UC Irvine
UC Irvine Previously Published Works

Title

Evaluating hospital infection control measures for antimicrobial-resistant pathogens using stochastic transmission models: Application to vancomycin-resistant enterococci in intensive care units

Permalink

Journal

Statistical Methods in Medical Research, 27(1)

ISSN

0962-2802

Authors

Wei, Yinghui
Kypraios, Theodore
O'Neill, Philip D
et al.

Publication Date

2018

DOI

10.1177/0962280215627299

Peer reviewed

Evaluating hospital infection control measures for antimicrobial-resistant pathogens using stochastic transmission models: Application to vancomycin-resistant enterococci in intensive care units

Yinghui Wei,[1] Theodore Kypraios,[2] Philip D O'Neill,[2]
Susan S Huang,[3] Sheryl L Rifas-Shiman[4] and Ben S Cooper[5,6]

Abstract

Nosocomial pathogens such as methicillin-resistant *Staphylococcus aureus* (MRSA) and vancomycin-resistant *Enterococci* (VRE) are the cause of significant morbidity and mortality among hospital patients. It is important to be able to assess the efficacy of control measures using data on patient outcomes. In this paper, we describe methods for analysing such data using patient-level stochastic models which seek to describe the underlying unobserved process of transmission. The methods are applied to detailed longitudinal patient-level data on vancomycin-resistant *Enterococci* from a study in a US hospital with eight intensive care units (ICUs). The data comprise admission and discharge dates, dates and results of screening tests, and dates during which precautionary measures were in place for each patient during the study period. Results include estimates of the efficacy of the control measures, the proportion of unobserved patients colonized with vancomycin-resistant *Enterococci*, and the proportion of patients colonized on admission.

[1]Centre for Mathematical Sciences, School of Computing, Electronics and Mathematics, University of Plymouth, UK
[2]School of Mathematical Sciences, University of Nottingham, UK
[3]Division of Infectious Disease and Health Policy Research Institute, University of California Irvine School of Medicine, Irvine, CA, USA
[4]Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, USA
[5]Mahidol Oxford Tropical Medicine Research Unit (MORU), Bangkok, Thailand
[6]Centre for Tropical Medicine and Global Health, Nuffield Department of Clinical Medicine, University of Oxford, Oxford, UK

Corresponding author:
Yinghui Wei, Centre for Mathematical Sciences, School of Computing, Electronics and Mathematics, University of Plymouth, Plymouth, PL4 8AA, UK.
Email: yinghui.wei@plymouth.ac.uk.

1 Introduction

The emergence of antimicrobial-resistant pathogens such as methicillin-resistant *Staphylococcus aureus* (MRSA) and vancomycin-resistant *enterococci* (VRE) during the past two decades is a major clinical and epidemiological problem throughout the world. Understanding these pathogens and assessing methods for their control continues to be an area of considerable importance (see literature[1–3]). In this paper, we describe methods for analysing longitudinal patient-level hospital data on nosocomial pathogens. The methods involve defining stochastic models which describe how the pathogen spreads between individual patients. Specifically, we build on the methods of Kypraios et al.[4] in which data on MRSA were analysed and extend this approach to data on VRE. The focus of the earlier paper was towards clinical results, and in particular, there was no detailed description of the statistical methods used. In contrast, here we fully describe these methods and also develop methods for model comparison and model assessment which were not in the earlier work.

The kind of data we consider typically consist of admission and discharge times for each patient on a hospital ward, along with the times and results of screening tests for the pathogen in question. There may also be other information, such as the times and types of infection control measures administered, or treatments. Such data are characterized by two aspects, namely that they are highly dependent (e.g. whether or not a patient is colonized with the pathogen in question depends on the status of other patients) and that the process which generates the data, namely transmission, is unobserved. This in turn implies that standard statistical methods for longitudinal data such as time-series methods or survival analysis methods are not usually natural in this setting. Instead, modelling methods are more appropriate.

Various modelling approaches to analysing longitudinal epidemiological data on hospital pathogens have been developed, typically applied to data on MRSA. These include fitting deterministic models,[5] maximum-likelihood methods applied to Markov chain models,[6–8] hidden Markov models,[9] and non-parametric approaches for Markov process models.[10] All of these methods involve simplifying assumptions of one kind or another which can be relaxed by using individual-level stochastic models and a Bayesian framework. The latter approach, using Markov chain Monte Carlo (MCMC) methods, is taken in literature[4,11] essentially by developing the ideas in literature.[12,13] This is the approach that we adopt in this paper. Since the underlying models essentially describe the unobserved process of transmission between patients, the MCMC methods involve augmenting the observed data with the transmission events themselves. The appeal of this approach is that it is highly flexible, it usually avoids the need for simplifying assumptions which are common in other methods, and can provide much richer inferential information from the data than maximum likelihood methods.

In this paper, we analyse data on VRE, which are bacteria from the genus *Enterococcus* which have acquired resistance to vancomycin. *Enterococci* also have intrinsic resistance to several classes of antibiotics. They usually live harmlessly within the gastrointestinal tract or on the skin, but can cause infections, especially among those with weakened immune systems. In consequence, they can produce disease in highly vulnerable hospital patients such as those in oncology wards or intensive care units (ICUs). Changes in resistance patterns over time frequently mean that VRE strains are resistant to most antimicrobial agents. VRE have been endemic in hospital settings in Europe and the USA since the start of the 21st century and are a major problem in healthcare settings throughout the world. In such environments, transmission of VRE between patients is believed to occur indirectly, largely via healthcare workers and in particular by contamination of hands or gloves.[14,15] Environmental contamination is also thought to be a possible transmission route.

For this reason, control strategies typically focus on activities such as hand-washing, barrier precautions (glove and gown use), and environmental cleaning.

The paper is structured as follows. The data structure and transmission models are described in sections 2 and 3, respectively. The MCMC methods are introduced in section 4 and applied to the VRE data in section 5. We finish with some discussion in section 6.

2 Data structure

The individual patient data are taken from a 17-month longitudinal study in a hospital in Boston, USA, involving over 8000 patients who were admitted to one of eight different ICUs during the study period. The data consist of the dates on which each patient was admitted and discharged from the ICU, the dates and results of any screening tests that took place, and any dates during which the patient was placed under precautionary measures due to their being supposed to be colonized with VRE. The precautionary measures consisted of glove and gown use by healthcare workers, and the use of isolation rooms, and were the same for all patients concerned. Such data contain many inherent dependencies: for instance, whether or not a patient becomes colonized could reasonably be supposed to depend on the number of other currently colonized patients in the ICU. For this reason, it is natural to analyse such data using a model-based approach which explicitly describes how the pathogen can spread between patients over time.

Each patient may undergo a number (possibly zero for some patients) of screening tests, the results of which are either positive (denoting colonization with the pathogen) or negative. Note that we take *colonization* to mean either asymptomatic carriage (i.e. the pathogen is present in the body but has not caused clinical infection) or infection, and we do not distinguish between these two states. Furthermore, for VRE the amount of pathogen shed by carriers or those infected is usually similar, which further motivates this modelling assumption. Each patient may also be placed in isolation during their stay in the ward, usually as a result of a positive test, or for other clinical reasons. The exact meaning of isolation depends on the study in question, examples including physical isolation in a single room, use of additional barrier precautions, etc. The models we describe below assume that only one kind of isolation is adopted, although it would be straightforward to relax this assumption.

3 Models

The data set of interest comprises eight different ICU wards. We analyse the data from each ICU separately. This is motivated by the fact that the ICUs have different kinds of patients (e.g. medical/surgical) and so there is no reason to suppose that parameters governing VRE transmission are common across all wards. Furthermore, relatively few patients were admitted to more than one of the ICUs.

In the following, we describe modelling approaches for data on patients in a single ICU. The study period begins and ends at times T_S and T_E, respectively. During the study period, the admission and discharge times of all patients on the ward are recorded. Patients already present in the ward at time T_S are deemed to have been admitted at T_S, and those in the ward at T_E are deemed to have been discharged at T_E. Patients might be re-admitted during the study period.

We consider three models for colonization, with the following common assumptions. First, at each point in time every patient in the ward is either susceptible or colonized and can be either in or out of isolation. Second, the test results are assumed to have perfect specificity (so that a positive test

result can only arise if the patient really is colonized) and sensitivity $p \times 100\,\%$ (so a colonized patient has probability p of testing positive). The assumption of perfect specificity is natural in the setting we consider, since screening tests are usually culture-based, and false positives can only arise via the rare event of contamination.[16] However, this assumption can be easily relaxed within our framework.

Third, if a patient has a positive test, then they are assumed to remain colonized for the following six months.[17–19] In particular, if a patient is re-admitted within six months of a positive test, then they are said to be *colonized on re-admission* at the time of re-admission, and assumed to remain colonized until next discharged. Conversely, if a patient is admitted to the ward with no positive test result during the previous six months, which in many cases is simply because they have never been previously admitted at all, then they are called a *new admission* at the time of admission. Each new admission has a probability φ of being *colonized on admission*, independently of all other patients and admissions. We refer to φ as the *importation probability*. New admissions are also formally regarded in the model as being new patients, even if they have been previously admitted. In particular, the term *patient* should be interpreted in this way in this section and the next. Although the assumption that the same individual can be formally treated as two patients is made largely for simplicity, it is often pragmatic in practice unless there are a large number of re-admissions in the data.

