Bayesian Smoothing Models for Estimating Hazard Functions of the Methicillin-Resistant *Staphylococcus Aureus* Colonization Data

Noor Azina Ismail* and A. N. Pettitt**

*Department of Applied Statistics

Faculty of Economics and Administration

University of Malaya, 50603 Kuala Lumpur, Malaysia

email: nazina@um.edu.my

**Centre in Statistical Sciences and Industrial Mathematics

School of Mathematical Sciences

Queensland University of Technology

GPO Box 2434, Brisbane, Qld 4001, Australia

email: a.pettitt@fsc.qut.edu.au

Abstract

A new method for predicting the time to colonization of patients is developed in this paper. The time to colonization of MRSA is modeled using a Bayesian smoothing approach for the hazard function. There are two prior models discussed in this paper: the first difference prior and the second difference prior. The second difference prior model gives smoother estimates of the hazard functions. The resulting hazard and survival probabilities obtained from these methods could be used to find the expected number of patients, $\hat{N}$ to be colonized in a fixed number of future days. Then $\hat{N}$ can be used as a control limit for a control chart.

1 Introduction

In hospitals, predictions are made based on the times to occurrence of an event and these are crucial to our understanding of the spread or evolution of the event. In this paper, we study the analysis involved in making predictions for the occurrence of Methicillin-Resistant *Staphylococcus Aureus* (MRSA) colonization in the Intensive Care Unit (ICU) at the Princess Alexandra (PA) hospital, Brisbane, Australia. Routinely, ICU personnel at this hospital had to rely on the Acute Physiology Age Chronic Health Evaluation (APACHE) III system [1, 2, 3] to predict hospital outcomes such as mortality. Colonization, as a hospital outcome, had never been considered.

Since this study focuses on MRSA colonization, instead of infection, it is important to distinguish between the two. Cultures from a patient were tested using the appropriate test discussed by Noor Azina [4]. If one of the cultures is positive, but the patient does not show any signs of illness, then such patient is considered to be colonized with MRSA [5, 6, 7]. A patient is said to be infected if there is evidence of fever, raised peripheral white cell count [8] and laboratory findings meeting the definition of nosocomial infection [7]. Many studies [6, 9, 7] reported that colonization preceeds infection, with more than 30% of patients who are colonized

with MRSA developing an MRSA infection such as wound infection, bacteremia, pneumonia or urinary tract infection.

Studies [6, 10, 11] found that two of the factors associated with an increased risk of acquiring MRSA are prolonged hospital stay and presence in an ICU and hence the study on how long before a patient in an ICU is colonized with MRSA is very crucial. In this paper, the time to colonization is the number of days a patient is free from colonization. From the data, only 4.5% ($\frac{123}{2758}$) of all patients in the ICU are colonized. For the rest of the patients, the time to colonization is censored in that if we had the time when the swabs indicated that the patient was last free of colonization and was first colonized, then the actual time of colonization would have occurred sometime between these two swab times. This situation occur since the date when a swab, later proved to be cultured-positive, is unavailable. Then for patients noted as being colonized with MRSA at any time, the time to MRSA colonization is taken to be the date of admission to the ICU to the date of notification of colonization, or the date of discharge from the ICU. Another point to note is that for censored data we assume that the reason for a patient being discharged from the ICU and free of colonization, is unrelated to their being colonized or not. Finally, we note that some patients die in the ICU and these patients are treated the same way as those who are discharged for the purposes of this study on colonization.

Section 2 gives the estimation of hazard and survival functions. In our data, time to colonization takes values from 1 to the maximum number of days spent in the ICU, J. Two nonparametric autoregressive prior models, involving first and second differences, are introduced in Section 3 to obtain smooth estimates of the functions and automatically estimate the hazard function for days for which there is no data. Breslow and Clayton [12] use the nonparametric autoregressive smoothing method based on a generalized linear mixed model (GLMM) with an autoregressive error component for smoothing the birth cohort effects in an age-cohort model of breast cancer

incidence. They use the model in which each point is predicted by linear extrapolation from its two immediate predecessors rather than just the most immediate one. Spiegelhalter et al [13] have implemented this model in BUGS. We use BUGS to calculate a posterior distribution over the parameter and unobserved data.

In Section 4, we discuss how the results obtained from nonparametric autoregressive smoothing are used to predict the number of patients with MRSA colonization. We show how to use the results to set up control limits for use for controlling the number of colonizations. We also investigate whether a Poisson or a binomial approximation is suitable to handle our problem. Section 5 highlights the findings in this paper and limitations of the study.

