MCMC 2: Lecture 5 Applications

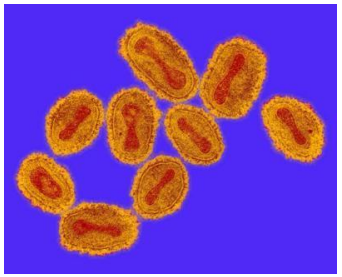
Phil O'Neill Theo Kypraios
School of Mathematical Sciences
University of Nottingham

Contents

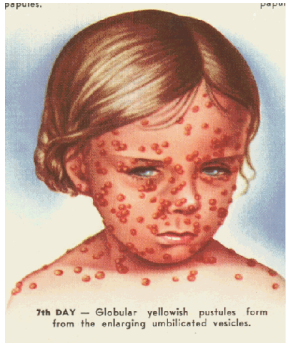
- 1. Smallpox: SEIR model with population structure
- 2. MRSA: whole-genome sequence data

Contents

- 1. Smallpox: SEIR model with population structure
- 2. MRSA: whole-genome sequence data



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 bioterrorism weapon

Smallpox outbreak

- Abakiliki: 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 WHO report (Thompson and Foege, 1968)

Smallpox 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

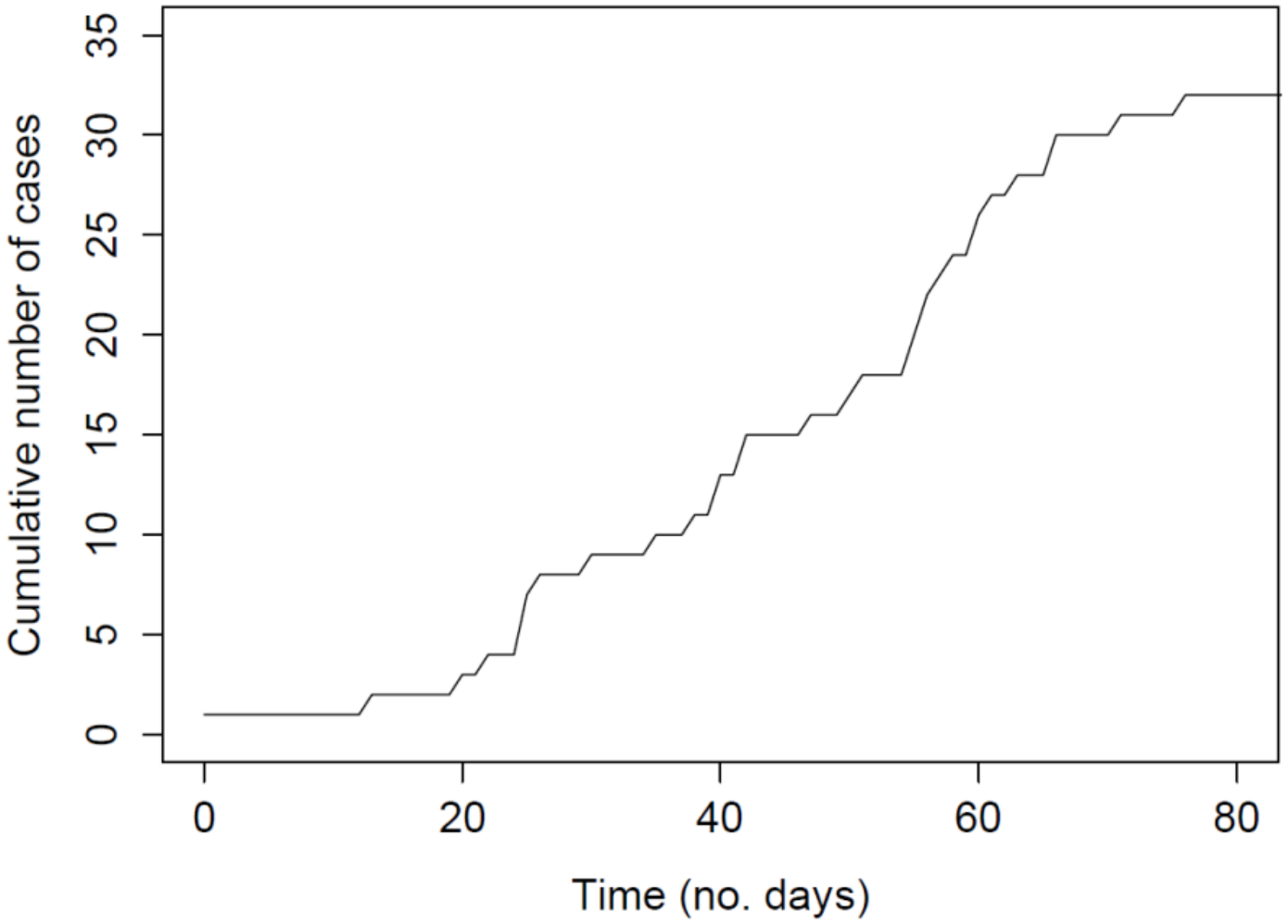
Smallpox data

Also know, for each of 9 compounds*

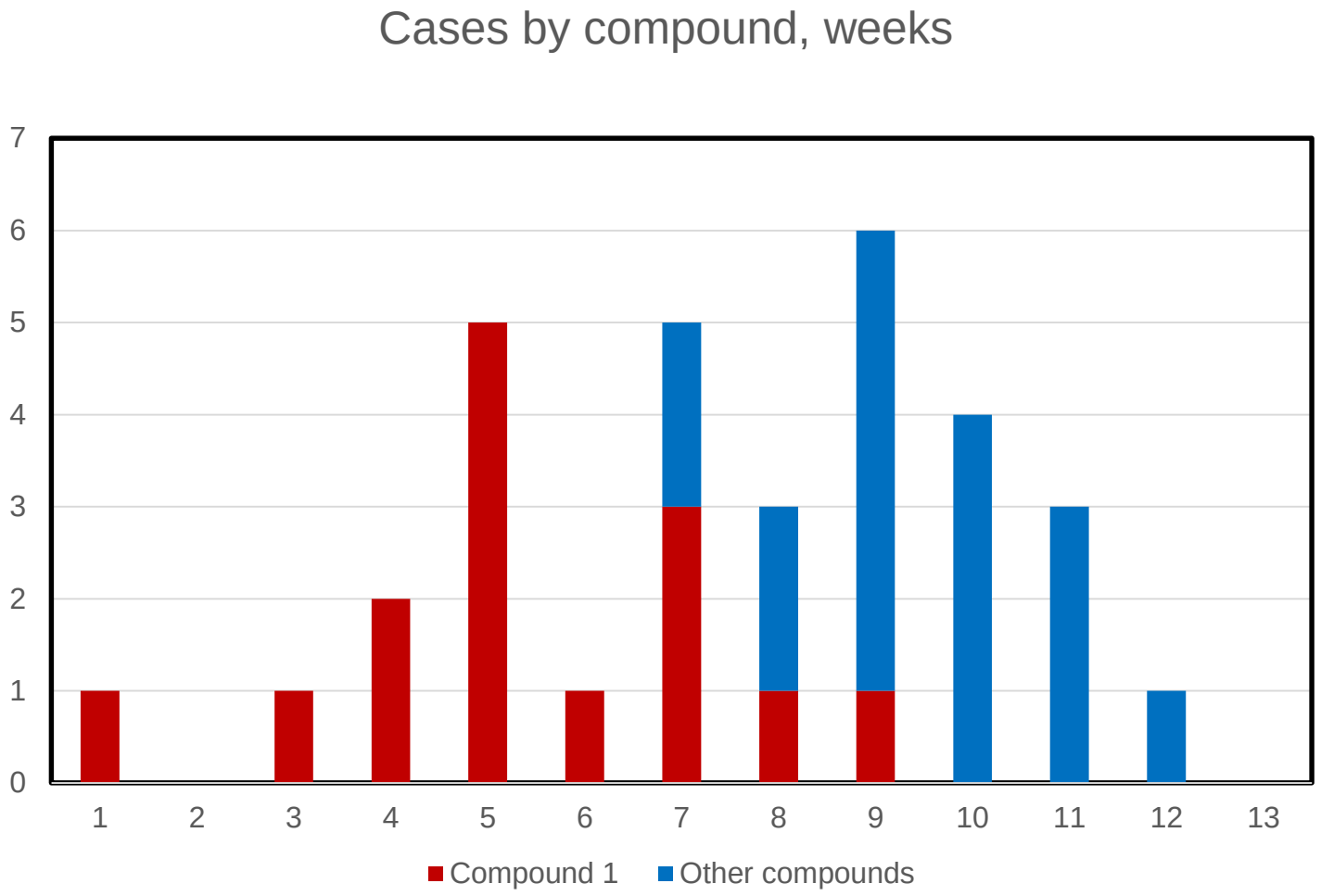
- Number of FTC and non-FTC individuals
- Vaccination status of individuals (with a few exceptions)

