1 Introduction

Recent decades have seen the global emergence of Methicillin-resistant *Staphylococcus aureus* (MRSA), causing substantial health and economic burdens on patients and health-care systems. Eradicating, treating and containing the pathogen is complicated as this strain is resistant to certain antibiotics such as Methicillin, Cloxacillin and other related antibiotics (Ayliffe and English, 2003). This epidemic has occurred at the same time that policies, promoting higher patient throughput in hospitals, have led to many services operating at, or near, full capacity. A result has been a limited ability to scale services according to fluctuations in patient admissions and available staff, and hospital over-crowding and under-staffing. Clements et al. (2008) state that over-crowding and under-staffing lead to failure of MRSA control programmes via decreased health-care worker hand-hygiene compliance, increased movement of patients and staff between hospital wards, decreased levels of cohorting, and overburdening of screening and isolation facilities. In turn, a high MRSA incidence leads to increased inpatient length of stay and bed blocking, exacerbating over-crowding and leading to a vicious cycle characterised by further infection control failure.

In order to design hospital systems it is therefore critical to evaluate the effect of systems changes on the incidence of MRSA infection and other similar adverse events. Hence, it is important for hospital administrators to know that the methods presented in this paper suggest that transmission rates can be estimated accurately with quite coarse data and therefore allow detection of improvements from system changes.

In an intensive care unit (ICU) setting, the pathogen is usually spread via the hands of temporarily contaminated health-care workers who are in contact with colonised patients. Unfortunately it has been shown that many health-care practitioners do not comply with hand hygiene protocol (Harbarth, 2006). Other within-ICU modes of transmission are generally negligible as the patients are not mobile. However, patients being transferred from another hospital may be colonised or infected. Furthermore, new patients may unknowingly be carriers of the pathogen.

In this paper we use approximate Bayesian computation (ABC) to infer the transmission rates between patients and health-care workers and vice-versa. In particular we utilise the sequential Monte Carlo (SMC) likelihood-free algorithm of Drovandi and Pettitt (2011a), but any ABC algorithm could be applied. This involves a re-analysis of the continuous-time multivariate

Markov process models including the colonised status of patients and health-care workers developed in Drovandi and Pettitt (2008) and McBryde et al. (2007). The models in Drovandi and Pettitt (2008) are used to help explain routinely collected incidence data on colonised patients.

Motivation for the ABC approach stems from the difficulty in computing the likelihood function of the multivariate Markov process. One approach to calculating the likelihood involves computing the matrix exponential to obtain the probability transition matrix. Unfortunately for the full trivariate Markov process involving the colonised patients, health-care workers and incidence count, there are a large number of states in the Markov process and the matrix exponential is an expensive computation (Moler and Van Loan, 2003). However, we apply a pseudo-equilibrium approximation as per Drovandi and Pettitt (2008) and McBryde et al. (2007) to eliminate the health-care worker stochastic variable to create a more tractable likelihood. We compare the likelihood-based and likelihood-free inferences on the transmission rates for both the pseudo and full models.

Furthermore, the likelihood-based approach requires: (a) a bound on the incidence variable to create a finite state process and, (b) a fixed number of patients and health-care workers in the ICU. The ABC approach relies solely on simulation from the model and therefore does not suffer from the same restrictions.

The paper is organised as follows. In section 2 the data available for analysis is described. An introduction to ABC is provided in section 3 together with some analytic results on a toy example to highlight the error due to the ABC approximation. In section 4 we detail the models developed to provide a biological explanation of the data. Section 5 refers to one form of likelihood computation for multivariate Markov process models and we give details on model simulation in section 6. The inference results are provided in section 7 followed by a concluding discussion in section 8.

2 Data

The data available for analysis consist of 3,329 patient records collected between the 8th of August 2001 and the 3rd of March 2004 (inclusive) in the ICU at the Princess Alexandra Hospital, Brisbane, Australia. These patients had a hospital length of stay of at least 48 hours. The patients were swabbed for MRSA on admission, discharge and twice weekly.

Whilst individual patient data are available, data often take the form of daily/weekly/monthly prevalence or incidence counts as such data are reported routinely by hospitals. Here we use weekly incidence counts, which are shown in Figure 1. Here a new colonisation is defined as a patient whose swab was negative on admission and positive during their stay in the ward.

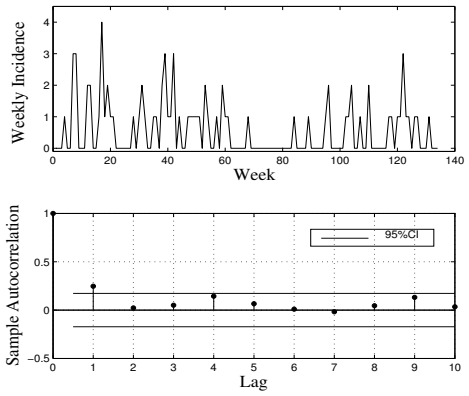

Figure 1: (top) Weekly incidence data collected from the ICU at the Princess Alexandra Hospital, Brisbane, Australia, between the 8th of August 2001 and the 3rd of March 2004. (bottom) Sample autocorrelation function of the weekly incidence data.

3 Approximate Bayesian Computation

A Bayesian statistician is interested in obtaining the posterior distribution, $\pi(\boldsymbol{\theta}|\boldsymbol{y})$, given by

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}),$$

where $\boldsymbol{\theta}$ is the parameter of interest, $\boldsymbol{y}$ is the observed data assumed to be drawn from $f(\cdot|\cdot)$ and $\pi(\cdot)$ is the prior. However, it is becoming increasingly apparent that there are many statistical models that do not emit a computationally tractable likelihood function.

Fortunately a class of simulation methodologies popularly termed approximate Bayesian computation (ABC) can produce statistically valid inferences about the posterior distribution when the likelihood function is not computationally tractable.

Although ABC approaches do not necessitate the evaluation of the likelihood function, it is paramount that simulating data from the statistical model is relatively fast to perform. The initial need for a likelihood function is alleviated by simulating data from the model and searching for parameter values that produce simulated data close to the observed data. This assessment of closeness typically involves comparing a set of summary statistics. Optimally, this set of statistics will be minimal sufficient statistics for the parameter of interest. In the absence of minimal sufficient statistics, the search for appropriate summary statistics is a trade-off between obtaining near sufficiency and a low dimensional set of summaries.

ABC methods are becoming an increasingly important component in the statisticians toolbox since they allow for inferences on certain statistical models that were previously problematic. Furthermore, the approach allows for progressively more realistic models to be developed. For example, ABC is now widely applied in population genetics examples using a coalescent model (see for example, Beaumont et al. (2002)). Other models include Ising type models (Grelaud et al., 2009), alpha-stable models (Peters et al., 2010), quantile distributions (Drovandi and Pettitt, 2011b) and Markov processes (Drovandi and Pettitt, 2011a). See also Blum and Tran (2010) for an ABC example on the types of models presented in this paper. The application areas of ABC is also widespread, from epidemiology (McKinley et al., 2009) through to finance (Peters et al., 2010).

