

# Developing an Antimicrobial Strategy for Sepsis in Malawi

-

Thesis submitted in accordance with the requirements of the Liverpool School of Tropical Medicine for the degree of Doctor in Philosophy by Joseph Michael Lewis

August 2019



# Contents

|   |           |
|---|-----------|
| <b>Preface</b>  | <b>9</b>  |
| <b>1 Introduction</b>   | <b>11</b> |
| 1.1 Chapter Overview . . . . .  | 13        |
| 1.2 Sepsis in sub-Saharan Africa . . . . .  | 13        |
| 1.3 ESBL-E in sub-Saharan Africa . . . . .  | 13        |
| 1.4 Conclusions . . . . .   | 13        |
| 1.5 Thesis overview . . . . .   | 13        |
| 1.6 Appendix . . . . .  | 13        |
| 1.7 References . . . . .  | 13        |
| <b>2 Methods</b>  | <b>15</b> |
| 2.1 Chapter Overview . . . . .  | 17        |
| 2.2 Study site . . . . .  | 17        |
| 2.3 Clinical Study . . . . .  | 17        |
| 2.4 Diagnostic Laboratory Procedures . . . . .                                    | 17        |
| 2.5 Molecular methods . . . . .   | 17        |
| 2.6 Bioinformatics . . . . .  | 17        |
| 2.7 Statistical Analysis . . . . .  | 17        |
| 2.8 Study Team . . . . .  | 17        |
| 2.9 Data Collection and Storage . . . . .   | 17        |
| 2.10 Ethical Approval, Consent and Participant Remuneration . . . . .             | 17        |
| <b>3 A clinical and microbiological description of sepsis in Blantyre, Malawi</b> | <b>19</b> |
| 3.1 Chapter overview . . . . .  | 20        |
| 3.2 Introduction and chapter aims . . . . .                                       | 20        |
| 3.3 Aims and Methods . . . . .  | 20        |
| 3.4 Results . . . . .   | 20        |
| 3.5 Discussion . . . . .  | 20        |

|          |   |           |
|----------|---|-----------|
| 3.6      | Conclusions and further work . . . . .                                      | 20        |
| 3.7      | Appendix . . . . .  | 20        |
| <b>4</b> | <b>Exploratory modelling of sepsis outcome</b>                              | <b>21</b> |
| <b>5</b> | <b>ESBL-E carriage in Malawian adults in health and disease</b>             | <b>23</b> |
| 5.1      | Chapter Overview . . . . .  | 24        |
| 5.2      | Introduction and chapter aims . . . . .                                     | 24        |
| 5.3      | Methods . . . . .   | 24        |
| 5.4      | Results . . . . .   | 24        |
| 5.5      | Discussion . . . . .  | 24        |
| 5.6      | Conclusions and further work . . . . .                                      | 24        |
| <b>6</b> | <b>Whole genome sequencing of ESBL <i>E. coli</i> carriage isolates</b>     | <b>25</b> |
| 6.1      | Chapter overview . . . . .  | 27        |
| 6.2      | Methods . . . . .   | 27        |
| 6.3      | Results . . . . .   | 27        |
| 6.4      | Discussion . . . . .  | 27        |
| 6.5      | Appendix . . . . .  | 27        |
| <b>7</b> | <b>Genomics I</b>   | <b>29</b> |
| <b>8</b> | <b>Continuous time Markov models to understand ESBL-E carriage dynamics</b> | <b>31</b> |
| 8.1      | Chapter Overview . . . . .  | 31        |
| 8.2      | Introduction and chapter aims . . . . .                                     | 31        |
| 8.3      | Methods . . . . .   | 31        |
| 8.4      | Results . . . . .   | 38        |
| 8.5      | Discussion . . . . .  | 38        |
| 8.6      | Conclusion and further work . . . . .                                       | 38        |
|          | <b>References</b>   | <b>41</b> |

# List of Tables



# List of Figures

|     |  |    |
|-----|--|----|
| 8.1 | Two state ESBL-E model . . . . .   | 34 |
| 8.2 | Parameter estimates from Markov models . . . . .                           | 39 |
| 8.3 | Predicted proprtion of ESBL-E positive samples, stratified by arm. . . . . | 40 |