* “compound” = housing built around a courtyard, houses several families

Smallpox data

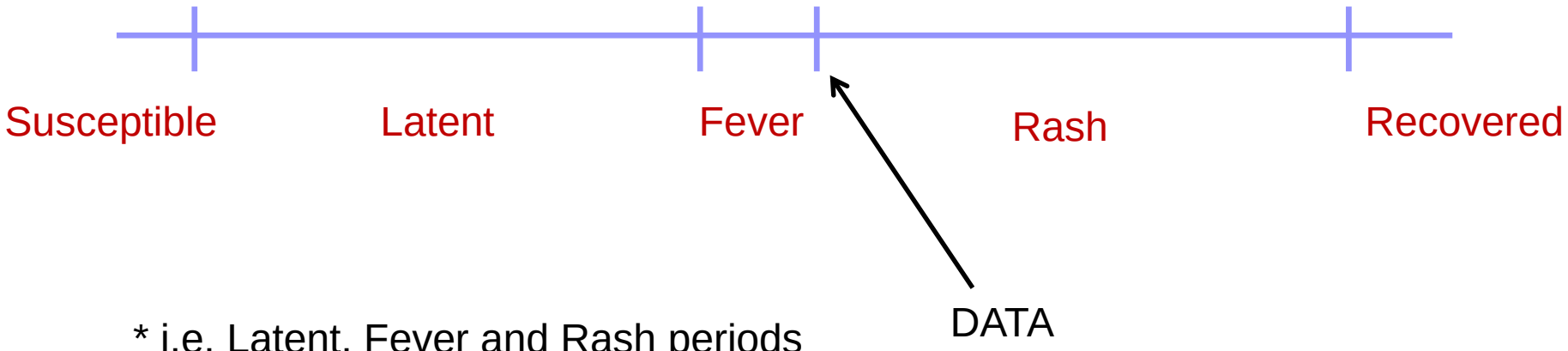


Smallpox data



Smallpox model

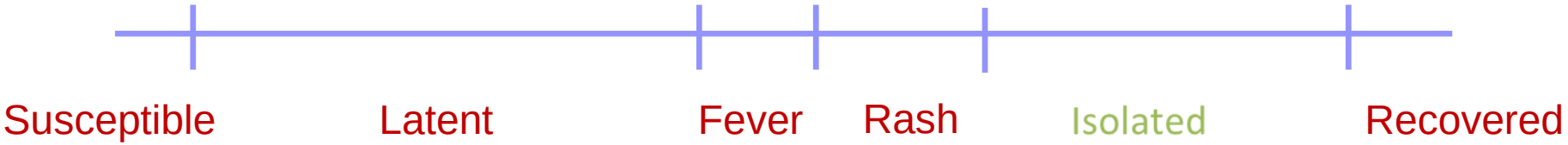
- SEIR-type model
- Each stage* assumed to have a known distribution (gamma with known parameters)



* i.e. Latent, Fever and Rash periods

Smallpox model

- Control measures introduced at time t_Q
- After this time, cases isolated swiftly



* i.e. Latent, Fever and Rash periods

Smallpox model

Model also has population structure:

9 compounds (251 people)

located inside town (32,000 people)

Smallpox model

- Three infection rate* parameters:
  - Within-compound, same faith λ_h
  - Within FTC λ_f
  - Within population λ_a
- Also: less infectious in Fever period (factor b)

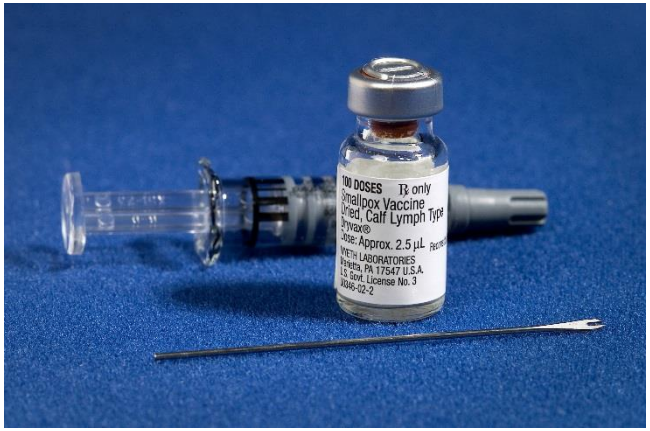
- *same meaning as β in SIR model

Smallpox model

- All-or-nothing vaccine model:

  P(vaccine works) = v

- for each vaccinated individual, independently

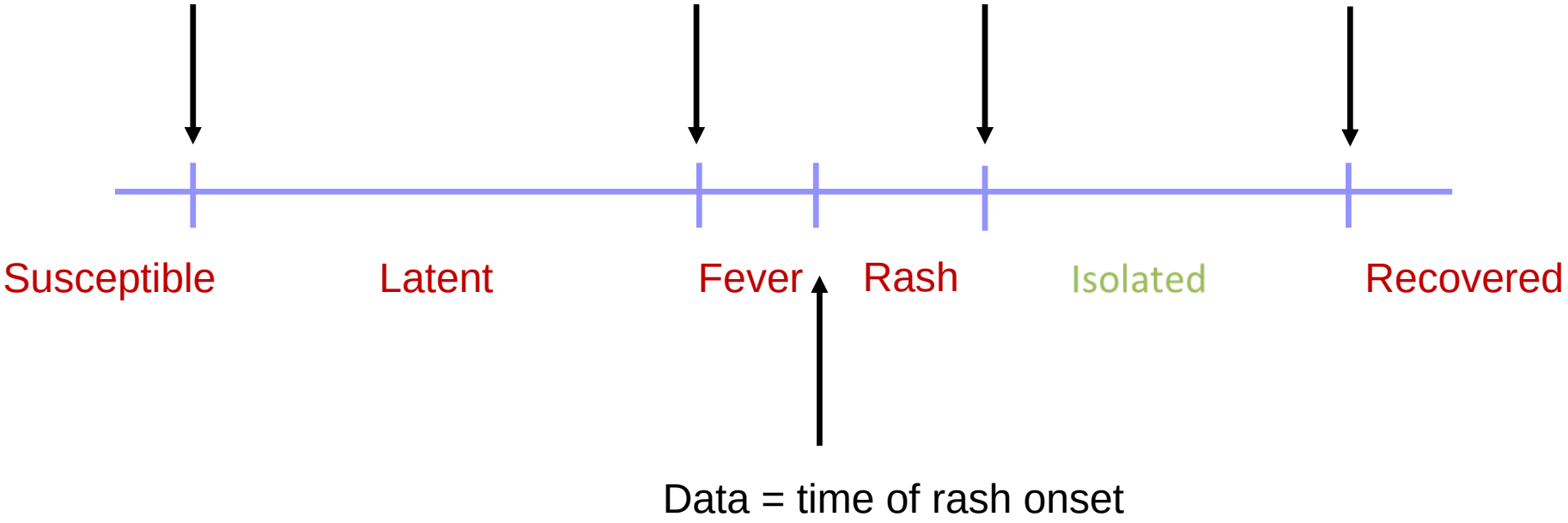


Inference problem

- Model has 6 parameters (λ_a , λ_f , λ_h , b, t_Q, v)
- Data consist of population structure, vaccination status, and rash times for each case
- As usual, the likelihood is intractable
- Proceed using data augmentation (as for SIR model in lectures)