2 Estimation of Hazard and Survival Functions

Let T denotes a positive random variable representing the time to colonization of individuals in a homogeneous population where T can take discrete values $1, 2, 3, \ldots$. We assume that if colonization and departure from the ICU happens on the same day for a patient, then that case is in the risk set for that number of days of 'survival' (or days free from colonization), as is the patient who is colonized on that day but does not leave the ICU. In other words, if a patient is noted to be colonized on day t where $t = 1, \ldots, T$, then the patient is included in the risk set for days $1, \ldots, t$ and the time to colonization is taken to be t. As an illustration, if a patient is noted to be colonized on day 1, he or she is included in the count, n_1, of the number of patients at risk on day 1 and, since the time to colonization starts from 1 and not 0, the time to colonization for this patient is taken to be 1.

Following Kalbfleisch and Prentice [14], the product moment (non-parametric) estimate of

the survival function $S(T+1)$, the probability of a patient not colonized on day $t+1$, is

$$\hat{S}(t+1) = \prod_{j=1}^{t} \left(1 - \frac{r_j}{n_j}\right) \quad t \geq 1, \tag{1}$$

and $\hat{S}(1) = 1$. We define n_j as the number of individuals at risk at t and r_j as the number of individuals colonized with MRSA.

The corresponding estimate of the hazard function, $h(t)$, the probability of a patient colonized with MRSA on day t given that they are not colonized before day t, can be estimated by the ratio of the number of patients colonized on day t to the number of patients at risk on day t. Giving the estimate

$$\hat{h}(t) = \frac{r_t}{n_t}. \tag{2}$$

The range of times to colonization in our data is from 1 to 35 days with an average of $3\frac{1}{2}$ days. Since r_t and n_t are known for each t, we can calculate the estimate of $h(t)$ (equation (2)), and estimates of $S(t)$ (equation (1)). For a day where $r_t = 0$, the hazard function estimate $\hat{h}(t)$ is not defined and it would be unreasonable to take it equal to zero. None of the patients was colonized after day 28 and so the hazard function is only calculated up to day 28. Table 1 shows values of the hazard function estimates and Figure 1 shows a plot of $\hat{h}(t)$ interpolating between values of $h(t)$ for days where there are no colonizations is not smooth. The hazard function is initially small, owing to the very small number of patients colonized with MRSA at their early stage of stay in the ICU. This is followed by a period of relatively high hazard of colonization after which the colonization hazard decreases and the cycle starts again. We now show how to employ the non-parametric autoregressive smoothing [12, 13] to make predictions on MRSA colonizations to improve estimates of the survival and the hazard functions.

Table 1: Summary of MRSA data, estimated hazard function and estimated survival function. The estimated standard error and 95% confidence interval for the survival function at each time point are also included. When $r_t = 0$ the hazard rate is found by interpolation where possible but extrapolation is not used.