# Preface

Placeholder



# Chapter 1

## Introduction

Placeholder



## 1.1 Chapter Overview

## 1.2 Sepsis in sub-Saharan Africa

### 1.2.1 Search strategy

### 1.2.2 Defining sepsis

### 1.2.3 Applicability of sepsis-3 definitions in sub-Saharan Africa

### 1.2.4 Sepsis epidemiology in sub-Saharan Africa

#### 1.2.4.1 Incidence

#### 1.2.4.2 Risk factors: the sepsis population in sub-Saharan Africa

#### 1.2.4.3 Outcomes

### 1.2.5 Sepsis aetiology in sub-Saharan Africa

#### 1.2.5.1 Bacterial zoonoses, Rickettsioses and arboviruses

#### 1.2.5.2 HIV opportunistic infections: PCP, histoplasmosis and cryptococcal disease

### 1.2.6 Sepsis management

#### 1.2.6.1 Early goal directed therapy

#### 1.2.6.2 Evidence to guide antimicrobial therapy in sSA

#### 1.2.6.3 Evidence to guide intravenous fluid therapy in sub-Saharan Africa

## 1.3 ESBL-E in sub-Saharan Africa

### 1.3.1 Search strategy

### 1.3.2 Introduction: definition and classification of ESBL-E

### 1.3.3 Global molecular epidemiology of ESBL-E: an overview

#### 1.3.3.1 1980s-1990s: First identification of ESBL in nosocomial pathogens

#### 1.3.3.2 1990s-2010s: Emergence and globalisation of CTX-M



## Chapter 2

# Methods

Placeholder





## 2.1 Chapter Overview

## 2.2 Study site

### 2.2.1 Malawi

### 2.2.2 Queen Elizabeth Central Hospital

### 2.2.3 Participating Laboratories

#### 2.2.3.1 Malawi-Liverpool-Wellcome Clinical Research Programme

#### 2.2.3.2 Malawi College of Medicine Tuberculosis Laboratory

#### 2.2.3.3 Wellcome Trust Sanger Institute

## 2.3 Clinical Study

### 2.3.1 Entry Criteria

### 2.3.2 Study Visits and Patient Sampling

#### 2.3.2.1 Enrollment assessment and first six hours

#### 2.3.2.2 Subsequent visits

#### 2.3.2.3 Blood, urine, and stool, sputum and CSF collection

#### 2.3.2.4 Imaging: chest x-ray and ultrasound scanning

### 2.3.3 Outcomes and sample size calculations

## 2.4 Diagnostic Laboratory Procedures

### 2.4.1 Point of care diagnostics

### 2.4.2 Laboratory diagnostics

#### 2.4.2.1 Haematology and biochemistry

#### 2.4.2.2 Aerobic blood and CSF culture

#### 2.4.2.3 Mycobacterial blood culture

#### 2.4.2.4 Sputum Xpert



## Chapter 3

# A clinical and microbiological description of sepsis in Blantyre, Malawi

Placeholder

### 3.1 Chapter overview

### 3.2 Introduction and chapter aims

### 3.3 Aims and Methods

### 3.4 Results

#### 3.4.1 Study population

#### 3.4.2 Symptoms and health-seeking behaviour

#### 3.4.3 Admission physiology and laboratory investigations

#### 3.4.4 Aetiology

#### 3.4.5 Treatment

#### 3.4.6 Outcome

#### 3.4.7 Determinants of mortality

### 3.5 Discussion

#### 3.5.1 Demographics and outcome: significant longer-term mortality

#### 3.5.2 Aetiology: TB dominates as a cause of sepsis

#### 3.5.3 Determinants of 28-day mortality: an expanded role for TB therapy?

#### 3.5.4 Limitations

### 3.6 Conclusions and further work

### 3.7 Appendix

## Chapter 4

# Exploratory modelling of sepsis outcome



## Chapter 5

# ESBL-E carriage in Malawian adults in health and disease

Placeholder

## **5.1 Chapter Overview**

## **5.2 Introduction and chapter aims**

## **5.3 Methods**

## **5.4 Results**

### **5.4.1 Study population**

### **5.4.2 Exposures during the study period**

### **5.4.3 ESBL-E colonisation**

### **5.4.4 Associations of ESBL colonisation**

## **5.5 Discussion**

### **5.5.1 Limitations**

## **5.6 Conclusions and further work**



## Chapter 6

# Whole genome sequencing of ESBL *E. coli* carriage isolates

Placeholder