Inference problem

- Augment with event times for each case:



Inference problem

- Augmented likelihood is similar to SIR model:

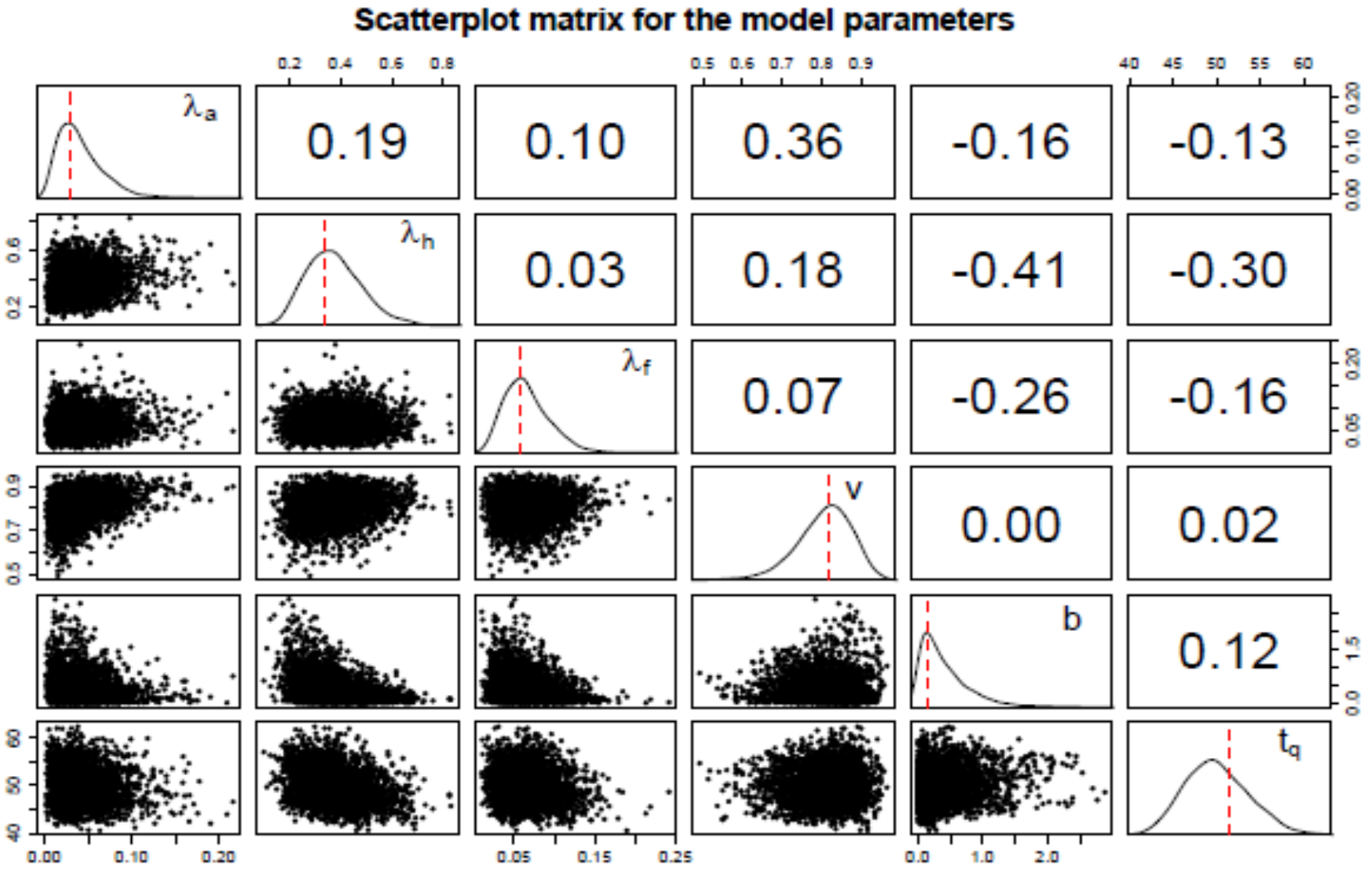
  L = infection process part

  x latent/fever/rash/isolation part

  x vaccination status part

- MCMC algorithm updates the model parameters and the unknown event times

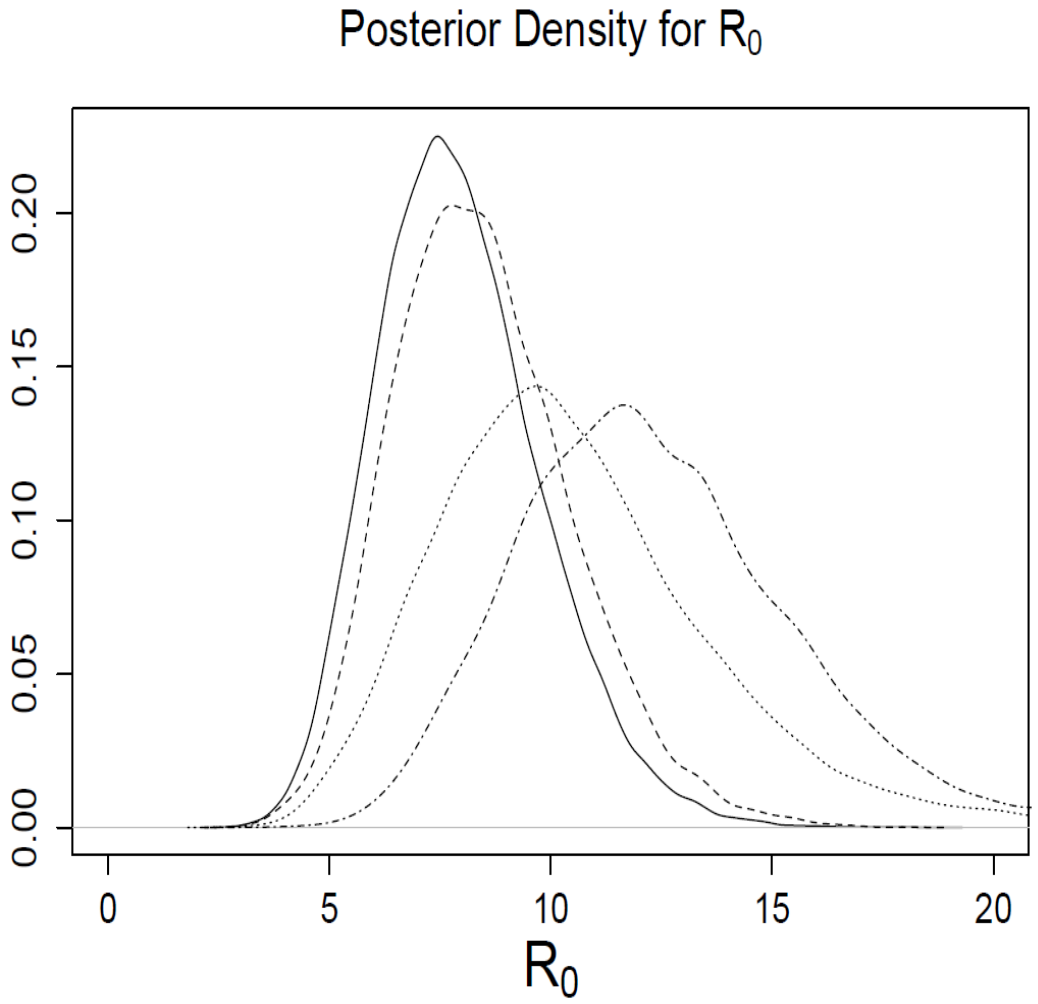
Results


Scatterplot matrix for the model parameters
λa
λh
λf
v
b
tq
0.19
0.10
0.36
-0.16
-0.13
0.03
0.18
-0.41
-0.30
0.07
-0.26
-0.16
0.00
0.02
0.12

