 2 | M | 17 May | - | 0 | " | 1 |
| 16 | 3-1/2 | F | 22 May | - | 0 | " | 1 |
| 17 | 1 | F | 25 May | - | 0 | " | 5 |
| 18 | 30 | F | 26 May | - | 0 | " | 2 |
| 19 | 4-1/2 | F | 30 May | - | 0 | " | 1 |
| 20 | 13 | M | 30 May | 1963 Feb. 1967 | 0 | " | 2 |
| 21 | 26 | F | 31 May | 1958 | 0 | No | 6 |
| 22 | 35 | M | 31 May | Last one in 1948 | + | Yes | 5 |
| 23 | 2 | F | 1 June | - | 0 | " | 2 |
| 24 | 2 | M | 2 June | - | 0 | " | 7 |
| 25 | 11 | F | 4 June | - | 0 | " | 4 |
| 26 | 1 | F | 4 June | - | 0 | " | 2 |
| 27 | 3 | M | 5 June | - | 0 | " | 2 |
| 28 | 40 | M | 7 June | 1956 | 0 | No | 8 |
| 29 | 28 | F | 10 June | - | 0 | Yes | 3 |
| 30 | 27 | M | 10 June | - | 0 | " | 9 |
| 31 | 9 | F | 15 June | - | 0 | " | 5 |
| 32 | 35 | M | 20 June | 1963 | + | " | 2 |

# 1. The data & history

## The recorded data

Compound

- Housing built around a courtyard
- Houses several families

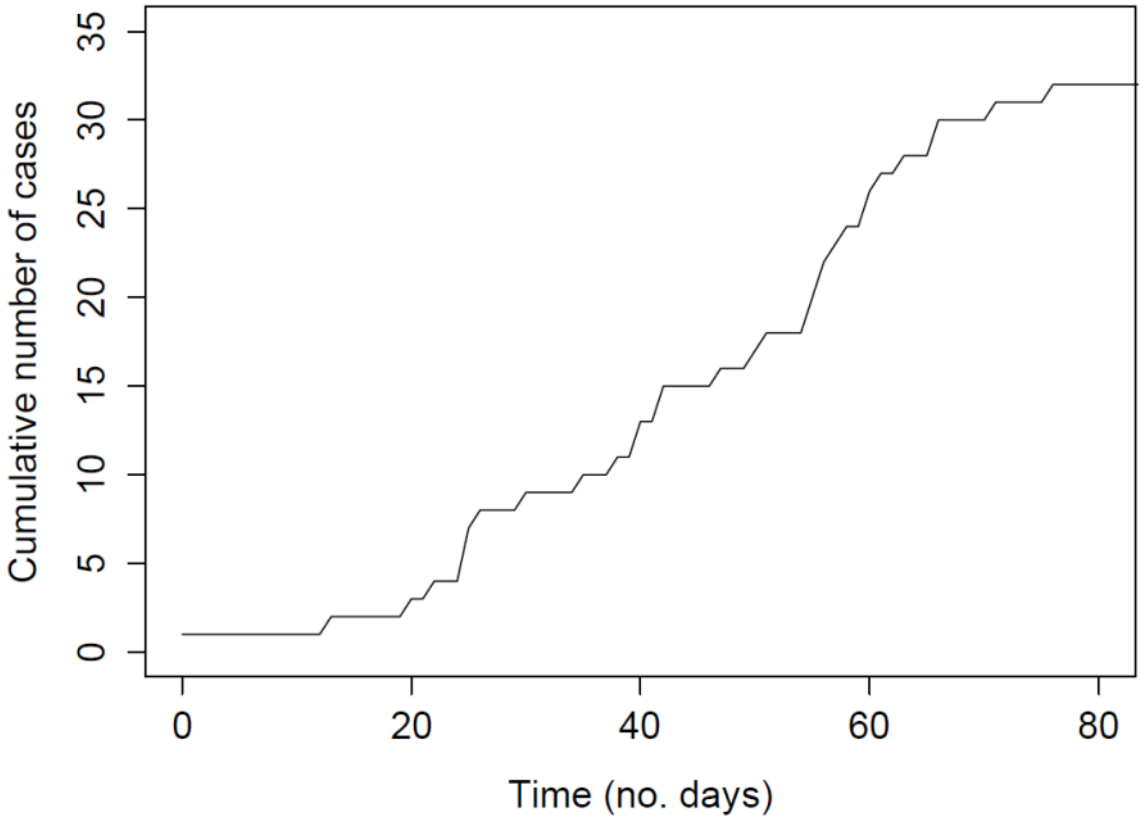
Additional data on each of 9 compounds:

- Number of FTC and non-FTC individuals
- Vaccination status of each individual*

* With a few exceptions

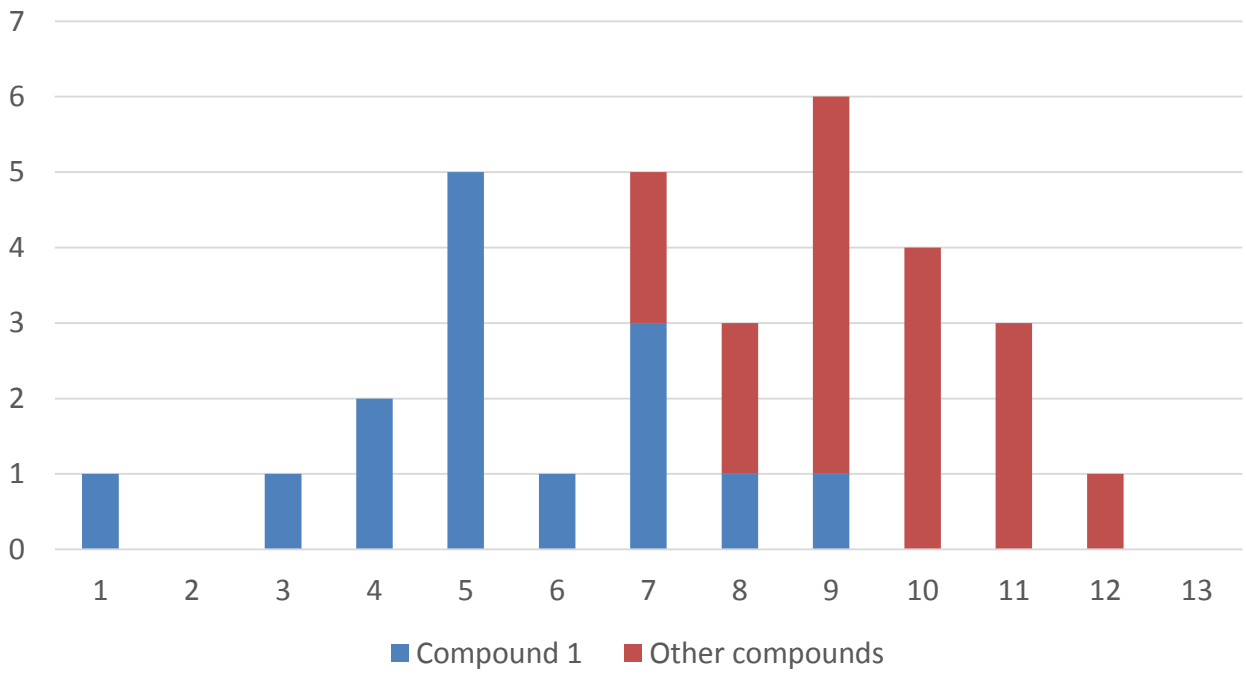
# 1. The data & history

## The recorded data



# 1. The data & history

## The recorded data



## The data in the epidemic modelling literature

First appears as an illustrative data set in Bailey and Thomas (1971):

- Only FTC individuals included in analysis (120)
- Only rash onset times

Modelling assumes

- Homogeneous mixing population (FTC)
- Simple/unrealistic transmission model

## The data in the epidemic modelling literature

- Numerous subsequent appearances in the literature (~ 20; 1972 – 2016), which…
- …all use Bailey and Thomas' version of the data and unrealistic models
- Ray and Marzouk (2008) include compounds but still only FTC individuals
- Eichner and Dietz (2003) considers the full data set

## The data in the epidemic modelling literature

Eichner and Dietz (2003)


"All the News That's Fit to Print"

The New York Times

Late Edition

VOL. CL. . No. 51,874 NEW YORK, WEDNESDAY, SEPTEMBER 12, 2001 75 CENTS

U.S. ATTACKED

HIJACKED JETS DESTROY TWIN TOWERS AND HIT PENTAGON IN DAY OF TERROR

A CREEPING HORROR

Buildings Burn and Fall as Onlookers Search for Elusive Safety

President Vows to Exact Punishment for 'Evil'

- Use realistic stochastic model
- Fit model using maximum likelihood, where…
- … the likelihood itself is an approximation
- Estimate key epidemiological parameters

## Motivation for current work

- Fit stochastic transmission model, avoiding any likelihood approximation
- Explore model adequacy
- Estimate key quantities
- Compare results to Eichner and Dietz

The University of Nottingham

1. The data and some history
2. Fitting epidemic models via DA-MCMC
3. Smallpox transmission model
4. Likelihood
5. Results
6. Concluding comments

# 2. Fitting models

As seen in this module, basic approach is:

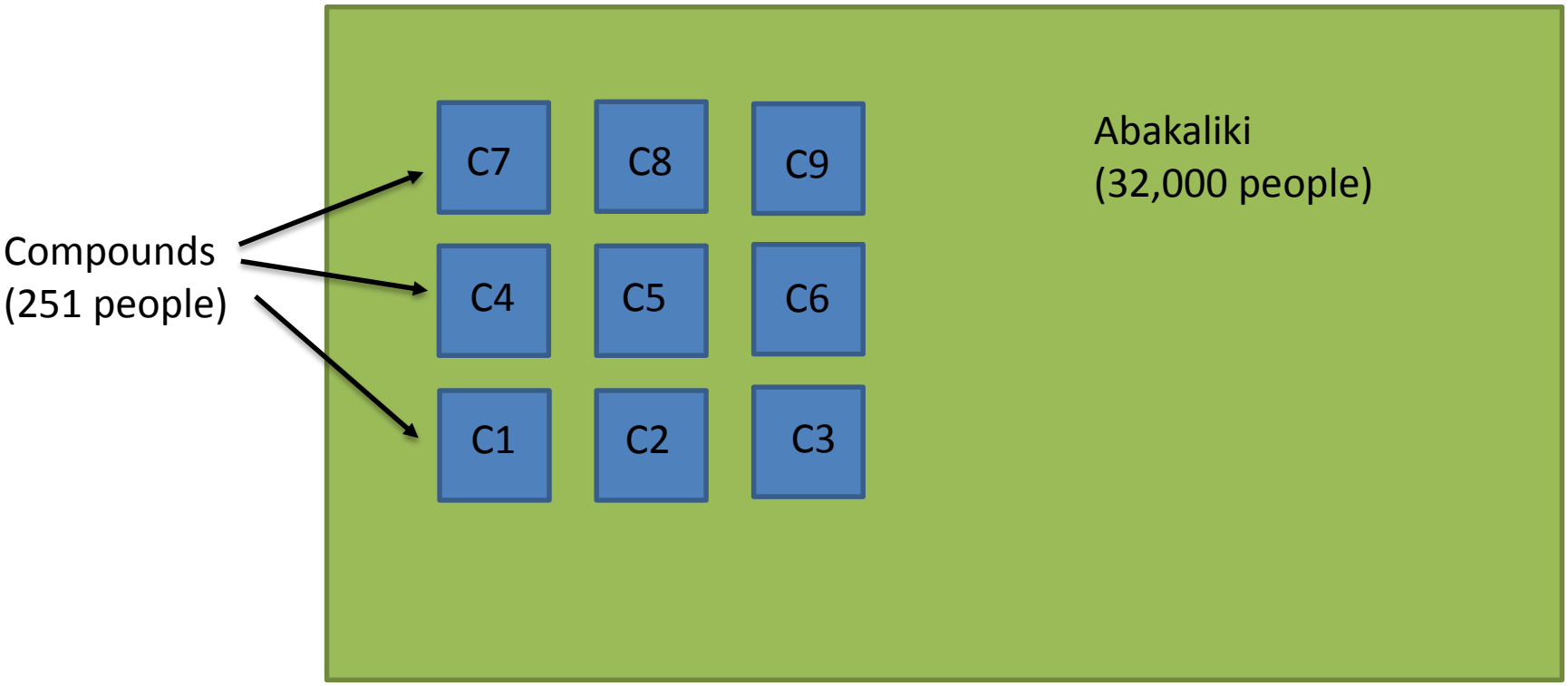
- Write down likelihood, augmented if necessary with any missing data
- Target density is likelihood times prior density
- Write MCMC algorithm to sample target
- Run algorithm and interpret results