To avoid the computation of the likelihood, ABC introduces an auxiliary variable, $\boldsymbol{x}$, which is the simulated data. The joint approximate posterior distribution of the parameter and the auxiliary data is given by

$$\pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y}, \epsilon) \propto f(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\phi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}, \epsilon), \tag{1}$$

(Luciani et al., 2009) where $f(\boldsymbol{x}|\boldsymbol{\theta})$ is the likelihood function evaluated at the simulated data and $\phi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}, \epsilon)$ is a weighting function that assesses the similarity between $\boldsymbol{x}$ and $\boldsymbol{y}$, giving high weight if they are close. Here ϵ is a tolerance specifying how close the simulated and observed data must be. Of interest then is the marginal approximate posterior distribution of the parameter which can be obtained by marginalising over the simulated data

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}, \epsilon) \propto \int_{\boldsymbol{x}} f(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\phi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}, \epsilon)d\boldsymbol{x}.$$

ABC algorithms are designed to sample from either or both of the marginal and joint targets (see Sisson et al. (2010) for more details).

Most weighting functions are of the form $\phi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}, \epsilon) = \phi(\boldsymbol{y}|\boldsymbol{x}, \epsilon)$ (Reeves and Pettitt, 2005), producing a simple hierarchical structure $\boldsymbol{y} \rightarrow \boldsymbol{x} \rightarrow \boldsymbol{\theta}$. There are several forms of the weighting function, but most involve a discrepancy function, $\rho(\boldsymbol{y}, \boldsymbol{x})$, that provides an overall measure of the difference between observed and simulated datasets. Typically, this function involves a comparison of a set of p summary statistics, $\boldsymbol{s}(\cdot) = \{s_1(\cdot), \ldots, s_p(\cdot)\}$, of the observed and simulated data

$$\rho(\boldsymbol{y}, \boldsymbol{x}) = \|\boldsymbol{s}(\boldsymbol{y}) - \boldsymbol{s}(\boldsymbol{x})\|,$$

for a suitably chosen norm. For example, the simplest and most commonly used weighting function is based on the indicator function

$$\phi(\boldsymbol{y}|\boldsymbol{x}, \epsilon) = 1(\rho(\boldsymbol{y}, \boldsymbol{x}) \leq \epsilon),$$

which is unity if the discrepancy function is less than or equal to a pre-defined tolerance ϵ. This is sometimes referred to as the uniform weighting function. The approach can be extended to B simulated datasets (e.g. Sisson et al. (2007)), which can produce general weights but we take $B = 1$ throughout. Another approach for assigning different weights to different parameter values is to use a non-uniform weighting function. Some options include the Epanechnikov and Gaussian weighting functions so that the weight varies smoothly with $\rho(\boldsymbol{y}, \boldsymbol{x})$. Apart from the section below, we use the uniform weighting function throughout.

3.1 The Effect of the Approximation

It is no surprise that there is some price to pay in terms of accuracy when the likelihood model is only used to simulate data. However, it can be shown that if sufficient statistics are used and $\epsilon = 0$ the true posterior is produced. In most applications of ABC, sufficient statistics are not available. As an alternative, one may use a careful selection of statistics that is believed to encompass most of the information in the observed data. Matching on these non-sufficient summary statistics leads to one source of error inherent in ABC.

Regardless of the use of summary or sufficient statistics, it is generally impractical (and completely impossible for continuous data) to attempt to match these statistics exactly as the acceptance probabilities are too low. To overcome this, ABC methods (Pritchard et al., 1999) introduce the tolerance,

ϵ (see Fu and Li (1997) and Weiss and von Haeseler (1998) for a non-Bayesian context), so that the summaries do not need to match exactly. This tolerance introduces a second source of error into the ABC approximation, but is necessary to ensure inferences can be obtained in reasonable time.

To examine these sources of error more closely, consider data $Y_i \overset{\text{iid}}{\sim} N(\mu, \phi)$, $i = 1, \ldots, n$, with an unknown mean, μ, and a known variance, ϕ. A sufficient statistic for μ is the sample mean, $\bar{y}$, which has the distribution $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \sim N(\mu, \phi/n)$. Consider further that the data is split into two independent components so that the original sufficient statistic becomes $\bar{y} = \lambda_1\bar{y}_1 + \lambda_2\bar{y}_2$, where $\boldsymbol{y_1}$ and $\boldsymbol{y_2}$ are the independent components and $\lambda_1 + \lambda_2 = 1$. Assume that the simulated data is also split into components of the same size, $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$. The non-sufficiency is introduced here by matching between observed and simulated data on the basis of $\boldsymbol{y_1}$ and $\boldsymbol{x_1}$. For analytic results consider the Gaussian weighting function $\phi(\boldsymbol{y}|\boldsymbol{x}, \epsilon) \propto e^{\frac{1}{2\epsilon}(\bar{y}_1 - \bar{x}_1)^2}$. Computing the marginal approximate posterior of μ by integrating out the simulated sample mean gives

$$\pi(\mu|\bar{y}_1, \epsilon) \propto N\left(\mu; \bar{y}_1, \frac{\phi}{\lambda_1 n} + \epsilon\right)\pi(\mu).$$

The true posterior is proportional to $N(\mu; \bar{y}, \frac{\phi}{n})\pi(\mu)$. It is evident that there are two sources of error; non-sufficiency of the summary statistic affects the mean of the likelihood and both aspects of the approximation inflate the variance of the likelihood.

3.2 Algorithms

3.2.1 Acceptance Sampling

The first ABC algorithms to appear were acceptance-based sampling algorithms and most involved applications in population genetics. The first genuine ABC method with summary statistics and a tolerance was developed by Pritchard et al. (1999). This basic algorithm, which involves draws from the prior predictive distribution, was popularised by Beaumont et al. (2002) and is presented in Algorithm 1.

3.2.2 Markov Chain Monte Carlo

The acceptance threshold can be improved in general if local moves are proposed in high posterior support regions, and thus can avoid wasteful proposals

Algorithm 1 Acceptance sampling ABC.

1: Draw $\boldsymbol{\theta} \sim \pi(\cdot)$
2: Simulate $\boldsymbol{x} \sim f(\cdot|\boldsymbol{\theta})$
3: If $\rho(\boldsymbol{y}, \boldsymbol{x}) \leq \epsilon$ then accept $\boldsymbol{\theta}$
4: Repeat lines 1, 2 and 3 until J samples are drawn

in regions of negligible posterior probability. This is particularly important in the ABC context, since generation of a good parameter value can still often lead to rejection due only to the variability in the simulated data. Therefore Marjoram et al. (2003) proposed a Markov chain Monte Carlo (MCMC) approach to ABC whereby a Markov chain is developed whose invariant distribution is the joint approximate posterior distribution of the space of simulated data and parameter in (1). Note that this approach also samples from the marginal target as a by-product via Monte Carlo integration. The proposal distribution is carefully selected so that evaluation of the likelihood is avoided. More specifically, the proposal for $(\boldsymbol{x}^*, \boldsymbol{\theta}^*)$ based on current values $(\boldsymbol{x}, \boldsymbol{\theta})$ is given by $q(\boldsymbol{x}^*, \boldsymbol{\theta}^*|\boldsymbol{x}, \boldsymbol{\theta}) = f(\boldsymbol{x}^*|\boldsymbol{\theta}^*)q(\boldsymbol{\theta}^*|\boldsymbol{\theta})$, which ensures that the Metropolis-Hastings ratio is free of likelihood evaluations.