## 6.1 Chapter overview

## 6.2 Methods

### 6.2.1 Bioinformatic pipeline

### 6.2.2 Global *E. coli* collection

### 6.2.3 Statistical analysis

## 6.3 Results

### 6.3.1 Samples and quality control

### 6.3.2 Phylogroup, MLST and core genome phylogeny of study isolates

### 6.3.3 Study isolates in a global context

### 6.3.4 Antimicrobial resistance determinants

#### 6.3.4.1 $\beta$ -lactam resistance

#### 6.3.4.2 Quinolone resistance

#### 6.3.4.3 Aminoglycoside resistance

#### 6.3.4.4 Chloramphenicol, co-trimoxazole, tetracycline and other resistance determinants

#### 6.3.4.5 Clustering and lineage association of AMR determinants

### 6.3.5 Plasmid replicons

### 6.3.6 Testing metadata associations: SNP distance, hierBAPS sequence clusters and ESBL-clusters

#### 6.3.6.1 Hierarchical BAPS clustering of core gene pseudosequences

#### 6.3.6.2 ESBL-clusters

#### 6.3.6.3 Assessing for healthcare-associated lineages

#### 6.3.6.4 Assessing for within-patient conservation of lineage or MGE

## 6.4 Discussion



## Chapter 7

# Genomics I



## Chapter 8

# Continuous time Markov models to understand ESBL-E carriage dynamics

### 8.1 Chapter Overview

whatevs bru

### 8.2 Introduction and chapter aims

### 8.3 Methods

In the broadest sense when constructing a model, our aim is to estimate the most likely values of the parameters of the model,  $\theta$ , given the data we have,  $x$ . The starting point for estimating likely parameter values, given a choice of model, is usually the *likelihood*: this is the probability of the data, given a set of parameter values. In standard probability notation, this is written as  $P(x|\theta)$ . In fact, this is not the quantity we are interested in; we would like to know  $P(\theta|x)$ : the probability of the parameter values, given the data. Both frequentist and Bayesian modelling approaches provide methods to estimate this quantity, but the starting point for both is the likelihood,  $P(x|\theta)$ , because it is usually much more straightforward to derive an expression for  $P(x|\theta)$  rather than  $P(\theta|x)$ . I will here derive a general likelihood for a two state intermittently observed process; in order to use this likelihood, it is necessary to make some assumptions about the data generating process. I have chosen to use a Markov

model, and I will then derive the likelihood for this model, describe how covariates will be incorporated, describe how the model was fit - the process taking us from the likelihood to the most likely parameter values - and finally how goodness of fit was assessed.

### 8.3.1 General form of likelihood

First, I derive a general expression for the likelihood of a two-state intermittently observed process without making any assumptions about the model structure or functional form. Assume we have  $N$  participants with any given participant  $n$  in a state  $S_n(t)$  at time  $t$ : either ESBL-E colonised ( $S_n(t) = 1$ ) or uncolonised ( $S_n(t) = 0$ ). For each participant  $n$  we have a number of measurements of  $S_n(t)$  at a number of time points. The number of measurements varies for each participant, and can be denoted by  $j_n$ , making the time of measurements  $t_{j_n}^n$  for participant  $n$ ; and so for participant  $n$  we know the  $j_n$  values  $S_n(t_{j_n}^n)$ .

To arrive at the likelihood for these observations, consider first the simplest situation that we have: the measurements of ESBL status at two time points,  $t_A$  and  $t_B$  for a single participant,  $n$ . The likelihood we wish to calculate, in words, is the probability of the participant being in the second observed state at time  $t_B$ , given they were in the first state at  $t_A$  and given the parameters of the model,  $\theta$ . Or, mathematically:

$$P(S_n(t_B)|S_n(t_A), \theta) \quad (8.1)$$

Assuming all the observations are independent, the probability of all of the states we have observed for this participant is the product of all the probabilities of the individual states:

$$\prod_{k=2}^{j_n} P(S_n(t_k^n)|S_n(t_{k-1}^n), \theta) \quad (8.2)$$

And the probability of observing the data we have is then simply the product of the probability of all the individual transitions:

$$\prod_{n=1}^N \prod_{k=2}^{j_n} P(S_n(t_k^n)|S_n(t_{k-1}^n), \theta) \quad (8.3)$$

This is the quantity that we wish to calculate: the likelihood for the observed data,  $P(x|\theta)$ .