The University of Nottingham

1. The data and some history
2. Fitting epidemic models via DA-MCMC
3. Smallpox transmission model
4. Likelihood
5. Results
6. Concluding comments

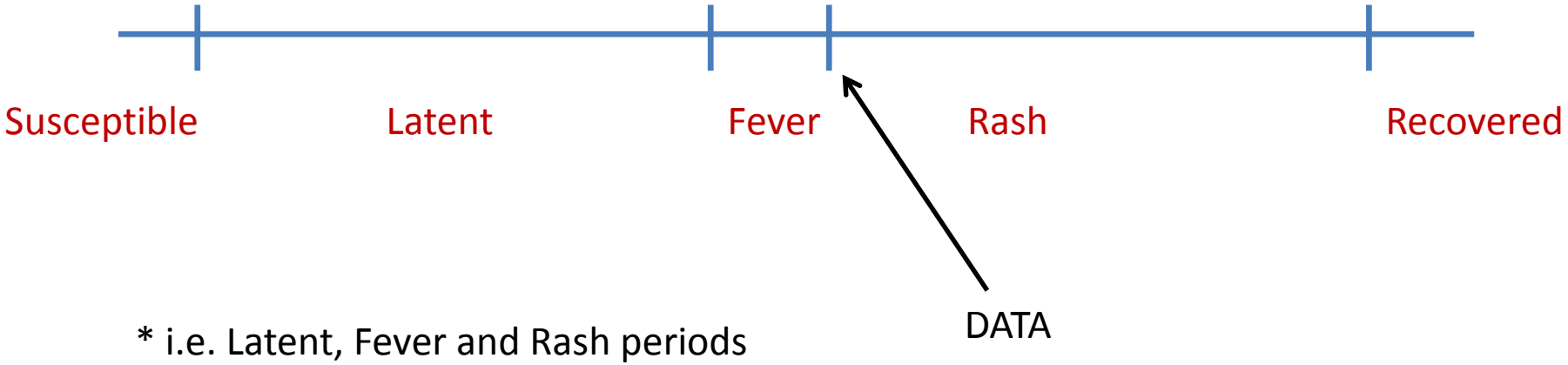
# Abakaliki smallpox model (Eichner-Dietz)

## Population structure: 9 compounds in town



## Abakaliki smallpox model (Eichner-Dietz)

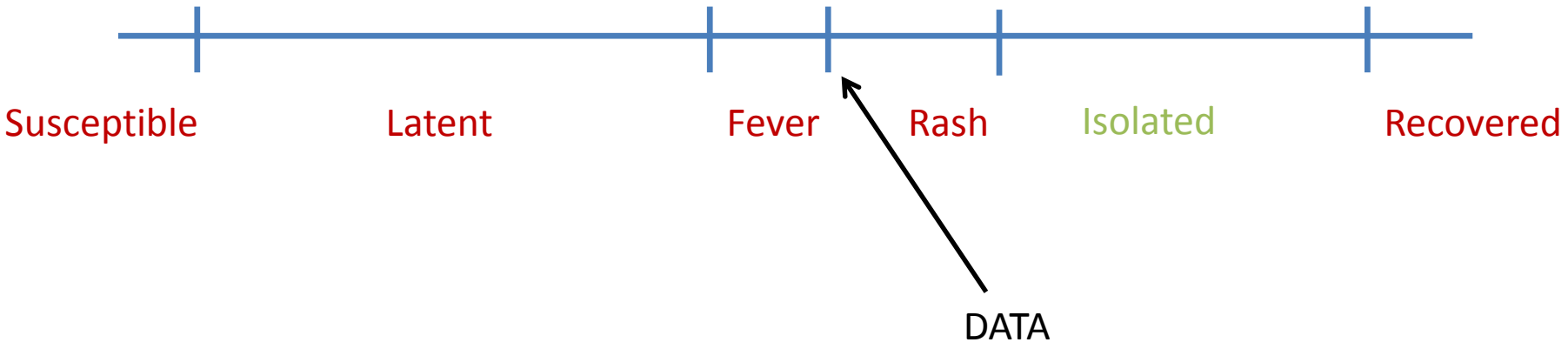
- SEIR-type model (E = “Exposed” = latent)
- Stage-times* are known Gamma distributions




* i.e. Latent, Fever and Rash periods

## Abakaliki smallpox model (Eichner-Dietz)

- Control measures introduced at time $t_Q$
  = isolation = reduced rash period



## Abakaliki smallpox model (Eichner-Dietz)

Three infection rate* parameters:

- Within-compound, same faith $\lambda_h$
- Within FTC $\lambda_f$
- Within population $\lambda_a$

Also: less infectious in Fever period (factor b)

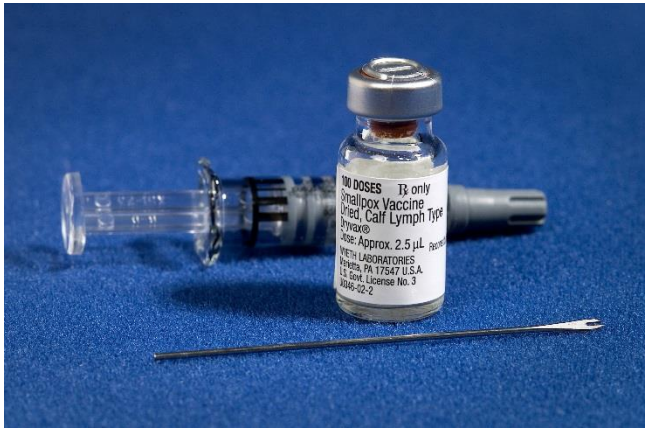
*same meaning as β in SIR model

## Abakaliki smallpox model (Eichner-Dietz)

All-or-nothing vaccine model:

P(vaccine works) = v

for each vaccinated individual, independently

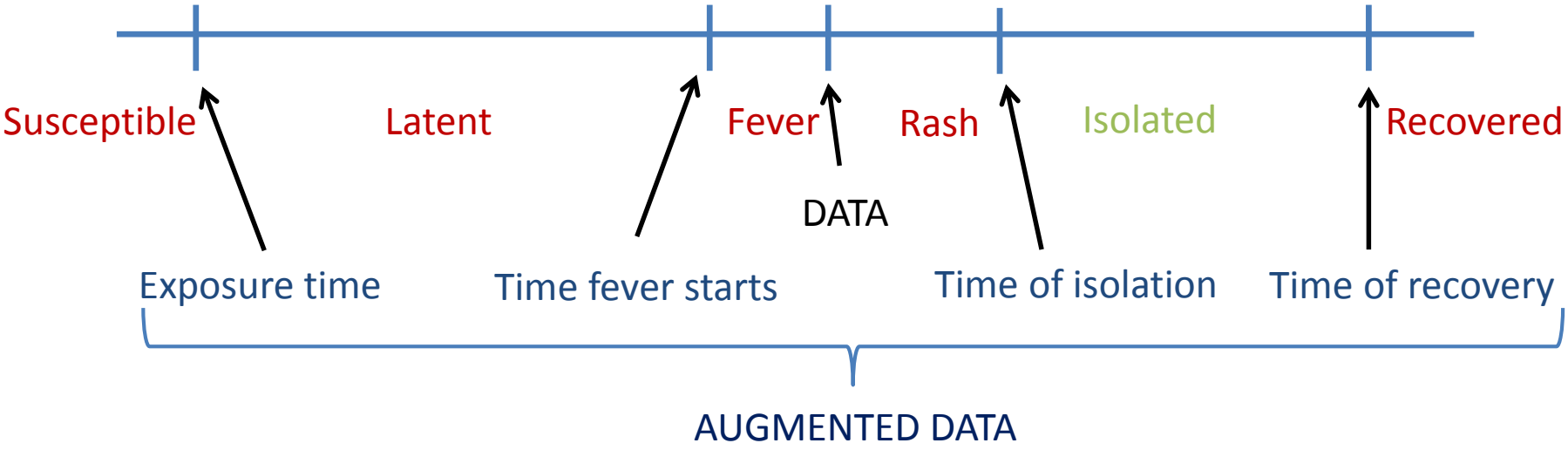


## Abakaliki smallpox model (Eichner-Dietz)

- Six parameter model ($\lambda_a$ , $\lambda_f$ , $\lambda_h$ , b, $t_Q$, v)
- E-D analysis is based on a likelihood approximation using back-calculation
- What happens if instead we use data-augmentation and MCMC?

## Abakaliki smallpox model – Data augmentation

- Augmented data = unknown event times…
- …and outcome for vaccinated individuals



The University of Nottingham

1. The data and some history
2. Fitting epidemic models via DA-MCMC
3. Smallpox transmission model
4. Likelihood
5. Results
6. Concluding comments

## Augmented likelihood

Augmented likelihood (of event times and types) given the model parameters is

L = (infection process part)

x (latent/fever/rash/isolation part)

x protection status part for vaccinees

## Augmented likelihood: infection process part

For (susceptible) individual k, define

$\Lambda_k(t)$ = infection pressure at time t acting on k

= hazard rate of infection for k

= sum of infection rates towards k

$\Lambda(t) = \Sigma_k \Lambda_k(t)$ = total pressure at time t

# Augmented likelihood: infection process part

Likelihood of infection process part is

$$\prod \Lambda_k(t_k) \quad x \quad \exp\left( - \int \Lambda(t) \; dt \right)$$

Likelihood of infection events

Likelihood of avoidance of infection

## Augmented likelihood: Latent/fever/… part

For each individual who becomes infected, multiply together the density functions for each stage (latent, fever, rash, isolation)

## Augmented likelihood: protection status part

Likelihood of protection statuses = $v^a (1-v)^b$

where

a = no. of vaccinated protected individuals

b = no. of vaccinated unprotected individuals

## Augmented likelihood: protection status part

Problem: there are a lot of protection statuses (outside compounds, about 30,000)

Solution: we can integrate most out of the likelihood; for example

$$\text{No. outside, vacc, prot} \sim \text{Binomial}(m,v)$$

where m = no. outside, vacc

## Augmented likelihood: computation

- Computing the likelihood is quite involved in practice
- Lots to keep track of
- Individuals who move complicate matters!

## Target density

$$\pi(\theta, \mathbf{y} \mid \mathbf{r}) \propto \pi(\mathbf{y}, \mathbf{r} \mid \theta)\, \pi(\theta)$$