The proposal distribution for the parameter, $q(\cdot|\cdot)$, is essentially arbitrary and can be tuned to achieve a desired acceptance probability. This MCMC approach can be found in full in Algorithm 2. This basic algorithm has since been extended (see for example, Bortot et al. (2007), Wegmann et al. (2009) and Sisson and Fan (2010)).

3.2.3 Sequential Monte Carlo

The next set of algorithms are based on SMC (Del Moral et al., 2006) approaches to ABC pioneered by Sisson et al. (2007). SMC methods are particularly suited to ABC, since a natural sequence of targets involves a non-increasing sequence of tolerances $\epsilon_1 \geq \epsilon_2 \geq \cdots \geq \epsilon_T$. More specifically, the sequence of joint targets is given by

$$\pi_t(\boldsymbol{\theta}, \boldsymbol{x}|\boldsymbol{y}, \epsilon_t) \propto f(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})1(\rho(\boldsymbol{y}, \boldsymbol{x}) \leq \epsilon_t), \text{ for } t = 1, \ldots, T.$$

The sequence of marginal targets follow naturally

$$\pi_t(\boldsymbol{\theta}|\boldsymbol{y}, \epsilon_t) \propto \int_{\boldsymbol{x}} f(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})1(\rho(\boldsymbol{y}, \boldsymbol{x}) \leq \epsilon_t)d\boldsymbol{x}, \text{ for } t = 1, \ldots, T.$$

Algorithm 2 MCMC ABC algorithm of Marjoram et al. (2003).

Start with $\boldsymbol{\theta^0}$ (e.g. a single draw from acceptance sampling ABC or use a burn-in)
for $i = 1$ **to** J **do**
  Draw $\boldsymbol{\theta^*} \sim q(\cdot|\boldsymbol{\theta^{i-1}})$
  Simulate $\boldsymbol{x^*} \sim f(\cdot|\boldsymbol{\theta^*})$
  Compute MH ratio $\alpha = \frac{\pi(\boldsymbol{\theta^*})q(\boldsymbol{\theta}|\boldsymbol{\theta^*})}{\pi(\boldsymbol{\theta})q(\boldsymbol{\theta^*}|\boldsymbol{\theta})} 1(\rho(\boldsymbol{y}, \boldsymbol{x^*}) \leq \epsilon)$
  if $U(0,1) < \alpha$ **then**
    $\boldsymbol{\theta^i} = \boldsymbol{\theta^*}$
  else
    $\boldsymbol{\theta^i} = \boldsymbol{\theta^{i-1}}$
  end if
end for

The SMC approach produces J weighted particles distributed according to each target in the sequence via a series of weighting, resampling and mutation steps.

In this paper we use the SMC ABC replenishment algorithm of Drovandi and Pettitt (2011a), which is described below. Consider that the particle set for target $t-1$ is given by $\{\boldsymbol{\theta^i_{t-1}}, \rho^i_{t-1}\}^J_{i=1}$. The algorithm determines the next tolerance, ϵ_t, dynamically by taking it as the $(1-\alpha)$th empirical quantile of the particles discrepancies, where α is a tuning parameter.

Since our mutation step makes use of an MCMC kernel, the particles values do not get updated from target $t-1$ to t but are simply re-weighted to reflect the new target (Chopin, 2002). The so called incremental weight calculation is given here by

$$\tilde{w}^i_t \propto \frac{1(\rho(\boldsymbol{x^i_{t-1}}, \boldsymbol{y}) \leq \epsilon_t)}{1(\rho(\boldsymbol{x^i_{t-1}}, \boldsymbol{y}) \leq \epsilon_{t-1})},$$

such that $W^i_t \propto \tilde{w}^i_t W^i_{t-1}$ where W^i_t is the normalised weight of the ith particle at target t. Clearly after the re-weighting step there will be $\leq$ J particles with non-zero weight, referred to hereafter as 'alive' particles. The number of particles with zero weight is controlled by α, for example if $\alpha = 0.5$ then there will be $J/2$ 'alive' particles. The advantage of this re-weighting step is that since we begin the algorithm with J perfect draws from the first target (by performing acceptance sampling with ϵ_1) the importance weights are either

proportional to one or equal to zero throughout and hence their values do not need to be maintained.

After the re-weighting step, to boost the particle population size back to J, we resample with replacement from the 'alive' particles, effectively duplicating some of the particles. To increase the diversity we use an MCMC kernel of invariant distribution involving the adaptively determined tolerance ϵ_t to move the resampled particles. Our MCMC proposal distribution, $q_t(\cdot|\cdot)$, is also determined adaptively in the spirit of Chopin (2002). More specifically the tuning parameters of q_t (for example the covariance matrix of a multivariate normal or t-distribution random walk) are inferred using sample moments of the 'alive' particles. Due to low MCMC acceptance rates inherent in ABC, we repeat the MCMC step R_t times such that

$$R_t = \frac{\log(c)}{\log(1 - p_{t-1}^{\text{acc}})},$$

where p_{t-1}^{acc} is the acceptance rate of the MCMC step of the previous iteration and c is a tuning parameter with a small value. It is clear that R_t is also determined dynamically.

It is the fully adaptive nature of the Drovandi and Pettitt (2011a) algorithm that makes it so attractive. The only tuning parameters consist of ϵ_1, ϵ_T, α and c. A reasonable choice for c is 0.01 and a sensible choice for α is 0.5, that is, to drop half the particles at each iteration. Furthermore, ϵ_1 can be chosen to achieve a particular acceptance rate in the initial acceptance sampling phase. Finally, the stopping rule for the algorithm could be when the MCMC acceptance rate becomes intolerably low, which determines ϵ_T. The main algorithm is presented in Algorithm 3 while more details of the MCMC step are given in Algorithm 4.

It appears quite clear that the MCMC proposal in the SMC setup, q_t, will be far more efficient than the usual MCMC proposal, q, for a fixed tolerance as q_t will contain a closer to optimal random walk standard deviation and can incorporate the correlations between parameters. This algorithm could be viewed as J interacting MCMC kernels running in parallel, and this helps prevent the algorithm from becoming stuck in areas of low posterior probability and will have an improved chance of representing multimodal targets.

The SMC ABC algorithm of Sisson et al. (2009) and Beaumont et al. (2009) use a forward kernel, instead of an MCMC kernel. To propagate the particle to the next target at t, a particle is first resampled from the

Algorithm 3 The SMC ABC replenishment algorithm of Drovandi and Pettitt (2011a).

1: Set J_a as the integer part of αJ
2: Perform the acceptance sampling algorithm with ϵ_1. This produces a set of particles $\{\boldsymbol{\theta^i}, \rho^i\}_{i=1}^{J}$
3: Sort the particle set by ρ and set $\epsilon_t = \rho^{J-J_a}$ and $\epsilon_{\max} = \rho^J$. If $\epsilon_{\max} \leq \epsilon_T$ then finish, otherwise go to 4
4: Compute the tuning parameters of the MCMC kernel $q_t(\cdot|\cdot)$ using the particle set $\{\boldsymbol{\theta^i}\}_{i=1}^{J-J_a}$
5: **for** $j = J - J_a + 1$ **to** J **do**
6: Resample $\boldsymbol{\theta^j}$ from $\{\boldsymbol{\theta^i}\}_{i=1}^{J-J_a}$
7: Apply the MCMC kernel with invariant distribution involving ϵ_t to particle $\boldsymbol{\theta^j}$ for R_t iterations. See Algorithm 4
8: **end for**
9: Compute R_t based on the overall MCMC acceptance rate of the previous iteration and go to 3

Algorithm 4 The MCMC step of the Drovandi and Pettitt (2011a) algorithm for the jth particle.