The final component of the model is the mechanism for potential colonization of patients who are not already colonized. We adopt the usual assumption from stochastic disease transmission modelling that susceptible individuals are colonized according to a Poisson process whose rate might depend on the numbers of existing colonized patients within the ward. In the context of nosocomial pathogens, the transmission between patients is usually indirect, typically via healthcare workers or the environment. We refer to patients who are colonized in this way as *colonized on the ward*.

Specifically, each susceptible patient is, independently of other susceptibles, assumed to be colonized at a time corresponding to the first point of a non-homogeneous Poisson process of rate $\lambda(t) \geq 0$ at time t. We consider three possible models for $\lambda(t)$. In order to investigate the effectiveness of control measures, in each model we differentiate between isolated and non-isolated patients.

3.1 Full model

This model is defined in Forrester et al.[11] and assumes that

$$\lambda(t) = \beta_0 + \beta_1 C(t) + \beta_2 Q(t)$$

where $Q(t)$ and $C(t)$ denote the number of colonized patients on the ward at time t who are in and out of isolation, respectively. Here β_0 represents a background rate of colonization which is unaffected by the current prevalence on the ward, for instance arising due to medical staff who service many wards and may act as vectors for the pathogen. Conversely, β_1 and β_2 represent rates attributed to colonized patients. This way of modelling transmission is a natural generalization of the usual mass-action assumption in standard epidemic models such as the SIR (susceptible-infective-removed) epidemic model (see e.g. O'Neill and Roberts[13]), which has itself been extensively used to successfully describe pathogen transmission in many different settings. Specifically, our approach assumes that each colonized individual with the same isolation status (in or out) makes an equal and independent contribution to the overall colonization rate. This is a

very natural assumption with a clear epidemiological interpretation and also avoids the need to specify more complex dependencies between individuals.

3.2 No-background transmission model

This model is a special case of the full model in which β_0 is assumed to be close to zero with high probability. In practice, this is achieved by assigning a suitable prior distribution to β_0, as discussed below. This model represents the hypothesis that transmission almost always occurs due to the presence of colonized patients in the ward. As well as being of interest in its own right, considering this model also allows us to assess the impact of assuming background transmission.

3.3 Non-linear model

This model assumes that it is only the presence of colonized patients, not their number, which affects the rate of new colonizations. In particular, it is not assumed that $\lambda(t)$ is linear in $C(t)$ and $Q(t)$, as in the full model. Such a model is suitable if the amount of pathogen in the ward from one patient is so great as to have a saturation effect. Specifically

$$\lambda(t) = \beta_0 + \beta_1 \chi_{\{C(t) > 0\}} + \beta_2 \chi_{\{Q(t) > 0\}}$$

where χ_A denotes the indicator function of the event A. In some sense this model provides an extreme alternative to the mass-action type assumptions in the previous models, since now there is no increase in colonization rate as the numbers of colonized individuals increase beyond one.

4 Inference methods
4.1 Notation and likelihoods

Let $\theta = (p, \varphi, \beta_0, \beta_1, \beta_2)$ denote the vector of model parameters. We denote all data (admission and discharge times, times and outcomes of test results, times of patient isolation) by y, although as explained above we only have a probability model for the colonization process and test sensitivity. Adopting a Bayesian framework, our objective is to explore the posterior density $\pi(\theta \mid y)$. Now the likelihood $\pi(y \mid \theta)$ is intractable in practice, since its evaluation involves integrating over all possible unobserved colonization times (cf. O'Neill and Roberts,[13] for the analagous situation for standard epidemic models). Consequently, we augment the parameter space to include all colonization times, c say.

Under the assumption of perfect test specificity, c includes the unobserved colonization time for each patient who ever has a positive test result. Additionally, individuals who do not have a positive test result may still be colonized, their undetected status arising either because they never had a test at all, or because any test was a false negative. The colonization times of such individuals are also included in c. Note also that c is also assumed to describe which patients are colonized on admission and which are colonized on re-admission.

By Bayes' Theorem,

$$\pi(\theta, c \mid y) \propto \pi(y, c \mid \theta)\pi(\theta)$$

where $\pi(\theta)$ denotes the prior density of θ. It remains to evaluate the augmented likelihood $\pi(y, c \mid \theta)$. Let n_A denote the number of new admissions and n_{CA} denote the number of these who are colonized on admission. Note that n_A is determined by the data y, but n_{CA} is determined by both y and c.

Define n_{TP} and n_{FN} as the number of true-positive and false-negative test results. Under the assumption of perfect specificity, n_{TP} is determined by y. Conversely, since sensitivity need not be perfect, n_{FN} is determined by y and c.

Define K as the set of patients who are colonized on the ward. For a typical patient in K, denoted j say, define $n_C(j)$ and $n_Q(j)$ as the numbers of patients who are respectively colonized and not isolated, and colonized and isolated, at the time of j's own colonization. Note that K, $n_C(j)$ and $n_Q(j)$ can all be determined from y and c, but none are known explicitly from just y.

The augmented likelihood for the full model takes the form

$$
\pi(y, c | \theta) \propto \alpha^{n_{CA}}(1-\alpha)^{n_A - n_{CA}} \times p^{n_{TP}}(1-p)^{n_{FN}}
$$
$$
\times \prod_{j \in K} \left(\beta_0 + n_C(j)\beta_1 + n_Q(j)\beta_2 \right)
$$
$$
\times \exp\left(-\int_{T_S}^{T_E} (\beta_0 S(t) + \beta_1 C(t)S(t) + \beta_2 Q(t)S(t)) \, dt \right)
\tag{1}
$$

where $S(t)$ denotes the number of susceptible patients on the ward at time t. The likelihood can be derived as follows. First, the terms involving α account for the probability of n_{CA} patients being colonized on admission and the remaining $n_A - n_{CA}$ being not colonized on admission. Second, the terms involving p give the probability of observing the true-positive and false-negative test results. Note that since we keep track of each individual patient, for both the α and p expressions we do not require any combinatorial terms; for instance, we know exactly which patients are colonized on admission, not just the number.

The remaining part of the likelihood arises from the colonization process. Recall that if a non-homogeneous Poisson process with rate $\lambda(t)$ at time t is observed during a time interval $[0, T]$, and m points occur at times $0 \le t_1 \le t_2 \le \ldots \le t_m \le T$, the likelihood is given by

$$
\prod_{i=1}^{n} \lambda(t_i^-) \exp\left(-\int_0^T \lambda(t) \, dt \right)
\tag{2}
$$

where $\lambda(t_i^-)$ denotes the rate just before time t_i, see for example Daley and Vere-Jones,[20] Chapter 7. The product term accounts for the points occurring and the exponential term accounts for the absence of points at other times. In our setting, each susceptible patient is colonized according to a Poisson process of rate $\Lambda(t) = \beta_0 + \beta_1 C(t) + \beta_2 Q(t)$ at time $t \ge 0$, and so the overall colonization process (i.e. the process which counts colonization events but does not specify which individuals are actually colonized) is a Poisson process of rate $S(t)\Lambda(t)$. Setting $\lambda(t) = S(t)\Lambda(t)$ in (2) gives the likelihood of this overall colonization process. However, we also know the identity of each colonized patient, and since each susceptible patient is equally likely to be colonized at any given time, the probability that a particular patient is colonized is $1/S(t)$. Incorporating this probability into the product term in (2) yields the product term in (1). Finally, the likelihood in (1) can also be derived and expressed by purely thinking in terms of individual patients, which results in the integral being expressed as a sum over patients, see Forrester et al.[11]

It is straightforward to modify (1) to obtain the corresponding likelihoods for the no-background and non-linear models. Finally, we assign independent Beta prior distributions to p and α, and independent exponential prior distributions to the colonization rate parameters β_0, β_1 and β_2.

4.2 Markov Chain Monte Carlo algorithm

In order to explore πð8, cj yÞ, we use a Metropolis-Hastings MCMC algorithm which updates the elements of 8 and the unknown colonization times c, as follows. First, it follows from (1) and our choice of prior distributions that both p and <P have Beta-distributed full conditional distributions, and hence can be updated according to Gibbs steps. Second, the colonization rate parameters f3$_0$, f3$_1$ and f3$_2$ can all be updated using a Gaussian random walk, where proposed negative values are rejected immediately since the proposed likelihood is zero.

Third, updates for the colonization times c are achieved via three possible steps, one of which is chosen uniformly at random during each iteration of the MCMC algorithm. Recall that patients who are colonized on re-admission are assumed to have a colonization time equal to their time of admission. Since such re-admissions are determined directly by the data, the colonization times in question are never updated in the MCMC algorithm.

Of the remaining patients, those with a positive test result are assumed to be definitely colonized, and such patients therefore always have a colonization time. Denote the set of such patients P and note that P is determined by the data y. All other remaining patients may or may not be colonized; denote the set of those currently colonized by N$_1$ and those not by N$_0$. Thus, P, N$_0$ and N$_1$ have no intersection and their union is the set of all patients other than those colonized on re-admission. Let n_0, n_1 and n_p denote the number of patients in N$_0$, N$_1$ and P, respectively.