**r** = data

**y** = augmented data

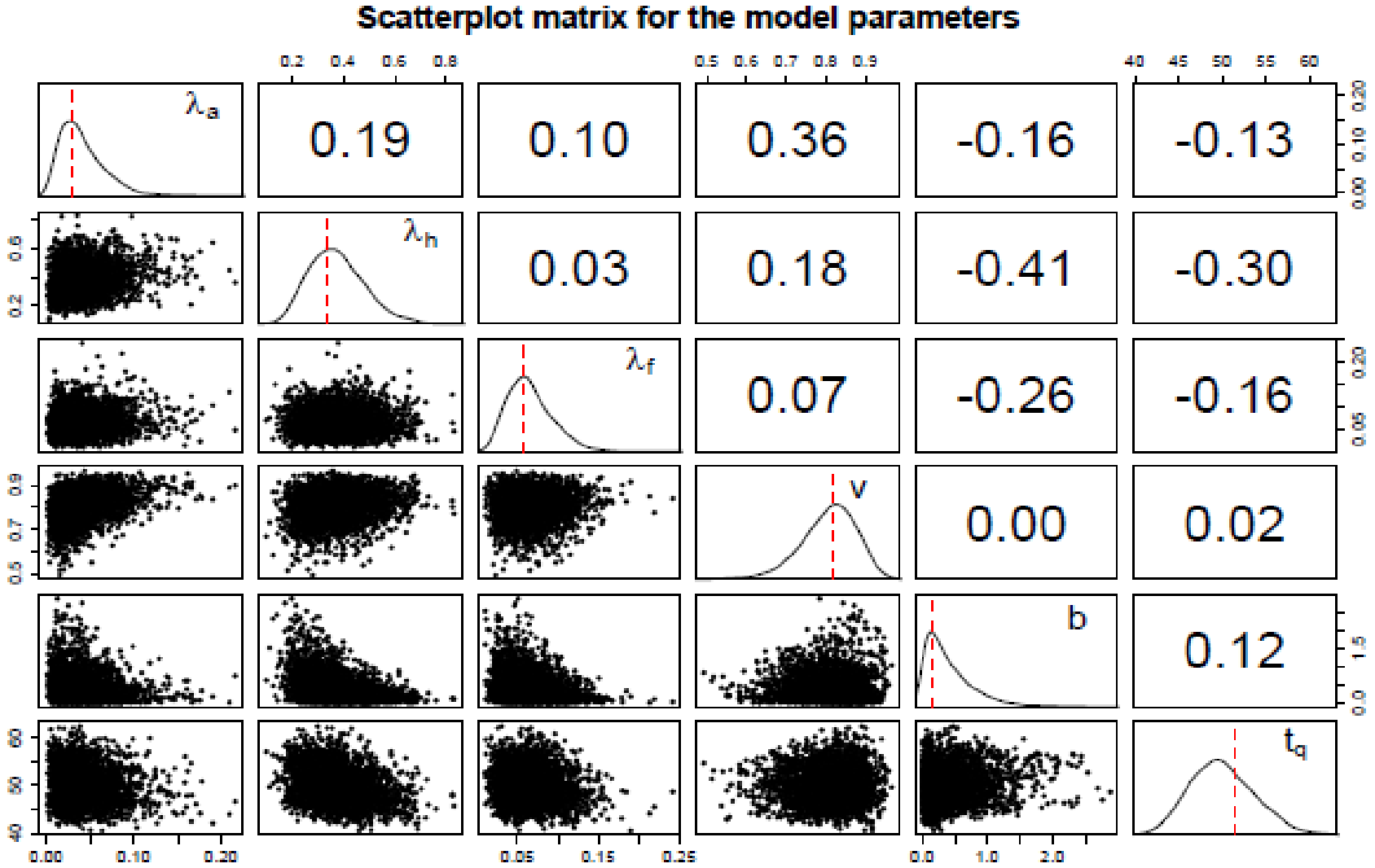
θ = model parameters

The University of Nottingham

1. The data and some history
2. Fitting epidemic models via DA-MCMC
3. Smallpox transmission model
4. Likelihood
5. Results
6. Concluding comments

The University of Nottingham

# 5. Results



# 5. Results

As well as model parameters we are interested in epidemiological quantities. For instance,

$R_0 \approx$ average number of secondary cases caused by one case in a large population

$$= (\mu_R + b\,\mu_F)(\lambda_a + \lambda_f + \lambda_h) \quad \text{(for FTC member)}$$

Average time infectious (Rash period, fever period; infectivity factor; before control measures)

Overall rate of infection

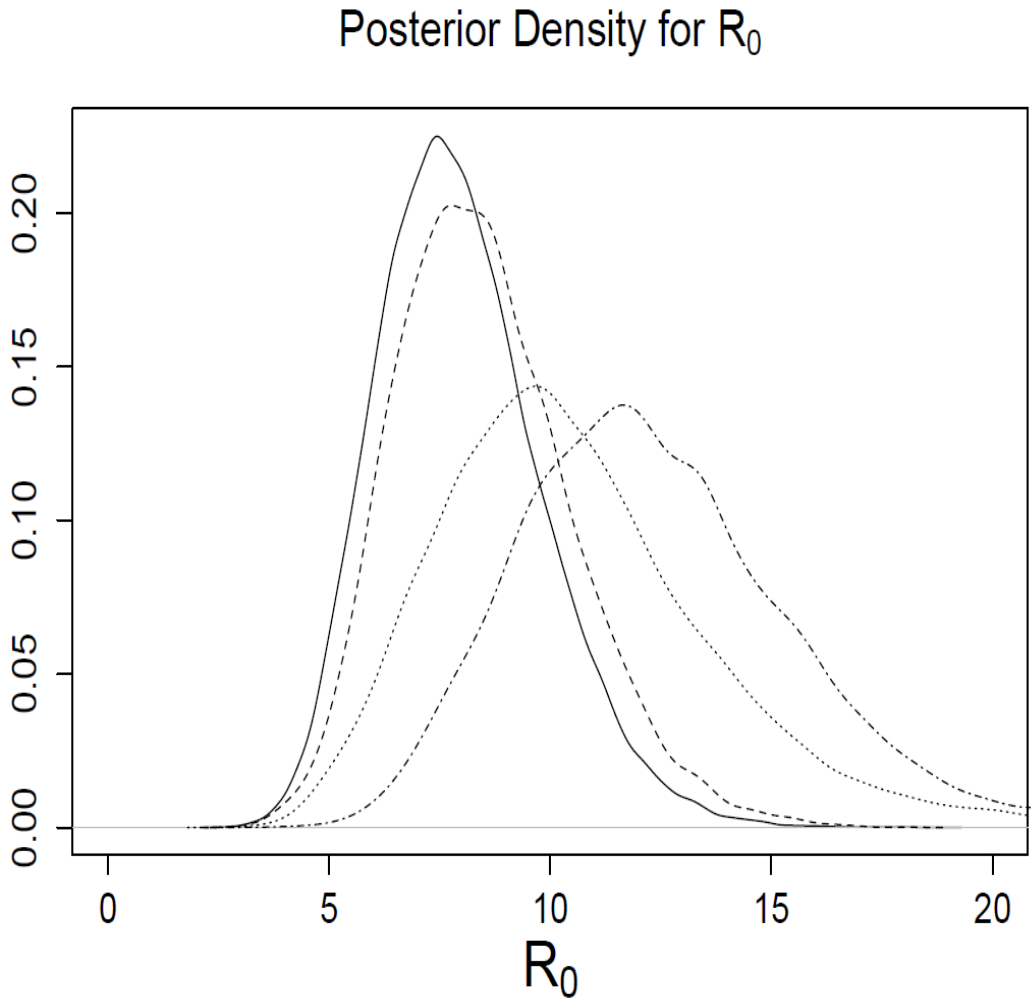
$R_0$ is called the basic reproduction number; $R_0 < 1$ to prevent epidemics

The University of Nottingham

# 5. Results

$E[R_0 | \text{ data}] \approx 8$

Dashed lines show different choices for latent period etc



# 5. Results

$$R_0 = (\mu_R + b\,\mu_F)(\lambda_a + \lambda_f + \lambda_h)$$

is an “overall” reproduction number.

Can also define specific reproduction numbers for transmission in compounds, FTC, outside compounds. For example

$$R_a = (\mu_R + b\,\mu_F)\lambda_a$$

is for individuals outside the compounds.