### 8.3.2 Markov model likelihood

In order to calculate the likelihood, we need to make some assumptions about the data generating process. In this case, I have chosen to use a Markov model. Markov models are defined by instantaneous transition probabilities, analogous to the hazard of death in a survival model, which is a simple two-state Markov system. Unlike a survival model (where it is not possible to move from the death state to alive), a general Markov model is defined by a transition hazard from each state to each other state in the system. These are traditionally expressed as a  $\mathbf{Q}$  matrix of instantaneous transition intensities (assuming a two-state system):

$$\mathbf{Q}(t) = \begin{pmatrix} q_{00}(t) & q_{01}(t) \\ q_{10}(t) & q_{11}(t) \end{pmatrix} \quad (8.4)$$

Where  $q_{ij}$  represents the instantaneous transition intensity from state  $i$  to state  $j$ . The rows of the  $\mathbf{Q}$ -matrix must sum to 1 (every participant has to be in one state or another), so if we define the hazard of ESBL-E acquisition to be  $\lambda$  and the hazard of ESBL-E loss to be  $\mu$  (8.1), the  $\mathbf{Q}$ -matrix becomes, in our case:

$$\mathbf{Q}(t) = \begin{pmatrix} -\lambda(t) & \lambda(t) \\ \mu(t) & -\mu(t) \end{pmatrix} \quad (8.5)$$

However, we are not interested in the  $\mathbf{Q}$ -matrix *per se* but rather the probability  $p_{ij}$  of starting in state  $i$  at time 0 and being in state  $j$  at time  $t$ ; this can be written in matrix notation as  $\mathbf{P}(t)$  and is related to  $\mathbf{Q}(t)$  by the differential equations:

$$\frac{d\mathbf{P}(t)}{dt} = \mathbf{Q}(t) \cdot \mathbf{P}(t) \quad (8.6)$$

Where  $\mathbf{Q}(t) \cdot \mathbf{P}(t)$  is the matrix product of  $\mathbf{Q}(t)$  and  $\mathbf{P}(t)$ . In order to evaluate  $\mathbf{P}(t)$ , therefore we need to solve this system of differential equations. However, there are limited situations in which these equations have analytic solutions. If the system has time constant or piecewise constant  $\mathbf{Q}$  matrix the matrix exponential is a solution:

$$\mathbf{P}(t) = e^{\mathbf{Q}t} \quad (8.7)$$

However, there is no reason to suspect particularly that the effect of covariates on ESBL-E carriage (e.g. antimicrobials) would be stepwise constant and so a more flexible model is needed. For general time-varying transition intensities, there is no analytic solution to the

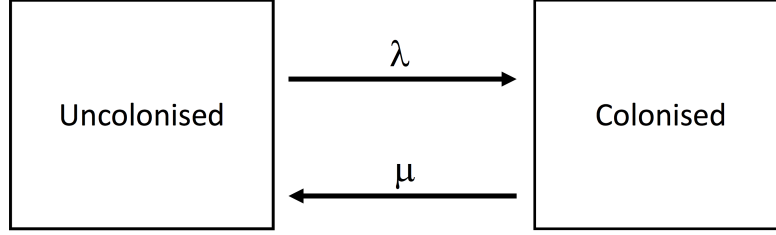


Figure 8.1: Two state ESBL-E model showing instantaneous hazard of ESBL-E acquisition ( $\lambda$ ) or loss ( $\mu$ ).

above equations. However, all is not lost: we can express the likelihood in terms of the differential equations defined by the equations above and solve them numerically in order to calculate the likelihood. The matrix notation above can be simplified, assuming that the system starts in state 1 or 0:

$$\frac{dP_0(t)}{dt} = -\lambda(t)P_0(t) + \mu(t)P_1(t) \quad (8.8)$$

$$\frac{dP_1(t)}{dt} = \lambda(t)P_0(t) - \mu(t)P_1(t) \quad (8.9)$$

Where  $P_i(t)$  is the probability of being in state  $i$  at time  $t$ . Numerical ordinary differential equation (ODE) solvers can quickly solve these equations to calculate, for example,  $P(S_n(t_B)|S_n(t_A), \theta)$  from the simplest example above: the probability that a participant  $n$  at time  $t_B$  is in a given state, given that they were in state  $S_n(t_A)$  at time  $t_A$ , and given the parameters  $\theta$ . This calculation can be completed for all measurements and participants, resulting in the likelihood of the system,  $P(x|\theta)$ .