The three possible steps involved in updating the colonization times are as follows. We define a_j, d_j and c_j as the admission, discharge and colonization times, respectively, of individual j.

1. Adding a colonization time: an individual j is selected randomly from N$_0$. With probability <P$_{0,j}$ is proposed to be colonized on admission, so c_j¼a_j; otherwise c_j is sampled uniformly from (a_j, d_j). Denoting e as the proposed new c with c_j added, the move is accepted with probability

$$\frac{n_0 ð d_j - a_j Þ \pi ð 8, e j y Þ}{ð 1 - <P_0 Þ ð n_1 þ 1 Þ \pi ð 8, c j y Þ} \wedge 1$$

if c_j 2 ða_j, d_j Þ, and

$$\frac{n_0 \pi ð 8, e j y Þ}{<P_0 ð n_1 þ 1 Þ \pi ð 8, c j y Þ} \wedge 1$$

if c_j¼a_j, where $x \wedge y$ denotes minðx, yÞ.

2. Deleting a colonization time: an individual j is randomly selected from N$_1$, and its colonization time c_j is removed from c with probability

$$\frac{ð 1 - <P_0 Þ n_1 \pi ð 8, e j y Þ}{ð n_0 þ 1 Þ ð d_j - a_j Þ \pi ð 8, c j y Þ} \wedge 1$$

if c_j 2 ða_j, d_j Þ, and

$$\frac{<P_0 n_1 \pi ð 8, e j y Þ}{ð n_0 þ 1 Þ \pi ð 8, c j y Þ} \wedge 1$$

if c_j¼a_j, where e denotes c with c_j removed.

3. Moving a colonization time: randomly select an individual j from N$_1$ [P. With probability <P$_0$, set the proposed new colonization time as e_j¼a_j; otherwise, e_j is sampled uniformly from (a_j, t_j),

where $t_j = d_j$ if $j \in N_1$ and t_j is the time of j's first positive test if $j \in P$. Thus the proposal density of ε_j is

$$q(\varepsilon_j \mid c_j) = q(\varepsilon_j) = \begin{cases} \dfrac{\pi_0}{n_1 + n_p} & \text{if } \varepsilon_j = a_j \\[2ex] \dfrac{1 - \pi_0}{(n_1 + n_p)(t_j - a_j)} & \text{if } \varepsilon_j \in (a_j, t_j) \end{cases}$$

The new colonization time ε_j is accepted with probability

$$\frac{\pi(\theta, \varepsilon_j \mid y) q(c_j \mid \varepsilon_j)}{\pi(\theta, c_j \mid y) q(\varepsilon_j \mid c_j)} \wedge 1$$

where ε denotes c with c_j replaced by ε_j.

4.3 Model assessment

Although our main focus is not to formally distinguish between competing models, it is nevertheless of interest to perform a comparison. There is no canonical method to do this in our setting, and different approaches have their own pros and cons. For simplicity, we used a form of deviance information criterion (DIC[21]). Since our setting involves missing data (namely, the unobserved colonization times), we require a form of DIC which takes this into account and moreover which can be readily calculated. We therefore used DIC_6 from Celeux[22] defined by

$$\mathrm{DIC}_6 = -4E_{\theta,c}[\log \pi(y, c \mid \theta)] + 2E_c[\log \pi(y, c \mid \hat{\theta}_\theta(y)) \mid y, \hat{\theta}_\theta(y)] \tag{3}$$

where $\pi(y, c \mid \theta)$ is given by (1) and we set $\hat{\theta}_\theta(y)$ to be the posterior mean of θ, as estimated from our MCMC algorithm. In practice, DIC_6 can be easily evaluated using MCMC output as described in Celeux.[22] Although we used DIC_6 for convenience, it has been explored in a similar context to ours by Worby[23] who found that it can be effective provided there are sufficient data available.

We also consider two forms of model assessment, i.e. methods for exploring how well each model fitted the observed data. The first of these is a posterior predictive p-value[24] to assess the model fit to data, specifically with the total number of detected colonizations as the discrepancy statistic to measure the difference between the observed and simulated data. To implement this method, the model parameters in question are first repeatedly sampled from their posterior distribution, i.e. from the MCMC output. In practice we take every 100th sample from the chain. For each sampled set of parameter values, the model is then simulated forward in time to produce a possible realization. We then calculate a p-value as the proportion of simulations in which the total number of detected colonizations is greater or equal to that in the observed data. A p-value close to 0.5 indicates a good fit to the data. In contrast, an extreme p-value ($p > 0.95$ or $p < 0.05$) indicates that the fitted model is not appropriate for the data.

The second method of model assessment is to consider the predictive distribution of the number of detected colonizations through time. This is implemented by again sampling parameters from the posterior distribution and simulating the model forward in time for each set of parameter values to produce a large number of realizations of the whole process. We then graphically compare the observed data with the mean and quantiles of the set of simulated realizations.

5 Application to data

5.1 Data

The data we consider were collected from eight 10-bed ICUs in a tertiary academic medical center in Boston, Massachusetts over a 17-month period. The ICUs comprised two medical wards (which we denote M1, M2), two general surgical wards (GS1, GS2), and four specialty surgery wards (SS1-SS4). Routine rectal admission and weekly screening was carried out using swab tests, with compliance around 90%. In addition, there was a protocol in which patients who were identified as being VRE-positive were placed under contact precautions which consisted of gown and glove use by healthcare workers, and use of single rooms.

Figure 1 shows the number of first detected colonizations and total current detected colonizations for one of the wards (medical ward M1) on a weekly basis during the observation interval. Descriptive statistics of the individuals in ICUs are shown in Table 1. The relatively high values for the standard deviation of length of stay reflect the fact that ICU patients typically have either quite short or quite long stays. The median length of stay was around 3–6 days. On most wards, patients have just over one swab test on average, and thus there are many patients in the data set with only one (usually admission) swab test. Finally, around 5–10% of all patients in the study were re-admitted during the study period.

5.2 Algorithm implementation

Before considering the data at hand, we first simulated data sets from each of the true models and then used our MCMC algorithm to estimate the model parameters. The results indicated that our methods could recover the true parameter values reasonably well. The MCMC algorithm was then implemented separately for the eight wards for each of the three models under consideration, and
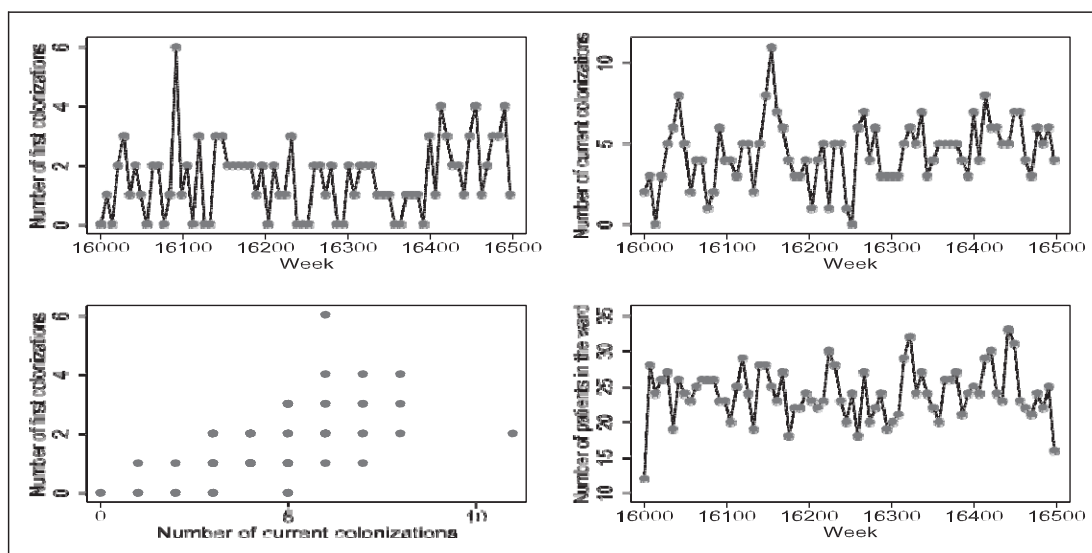

Figure 1. Ward M1. Number of first detected colonizations, current total detected colonizations, and number of patients on a weekly basis during the observation period. We also show the relationship between the number of first colonizations and number of current colonizations.

Table 1. Summary statistics for VRE data.