# 5. Results

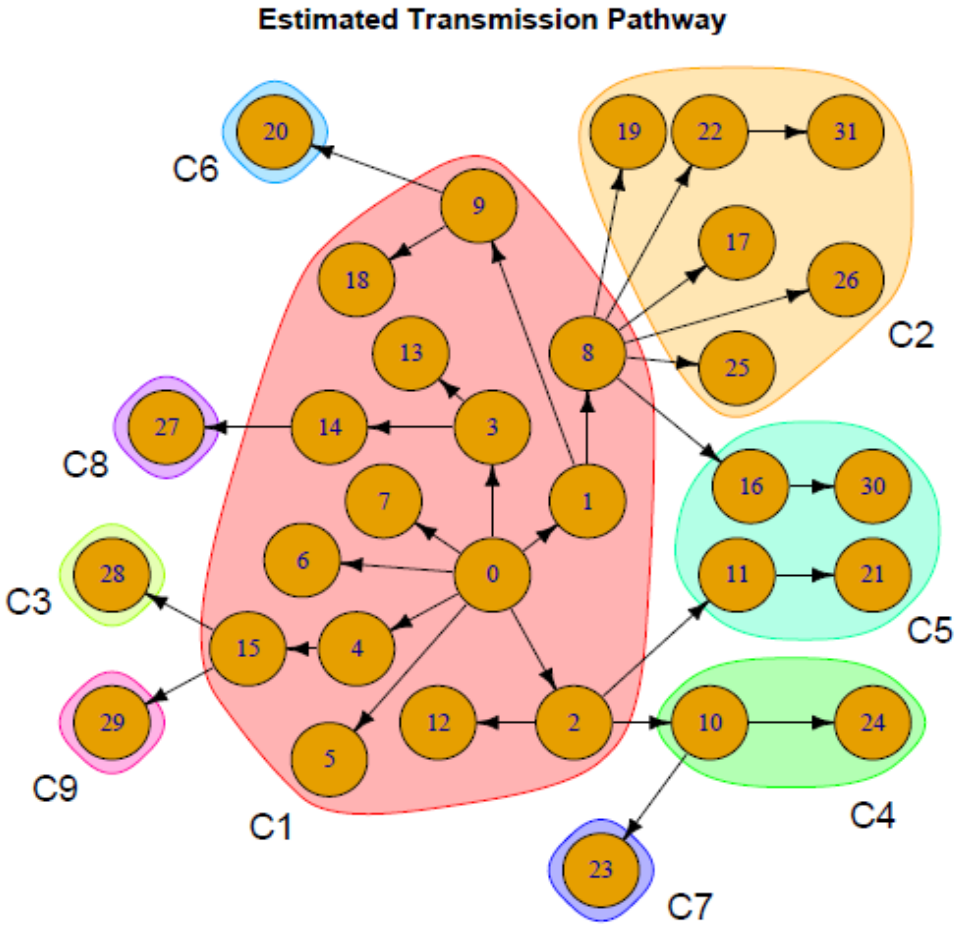
## Impact of control measures

| Before control measures | After control measures |
|---|---|
| $R_0 > 1$ | $R_0 > 1$ |
| $R_h > 1$ | $R_h > 1$ |
| $R_f > 1$ | $R_f < 1$ |
| $R_a < 1$ | $R_a < 1$ |

In simulations, epidemic never takes off in whole population; always subcritical

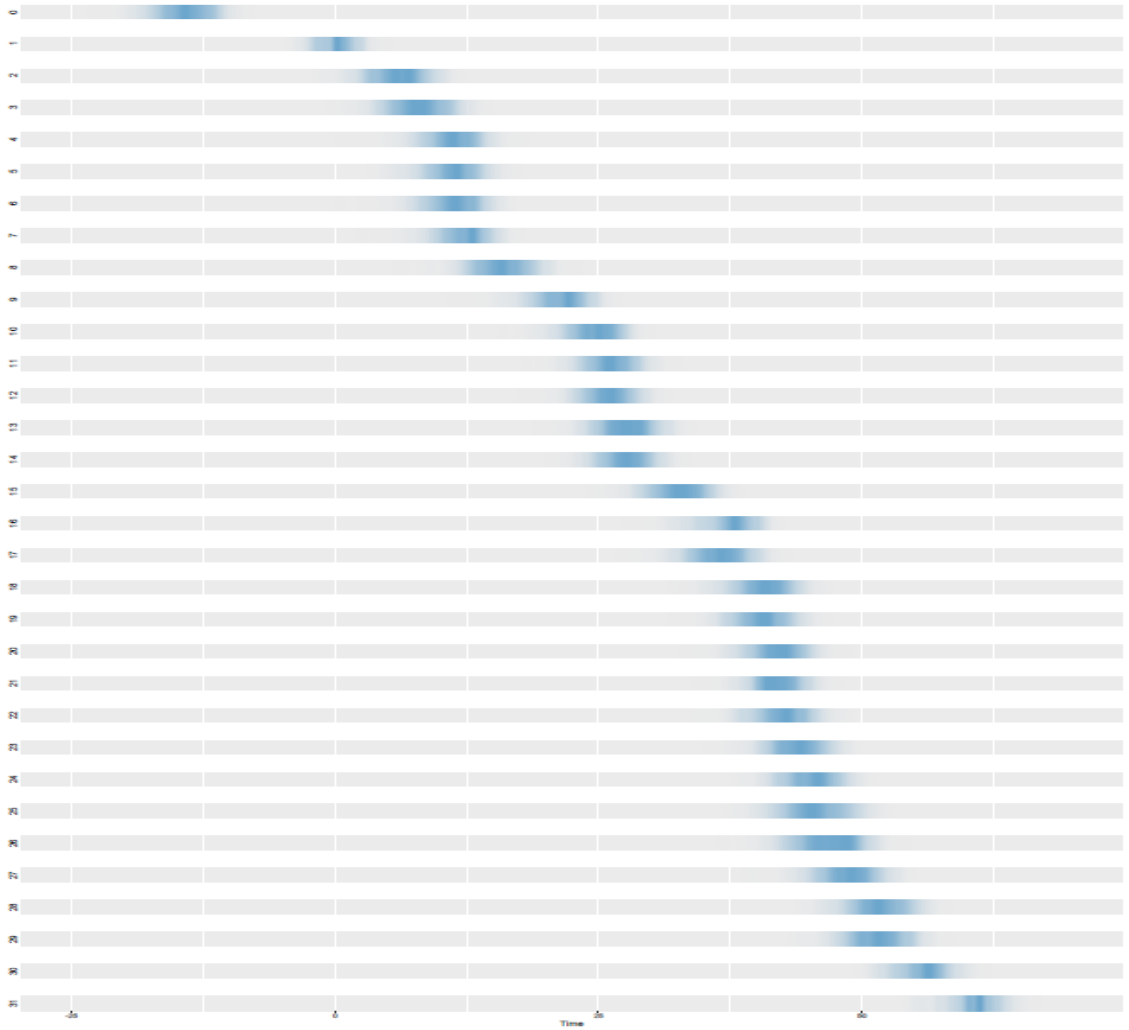
# 5. Results

Who infects whom



# 5. Results

## Infection times



## Model adequacy

- We use forward simulation of the model to assess model adequacy
- Parameter values are drawn from posterior distribution (i.e. from MCMC output)