1: **for** $k = 1$ **to** R_t **do**
2: Propose $\boldsymbol{\theta^*} \sim q_t(\cdot|\boldsymbol{\theta^j})$ and simulate $\boldsymbol{x^*} \sim f(\cdot|\boldsymbol{\theta^*})$
3: Compute MH rato $\alpha = \frac{\pi(\boldsymbol{\theta^*})q(\boldsymbol{\theta^j}|\boldsymbol{\theta^*})}{\pi(\boldsymbol{\theta^j})q(\boldsymbol{\theta^*}|\boldsymbol{\theta^j})} 1(\rho(\boldsymbol{y}, \boldsymbol{x^*}) \leq \epsilon_t)$
4: **if** $U(0,1) < \alpha$ **then**
5: Set $\boldsymbol{\theta^j} = \boldsymbol{\theta^*}$ and $\rho^j = \rho(\boldsymbol{y}, \boldsymbol{x^*})$
6: **end if**
7: **end for**

current population of particles at $t-1$ proportional to their weights, $\boldsymbol{\theta^*} \sim \{\boldsymbol{\theta_{t-1}^i}, W_{t-1}^i\}_{i=1}^J$. To ensure diversity, the particle's parameter is perturbed according to a Markov kernel, $\boldsymbol{\theta^{**}} \sim K_t(\cdot|\boldsymbol{\theta^*})$, and then data is simulated from the model, $\boldsymbol{x^{**}} \sim f(\cdot|\boldsymbol{\theta^{**}})$. If the distance between $\boldsymbol{x^{**}}$ and $\boldsymbol{y}$ is within the current tolerance, ϵ_t, the parameter value $\boldsymbol{\theta_t^i} = \boldsymbol{\theta^{**}}$ is accepted, otherwise the process is repeated until the condition is satisfied. The particle is re-weighted according to

$$W_t^i \propto \frac{\pi(\boldsymbol{\theta_t^i})}{\sum_{j=1}^N W_{t-1}^j K_t(\boldsymbol{\theta_t^i}|\boldsymbol{\theta_{t-1}^j})},$$

which essentially amounts to an approximation of the importance distribution based on the particle set. Unlike our algorithm the importance weights will be non-uniform hence their values will need to be maintained. However this algorithm will not suffer from duplicated particles but the effective sample size (ESS) will be less than J, whereas the ESS in our algorithm is always equal to J. We implement both of these algorithms for one of the models described below and present the results in section 7.1.1.

3.3 Learning the Optimal MCMC Kernel

The choice of an MCMC kernel has the additional benefit that the method of Fearnhead and Taylor (2010) can be applied. Fearnhead and Taylor (2010) propose to allow each particle to have its own MCMC kernel and compare the performance of each kernel using a criterion such as the expected squared jumping distance (ESJD). The SMC algorithm then attempts to learn the optimal MCMC kernel in the original subset of MCMC kernels by weighting the kernels (based on the chosen criterion) after each MCMC step. Resampling these MCMC kernels proportional to their weights eliminates poor kernels and duplicates kernels that are performing well.

The MCMC proposals we consider are of the form, $\boldsymbol{\theta}^* \sim N(\boldsymbol{\theta_t^i}, g^2\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\pi_t}})$ then $\boldsymbol{x}^* \sim f(\cdot|\boldsymbol{\theta}^*)$, where $\boldsymbol{\theta_t^i}$ is particle i at target t, $(\boldsymbol{\theta}^*, \boldsymbol{x}^*)$ is a proposal (parameter, simulated data) and $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\pi_t}}$ estimates the covariance matrix of the parameter at the current target π_t using the surviving particles. Therefore the aim is to determine the optimal scaling, g, of the random walk proposal, however it is straightforward to extend the set of MCMC proposals as per Fearnhead and Taylor (2010).

The performance criterion we use is the same as that in Fearnhead and Taylor (2010) and is based on the ESJD, which is given by

$$\Lambda(\boldsymbol{\theta_t^i}, \boldsymbol{\theta^*}) = (\boldsymbol{\theta_t^i} - \boldsymbol{\theta^*})^T \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\pi_t}}^{-1} (\boldsymbol{\theta_t^i} - \boldsymbol{\theta^*}).$$

The jumping distance is then multiplied by the acceptance probability of the proposal to form the final criterion. The acceptance probability is zero or one in this context (since we use a uniform prior, a symmetric proposal and the indicator ABC weighting function). However, as opposed to only one MCMC iteration in Fearnhead and Taylor (2010), we apply the MCMC step R_t times. We use these R_t proposals of parameter and simulated data, $(\boldsymbol{\theta_1^*}, \boldsymbol{x_1^*}), \ldots, (\boldsymbol{\theta_{R_t}^*}, \boldsymbol{x_{R_t}^*})$, of the MCMC kernel to obtain an estimate of the kernel's performance by summing the criterion over all iterates. The resulting criterion for this application is given by

$$\tilde{\Lambda}(\boldsymbol{\theta_t^i}, \boldsymbol{\theta_{R_t}^*}) = \sum_{k=1}^{R_t} 1(\rho(\boldsymbol{x_k^*}, \boldsymbol{y}) \leq \epsilon_t) \Lambda(\boldsymbol{\theta_t^i}, \boldsymbol{\theta_k^*}),$$

keeping in mind that $\boldsymbol{\theta_t^i}$ will change throughout the R_t proposals every time an acceptance occurs. Note that it may also be possible to learn the optimal value of g through these R_t iterates but we take this more conservative approach. In order to weight the performance of each MCMC kernel we use a simple linear weighting function, $f(\tilde{\Lambda}) = \tilde{\Lambda}$.

Incorporating this into the algorithm is relatively straightforward. After the MCMC step $f(\tilde{\Lambda})$ must be computed for each particle. The MCMC kernels are then resampled proportional to their weights. It is also possible to add noise to the resampled values in order to prevent the distribution of g from converging too quickly to a point distribution, but we do not apply this here. See section 7.1.3 for application of this approach to our ABC algorithm.

4 Modelling and Parameters

4.1 Deterministic Model

A two compartment model was developed by McBryde et al. (2007) that models the number of colonised patients and health-care workers at time t, given by $Y_p(t)$ and $Y_h(t)$, respectively. The uncolonised states are not required

in the model as it is assumed that the numbers of patients and health-care workers are fixed at N_p and N_h, respectively.

Given that the ward size is fixed, it is assumed a discharged patient is immediately replaced with a new patient. Colonised and uncolonised patients are discharged at per-capita rates of μ' and μ respectively. The new patient to replace the discharged patient has a probability σ of being colonised.

Colonised health-care workers can become uncolonised due to a hand-washing event that occurs at a rate of κ per number of colonised health-care workers. There is assumed to be no spontaneous decolonisation of patients, so patients that are colonised before or during their stay remain so.