Results

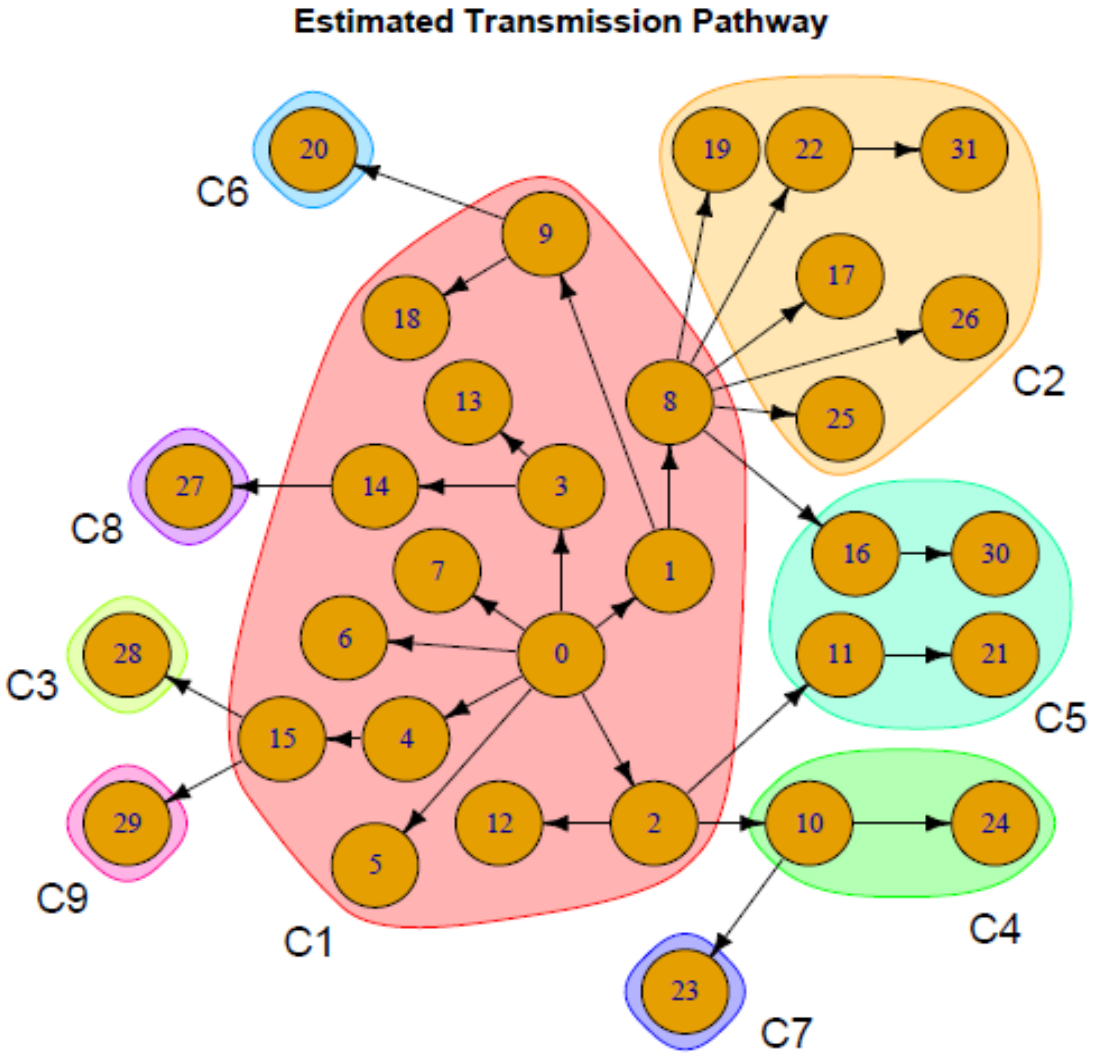
$E[R_0|\text{ data}] \approx 8$

Dashed lines show different choices for latent period etc



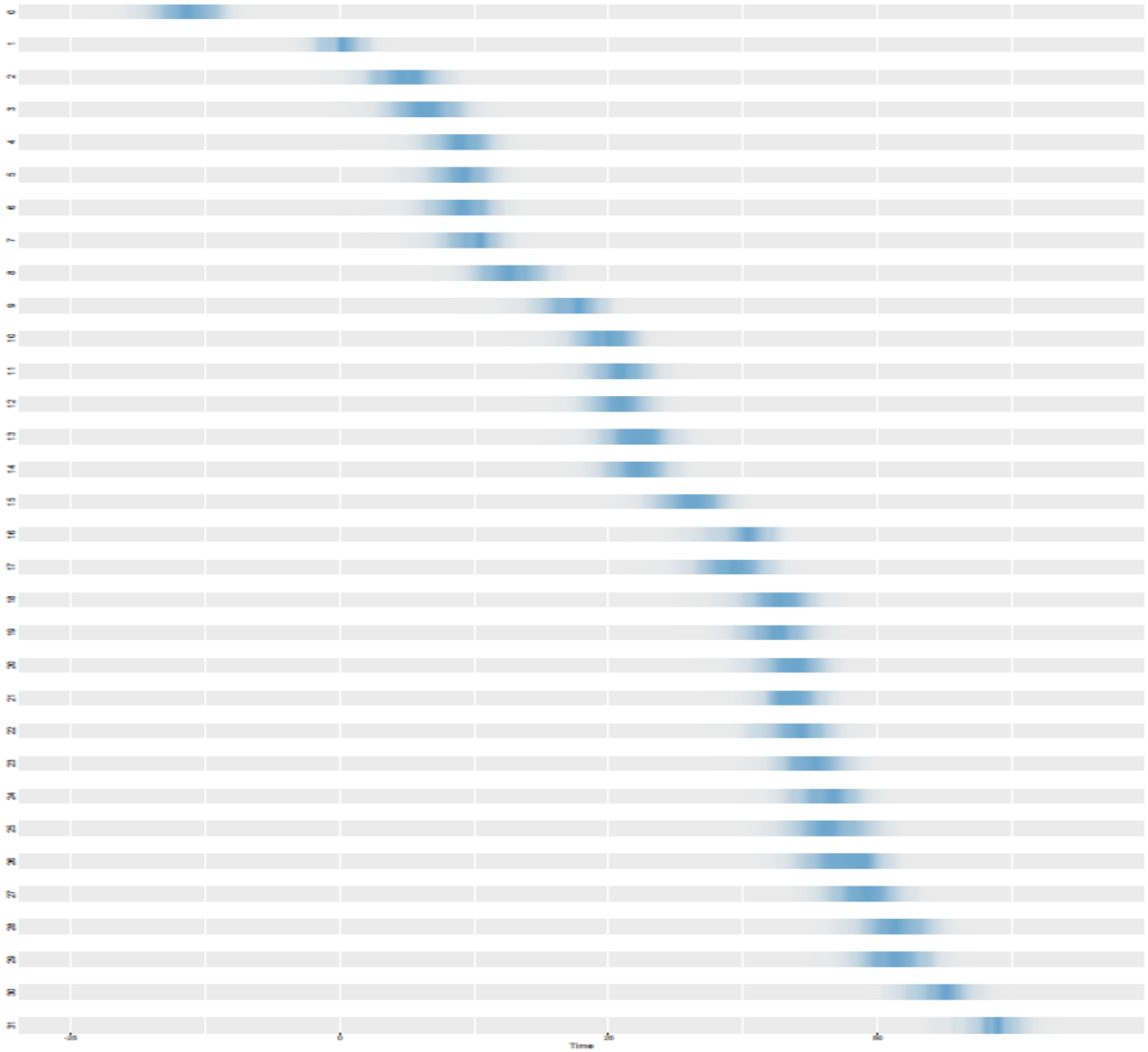
Results

Who infects whom



Results

Infection times



Model adequacy

- We use forward simulation to check the model fit
- The model parameters used in the simulation come from the posterior distribution, i.e. from the MCMC output