Ward	Number of patients	Length of stay Mean (SD)	Percent in contact precautions	Number of swab tests per person Mean (SD)	Number of VRE þ swab tests per person Mean (SD)
M1	1293	3.4(4.7)	11.4	1.2(0.8)	0.2(0.8)
M2	1018	4.4(6.4)	19.1	1.4(1.1)	0.7(1.6)
GS1	1227	3.4(5.2)	12.4	1.1(0.9)	0.3(0.9)
GS2	1030	4.0(8.3)	10.7	1.0(1.3)	0.1(0.7)
SS1	706	5.8(11.4)	12.5	1.4(1.6)	0.7(2.2)
SS2	888	4.9(9.7)	7.5	1.3(1.4)	0.3(1.8)
SS3	1097	3.8(6.4)	6.0	1.3(0.9)	0.1(0.6)
SS4	1263	3.6(5.2)	5.1	1.1(1.0)	0.1(0.5)

the DIC values computed. The model parameters were assigned uninformative independent prior distributions. Specifically, $\langle P, p \,\tilde{}\, U(0, 1)$ and $\beta_0, \beta_1, \beta_2 \,\tilde{}\, Exp(10^{-6})$ where $Exp(A)$ denotes an exponential distribution with mean A^{-1}, with the exception that $\beta_0 \,\tilde{}\, Exp(10^6)$ in the no-background model. The algorithm was implemented by using Microsoft Visual C++ 2010 Express software with double precision and the GSL Scientific Library. The C++ code is available upon request from the first author.

5.3   Results

5.3.1   Parameter estimates

Posterior medians of the model parameters are shown in Table 2. Estimates of both the importation probability $\langle P$ and the test sensitivity p appear robust to the choice of model, and both parameters are estimated with reasonable precision. The importation probability is generally in the range 5–15% other than medical ward M2 where it exceeds 20%. The test sensitivity is around 70% or higher, with the striking exception of general surgery ward GS2, where it is around 50%. These values for $\langle P$ and p are both reasonable in clinical terms.

The transmission rate parameter estimates exhibit clear variation between the models, which is to be expected due to the differences of the model assumptions themselves. The posterior standard deviations are typically of the same order as the posterior median values, illustrating that there is reasonable uncertainty in the posterior estimation of β_0, β_1 and β_2. This uncertainty is essentially a consequence of the prevalence of VRE (discussed below): the majority of patients never have a positive test which makes it harder to estimate transmission rate parameters accurately.

The background rate β_0 is (in the full and non-linear models) invariably estimated to be higher than β_1 and β_2, often around twice as much or more. This can be interpreted as saying that the background colonization rate is roughly equivalent to that due to two or three colonized patients on the ward. Regarding estimation of such rates, note that any colonization event which occurs whilst the ward contains no colonized patients must be attributable to the background rate, whereas colonization events that occur with (as is typical) one or two colonized patients in the ward could be due to background or transmission events. Thus, it is possible that, if we explicitly modelled the 'source' of each colonization event, more colonizations than not would be background colonization events.

Table 2. Posterior median and standard deviation of all model parameters.

Ward	$\langle P \rangle$	p	β_0	β_1	β_2
			Full model		
M1	0.12 (0.02)	0.78 (0.03)	0.0084 (0.004)	0.0023 (0.002)	0.0025 (0.003)
M2	0.23 (0.02)	0.81 (0.02)	0.0093 (0.006)	0.0028 (0.002)	0.0029 (0.002)
GS1	0.12 (0.01)	0.78 (0.03)	0.0075 (0.005)	0.0057 (0.004)	0.0034 (0.003)
GS2	0.05 (0.01)	0.49 (0.05)	0.0082 (0.005)	0.0034 (0.003)	0.0037 (0.003)
SS1	0.13 (0.02)	0.84 (0.02)	0.0038 (0.004)	0.0014 (0.002)	0.0050 (0.003)
SS2	0.05 (0.02)	0.71 (0.05)	0.0088 (0.006)	0.0065 (0.004)	0.0059 (0.004)
SS3	0.10 (0.02)	0.66 (0.05)	0.0067 (0.004)	0.0029 (0.003)	0.0048 (0.004)
SS4	0.04 (0.01)	0.68 (0.06)	0.0042 (0.002)	0.0023 (0.003)	0.0069 (0.006)
			No-background model		
M1	0.13 (0.01)	0.77 (0.03)	$10^{-4}\ (2 \times 10^{-4})$	0.0066 (0.003)	0.0054 (0.003)
M2	0.23 (0.02)	0.81 (0.02)	$10^{-4}\ (1 \times 10^{-4})$	0.0050 (0.003)	0.0056 (0.003)
GS1	0.13 (0.02)	0.77 (0.03)	$10^{-4}\ (1 \times 10^{-4})$	0.0099 (0.003)	0.0052 (0.003)
GS2	0.05 (0.01)	0.49 (0.05)	$10^{-4}\ (1 \times 10^{-4})$	0.0071 (0.004)	0.0079 (0.003)
SS1	0.13 (0.02)	0.84 (0.02)	$10^{-4}\ (1 \times 10^{-4})$	0.0015 (0.002)	0.0069 (0.002)
SS2	0.05 (0.02)	0.71 (0.04)	$10^{-4}\ (1 \times 10^{-4})$	0.0087 (0.004)	0.0104 (0.005)
SS3	0.11 (0.02)	0.65 (0.05)	$10^{-4}\ (1 \times 10^{-4})$	0.0072 (0.003)	0.0060 (0.004)
SS4	0.05 (0.01)	0.65 (0.06)	$10^{-4}\ (2 \times 10^{-4})$	0.0073 (0.004)	0.0074 (0.006)
			Non-linear model		
M1	0.12 (0.02)	0.77 (0.03)	0.0081 (0.004)	0.0040 (0.004)	0.0040 (0.004)
M2	0.23 (0.02)	0.81 (0.02)	0.0092 (0.006)	0.0050 (0.005)	0.0071 (0.006)
GS1	0.12 (0.01)	0.78 (0.03)	0.0113 (0.005)	0.0047 (0.005)	0.0048 (0.005)
GS2	0.05 (0.01)	0.49 (0.05)	0.0085 (0.005)	0.0061 (0.006)	0.0039 (0.004)
SS1	0.14 (0.02)	0.84 (0.02)	0.0048 (0.004)	0.0032 (0.004)	0.0069 (0.005)
SS2	0.05 (0.01)	0.71 (0.04)	0.0117 (0.007)	0.0081 (0.007)	0.0052 (0.006)
SS3	0.11 (0.02)	0.66 (0.05)	0.0071 (0.004)	0.0039 (0.004)	0.0057 (0.005)
SS4	0.04 (0.01)	0.68 (0.06)	0.0041 (0.002)	0.0028 (0.004)	0.0035 (0.004)

5.3.2 Unobserved carriage

Unobserved carriage arises in two ways, namely false-negative test results, or colonization which occurs before any test is carried out. Thus unobserved carriage is a function of both the test sensitivity and the frequency of testing. Our transmission models explicitly include all unobserved colonization events, and so in particular it is possible for individuals to be colonized but undetected. The inference methods enable us to estimate the extent of this unobserved colonization directly. Table 3 shows the observed mean monthly prevalence of VRE, calculated directly from the positive test results in the data, and the median posterior prevalence as predicted by the full model (other models gave similar results). It appears that the observed prevalence accounts for just over half of the predicted prevalence.

It is possible to derive more complex quantities which describe the extent of unobserved carriage. In particular, we define P_{hidden} as the proportion of all colonized-patient-days which are unobserved in the sense that a patient is colonized but not under precautions, and P_{wait} as the proportion of all colonized-patient-days during which a patient is pending knowledge of a positive test result and has yet to be placed under precautions. Denote p_j as the precaution time of individual j, with $p_j = 1$ for

Table 3. Observed and predicted monthly VRE prevalence.

Ward	Observed mean (SD)	Predicted median (SD)
M1	14.8 % (2.9%)	23.3% (3.7 %)
M2	28.6 % (5.2%)	41.3% (5.0 %)
GS1	18.2 % (2.8%)	22.8% (4.1 %)
GS2	9.9 % (2.9%)	15.6% (4.0 %)
SS1	18.9 % (7.6%)	26.4% (8.6 %)
SS2	11.4 % (6.5%)	16.9% (5.5 %)
SS3	9.0 % (3.1%)	16.7% (2.8 %)
SS4	4.5 % (3.0%)	6.9% (4.7 %)

an individual never placed under precautions, recall that t_j is the time of j's first positive test and set t_j ¼ 1 if j never has a positive test. It follows that

$$P_{hidden} \text{¼} \frac{\sum_{j:c_j 2\frac{1}{2}a_j,d_j]}(p_j \wedge d_j \text{Þ} - c_j \underline{\ } 0)}{\sum_{j:c_j 2\frac{1}{2}a_j,d_j]}(d_j - c_j\text{Þ}) \quad P^{wait} \text{¼} \frac{\sum_{j:p_j 2\frac{1}{2}t_j,d_j]}}{\sum_{j:c_j 2\frac{1}{2}a_j,d_j]}} \frac{p_j \quad t_j}{d_j \quad c_j} \text{Þ}$$

where $\wedge$ and $\underline{\ }$ denote min and max, respectively. To see that P_{wait} cannot exceed 1, note that any j in the numerator sum has $t_j \leq 1$, and under the assumption of perfect specificity a_j :s c_j :s t_j, and thus j also appears in the denominator sum.