Time to colonization t	Number at risk n_t	Number of colonizations r_t	Hazard function $\hat{h}(t)$	Survival function $\hat{S}(t+1)$	Standard Deviation of $\hat{S}(t+1)$	95% Confidence Interval
1	2758	1	0.0004	0.9996	0.0004	(0.9989, 1.0004)
2	2521	7	0.0028	0.9969	0.0011	(0.9946, 0.9991)
3	1136	17	0.0150	0.9819	0.0038	(0.9744, 0.9895)
4	723	17	0.0235	0.9589	0.0066	(0.9455, 0.9721)
5	512	8	0.0156	0.9439	0.0084	(0.9271, 0.9606)
6	376	7	0.0186	0.9263	0.0105	(0.9052, 0.9474)
7	298	9	0.0302	0.8983	0.0137	(0.8709, 0.9258)
8	237	13	0.0549	0.8491	0.0186	(0.8119, 0.8862)
9	190	6	0.0316	0.8222	0.0210	(0.7803, 0.8642)
10	161	5	0.0311	0.7967	0.0232	(0.7503, 0.8431)
11	144	2	0.0139	0.7856	0.0242	(0.7373, 0.8340)
12	120	8	0.0667	0.7333	0.0288	(0.6757, 0.7909)
13	98	5	0.0510	0.6959	0.0318	(0.6322, 0.7595)
14	77	4	0.0520	0.6597	0.0349	(0.5899, 0.7296)
15	66	3	0.0455	0.6297	0.0374	(0.5550, 0.7045)
16	58	1	0.0172	0.6189	0.0383	(0.5423, 0.6954)
17	52	2	0.0385	0.5951	0.0403	(0.5144, 0.6757)
18	41	3	0.0731	0.5515	0.0445	(0.4624, 0.6405)
19	36	0	0.0513	0.5434	0.0453	(0.4528, 0.6340)
20	34	1	0.0294	0.5353	0.0461	(0.4431, 0.6275)
21	30	1	0.0333	0.5175	0.0479	(0.4217, 0.6132)
22	26	0	0.0389	0.5089	0.0493	(0.4103, 0.6075)
23	24	0	0.0444	0.5003	0.0507	(0.3989, 0.6017)
24	20	1	0.0500	0.4916	0.0520	(0.3876, 0.5956)
25	16	0	0.0792	0.4711	0.0561	(0.3589, 0.5833)
26	15	0	0.1083	0.4506	0.0602	(0.3302, 0.5710)
27	13	0	0.1375	0.4301	0.0643	(0.3016, 0.5587)
28	12	2	0.1667	0.4096	0.0684	(0.2729, 0.5464)
29	9	0	NA	NA	NA	NA
30	9	0	NA	NA	NA	NA
31	8	0	NA	NA	NA	NA
32	8	0	NA	NA	NA	NA
33	5	0	NA	NA	NA	NA
34	3	0	NA	NA	NA	NA
35	1	0	NA	NA	NA	NA

Note that $S(1) = 1$.

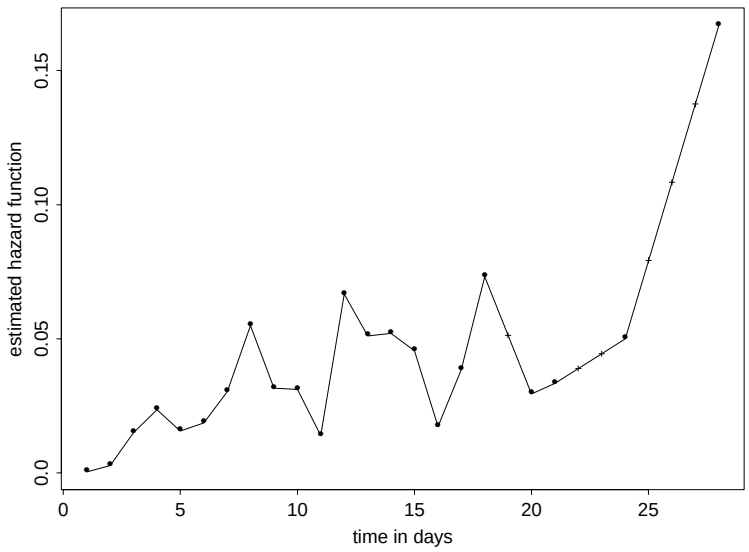

Figure 1: Estimates of colonization rates for the MRSA data. Interpolated values are shown with +.

3 Nonparametric Smoothing Method

We first define the Bayesian smoothing model. We assume that the number of patients colonized with MRSA, r_t, has the binomial distribution with index n_t and parameter h_t. n_t is the number of patients at risk on day t, so $r_t \sim \text{Bin}(n_t, h_t)$ where $h_t = P(\text{colonized on day } t | \text{not colonized before day } t)$ and $0 < h_t < 1$.

Now let us choose the logit transformation of the hazard function, h_t, for MRSA colonization, as the link function so that

$$\text{logit}(h_t) = \log\left[\frac{h_t}{1-h_t}\right] = \lambda_t,$$

and since $0 < h_t < 1$, the range for λ_t is $-\infty < \lambda_t < \infty, t = 1, 2, \ldots$.