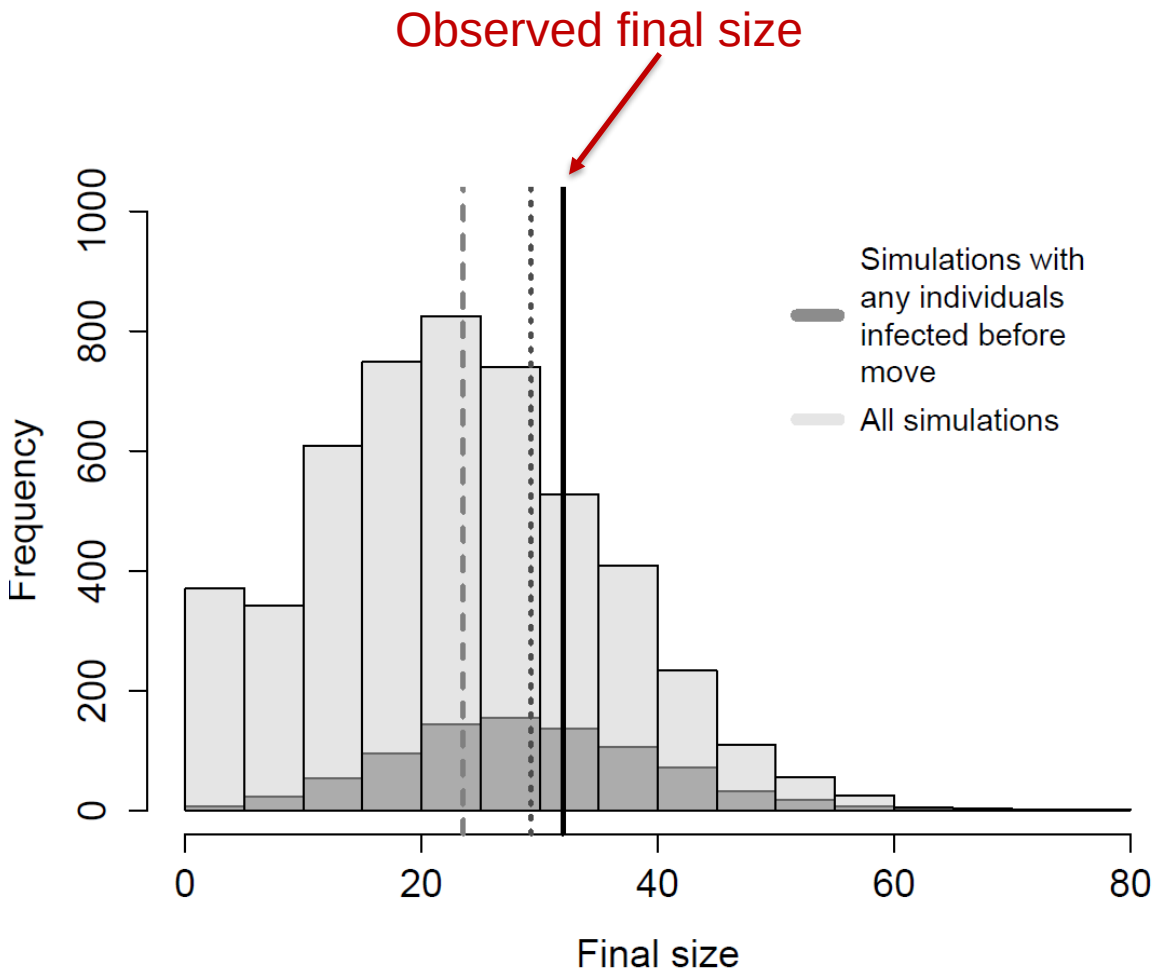
Model adequacy

Final size =

Number of cases

Better fit if

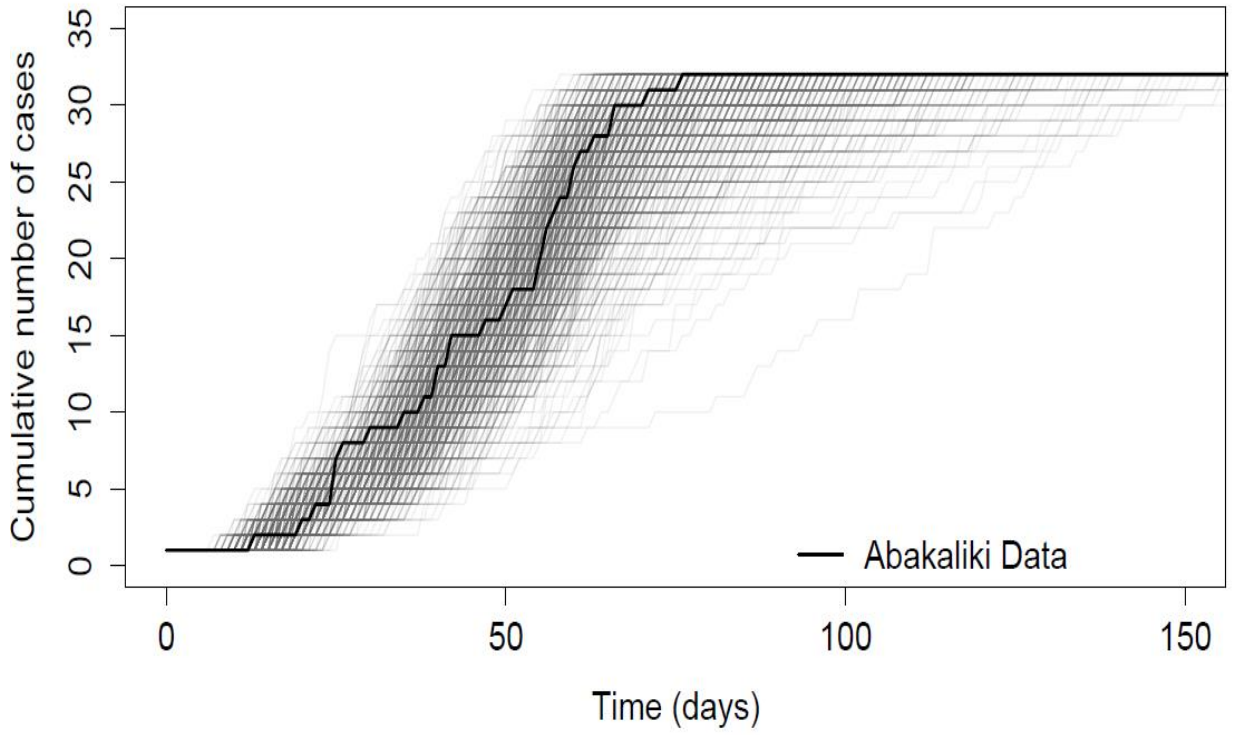
movers infected



Model adequacy

Time course of epidemic

(conditioned on final size)



Epidemics 19 (2017) 13–23

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 9 December 2016

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.

Contents

- 1. Smallpox: SEIR model with population structure
- 2. MRSA: whole-genome sequence data

MRSA: Data

- The data are taken from two Intensive Care Unit wards in a Thai hospital over a 3-month period.
- The pathogen of interest is Methicillin Resistant *Staphylococcus Aureus* (MRSA).
- Data tell us about MRSA colonisation status of patients.

MRSA: Data

- Patients underwent screening tests for Methicillin Resistant *Staphylococcus Aureus* (MRSA).
- For some patients, MRSA isolates were sequenced.