Figure 2 gives posterior median values of P_{hidden} and P_{wait} for all three models. The values appear to be robust across the models themselves. The apparently high values of P_{hidden} are not surprising in view of the predicted prevalence values; in particular, each individual who is colonized but never placed under precautions contributes the entire duration of their colonization to the numerator of P_{hidden}. Conversely, the low P_{wait} values indicate that patients who were detected were placed under precautions swiftly.

5.3.3 Efficacy of precaution measures
The efficacy of the precaution measures in reducing transmission may be evaluated by comparing the rate of $f3_1$ and $f3_2$, with $f3_1$ greater than $f3_2$ indicating a positive benefit. We specifically consider the posterior probability $P\eth f3_1 4 f3_2 j y\text{Þ}$ and the posterior median value of $f3_1=f3_2$, as given in Figure 3. From these values, there is no compelling evidence to support the effectiveness of the precaution measures. The probability $P\eth f3_1 4 f3_2 j y\text{Þ}$ varies widely between wards, only once exceeding 0.8 (ward GS1) and with most values in the range 0.45-0.65. One exception is the specialty surgery ward SS1 which has values of 0.16 and 0.08, providing some evidence to suggest that precaution measures had a negative impact under the assumptions of the models in question (full model and no-background model).

The $\log\eth f3_1=f3_2\text{Þ}$ values give very similar conclusions, and in particular posterior point values obtained by pooling estimates across all wards (using the inverse variance method as described in Sutton et al.[25] (section 5.2) are all close to unity for each model. However, the posterior credible intervals for $\log\eth f3_1=f3_2\text{Þ}$ indicate a high degree of posterior uncertainty for this quantity, which is most likely inherited from the high posterior variances for $f3_1$ and $f3_2$.

5.3.4 Model assessment
The posterior predictive p-values are given in Figure 4. There is no strong evidence to suggest that the fitted models are inappropriate for the data, with one exception of ward SS3. Based on the p-values, there is no clear indication that any model is preferred over the others.
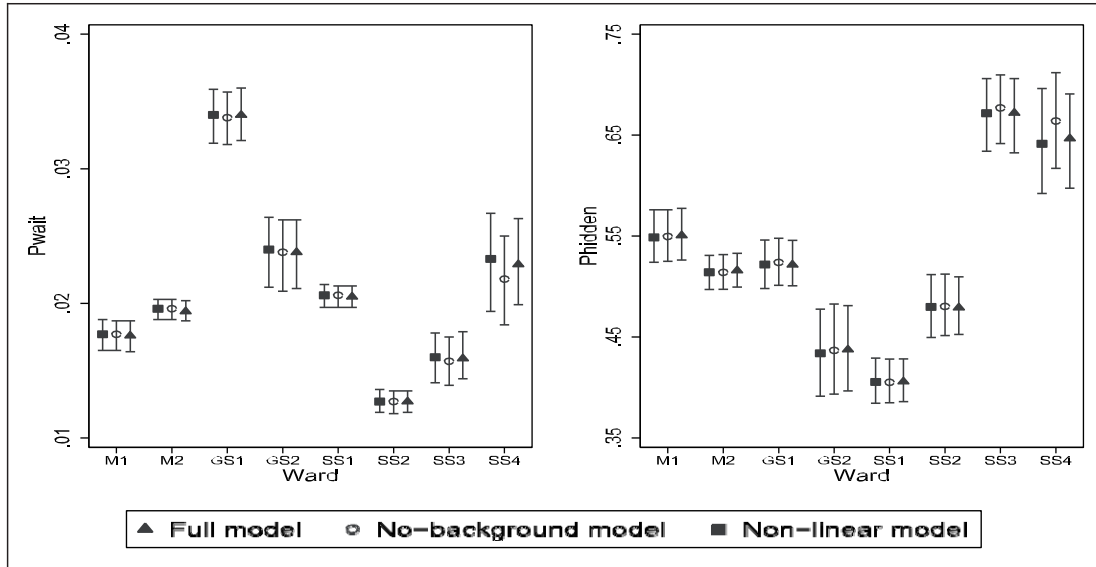
Figure 2. Posterior median and 95% credible intervals of the percentage of colonized-patient-days attributed to undetected colonized patients (P_{hidden}), and the percentage of colonized-patient-days attributed to colonized patients who had been swabbed but were waiting for results (P_{wait}).

Figures 5 to 7 show the comparison between simulated and the observed epidemic data based on the number of detected colonizations at 14-day intervals. The figures indicate that the number of detected colonizations are reasonably well approximated by the simulated realizations, although the models can fail to capture the occasional more extreme fluctuations in the data. Overall, there is no clear indication that one model is to be preferred.

5.3.5   Model comparison
DIC values are given in Table 4. It is notable that the no-background model is preferred in the medical and general wards, although in some cases the full model DIC values are very similar. This suggests that the risk of new colonization events increases with the number of existing colonized patients. In contrast, for all of the specialty surgery wards the non-linear model is favoured. This suggests that it is the presence of colonized patients, rather than the number, that is the driving force behind new colonizations in these wards.

5.3.6   Sensitivity analysis
In our primary analysis, we have used $Exp(10^{-6})$ (mean 10^6, variance 10^{12}) as the prior distribution for each of β_0, β_1 and β_2. In particular, this means that the prior density for these parameters is relatively flat in the region of interest, and so we expect the prior to have little impact on the posterior distributions. To explore this assumption, we conducted a sensitivity analysis by first using $Exp(10^{-3})$ prior distributions and then $Exp(10^{-12})$ prior distributions. In both cases, the posterior estimates of model parameters were found to be very similar to those in the original analysis.

We have assumed that patients remained colonization for six months after they became colonized. We assessed the impact of this assumption on the posterior distribution by assuming
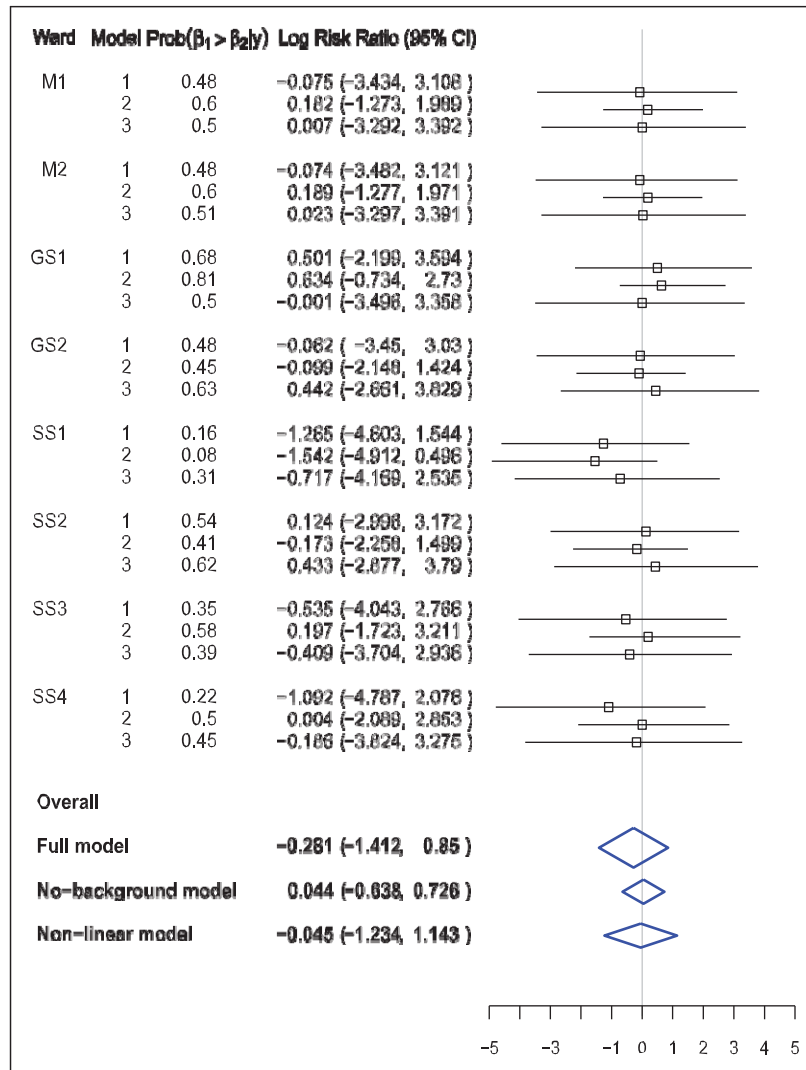
Ward	Model	Prob($\beta_1 > \beta_2$|y)	Log Risk Ratio (95% CI)
M1	1	0.48	−0.075 (−3.434, 3.108)
	2	0.6	0.182 (−1.273, 1.989)
	3	0.5	0.007 (−3.292, 3.392)
M2	1	0.48	−0.074 (−3.482, 3.121)
	2	0.6	0.189 (−1.277, 1.971)
	3	0.51	0.023 (−3.297, 3.391)
GS1	1	0.68	0.501 (−2.199, 3.594)
	2	0.81	0.834 (−0.734, 2.73)
	3	0.5	−0.001 (−3.498, 3.358)
GS2	1	0.48	−0.062 (−3.45, 3.03)
	2	0.45	−0.099 (−2.148, 1.424)
	3	0.63	0.442 (−2.661, 3.829)
SS1	1	0.16	−1.265 (−4.603, 1.544)
	2	0.08	−1.542 (−4.912, 0.496)
	3	0.31	−0.717 (−4.169, 2.535)
SS2	1	0.54	0.124 (−2.898, 3.172)
	2	0.41	−0.173 (−2.259, 1.499)
	3	0.62	0.433 (−2.877, 3.79)
SS3	1	0.35	−0.535 (−4.043, 2.766)
	2	0.58	0.197 (−1.723, 3.211)
	3	0.39	−0.409 (−3.704, 2.936)
SS4	1	0.22	−1.092 (−4.787, 2.078)
	2	0.5	0.004 (−2.089, 2.853)
	3	0.45	−0.186 (−3.824, 3.275)