Finally, the colonised populations can increase by a transmission between a colonised patient and an uncolonised health-care worker or vice-versa. The parameter c is a common contact rate. The probability of a transmission between a colonised patient and an uncolonised health-care worker and vice-versa is given by p_{ph} and p_{hp} respectively. The rates of transmission are also dependent on the relevant population sizes. The above assumptions lead to the following deterministic system, which is shown as a compartment diagram in Figure 2

$$\begin{aligned} \frac{dY_p}{dt} &= cp_{hp}(N_p - Y_p)Y_h - \mu'(1-\sigma)Y_p + \mu\sigma(N_p - Y_p), \\ \frac{dY_h}{dt} &= cp_{ph}(N_h - Y_h)Y_p - \kappa Y_h. \end{aligned} \tag{2}$$

4.2 Assumptions and Fixed Parameters

We will assume further that the patient to health-care ratio is unity so that $N_h = N_p = N$. While in the study the ward size did vary slightly we will hold the ward size constant at its average value, $N = 15$. The parameter values obtained at the time of the study were $\sigma = 0.03$, $\mu' = 1/10.6$ per day, $\mu = 1/4$ per day and $p_{ph} = 0.13$. The parameters c, p_{hp} and κ need to be inferred from the data. However, Cooper et al. (1999) provide a formula to express κ in terms of known parameters by using the following relation between the pre-contact hand hygiene compliance, h (proportion of patient contacts that were preceded by hand-washing), and the hand hygiene rate

$$h = \frac{\kappa}{\kappa + cN}. \tag{3}$$

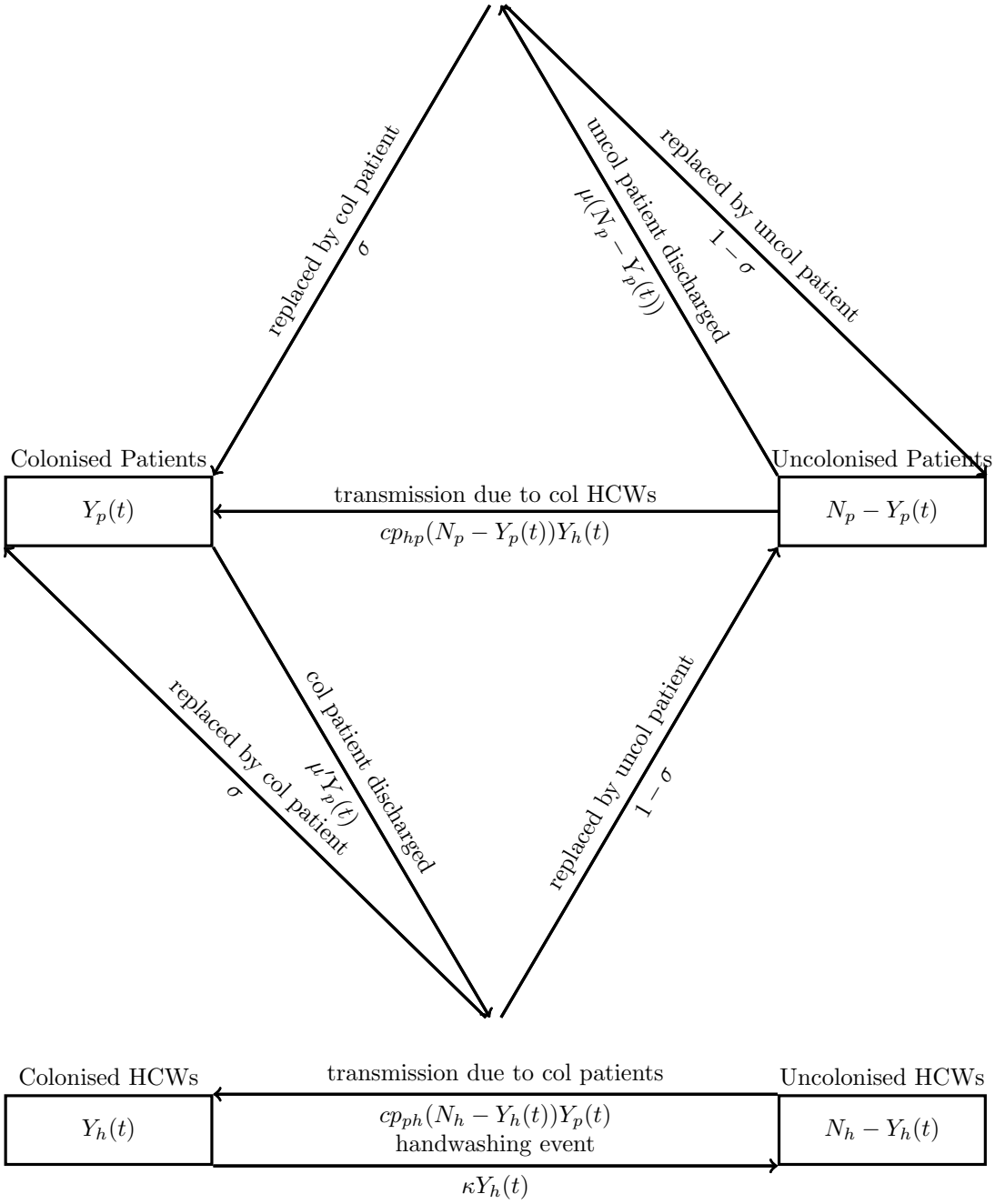

Figure 2: Compartment diagram for the model developed by McBryde et al. (2007).

See Drovandi and Pettitt (2008) for a theoretical justification. The hand hygiene compliance was measured to be $h = 0.59$ at the time of study. In section 7.1.2 we incorporate the uncertainty associated with these parameters into the analysis.

4.3 Markov Process Modelling

4.3.1 Trivariate Model

Since the population size N is small, it is appropriate to consider the stochastic variation in Y_p and Y_h as discrete, random counts (Bailey, 1990). At this stage we introduce a third variable, $N(t)$, which we define to be the incidence, that is, the number of new patient colonisations up to time t. We define a new colonised patient as one who becomes colonised as a result of transmission assumed to take place within the ICU. Using the model described in equation (2), we can create the analogous discrete Markov process. Given that the current values of the states are $Y_p(t) = i$, $Y_h(t) = j$ and $N(t) = k$, and a small time increment, Δ_t, that allows at most one event to occur, the probabilities of various combinations of the states at time $t + \Delta_t$ are given by

$$\begin{aligned}
P(Y_p = i+1, Y_h = j, N = k) &= \mu\sigma(N-i)\Delta_t + o(\Delta_t),\\
P(Y_p = i+1, Y_h = j, N = k+1) &= \phi_1(N-i)j\Delta_t + o(\Delta_t),\\
P(Y_p = i-1, Y_h = j, N = k) &= \mu'(1-\sigma)i\Delta_t + o(\Delta_t),\\
P(Y_p = i, Y_h = j+1, N = k) &= \phi_2(N-j)i\Delta_t + o(\Delta_t),\\
P(Y_p = i, Y_h = j-1, N = k) &= \frac{h\phi_2 N}{p_{ph}(1-h)}j\Delta_t + o(\Delta_t),
\end{aligned} \tag{4}$$

where $\phi_1 = cp_{hp}$ and $\phi_2 = cp_{ph}$. The probability of remaining in the current state is given by one minus the sum of the above probabilities. All other transitions occur with probability $o(\Delta_t)$. The likelihood-based inference has the restriction on the incidence that $N(t) \leq M$ for all t. $N(t)$ is reset to zero following each incidence recording.