Therefore, the likelihood, which is the product of independent binomials [14], written in

terms of λ_t is given by

$$p(\underset{\sim}{r}|\underset{\sim}{\lambda}) = \prod_{t=1}^{T} \binom{n_t}{r_t} \left(\frac{e^{\lambda_t}}{1+e^{\lambda_t}}\right)^{r_t} \left(\frac{1}{1+e^{\lambda_t}}\right)^{n_t - r_t}, -\infty < \lambda_t < \infty.$$

Since the model has been specified, there remains to specify the prior distribution for the unknown parameters, λ_t. The joint distribution of $\underset{\sim}{r}$ and $\underset{\sim}{\lambda}$ is

$$p(\underset{\sim}{r}, \underset{\sim}{\lambda}) = \left[\prod_{t=1}^{T} p(r_t|\lambda_t)\right] p(\underset{\sim}{\lambda}).$$

If there are data only on certain days D then the likelihood $\times$ prior is

$$p(\underset{\sim}{r}, \underset{\sim}{\lambda}) = \left[\prod_{t\in D} p(r_t|\lambda_t)\right] p(\underset{\sim}{\lambda}) \tag{3}$$

and inference for $\lambda_t, t \notin D$ is obtained from $p(\underset{\sim}{\lambda}|\underset{\sim}{r})$ as normal. Suppose only one day is missing, t_0, say. The likelihood is $\prod_{t\neq t_0} p(r_t|\lambda_t)$. The prior is $p(\lambda_{t_0}|\lambda_{\setminus t_0})p(\lambda_{\setminus t_0})$ where $\setminus t_0$ indicates all t such that $t \neq t_0$, so $p(\lambda_{t_0}|\lambda_{\setminus t_0}, \underset{\sim}{r})$ depends only on the full conditional prior $p(\lambda_{t_0}|\lambda_{\setminus t_0})$. The choice of the prior distribution is investigated in the next sections.

Before this choice is made, we discuss the model and analysis in more detail. We deal with a method which is called Bayesian analysis of the generalized linear mixed model (GLMM). The model described above is a generalized linear model (GLM). Our interest is to extend this model to include random effects. GLMs that include random effects are known as GLMM [12, 15]. GLMMs enable the accommodation of nonnormally distributed and correlated responses. When the full joint distribution for all quantities is defined (as for equation (3)), inference can be made based on Bayes theorem to calculate the posterior distribution of any unobserved quantities of interest, conditional on all observed data. However, one problem in making inference about GLMMs has been computational [16, 17]. A variety of techniques is available for fitting such models, such as penalized quasi-likelihood [12], best linear unbiased predictor (BLUP) [18],

maximum hierarchical likelihood [19], a method based on simulated moments [17] and many more. We choose to take an approach of simulating 'exact' solutions where the accuracy of the solution depends only on the computational resource. This approach is known as the Markov chain Monte Carlo (MCMC) method and specifically Gibbs sampling. Using MCMC methods, the prior distribution for λ is assumed to be multivariate normal with variance-covariance matrix Σ.

The random effects can also be dependent. This is especially true in our case because we have to rely on small samples to get precise estimates of the hazard function. As a consequence, we have a situation where there is a tendency for the effects of two levels which are close together to be similar. The most commonly used prior model that fits this description is the first difference model, also known as the random walk model. We also discuss a similar model known as the stochastic trend model.

3.1 The Prior Models

The simple random walk model is commonly used in Bayesian forecasting [20]. This model is also used effectively in numerous applications, particularly in short term forecasting, production planning and stock control [21]. For a random walk, each effect is derived from the immediately preceding effect. Another prior distribution that is considered in this study is the stochastic trend model. The stochastic-trend, also known as the second-difference prior, is where each effect is obtained by extrapolation from its two immediate predecessors. These models can be illustrated as a directed acyclic graph (DAG) as discussed in Clayton [15]. A full discussion of conditional independence graphs can be found in Whittaker [22]. In short, the graphical model expresses the assumption that the conditional mean for each effect is the mean of the immediately neighbouring effect.

The prior distribution for $\underset{\sim}{\lambda}$ in equation (3) can always be written as

$$\begin{aligned} p(\underset{\sim}{\lambda}|\tau) &= p(\lambda_1, \ldots, \lambda_T|\tau) \\ &= p(\lambda_1|\tau)p(\lambda_2|\lambda_1, \tau)p(\lambda_3|\lambda_1, \lambda_2, \tau)\ldots p(\lambda_T|\lambda_1, \ldots, \lambda_{T-1}, \tau) \end{aligned} \tag{4}$$

The prior, however, does not specify a distribution of $\lambda_1, p(\lambda_1|\tau)$, for the random walk model. On the other hand, the distributions of λ_1 and λ_2, $p(\lambda_1|\tau)$ and $p(\lambda_2|\tau)$, are unspecified for the stochastic-trend model.