MRSA: Data

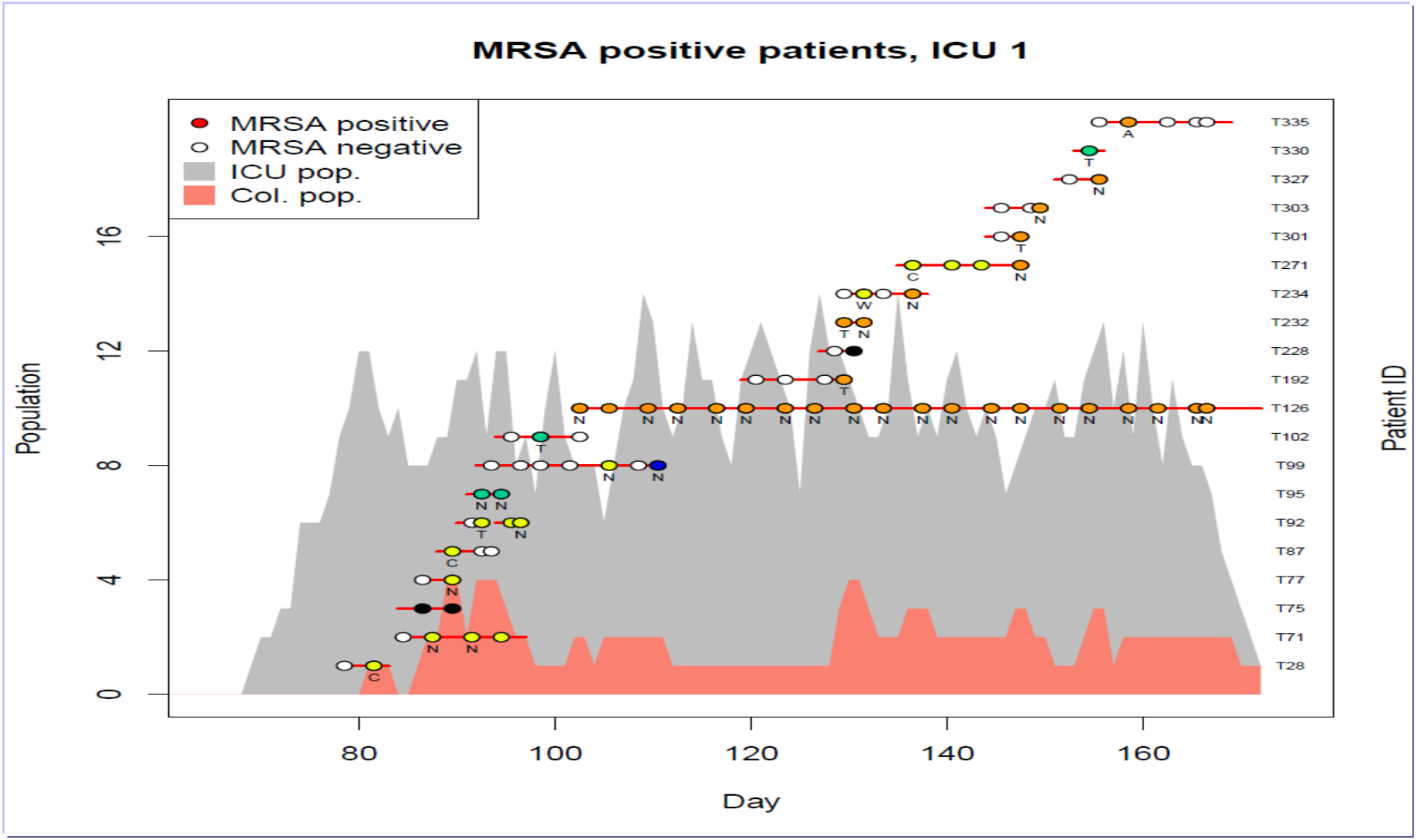
Individual-level data for each patient:

- Admission and discharge time
- Dates and outcomes of any screening tests
- Daily prescribing data (antibiotics etc.)
- Some sequenced isolates
- Other clinical information

MRSA: Data

	ICU 1	ICU 2
Ward type	Pediatric	Surgery
# patient episodes	170	114
# patients	169	98
# episodes with ≥ 1 +ve swab	20	29
Total # +ve swabs	51	89
Total # +ve swabs sequenced	43	40
Mean stay (days)	4.6	7.8

MRSA: Data


MRSA positive patients, ICU 1
MRSA positive
MRSA negative
ICU pop.
Col. pop.
Population
0
4
8
12
16
80
100
120
140
160
Day
Patient ID
T335
T330
T327
T303
T301
T271
T234
T232
T228
T192
T126
T102
T99
T95
T92
T87
T77
T75
T71
T28

MRSA: Model

Model for indirect transmission on ward:

- Each patient independently has probability p of being colonised on admission
- Positive patients are identified by diagnostic test with probability z (sensitivity)

MRSA: Model

Model for indirect transmission on ward:

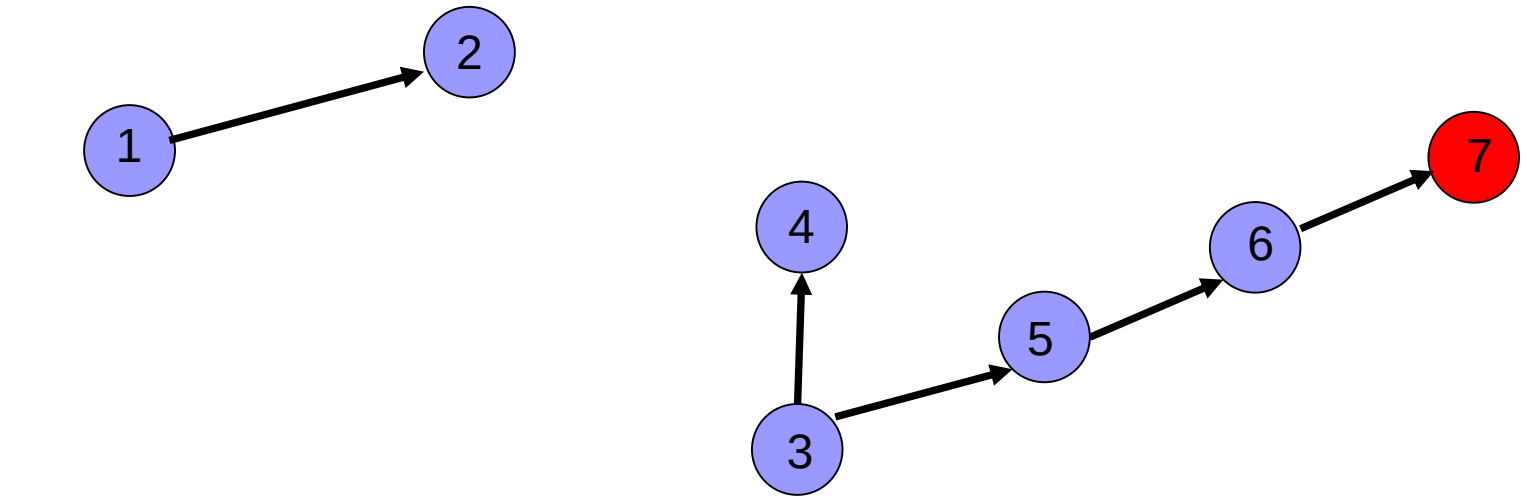
- Simple SI model for transmission whilst on the ward
- Here, I = “Colonised”, meaning has detectable levels of the pathogen
- No recovery in this model
- Keep track of who-colonises-whom