4.3.2 Approximating Bivariate Model

McBryde et al. (2007) introduce an approximation to the differential equation formulation (2) that can be used to eliminate the stochastic health-care

worker variable. The rate of change of the colonised health-care worker population is assumed to be zero. This is called a pseudo-equilibrium approximation, denoted by $\bar{Y}_h$. Setting $\frac{dY_h}{dt} = 0$ and making the substitution for κ given in equation (3), we obtain the following steady state result

$$\bar{Y}_h = \frac{NY_p}{\frac{hN}{p_{ph}(1-h)} + Y_p}.$$

We can use this relation in the Markov processes and the trivariate process (4) reduces to a bivariate process. Given that the current values of the states are $Y_p(t) = i$ and $N(t) = k$, and a small time increment, Δ_t, the probabilities of various combinations of the states at time $t + \Delta_t$ are given by

$$\begin{aligned} P(Y_p = i+1, N = k) &= \mu\sigma(N-i)\Delta_t + o(\Delta_t), \\ P(Y_p = i+1, N = k+1) &= \phi_1(N-i)\bar{Y}_h\Delta_t + o(\Delta_t), \\ P(Y_p = i-1, N = k) &= \mu'(1-\sigma)i\Delta_t + o(\Delta_t). \end{aligned}$$

We see that this model has only one parameter for inference, ϕ_1.

5 Computing the Likelihood

One approach to computing the likelihood of Markov processes involves the matrix exponential, which arises from continuous-time Markov chain theory (Grimmett and Stirzaker, 2001). Such a computation can be expensive in particular for multivariate processes as the number of states in the Markov chain can be large (see Sidje (1998) and Moler and Van Loan (2003) for the computational difficulties of computing the matrix exponential for large matrices). Additionally, computing the likelihood involves a marginalisation step as only one of the variables is actually observed. See the supplementary material of Drovandi and Pettitt (2008) for more details on computing the likelihood. Overall computing the likelihood is both computationally and algorithmically more difficult than simulating data from the model, implying that a simulation based approach is faster and easier to implement.

6 Model Simulation

Given values for the parameters, it is relatively straightforward to simulate data from a Markov process model using Gillespie's algorithm (Doob, 1945;

Gillespie, 1977). This algorithm involves simulating the time until the next event with an exponential distribution and choosing an event type based on their relative hazards. Simulating for incidence has an additional step in that the incidence must be reset to zero just after an incidence observation is collected. The starting values for the colonised patients and health-care workers is simulated from the appropriate stationary distribution.

7 Results

For each run of the SMC ABC replenishment algorithm below we used $J =$ 1,000, $\alpha = 0.5$ and $c = 0.01$. Furthermore the MCMC move kernel was a multivariate normal random walk with covariance matrix estimated empirically using the particles. The summary statistic used here was the mean of the weekly incidence data and the discrepancy function was the absolute value of the difference between the simulated and observed summaries.

7.1 Results of Inference for the Bivariate Model

7.1.1 Comparison of Likelihood-free and Likelihood-based Inference

We compared the ABC approximation to that when the likelihood function is available for the pseudo-equilibrium model. We ran the SMC ABC replenishment algorithm setting the initial and target tolerance of $\epsilon_1 = 8$ and $\epsilon_T = 0.04$ respectively (we found that reducing the tolerance further had negligible impact on the approximate posterior). Secondly we performed a normal random walk MCMC likelihood-based algorithm with a proposal standard deviation of 0.01 (tuned to ensure an acceptance rate of roughly 50%). We performed 11,000 iterations of this algorithm, discarding the first 1,000 as burn-in and thinning out the resulting sample by a factor of 10, producing more or less 1,000 independent draws. For this likelihood-based inference, we set $M = 4$ as the maximum allowable incidence value.

The resulting inferences are comparable. The ABC analysis produced a median with a 95% credible interval of 0.041(0.031,0.054) while the equivalent results for the likelihood-based analysis were 0.040(0.030,0.052). A comparison of these posterior distributions is shown in Figure 3 (solid and dashed lines). We also implemented the algorithm of Beaumont et al. (2009) with a tolerance schedule of $\{8, 6.95, 5.05, 2.48, 0.57, 0.28, 0.15, 0.075, 0.04\}$, which

was the adaptive set of tolerances determined by the replenishment algorithm. From a single run, both algorithms required approximately 25,000 model simulations, where the replenishment algorithm finished with 944 unique particles while the algorithm of Beaumont et al. (2009) resulted in an ESS of 926. Both algorithms produced very similar posterior distributions, and were validated against the posterior obtained using acceptance sampling, which produces perfect draws from the ABC posterior. The acceptance sampling algorithm was ten times less efficient, requiring roughly 250,000 model simulations for 1,000 samples.

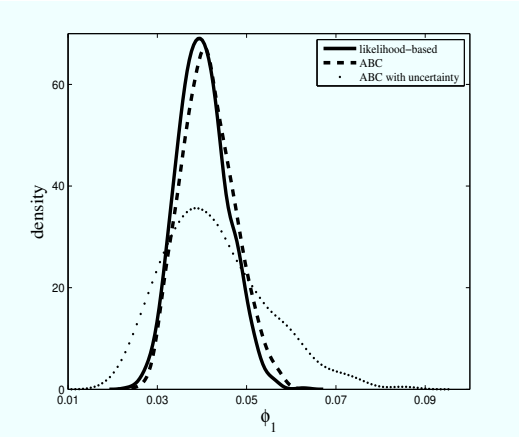

Figure 3: Posterior distributions for ϕ_1 of the pseudo-equilibrium model with likelihood-based inference (solid line), ABC (dashed line) and ABC including uncertainty in the fixed parameters (dotted line).

The closeness of the likelihood-free and likelihood-based inferences implies that the sum of the new cases over the time period is almost a sufficient statistic and fully efficient in terms of the likelihood analysis. This suggests that inferences will not be sensitive to the time grouping of the data (e.g. daily, weekly or monthly). This is consistent with the results of the time grouping sensitivity analysis in Drovandi and Pettitt (2008).

7.1.2 Incorporating Uncertainty in Fixed Parameters

The parameters assumed to be fixed in the model were all estimated at the time of the study and hence there is an element of uncertainty in their values. Out of the 3,329 patients, 100 were colonised upon entry. Over the time

period of the study, 235 and 89 patients were discharged in an uncolonised and colonised state respectively. The parameter p_{ph} was estimated based on 17 positive hand cultures out of 129 patient visits by health-care workers. Finally, $h = 0.59$ was estimated from 395 compliances out of 668 contacts.

Here we assigned an uninformative $U(0, 1)$ prior for the proportion parameters, namely h, p_{ph} and σ. We assume that one binomial observation (described above) is collected for each of these parameters, creating beta posterior distributions with parameter (α, β) given by $(396, 274)$, $(18, 113)$ and $(101, 3230)$ for h, p_{ph} and σ, respectively.

For the rate parameters, μ and μ', an uninformative improper prior that is proportional to the reciprocal of the parameter is assigned. Assuming that patients are discharged according to a Poisson process, gamma posteriors with parameter (a, b) given by $(235, 939)$ and $(89, 939)$ are obtained for μ and μ', respectively.