We need to choose the unspecified distribution appropriately so that the joint distribution $p(\lambda_1, \ldots, \lambda_\tau|\tau)$ has the property of symmetry with respect to λ_1 and λ_T, λ_2 and λ_{T-1}, etc. So, for example, $(\lambda_1|\lambda_2, \ldots, \lambda_T) \sim (\lambda_T|\lambda_{T-1}, \ldots, \lambda_1)$. The unspecified distribution is chosen to be normally distributed with a very small precision $(0.000001 \times \tau)$. We assume the prior distribution of λ to be normal with $\sigma = \tau^{-1/2}$. We retain the τ term in order to provide the appropriate likelihood for τ. We choose τ to be distributed as gamma(0.001, 0.001). This choice is noninformative but proper [23], and it was used before by Spiegelhalter et al [13].

In summary, our random walk prior model can be written as

$$\begin{aligned} \lambda_1 &\sim N(0, \frac{10^6}{\tau}) \\ \lambda_t &\sim N(\lambda_{t-1}, \tau^{-1}) \quad t = 2, 3, \ldots \\ \tau &\sim \text{gamma}(0.001, 0.001), \end{aligned}$$

and the second difference model can be written as

$$\begin{aligned} \lambda_1 &\sim N(0, \frac{10^6}{\tau}) \\ \lambda_2|\lambda_1 &\sim N(0, \frac{10^6}{\tau}) \\ \lambda_t &\sim N(2\lambda_{t-1} - \lambda_{t-2}, \tau^{-1}) \quad t = 3, 4, \ldots \\ \tau &\sim \text{gamma}(0.001, 0.001). \end{aligned}$$

For data analysis, λ_T would normally correspond to the largest observation but could be defined for T beyond this time for extrapolating the hazard function estimate in a smooth way.

3.2 Estimated Hazard and Survival Functions

We have already specified prior distributions for $\underset{\sim}{\lambda}$. The choice of starting values should not be important since the Gibbs sampler should be run long enough for it to overcome the influence of its initial states [24]. In our analysis, the initial run is 5000.

Full conditional distributions are derived from the joint distribution of the variables. The full conditional distribution for a quantity is the distribution of that quantity given current or known values for all other quantities. Consequently, to construct the full conditional distribution of λ_t, we need only pick out the terms in equation (3) which involve λ_t. Similarly, the full conditional distribution for τ depends only on the terms in equation (3) involving τ. Gibbs sampling algorithms begin by specifying starting values and then iteratively drawing samples from the full conditional distributions of unobserved parameters. The BUGS program then identifies the relevant terms for each full conditional distribution and works out how to compute the necessary inferential statements. So we only have to describe the proposed model and identify which parts have been observed. This model and the prior distributions have been specified in the previous sections.

Separate BUGS programs were written for the first difference and the second difference model. The programs also calculate the survival functions and are available from the first author. The input data consist of information about the number of patients colonized, r_t, and the number of patients at risk, n_t, per day t. A preliminary examination of convergence is carried out before the estimated parameter of interest is selected. This could be done by using the *diag* command in BUGS. *diag* calculates a test statistic Z that compares the early and the later part

of a run and sees whether it differs substantially or not. This test is a rough approximate to that suggested by Geweke [25]. The calculation of the test statistic Z is described in detail by Spiegelhalter et al [23].

Table 2 shows the mean values for the hazard functions, their standard deviations, the 95% credible intervals and the Geweke Z scores, obtained using the first and second difference models. Note that the results obtained in this paper are written in four decimal places except for Z values. The standard deviations and 95% credible intervals are automatically generated by BUGS. The Z scores in the fifth and the ninth column of Table 2 are in the range of -2 and 2, indicating that the estimates of the hazard functions for the first and the second difference models are acceptable.

Figures 2 and 3 show the plots for the hazard functions against the time to colonization for the first difference prior and the second difference prior, respectively. We note that the hazard function can be estimated for time beyond the largest observation by generating the λ_ts for large t. Information for λ_t comes from λs with earlier values of t rather than nonexistent data. The 95% credible intervals are also included in the plots. We note that very large credible intervals are obtained as the time to colonization becomes large.

We plot the hazard functions calculated from the crude data in Figure 4, together with the hazard functions obtained from the Bayesian smoothing methods (the first and second difference prior). As noted earlier, the hazard functions obtained from crude data are not smooth. The plot produced by the second difference prior model is smoother than the plot produced by the first difference prior model. The values of the hazard function for the three models are similar especially when the time to colonization is less than and equal to 13 days. We suggest the hazard function increases with time for days less than and equal to 13. From the data, we found that only 4% of the total patients in this study stayed longer than 13 days, so the values of the

Table 2: Summary of the posterior inference of the hazard functions, their standard deviations, 95% credible intervals and the Z scores. The results were shown for both inferences obtained from the first and second difference prior models.