MRSA: Model

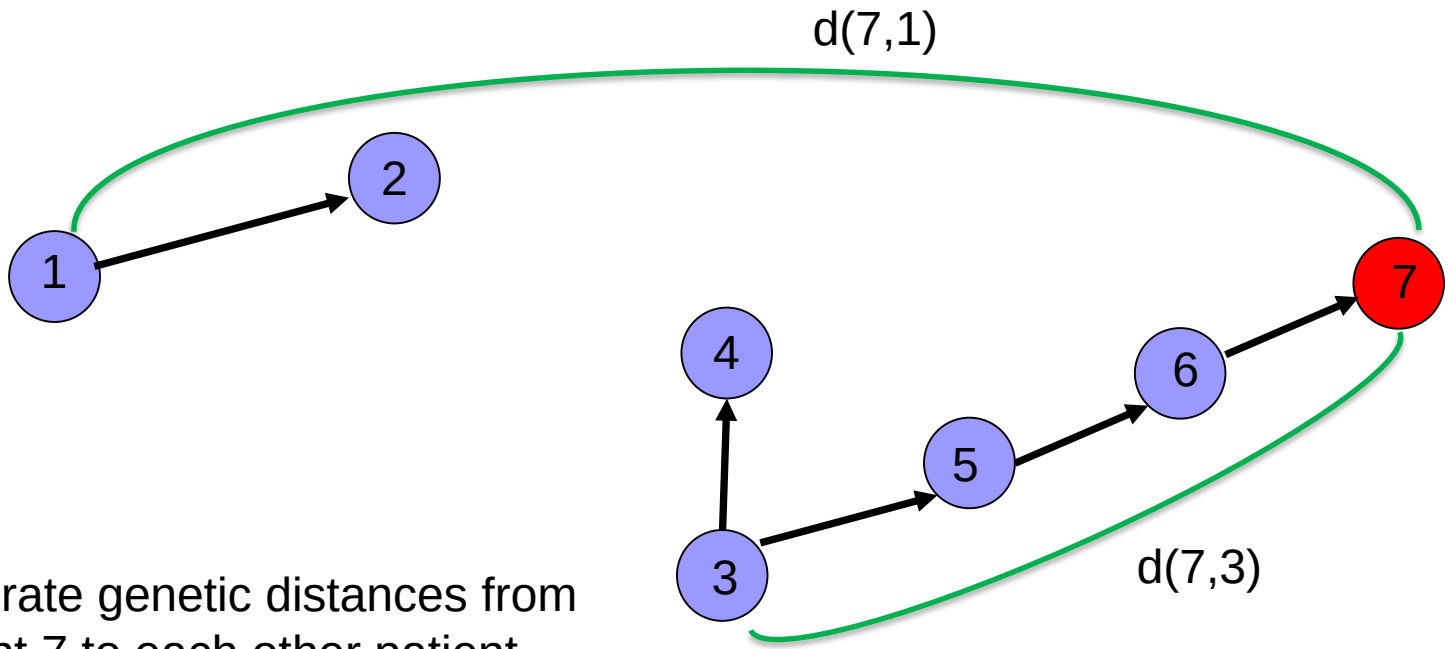
Model for genetic distances:

- For each newly-colonised patient we sample a genetic distance to all other previously-colonised patients
- The distribution of each sample depends on the relationship between patients in the transmission tree

MRSA: Model


1
2
3
4
5
6
7
Transmission link
Previously colonised patient
Newly colonised patient

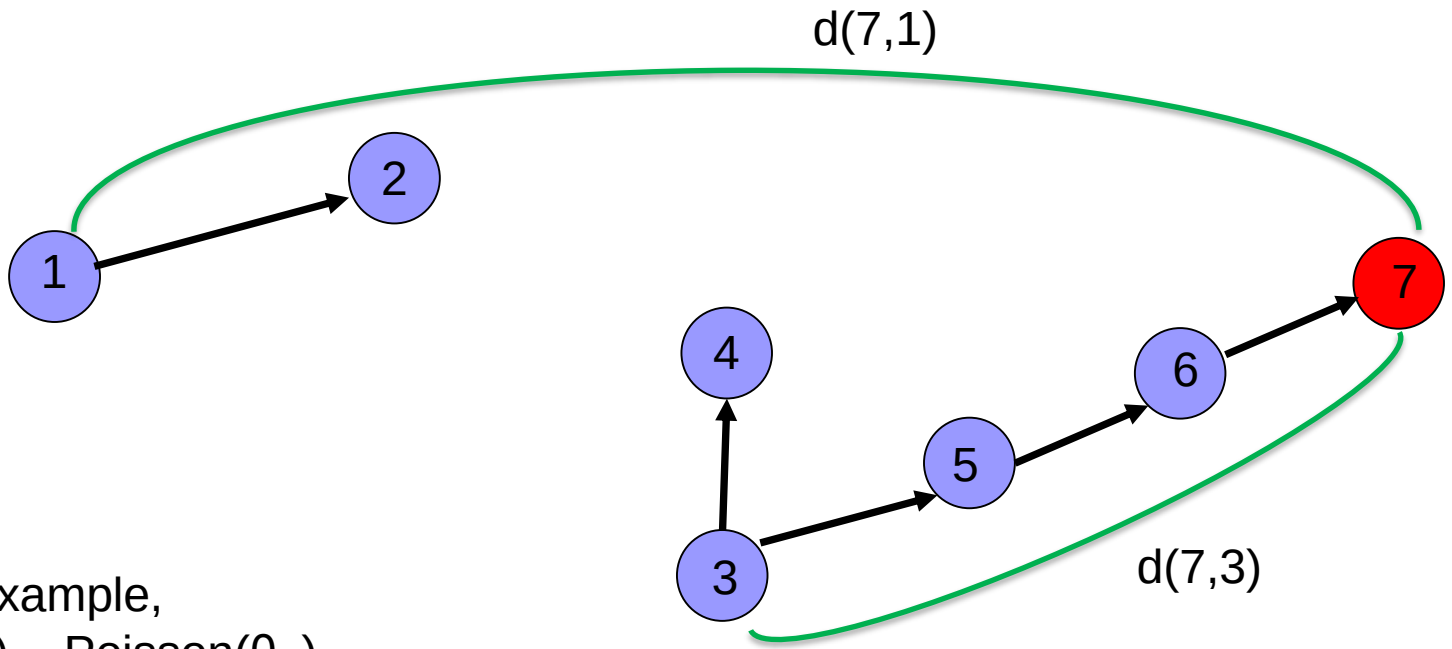
MRSA: Model



Generate genetic distances from patient 7 to each other patient

Genetic distances are drawn from distributions that depend on transmission relationship

MRSA: Model



For example,
$d(7,1) \sim \text{Poisson}(\theta_G)$

(7 and 1 not directly connected)

$d(7,3) \sim \text{Poisson}\ (d(3,5)+d(5,6)+d(6,7))$

(7 and 3 connected in a transmission chain)

MRSA: Model

Typically, genetic distance d(i,j) depends on

- Whether i and j are directly connected
- If connected, number of links in chain
- If connected, distances along chain

MRSA: inference problem

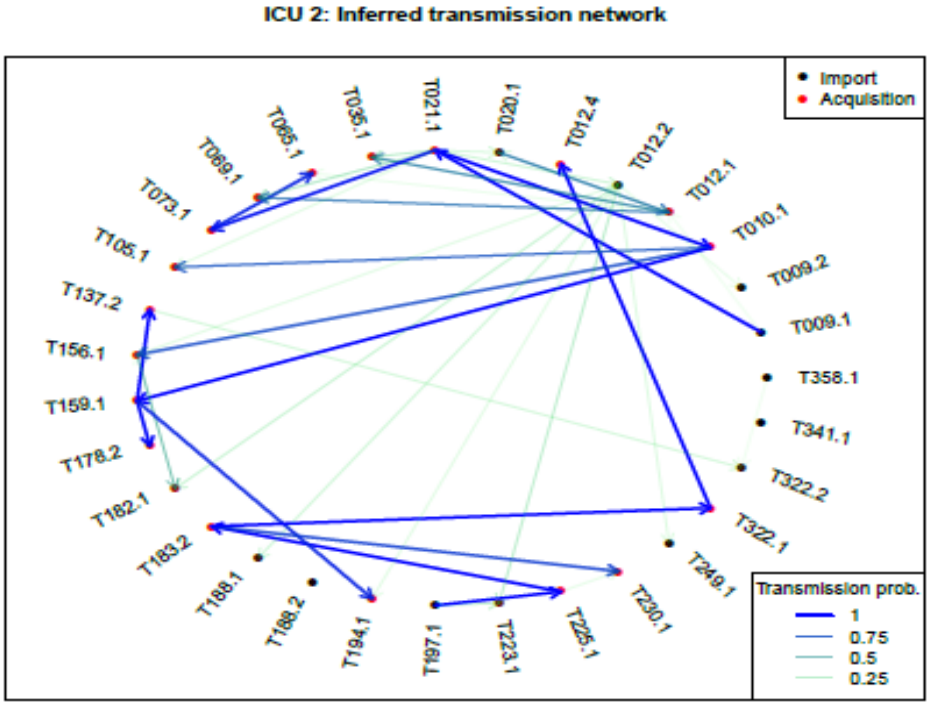
- Once again, likelihood is intractable
- Data augmentation: include the colonisation times and also who-colonises-whom
- This leads to a tractable augmented likelihood