Overall

Full model	−0.281 (−1.412, 0.85)
No−background model	0.044 (−0.638, 0.726)
Non−linear model	−0.045 (−1.234, 1.143)

−5 −3 −1 0 1 2 3 4 5

Figure 3. Assessment of precaution measures efficacy via $P(\beta_1 < \beta_2 | y)$ and posterior median of $\log(\beta_1/\beta_2)$ with the 95% credible intervals. Each diamond box represents the overall precaution measures efficacy estimate from the meta-analysis and its 95% confidence interval. Model indicators: 1 – Full model; 2 – No-background model; 3 – Non-linear model.

that patients remained colonized for three months. Again, the results obtained were very similar to those in the original analysis.

6 Discussion

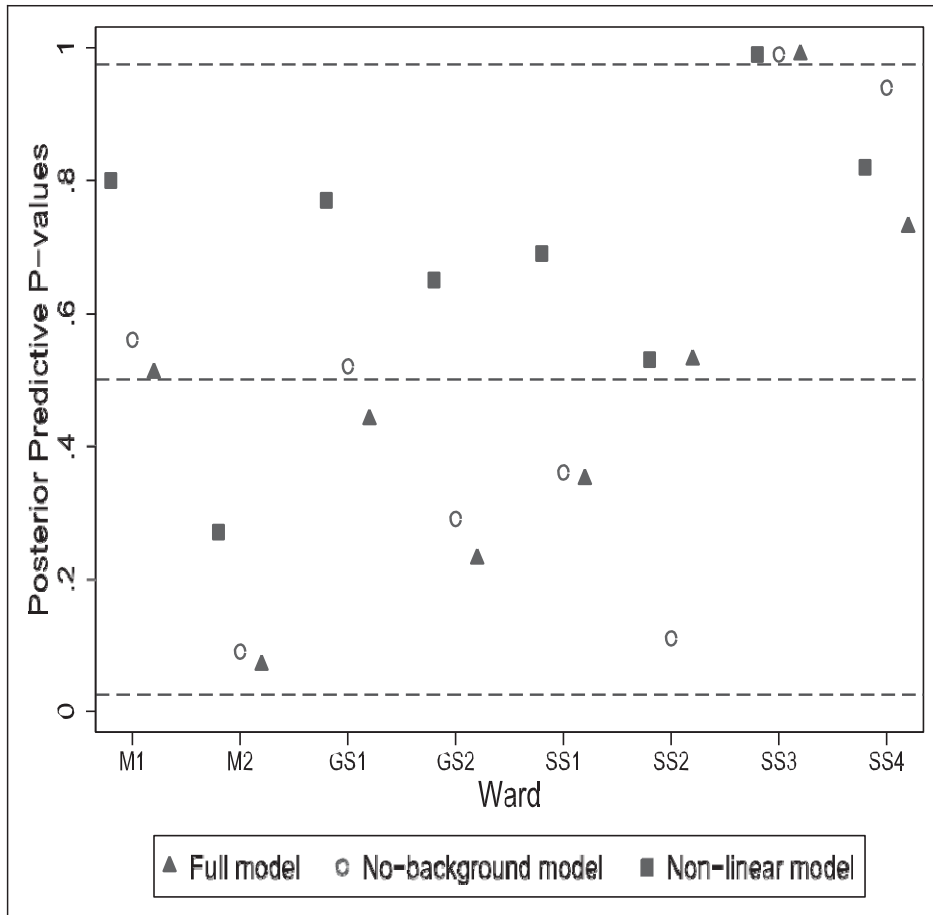We finish by discussing the results of the analysis of the VRE data set, and giving general comments on our methods.

Figure 4. Bayesian predictive posterior p-values based on number of detected colonizations. The horizontal lines represent $p ¼ 0.025, 0.5, 0.975$ from the bottom to the top.

6.1 VRE study

Our results show that almost half of VRE prevalence was unobserved. This was true despite high compliance admission rectal swabs for these ICU patients, suggesting that a single swab is insufficient for adequate sensitivity, as has been previously suggested.[26] In addition, despite the fact that weekly rectal screening for VRE was the standard of care in these ICUs, the short average length-of-stay meant that most patients only received one swab. Relatively high prevalence values are not unknown; for instance[27] reports that about 70% of swabs were found to be positive during a study in a UK haematology ward, while the prevalence estimated by Austin et al.[5] also reaches 70%. Furthermore, neither of these studies took into account imperfect swab test sensitivity as we have done.

For the full and non-linear models we found that the background rate of colonization ($f3_0$) was typically two to three times higher of that the colonization rates attributed to colonized patients ($f3_1, f3_2$), meaning that the background rate was roughly equivalent to that due to two or three
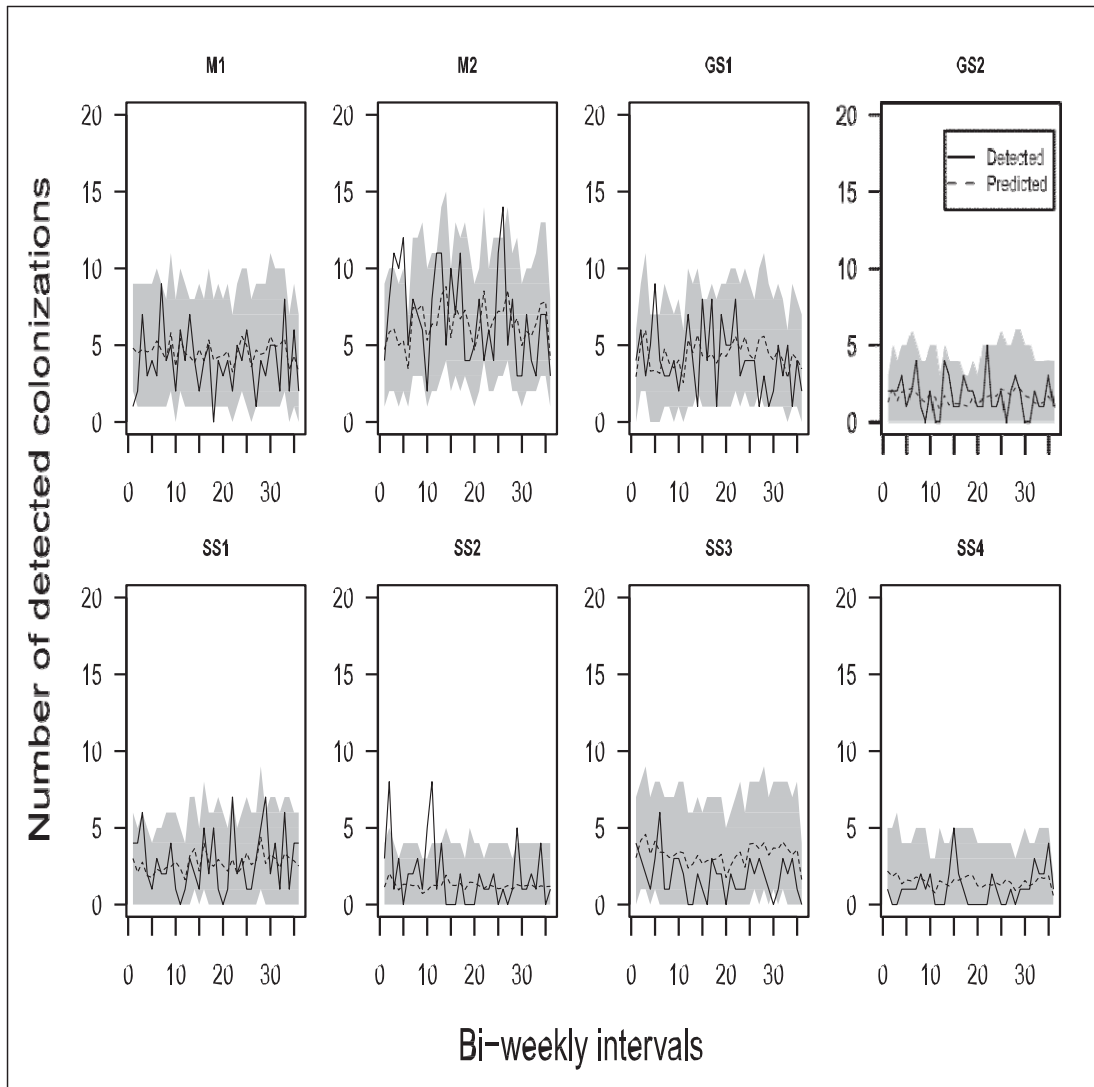
Figure 5. Full model. Number of detected colonizations at 14-day intervals. Solid lines – number of detected colonizations in data; dashed lines – mean number of detected colonizations based on 2000 simulations from the predictive distribution; shaded area – 2.5% and 97.5 % percentiles from the predictive distribution.