We incorporated the uncertainty in our ABC analysis using these probability distributions as priors for these parameters but these prior distributions are not updated with the incidence data. The plot of the posterior when including the uncertainty is shown by the dotted line in Figure 3. It is evident that the posterior variance of ϕ is substantially increased when incorporating the uncertainty, producing a 95% credible interval of (0.024,0.070). However point estimation based on the median would be similar under the three posterior distributions shown in Figure 3.

7.1.3 Optimal MCMC Random Walk Scaling

Here we revisited the one parameter likelihood-free problem in section 7.1.1 and applied the method of Fearnhead and Taylor (2010) in the ABC context based on the specifications given in section 3.3.

The set of allowable scalings was initially generated from a $U(0, 10)$ distribution. The final distribution of g following the end of the SMC ABC algorithm are shown in Figure 4. The performance of the random walk MCMC kernel appears to be relatively insensitive to the choice of g, but values below 1 and above 5 appear to be quite poor scalings. Note that the choice of $g = 1$ in section 7.1.1 would appear to be sub-optimal in terms of sampler performance. However, it is computationally less intensive as we found that larger values of g produced lower acceptance rates.

To investigate the validity of the distribution of g obtained from the algorithm, we ran the MCMC ABC algorithm of Marjoram et al. (2003) with

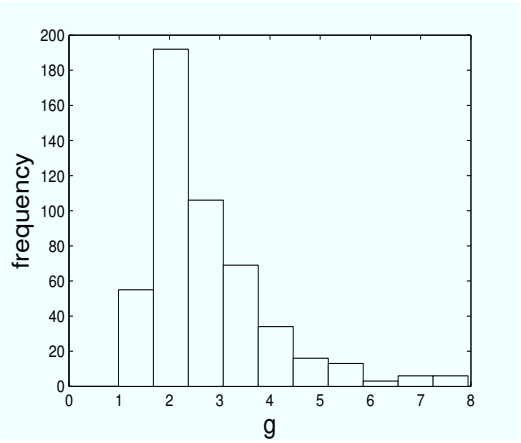

Figure 4: Distribution of the scaling parameter g produced from the SMC ABC algorithm that utilises the approach of Fearnhead and Taylor (2010) for the one parameter bivariate model.

a variety of different scalings and investigated the autocorrelation function up to 20 lags. We found that the optimal value of g could be anywhere between 1.25 and 4 with small values of g resulting in particularly poor mixing, which is in some agreement with the distribution obtained in Figure 4.

7.2 Results of Inference for the Trivariate Model

Next we considered obtaining inferences for ϕ_1 and ϕ_2 of the trivariate model. We placed independent $U(0,1)$ priors on the parameters. The prior for ϕ_2 may be justified since we expect this parameter to be around the same order as ϕ_1. For the ABC analysis we set $\epsilon_1 = 7.5$ and $\epsilon_T = 0.04$. For the likelihood-based inference we used a random walk Metropolis-Hastings algorithm with a standard deviation of 0.01 and 0.1 for the proposal of ϕ_1 and ϕ_2 respectively. We ran this algorithm for 11,000 iterations discarding the first 1,000 as burn-in. The acceptance rate was about 50%. We thinned the output by a factor of 10, producing 1,000 simulations from the true posterior.

The posterior distribution for ϕ_1 produced by the ABC analysis is very similar to the ABC posterior for ϕ_1 of the pseudo model with the median and 95% credible interval for the former being 0.041(0.031,0.053). See Figure 5 (left) for posterior densities for ϕ_1 of the trivariate model based on likelihood-free and likelihood-based inferences. The ABC posterior dis-

tribution for ϕ_2 was almost flat over (0,1), indicating that with this choice of summary statistic the data are not informative at all about this parameter. However, small values of ϕ_2 are slightly less preferred as can be seen by comparing the empirical posterior cumulative distribution function of ϕ_2 with the cumulative distribution function of a uniform random variate in Figure 5 (right). We also tried adding in extra summary statistics such as the variance of the incidence data but this did not improve the precision of our estimates.

The likelihood-based analysis also produced an imprecise estimate for ϕ_2 and is similar to the ABC posterior of this parameter. However, the availability of the likelihood allowed us to rule out very small values of ϕ_2, where the posterior distribution has an approximate range of (0.015,1). The inference on ϕ_2 is generally still very imprecise, however. To investigate the poor identification of this parameter, we produced many simulations from the model with ϕ_1 fixed at 0.04 and various values of ϕ_2. We found that the trajectories of the incidence, patient prevalence and health-care worker prevalence were very similar in terms of the mean and spread at each time point for values of ϕ_2 greater than 0.01. Small values of ϕ_2 led to slightly lower means but slightly greater variances at each time point for the patient incidence and prevalence variables. Therefore small values of ϕ_2 allow for larger incidence values to be realised but are still consistent with the summary statistic (sum of the incidence counts) as the mean is lower. The slightly informative inference from the likelihood-based approach may be misleading as the arbitrary upper limit of the incidence is set to $M = 4$ to keep computation tractable and a small value of ϕ_2 can produce incidence counts substantially larger than this. Therefore the inference on ϕ_2 may be sensitive to the choice of M.

The parameter ϕ_2 has a negligible impact on the colonised health-care worker population. This provides support as to why the pseudo-equilibrium model is a reasonable approximation in this scenario.

8 Discussion

Here we have presented a likelihood-free methodology to estimate the transmission rates of nosocomial MRSA, which has become endemic in many hospitals (Tiemersma et al., 2004; Evans and Brachman, 1998; Harbarth, 2006). We considered a pseudo-equilibrium simplification so that inferences

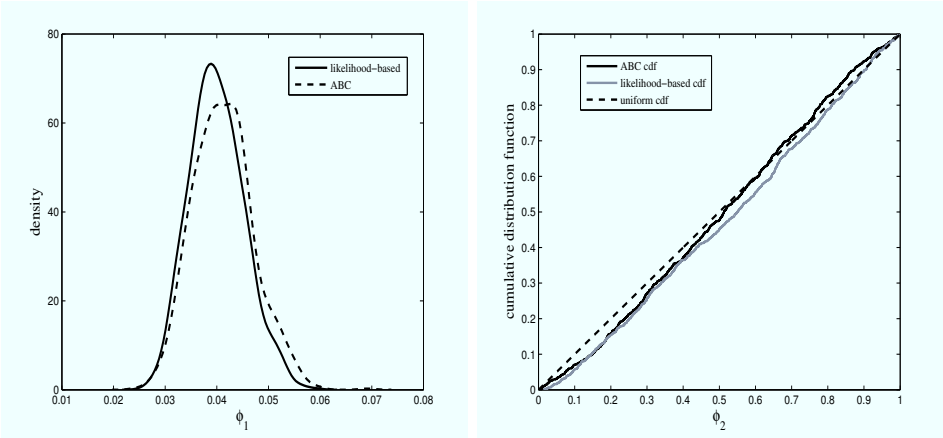

Figure 5: Posterior summaries of ϕ_1 and ϕ_2 of the trivariate model. For ϕ_1 (left) posterior densities are shown for likelihood-based inference (solid line) and ABC (dashed line). For ϕ_2 (right) the posterior empirical cumulative distribution functions are provided based on ABC (black) and likelihood-based inference (grey). The dashed straight line on the graph on the right denotes the theoretical cumulative distribution function of a $U(0,1)$ random variate.

could be obtained more easily for one of the parameters, compromising little on model fit. There was an agreement with the inferences of the likelihood-based and likelihood-free approaches, with the likelihood-free method easier to implement and faster to run.