In order to use this model for inference, two questions must be addressed: first, how to incorporate time-varying covariates; and second, how to practically fit the model. I address each of these questions below.

### 8.3.3 Incorporating covariates: a proportional hazard model

I have chosen to incorporate covariates using a proportional-hazards model, following both Marshall and Jones[ref] and the *msm* package in R. In this model the transmission intensities

become:

$$\lambda(t) = \lambda_0 \exp(\beta_0 x_0(t) + \beta_1 x_1(t) + \dots \beta_m x_m(t)) \quad (8.10)$$

$$\mu(t) = \mu_0 \exp(\alpha_0 x_0(t) + \alpha_1 x_1(t) + \dots \alpha_m x_m(t)) \quad (8.11)$$

Where the  $x_k, k = 1, 2, \dots, m$  are the  $m$  time-varying covariates in the model and the coefficients  $\alpha_k$  and  $\beta_k$  are the coefficients of these covariates; these have a straightforward interpretation in that the exponential,  $e^{\alpha_k}$  or  $e^{\beta_k}$  can be interpreted as a hazard ratio, as per a simple survival model.

An assumption then needs to be made about the functional form of  $x_m$ . In a stepwise-constant covariate model in which an exposure occurs between  $t_A$  and  $t_B$ ,  $x(t)$  would take the value 1 for all  $t_A \leq t \leq t_B$  and 0 at other times, meaning that the effect of the exposure does not persist once it ceases. Though this may be plausible for some exposures, it seems possible that antimicrobial exposure (for example) might have a longer lasting effect (perhaps mediated through the microbiota); in order to explore this possibility, it is necessary to decide on a flexible, plausible, functional form that such an effect might take. I have decided to use an exponential function, such that:

$$x_k(t) = \begin{cases} 0 & \text{if } t < t_A \\ 1 & \text{if } t_A \leq t \leq t_B \\ \exp \frac{-(t - t_B)}{\gamma_k} & \text{if } t > t_B \end{cases} \quad (8.12)$$

Where the parameter  $\gamma_k$  is a model parameter for each of the covariates, to be estimated from the data, and is related to the half life,  $t_{\frac{1}{2}}^k$  of the decay of the effect of the exposure by:

$$t_{\frac{1}{2}}^k = \gamma_k \ln(2) \approx 0.69\gamma_k \quad (8.13)$$

From the definition of the half life of an exponential decay process. This parameterisation has the advantage that the data can fit the size of the parameters  $\gamma_k$ ; if the data are more inkeeping with a stepwise effect of the covariates, then a small ( $\ll 1$ )  $\gamma$  would approximate a step function and this could be fit by the model. Alternatively a larger would result in the effect of the covariate persisting after exposure, but decaying over time. This allows us to test the hypothesis that antimicrobial exposure (for example) has an effect that persists once exposure finishes, by both the magnitude of the fitted  $\gamma_k$ , and comparing stepwise-constant covariate models to models incorporating the  $\gamma_k$  parameters.

### 8.3.4 Building and fitting models

The Bayesian probabilistic programming language *Stan* incorporates an ordinary differential equation solver, and will allow the fitting of the model in a Bayesian framework. In this framework, Bayes' rule allows us to estimate our probability distribution of interest,  $P(\theta|x)$ , called the *posterior* in the Bayesian framework, as long as we provide a *prior*, encoding our prior beliefs about the values of the parameters. *Stan* then uses the No-U-Turn Sampler (NUTS) implementation of Markov-chain Monte-Carlo (MCMC) sampling to sample from the posterior to provide  $P(\theta|x)$ . It can be shown that, given infinite chain length, MCMC estimates are guaranteed to be unbiased samples from the posterior; when they are providing unbiased samples the chains have said to converged. Unfortunately there is no diagnostic test that guarantees convergence, rather tests that are necessary but not sufficient to ensure convergence: running multiple chains from different starting points with examination of traceplots to show within and between mixing of chains, and the  $\hat{R}$  statistic, which measures mixing of the two halves of an MCMC chain. At convergence,  $\hat{R}$  should be close to 1. In addition, divergences - failure in the NUTS sampler - can be indicative of difficult topography in the posterior at the area where the divergences occur, and suggest that parameter estimates may be biased, and are flagged by *Stan*. All of these tests were used to diagnose convergence.