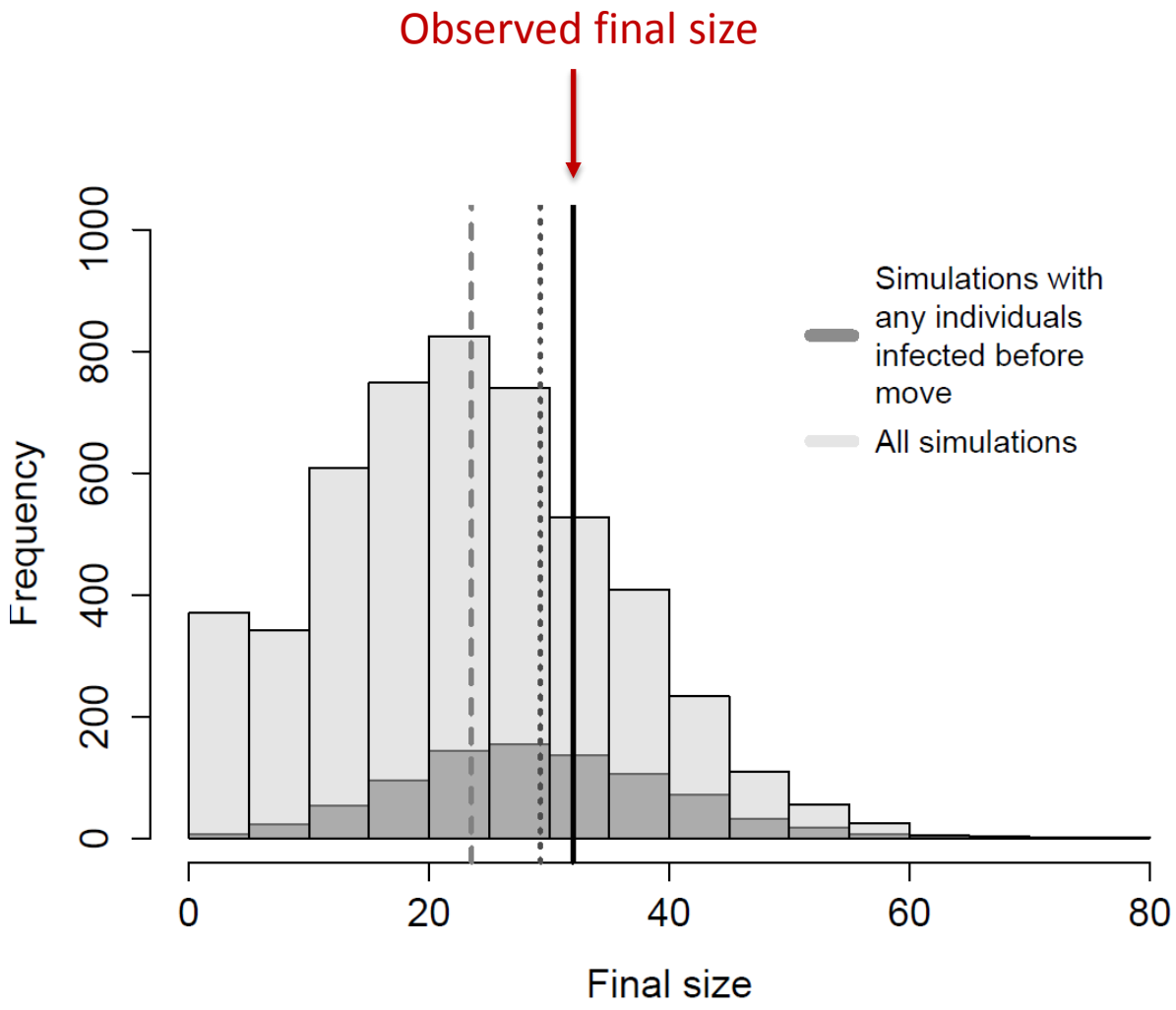
## Model adequacy

Final size =

Number of cases

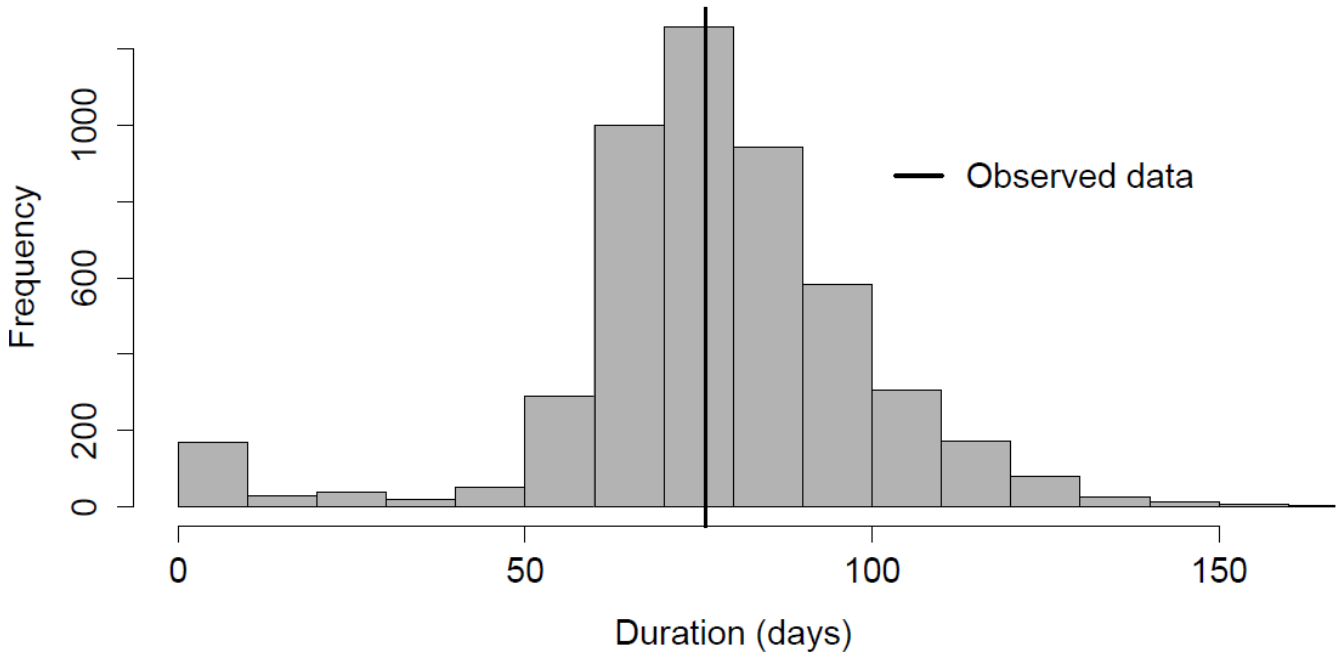
Better fit if

movers infected



# Model adequacy

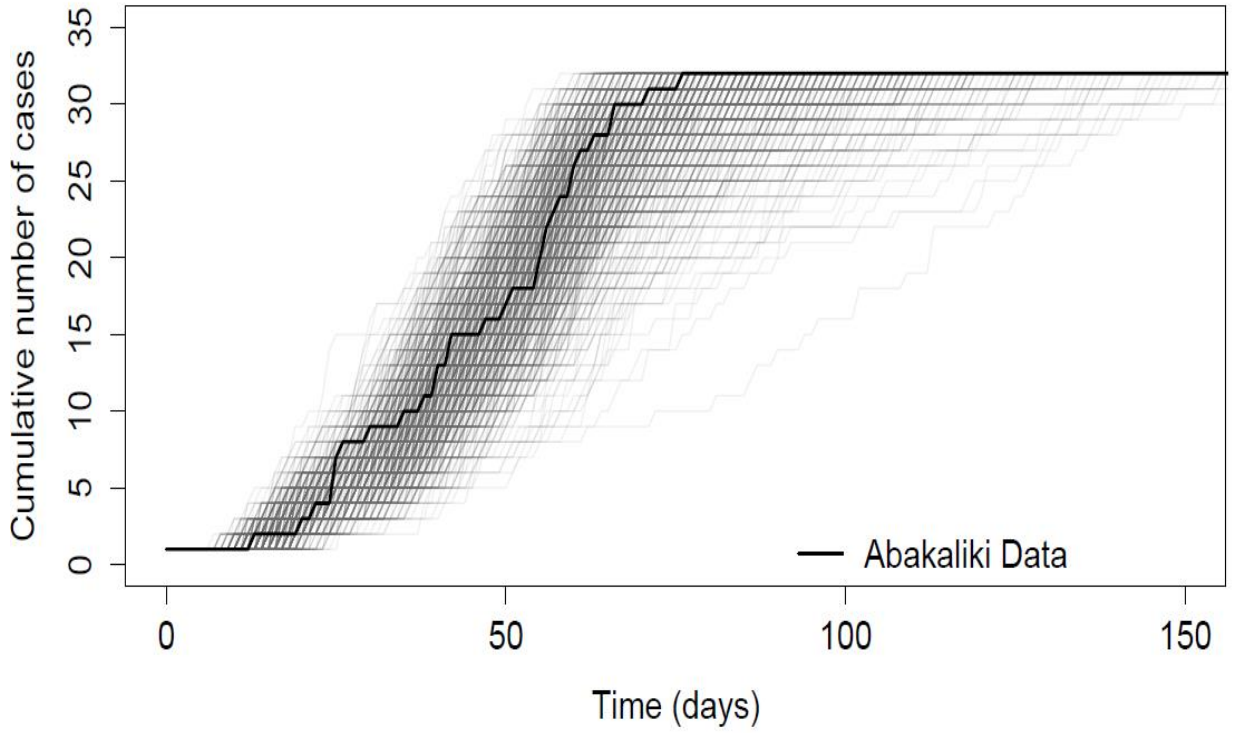
## Duration



## Model adequacy

Time course of epidemic

(conditioned on final size)



The University of Nottingham

1. The data and some history
2. Fitting epidemic models via DA-MCMC
3. Smallpox transmission model
4. Likelihood
5. Results
6. Concluding comments

## Some conclusions

- MCMC methods covered in module extended to a more complex model
- The approach provides plenty of useful information, not just estimates

The University of Nottingham

# 6. Concluding comments

G Model
EPIDEM-230; No. of Pages 11
ARTICLE IN PRESS

Epidemics xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Epidemics

journal homepage: www.elsevier.com/locate/epidemics

## Modelling and Bayesian analysis of the Abakaliki smallpox data

Jessica E. Stockdale, Theodore Kypraios, Philip D. O'Neill *

*School of Mathematical Sciences, University of Nottingham, United Kingdom*

ARTICLE INFO

*Article history:*
Received 27 July 2016
Received in revised form 28 October 2016
Accepted 7 November 2016
Available online xxx

*Keywords:*
Smallpox
Bayesian inference
Markov chain Monte Carlo
Stochastic epidemic model
Abakaliki

ABSTRACT

The celebrated Abakaliki smallpox data have appeared numerous times in the epidemic modelling literature, but in almost all cases only a specific subset of the data is considered. The only previous analysis of the full data set relied on approximation methods to derive a likelihood and did not assess model adequacy. The data themselves continue to be of interest due to concerns about the possible re-emergence of smallpox as a bioterrorism weapon. We present the first full Bayesian statistical analysis using data-augmentation Markov chain Monte Carlo methods which avoid the need for likelihood approximations and which yield a wider range of results than previous analyses. We also carry out model assessment using simulation-based methods. Our findings suggest that the outbreak was largely driven by the interaction structure of the population, and that the introduction of control measures was not the sole reason for the end of the epidemic. We also obtain quantitative estimates of key quantities including reproduction numbers.



### 1. Introduction

In 1967, an outbreak of smallpox occurred in the Nigerian town of Abakaliki. The vast majority of cases were members of the Faith Tabernacle Church (FTC), a religious organisation whose members

illustrate new data analysis methodology, but in virtually all cases most aspects of the data are ignored apart from the population of 120 FTC individuals and the case detection times, while the models used are not particularly appropriate for smallpox (see for example Becker, 1976; Yip, 1989; O'Neill and Roberts, 1999; O'Neill and