The likelihood of the pseudo-equilibrium model is relatively straightforward to compute. However, likelihood evaluation of the full trivariate model is substantially more complicated as it involves calculating the exponential of a large matrix. Its computation is time consuming in a high level language such as Matlab® (MathWorks, 2008) and can be unstable in some regions of the parameter space. There would be extra motivation for an ABC approach if the time intervals between observations were not constant, since a matrix exponential would be required for each unique time difference. Furthermore, the likelihood would become even less tractable if the ward size increased and varied, and the incidence count was larger (see Drovandi et al. (2011) for an ABC example on macroparasite population evolution, where larger populations exist). Finally, in the likelihood-based analysis, an arbitrary upper limit for the incidence, M, required specification.

References

Ayliffe, G. A. J. and English, M. P. (2003). *Hospital Infection: From Miasmas to MRSA*. Cambridge University Press, Cambridge.

Bailey, N. T. J. (1990). *The Elements of Stochastic Processes with Applications to the Natural Sciences*. Wiley-Interscience.

Beaumont, M. A., Cornuet, J.-M., Marin, J.-M., and Robert, C. P. (2009). Adaptive approximate Bayesian computation. *Biometrika*, 96(4):983–990.

Beaumont, M. A., Zhang, W., and Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics*, 162(4):2025–2035.

Blum, M. G. B. and Tran, V. C. (2010). HIV with contact tracing: a case study in approximate Bayesian computation. *Biostatistics*, 11(4):644–660.

Bortot, P., Coles, S. G., and Sisson, S. A. (2007). Inference for stereological extremes. *Journal of the American Statistical Association*, 102(477):84–92.

Chopin, N. (2002). A sequential particle filter method for static models. *Biometrika*, 89(3):539–551.

Clements, A., Halton, K., Graves, N., Pettitt, A., Morton, A., Looke, D., and Whitby, M. (2008). Overcrowding and understaffing in modern health-care systems: key determinants in Methicillin-resistant *Staphylococcus aureus* transmission. *The Lancet Infectious Diseases*, 8(7):427–434.

Cooper, B. S., Medley, G. F., and Scott, G. M. (1999). Preliminary analysis of the transmission dynamics of nosocomial infections: stochastic and management effects. *Journal of Hospital Infection*, 43:131–147.

Del Moral, P., Doucet, A., and Jasra, A. (2006). Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436.

Doob, J. L. (1945). Markoff chains–denumerable case. *Transactions of the American Mathematical Society*, 58(3):455–473.

Drovandi, C. C. and Pettitt, A. N. (2008). Multivariate Markov process models for the transmission of Methicillin-resistant *Staphylococcus aureus* in a hospital ward. *Biometrics*, 64(3):851–859.

Drovandi, C. C. and Pettitt, A. N. (2011a). Estimation of parameters for macroparasite population evolution using approximate Bayesian computation. *Biometrics*, 67(1):225–233.

Drovandi, C. C. and Pettitt, A. N. (2011b). Likelihood-free Bayesian estimation of multivariate quantile distributions. *Computational Statistics and Data Analysis*, 55(9):2541–2556.

Drovandi, C. C., Pettitt, A. N., and Faddy, M. J. (2011). Approximate Bayesian computation using indirect inference. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 60(3):503–524.

Evans, A. S. and Brachman, P. S. (1998). *Bacterial Infections of Humans: Epidemiology and Control*. Plenum Medical Book Co., New York.

Fearnhead, P. and Taylor, B. M. (2010). An adaptive sequential Monte Carlo sampler. *arXiv:1005.1193v2*.

Fu, Y. X. and Li, W. H. (1997). Estimating the age of the common ancestor of a sample of DNA sequences. *Molecular Biology and Evolution*, 14(2):195–199.

Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361.

Grelaud, A., Robert, C. P., Marin, J.-M., Rodolphe, F., and Taly, J.-F. (2009). ABC likelihood-free methods for model choice in Gibbs random fields. *Bayesian Analysis*, 4(2):317–336.

Grimmett, G. R. and Stirzaker, D. R. (2001). *Probability and Random Processes*. Oxford University Press, New York, third edition.

Harbarth, S. (2006). Control of endemic Methicillin-resistant *Staphylococcus aureus* - recent advances and future challenges. *Clinical Microbiology and Infectious Diseases*, 12:1154–1162.

Luciani, F., Sisson, S. A., Jiang, H., Francis, A. R., and Tanaka, M. M. (2009). The epidemiological fitness cost of drug resistance in *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences*, 106(34):14711–14715.

Marjoram, P., Molitor, J., Plagnol, V., and Tavaré, S. (2003). Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 100(26):15324–15328.

MathWorks (2008). *Matlab & Simulink Release Notes for Release 2008a*. The MathWorks, Inc.

McBryde, E. S., Pettitt, A. N., and McElwain, D. L. S. (2007). A stochastic mathematical model of Methicillin-resistant *Staphylococcus aureus* transmission in an intensive care unit: predicting the impact of interventions. *Journal of Theoretical Biology*, 245(3):470–481.

McKinley, T., Cook, A. R., and Deardon, R. (2009). Inference in epidemic models without likelihoods. *The International Journal of Biostatistics*, 5(1):24.

Moler, C. and Van Loan, C. (2003). Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Review*, 45(1):3–49.

Peters, G. W., Sisson, S. A., and Fan, Y. (2010). Likelihood-free Bayesian inference for alpha-stable models. *doi:10.1016/j.csda.2010.10.004*.

Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A., and Feldman, M. W. (1999). Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular Biology and Evolution*, 16(12):1791–1798.

Reeves, R. W. and Pettitt, A. N. (2005). A theoretical framework for approximate Bayesian computation. In Francis, A. R., Matawie, K. M., Oshlack, A., and Smyth, G. K., editors, *Proceedings of the 20th International Workshop on Statistical Modelling*, pages 393–396, Sydney, Australia.

Sidje, R. B. (1998). Expokit: a software package for computing matrix exponentials. *ACM Transactions on Mathematical Software (TOMS)*, 24(1):130–156.

Sisson, S. A. and Fan, Y. (2010). *To appear in MCMC handbook*, chapter Likelihood-free Markov chain Monte Carlo. Chapman & Hall.

Sisson, S. A., Fan, Y., and Tanaka, M. M. (2007). Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 104(6):1760–1765.

Sisson, S. A., Fan, Y., and Tanaka, M. M. (2009). Correction for Sisson et al., sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 106(39):16889–16890.

Sisson, S. A., Peters, G. W., Briers, M., and Fan, Y. (2010). A note on target distribution ambiguity of likelihood-free samplers. *arXiv:1005.5201v1*.

Tiemersma, E. W., Bronzwaer, S. L. A. M., Lyytikainen, O., Degner, J. E., Schrijnemakers, P., Bruinsma, N., Monen, J., Witte, W., and Grundmann, H. (2004). Methicillin-resistant *Staphylococcus aureus* in Europe, 1999-2002. *Emerging Infectious Diseases*, 10:1627–1634.

Wegmann, D., Leuenberger, L., and Excoffier, L. (2009). Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. *Genetics*, 182:1207–1218.

Weiss, G. and von Haeseler, A. (1998). Inference of population history using a likelihood approach. *Genetics*, 149(3):1539–1546.