Two decisions must be made in order to fit the model: covariates must be chosen to include and priors specified. Models were built sequentially, starting from the simplest possible, then adding complexity:

- *Model 1:* Composite antibacterial variable (includes all antibacterials) and hospitalisation variable as explanatory variables, both included with stepwise constant effect and no post exposure effect.
- *Model 2:* As per model 1 except antibacterial exposure modelled with decaying post-exposure effect.
- *Model 3:* Hospitalisation, TB therapy and co-trimoxazole exposure all modelled as stepwise constant covariates. All other antibacterials included in a composite variable with decaying post-exposure effect.
- *Model 4:* Hospitalisation, TB therapy and co-trimoxazole exposure all modelled as stepwise constant covariates; ceftriaxone, ciprofloxacin and amoxicillin exposure included in a composite variable with decaying post-exposure effect, with  $\gamma$  allowed to vary for each agent.

Weakly informative priors were used. A normally distributed prior centered at 0 with standard deviation 2 was used for all the  $\alpha$  and  $\beta$  parameters. A parameter value of 2 corresponds to

a hazard ratio of 7.4; it would be surprising if any effect is greater than this so this could be argued to be a weakly informative prior. Normally distributed priors centered at 0 with standard deviation 0.2 were used for the  $\mu$  and  $\lambda$  parameters; in a model with no covariates, the inverse of these parameters are the mean times that an individual would remain in the colonised or uncolonised states, respectively, so a value of 0.2 corresponds to a mean state occupancy time of 50 days. A normally distributed prior centred at 0 and with standard deviation 50 days was used for all  $\gamma$  parameters.

The *Stan* code for the models is given in the appendix to this chapter. Four chains were run in each case, with a warmup of 500 iterations and run for 1000 iterations in total. Convergence was assessed using the diagnostics described above. *Stan* v2.19 was used to sample from the posterior, accessed via Rstan v2.19.2, and run on the Wellcome Sanger Institute computing cluster under Linux red hat v7.6, running R v3.5.3, and using 4 cores per run. Posterior samples were brought to my local machine and further analyses undertaken with R3.6.0.

### 8.3.5 Assessing goodness of fit

Model goodness of fit was assessed in two ways; first, by graphical posterior predictive checks. The posterior parameter estimates for each MCMC draw were used to estimate a predicted probability for each data point then sampling from a bernoulli distribution generated predicted state occupancy. These were plotted against actual state occupancy, stratified by arm, to visualise the goodness of fit of the model, and to compare between models. Second, models were compared using leave-one-out cross validation, as implemented in the *loo* v2.1.0 package in R. This estimates the out-of sample predictive ability of the model by estimating a quantity called the expected log pointwise predictive density (*ELPD*) essentially the log of the likelihood for a new, unseen dataset conditional on the current data. This quantity is estimated using leave-one-out cross validation to produce and estimate of the *ELPD* - *ELPD<sub>loo</sub>*. The standard error of *ELPD<sub>loo</sub>* for a model is also calculated and so two models can be compared by comparing the *ELPD<sub>loo</sub>* difference and standard error; if the difference is greater than twice the standard error (i.e. a 95% confidence interval, assuming normality) we can be confident that one model would be expected to have greater out-of-sample predictive ability than the other. Because this technique estimates out-of-sample predictive ability it naturally incorporates a penalty for including multiple parameters and hence overfitting, as an overfit model would be expected to have worse out of sample predictive ability and hence lower *ELPD<sub>loo</sub>*.

This process was repeated, replacing the ESBL-E state first with *E. coli* (coding presence of *E.coli* at any time point as 1 and absence as 0) and then with *Klebsiella pneumoniae*, and then *add genomics chat here*

## 8.4 Results

All models seemed to converge within the 1000 iterations;  $\hat{R}$  was less than 1.1 for all parameters and all traceplots showed good mixing of chains (see appendix to this chapter). There was a computational cost to increasing the number of parameters: model one took 3.5hr to fit, model two 13.7hr, model three 17.1hr and model four xxxx. The parameter estimates for the models are shown in fig xxx