colonized patients on the ward. Background contamination is believed to be a possible route of VRE transmission[15,28,29] and these findings give no evidence to the contrary. However, some care is needed in interpreting $f3_0$ within our models, since it effectively models any colonization potential which is not associated with colonized patients in the ward, and in particular is not restricted to environmental contamination alone. We also found some pattern in the DIC values in table 4, suggesting different kinds of transmission in the specialty surgery wards compared to the other wards.
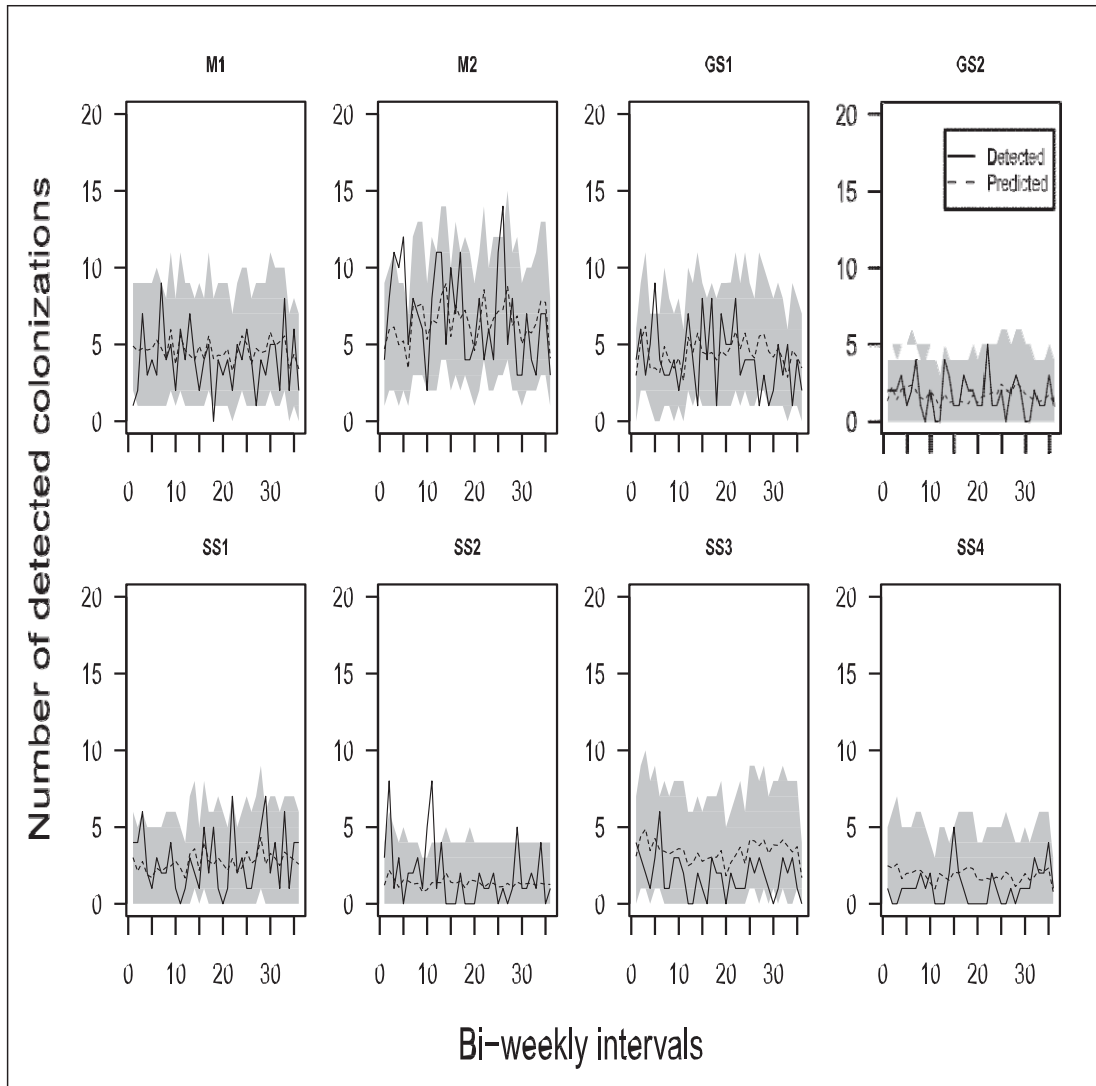
Figure 6. No-background model. Number of detected colonizations at 14-day intervals. Solid lines – number of detected colonizations in data; dashed lines – mean number of detected colonizations based on 2000 simulations from the predictive distribution; shaded area – 2.5% and 97.5 % percentiles from the predictive distribution.

We found no evidence to suggest that the use of barrier precautions was effective in reducing colonization. Broadly speaking, the colonization rates attributable to patients placed under precautions were not significantly different to the colonization rates attributable to patients who were not under precautions. One plausible explanation is that the barrier precautions really were not discernably more effective than standard procedures, especially given that at any one time there might only be one or two colonized patients on the ward, making any true difference hard to detect.
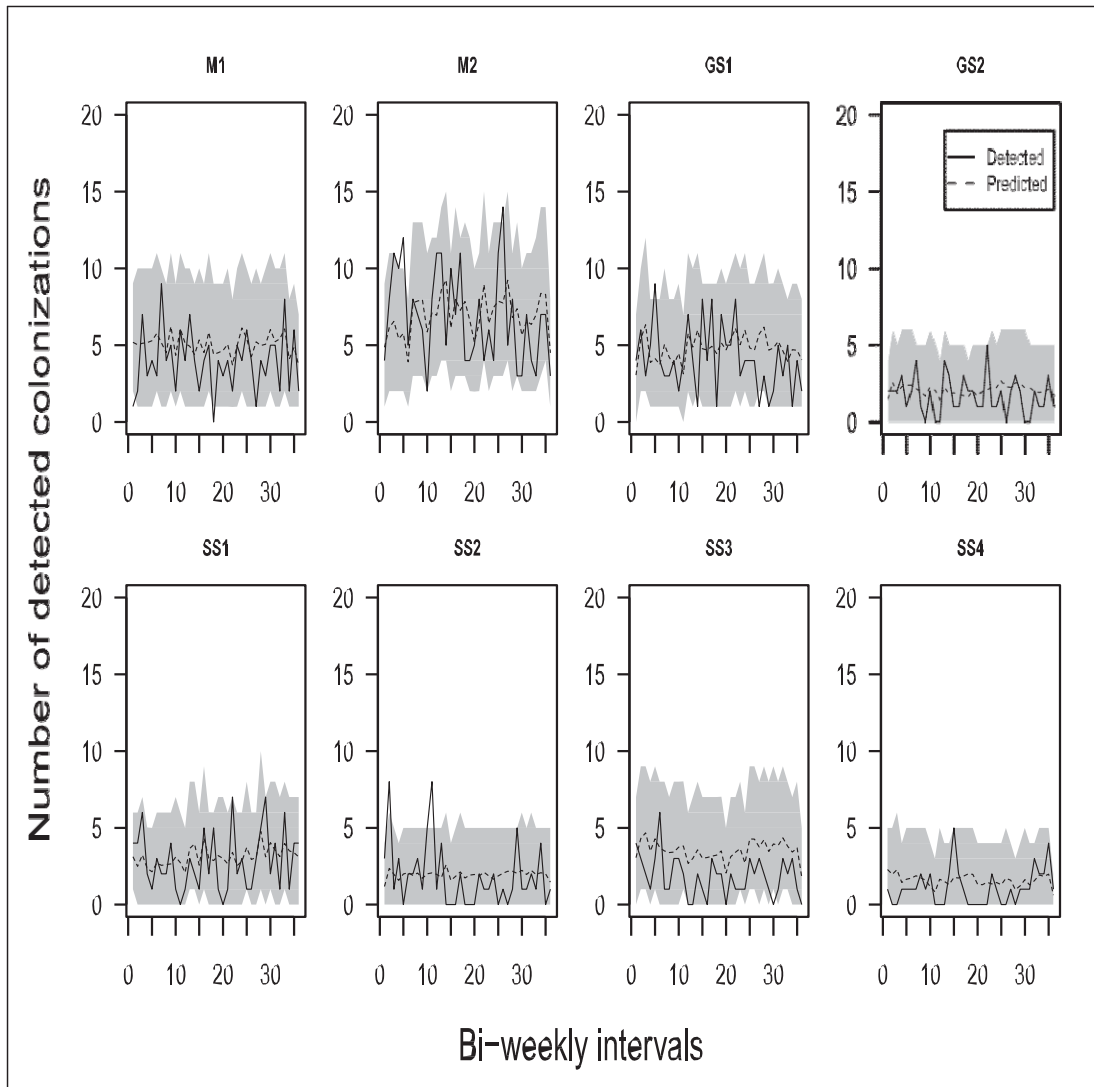
Figure 7. Non-linear model. Number of detected colonizations at 14-day intervals. Solid lines – number of detected colonizations in data; dashed lines – mean number of detected colonizations based on 2000 simulations from the predictive distribution; shaded area – 2.5% and 97.5 % percentiles from the predictive distribution.