MRSA: inference problem

- For MCMC algorithm, the challenging part is moving around the space of possible transmission trees, i.e. updating who-colonises-whom

MRSA: Results

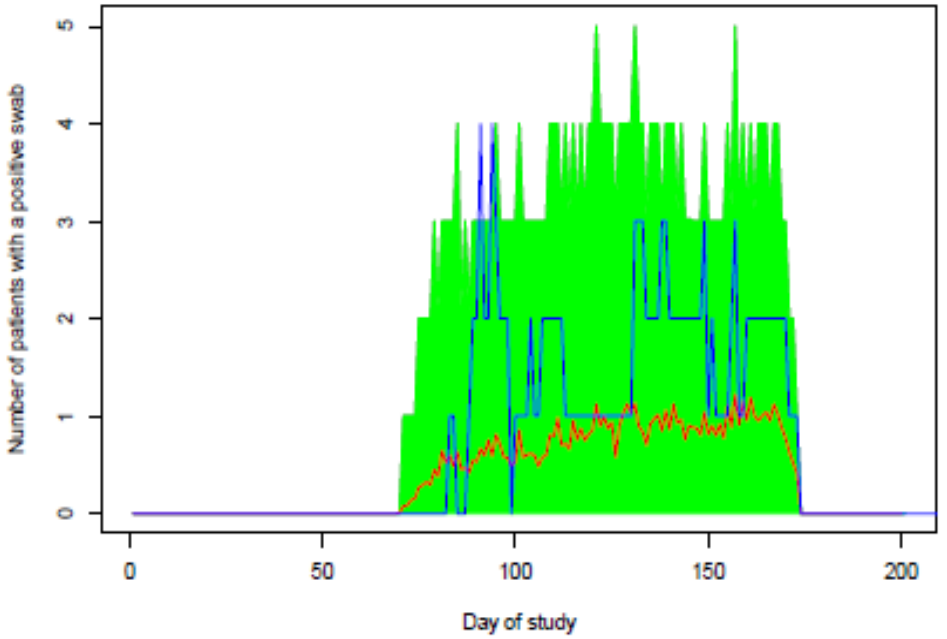
Can infer who-colonised-whom (and also who was colonised on admission) along with quantified degree of uncertainty



MRSA: Model adequacy

Can look at epidemiological aspects such as number of patients with a positive swab

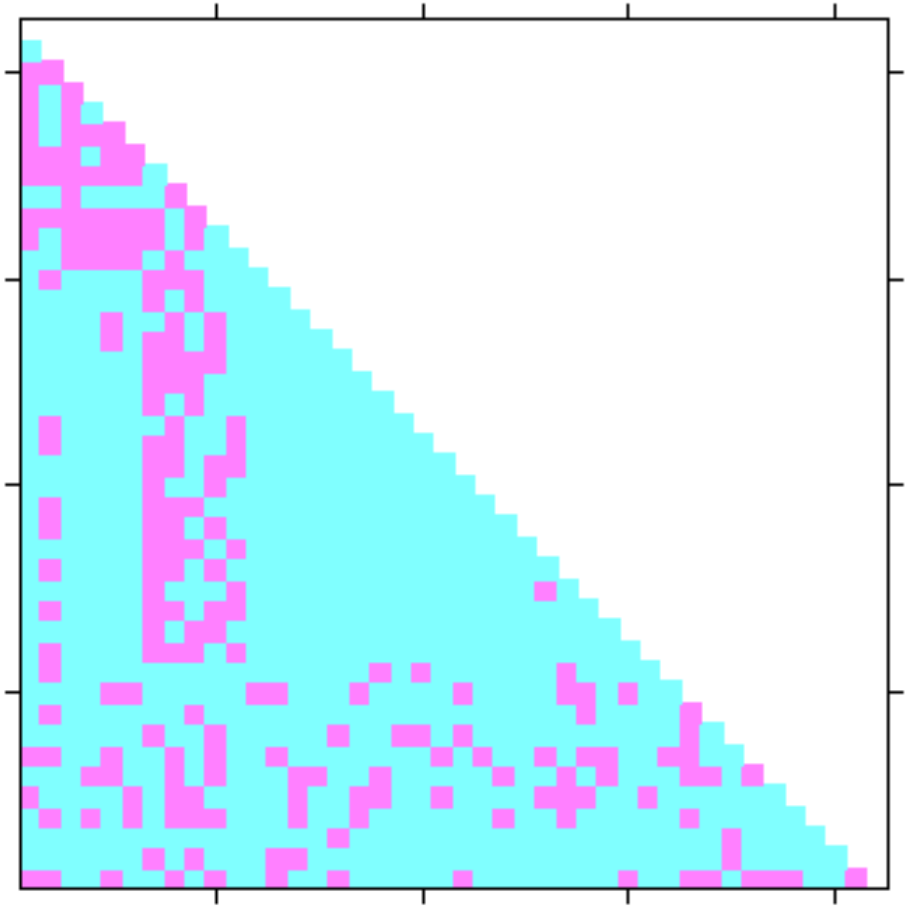
Blue = data
Red = mean of simulations
Green = 95% probability intervals



MRSA: Model adequacy

Can look at genetic aspects, e.g. are predicted genetic distances close to those observed?

Blue = data within 95% prediction
Pink = data outside 95% prediction



The Annals of Applied Statistics
2016, Vol. 10, No. 1, 395–417
DOI: 10.1214/15-AOAS898

RECONSTRUCTING TRANSMISSION TREES FOR COMMUNICABLE DISEASES USING DENSELY SAMPLED GENETIC DATA[1]

BY COLIN J. WORBY*,†,[2], PHILIP D. O'NEILL*, THEODORE KYPRAIOS*, JULIE V. ROBOTHAM‡, DANIELA DE ANGELIS‡,§,[3], EDWARD J. P. CARTWRIGHT¶, SHARON J. PEACOCK‖,[4] AND BEN S. COOPER**,††,[5]

*University of Nottingham**, *Harvard TH Chan School of Public Health*†, *Public Health England*‡, *MRC Biostatistics Unit*§, *Ipswich Hospital NHS Trust*¶, *University of Cambridge*‖, *University of Oxford*** *and Mahidol-Oxford Tropical Medicine Research Unit*††

Whole genome sequencing of pathogens from multiple hosts in an epidemic offers the potential to investigate who infected whom with unparalleled resolution, potentially yielding important insights into disease dynamics and the impact of control measures. We considered disease outbreaks in a setting with dense genomic sampling, and formulated stochastic epidemic models to investigate person-to-person transmission, based on observed genomic and epidemiological data. We constructed models in which the genetic distance between sampled genotypes depends on the epidemiological relationship between the hosts. A data-augmented Markov chain Monte Carlo algorithm was used to sample over the transmission trees, providing a posterior probability for any given transmission route. We investigated the predictive performance of our methodology using simulated data, demonstrating high sensitivity and specificity, particularly for rapidly mutating pathogens with low transmissibility. We then analyzed data collected during an outbreak of methicillin-resistant *Staphylococcus aureus* in a hospital, identifying probable transmission routes and estimating epidemiological parameters. Our approach overcomes limitations of previous methods, providing a framework with the flexibility to allow for unobserved infection times, multiple indepen-

Concluding comments

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