### 8.4.1 Exploring the effect of antimicrobial exposure on carriage

### 8.4.2 Exploring bacterial species differences in carriage dynamics

### 8.4.3 Exploring genotypic differences in carriage dynamics

## 8.5 Discussion

## 8.6 Conclusion and further work

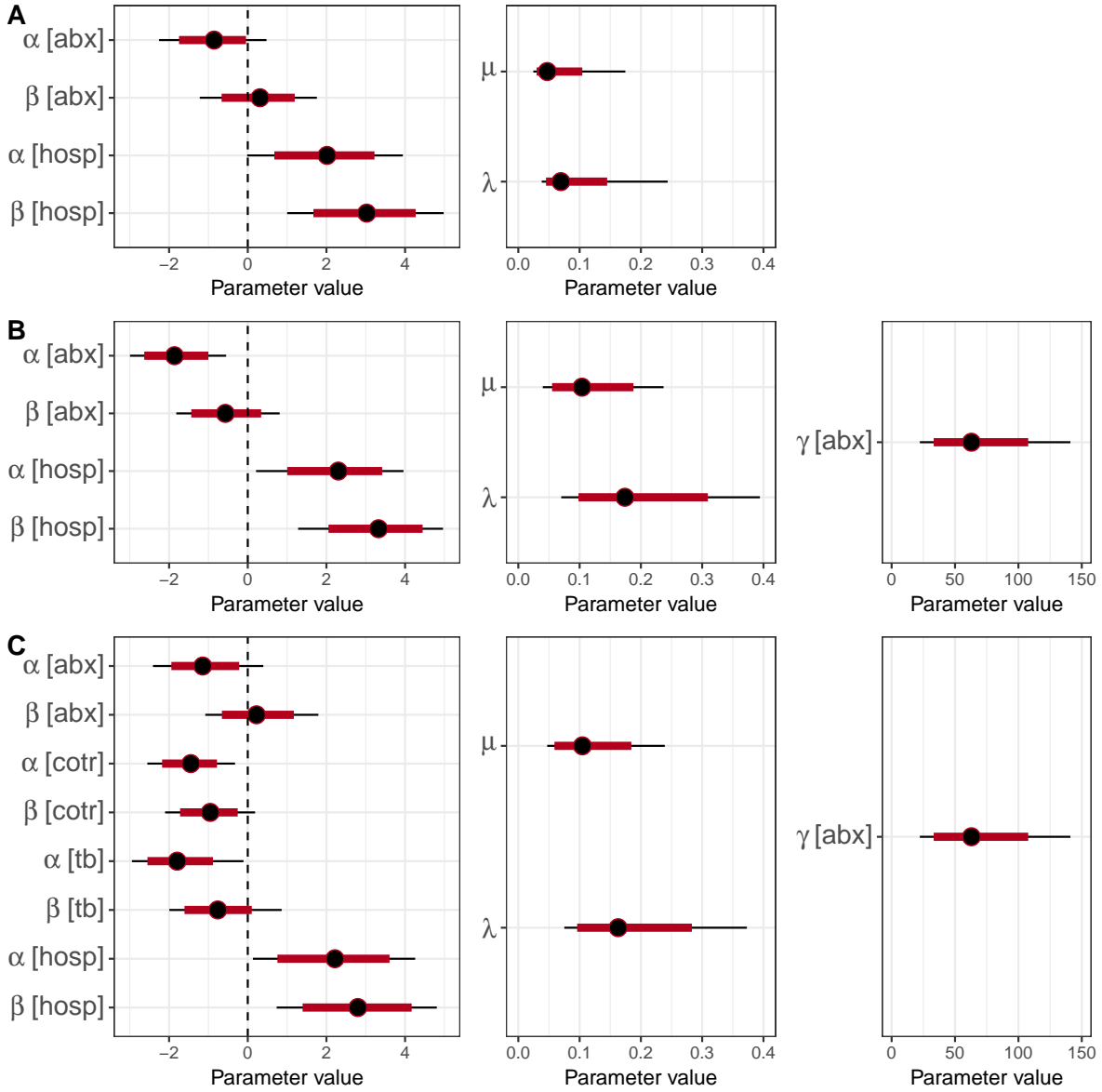


Figure 8.2: Parameter estimates from increasingly complex Markov models to predict ESBL carriage. Black lines are 95% and red lines 80% credible intervals. A: Model 1 includes stepwise constant covariates only, antimicrobial exposure and hospitalisation.  $\lambda$  is the baseline hazard and  $\beta$  the log hazard ratio of ESBL-E acquisition,  $\mu$  the baseline hazard and  $\alpha$  the log hazard ratio of ESBL-E loss. B: Model 2 adds a time-varying decay of the effect of antimicrobial exposure, parameterised by  $\gamma$  as described in the text. C: Model 3 adds stepwise constant covariates for TB therapy (tb) and cotrimoxazole (cotri) with all other antimicrobial exposure captured in the abx variable, which has a time varying decay of effect as before. In almost all cases 95% credible intervals of  $\alpha$ [hosp] and  $\beta$ [hosp] do not cross