6.2 Methods

We have described model-based methods for analysing patient-level data on nosocomial pathogens. The methods work at an individual level by explicitly keeping track of the status of each patient and assuming a mechanism by which patients may become colonized with the pathogen in question. One appealing aspect of this kind of modelling is that it is often straightforward to adapt it to

Table 4. DIC values.

	Full model	No-background	Non-linear
M1	1696.88	1688.15	1716.30
M2	1990.56	1927.34	2104.64
GS1	1817.78	1815.50	1831.92
GS2	1260.86	1196.64	1343.36
SS1	948.46	955.94	939.00
SS2	1186.40	1193.44	1171.1
SS3	1498.58	1499.42	1460.34
SS4	417.64	567.82	356.66

Note: Bold values indicate the preferred model.

include additional model features, as motivated by both the available data and the scientific questions of interest.

Our methods involved data augmentation, essentially providing estimates of quantities that are not directly observed. Although this was primarily motivated by the fact that such missing data enable us to calculate a likelihood, the fact that the missing data (namely, colonization times) correspond to real-life events mean that they can be exploited to provide further information. In our application we used them to estimate quantities that describe the impact of unobserved carriage.

We have considered three possible models for the colonization process, based on widely used natural assumptions for how the numbers of susceptible and colonized individuals affect the rate of new colonizations. Our purpose in doing so was to both explore the impact of particular assumptions regarding colonization, and to see how robust our general scientific conclusions were (e.g. regarding the efficacy of control measures) to the particular choice of model. However, it is clearly possible to develop other models. One example is to model colonization rates in the form $\beta_i C \delta t Þ^{\delta_i}$, $i = 1, 2$, with $0 \leq \delta_i \leq 1$ parameters to estimate (see e.g. O'Neill and Wen[30]), so that $\delta_i = 1$ gives the full model assumption and $\delta_i = 0$ gives the non-linear model assumption. However, estimating δ_i accurately would require a substantial amount of data, particularly if there are typically only a few colonized patients on the ward at any one time. Another option is to use something closer to non-parametric modelling, in which the colonization rate due to k colonized patients was β_k, either with or without constraints (e.g. requiring that β_k increases with k).

Our approach considers each ward separately, motivated by the fact that there is no a priori reason to assume common transmission parameters across wards. A natural way to relax this assumption would be to use a hierarchical model in which within-ward parameters are themselves sampled from some hyper-distribution, and all the data from all wards are analysed simultaneously, although in practice the amount of data augmentation is likely to make the MCMC algorithm infeasible. Alternatively, if data on staff movements between wards were available, it would then be possible to model between-ward interactions in a more explicit manner.

Acknowledgements

Funding

References

1. De Angelis G, Cataldo MA, De Waure C, et al. Infection control and prevention measures to reduce the spread of Vancomycin-resistant Enterococci in hospitalized patients: a systematic review and meta-analysis. *J Antimicrob Chemother* 2014; 69: 1185–1192.

2. Iosifidis E, Evdoridou I, Agakidou E, et al. Vancomycin-resistant Enterococcus outbreak in a neonatal intensive care unit: epidemiology, molecular analysis and risk factors. *Am J Inf Control* 2013; 41: 857–861.

3. Lee SC, Wu MS, Shih HJ, et al. Identification of Vancomycin-resistant Enterococci clones and inter-hospital spread during an outbreak in Taiwan. *BMC Infect Dis* 2013; 13: 163.

4. Kypraios T, O'Neill PD, Huang SS, et al. Assessing the role of undetected colonization and isolation precautions in reducing Methicillin-Resistant Staphylococcus transmission in intensive care units. *BMC Infect Dis* 2010; 10: 29.

5. Austin DJ, Bonten MJM, Weinstein RA, et al. Vancomycin-resistant enterococci in intensive-care hospital settings: transmission dynamics, persistence, and the impact of infection control programs. *Proc Natl Acad Sci* 1999; 96: 6908–6913.

6. Pelupessy I, Bonten MJM and Diekmann O. How to assess the relative importance of pathogens within hospital settings. *Proc Nat Acad Sci* 2002; 99: 5601–5605.

7. Bootsma MCJ, Bonten MJM, Nijssen S, et al. An algorithm to estimate the importance of bacterial acquisition routes in hospital settings. *Am J Epidem* 2007; 166: 841–851.

8. Drovandi CC and Pettitt AN. Multivariate Markov process models for the transmission of Methicillin-resistant Staphylococcus Aureus in a hospital ward. *Biometrics* 2008; 64: 851–859.

9. McBryde ES, Pettitt AN, Cooper BS, et al. Characterizing an outbreak of vancomycin-resistant enterococci using hidden Markov models. *J R Soc Int* 2007; 4: 745–754.

10. Wolkewitz M, Dettenkofer M, Bertz H, et al. Statistical epidemic modeling with hospital outbreak data. *Stat Med* 2008; 27: 6522–6531.

11. Forrester ML, Pettitt AN and Gibson GJ. Bayesian inference of hospital-acquired infectious diseases and control measures given imperfect surveillance data. *Biostatistics* 2007; 8: 383–401.

12. Gibson GJ and Renshaw E. Estimating parameters in stochastic compartmental models. *IMA J Math Appl Med Biol* 1998; 15: 19–40.

13. O'Neill PD and Roberts GO. Bayesian inference for partially observed stochastic epidemics. *J R Stat Soc* 1999; 162: 121–129.

14. Boyce JM, Mermal LA, Zervos MJ, et al. Controlling vancomycin-resistant enterococci. *Infect Control Hosp Epidemiol* 1995; 16: 634–637.

15. Bonten MJM, Hayden MK, Nathan J, et al. Epidemiology of colonization of patients and environment with vancomycin-resistant enterococci. *Lancet* 1996; 348: 1615–1619.

16. Perry JD, Davies A, Butterworth LA, et al. Development and evaluation of a chromogenic agar medium for methicillin-resistant Staphylococcus aureus. *J Clin Microbiol* 2004; 42: 4519–4523.

17. Huang SS, Rifas-Shaiman SL, Warren DK, et al. Improving MRSA surveillance and reporting in ICUs. *J Inf Dis* 2007; 195: 330–338.

18. Robicsek A, Beaumont JL and Peterson LR. Duration of colonization with methicillin-reistant Staphylococcus aureus. *Clin Infect Dis* 2009; 48: 910–913.

19. Scanvic A, Denic L, Gaillon S, et al. Duration of colonization by methicillin-resistant Staphylococcus aureus after hospital discharge and risk factors for prolonged carriage. *Clin Infect Dis* 2001; 32: 1393–1398.

20. Daley DJ and Vere-Jones D. *An introduction to the theory of point processes volume 1: elementary theory and methods*. New York: Springer-Verlag, 2003.

21. Spiegelhalter DJ, Best NG, Carlin BP, et al. Bayesian measures of model complexity and fit. *J R Stat Soc* 2002; 64: 583–616.

22. Celeux G, Forbes F, Robert CP, et al. Deviance information criteria for missing data models. *Bayesian Anal* 2006; 1: 651–674.

23. Worby CJ. *Statistical inference and modelling for nosocomial infections and the incorporation of whole genome sequence data*. PhD Thesis, University of Nottingham, UK, 2013. http://eprints.nottingham.ac.uk/13154/ (accessed 29 December 2015).

24. Gelman A, Meng XL and Stern H. Posterior predictive assessment of model finiteness via realized discrepancies. *Statistica Sinica* 1996; 6: 733–807.

25. Sutton AJ, Abrams KR, Jones DR, et al. *Methods for meta-analysis in medical research*. Chichester: Wiley, 2000.

26. D'Agata EM, Gautum S, Green WK, et al. High rate of false-negative results of the rectal swab culture method in detection of gastrointestinal colonization with Vancomycin-resistant enterococci. *Clin Infect Dis* 2002; 34: 167–172.

27. Bradley SJ, Kaufmann ME, Happy C, et al. The epidemiology of glycopeptide-resistant enterococci on a haematology unit – analysis by pulsed-field gel electrophoresis. *Epidemiol Infect* 2002; 129: 59–64.

28. Falk PS, Winnike J, Woodmansee C, et al. Outbreak of vancomycin-resistant enterococci in a burn unit. *Infect Control Hosp Epidemiol* 2000; 21: 575–582.

29. Martinez JA, Ruthazer R, Hansjosten K, et al. Role of environmental contamination as a risk factor for acquisition of vancomycin-resistant enterococci in patients treated in a medical intensive care unit. *Arch Intern Med* 2003; 163: 1905–1912.

30. O'Neill PD and Wen CH. Modelling and inference for epidemic models featuring non-linear infection pressure. *Math Biosci* 2012; 238: 36–48.