zero, suggesting that hospitalisation acts to both increase rate of ESBL-E acquisition and loss; for antimicrobial exposure, on the other hand, only the 95% for antimicrobial  $\beta$  values consistently do not cross zero, suggesting that the effect of antimicrobial exposure is to reduce the rate of ESBL-E loss.

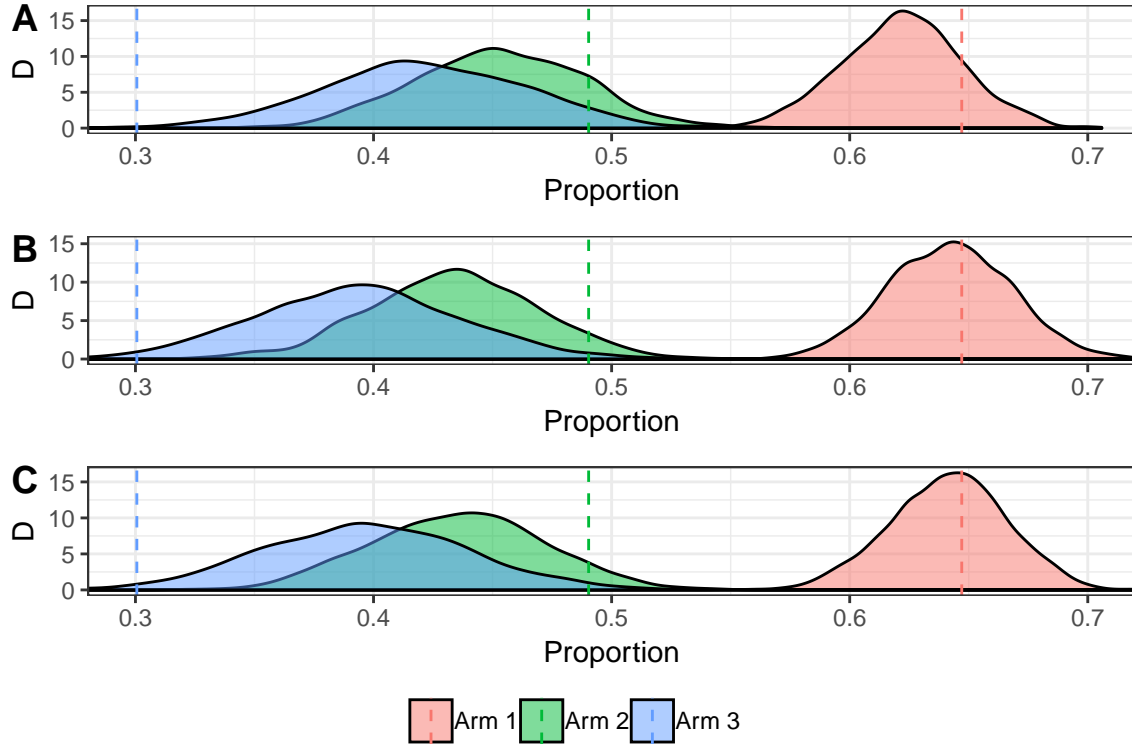


Figure 8.3: Posterior predictive checks: kernel density estimate,  $D$ , of predicted proportion of ESBL-E positive samples, stratified by arm for Model 1 (A), Model 2 (B), Model 3 (C), generated by sampling from a Bernoulli distribution using the predicted probability for each sample ( $n=993$ ) for each draw from the posterior, excluding warmup draws ( $n = 2000$ ). True proportion of ESBL-E positive samples are shown for each arm by dotted vertical line. In all cases, predictions are poor for arm 2 and 3 samples, but the addition of a term conferring an effect of antimicrobials that persists once exposure has stopped (quantified by  $\gamma$ ) improves fit, especially in arm 1 participants: compare Model 1 (A) with stepwise constant covariates to Model 2 (B) with persistent antimicrobial effect.



# References