	First Difference				Second Difference			
Day t	$\hat{h}(t)$	SD	95 % Credible Interval	Z score	$\hat{h}(t)$	SD	95% Credible Interval	Z score
1	0.0013	0.0006	(0.0004, 0.0029)	-0.45	0.0016	0.0007	(0.0005, 0.0031)	1.38
2	0.0033	0.0010	(0.0016, 0.0055)	0.02	0.0039	0.0009	(0.0023, 0.0058)	1.18
3	0.0127	0.0029	(0.0077, 0.0192)	1.20	0.0095	0.0020	(0.0063, 0.0142)	-1.09
4	0.0208	0.0047	(0.0128, 0.0313)	0.34	0.0165	0.0035	(0.0110, 0.0247)	-1.29
5	0.0174	0.0047	(0.0095, 0.0279)	-0.34	0.0215	0.0037	(0.0150, 0.0296)	-0.52
6	0.0204	0.0058	(0.0107, 0.0334)	-0.89	0.0263	0.0046	(0.0179, 0.0358)	0.46
7	0.0303	0.0081	(0.0170, 0.0484)	-0.17	0.0322	0.0055	(0.0221, 0.0437)	0.21
8	0.0460	0.0117	(0.0264, 0.0722)	-0.34	0.0371	0.0067	(0.0253, 0.0517)	-0.24
9	0.0344	0.0102	(0.0174, 0.0573)	1.39	0.0376	0.0071	(0.0247, 0.0525)	0.37
10	0.0312	0.0102	(0.0148, 0.0544)	0.57	0.0375	0.0077	(0.0234, 0.0535)	0.48
11	0.0280	0.0101	(0.0119, 0.0506)	0.16	0.0392	0.0084	(0.0239, 0.0571)	0.23
12	0.0517	0.0162	(0.0262, 0.0884)	0.60	0.0438	0.0096	(0.0272, 0.0650)	-0.54
13	0.0504	0.0164	(0.0244, 0.0883)	-0.72	0.0461	0.0109	(0.0277, 0.0700)	-0.86
14	0.0482	0.0172	(0.0216, 0.0881)	-0.50	0.0456	0.0114	(0.0263, 0.0707)	-0.79
15	0.0418	0.0165	(0.0167, 0.0810)	-0.07	0.0427	0.0112	(0.0234, 0.0674)	-0.47
16	0.0331	0.0147	(0.0116, 0.0681)	-0.25	0.0390	0.0113	(0.0201, 0.0642)	-0.03
17	0.0368	0.0167	(0.0128, 0.0768)	-0.66	0.0362	0.0117	(0.0174, 0.0632)	0.23
18	0.0403	0.0197	(0.0131, 0.0891)	-0.19	0.0333	0.0120	(0.0147, 0.0616)	0.45
19	0.0266	0.0143	(0.0072, 0.0617)	-0.52	0.0296	0.0117	(0.0116, 0.0576)	0.65
20	0.0256	0.0145	(0.0065, 0.0615)	-0.26	0.0267	0.0116	(0.0096, 0.0530)	0.70
21	0.0232	0.0140	(0.0053, 0.0580)	0.04	0.0243	0.0112	(0.0080, 0.0504)	0.65
22	0.0190	0.0125	(0.0036, 0.0509)	-0.20	0.0226	0.0110	(0.0068, 0.0486)	0.55
23	0.0186	0.0127	(0.0325, 0.0511)	-0.25	0.0218	0.0111	(0.0059, 0.0485)	0.37
24	0.0215	0.0153	(0.0373, 0.0612)	-0.57	0.0219	0.0117	(0.0055, 0.0510)	0.13
25	0.0201	0.0151	(0.0032, 0.0583)	-0.88	0.0225	0.0127	(0.0054, 0.0543)	-0.11
26	0.0209	0.0159	(0.0033, 0.0617)	-0.84	0.0234	0.0140	(0.0049, 0.0578)	-0.35
27	0.0248	0.0191	(0.0038, 0.0751)	-0.96	0.0246	0.0158	(0.0047, 0.0622)	-0.51
28	0.0336	0.0271	(0.0050, 0.1055)	-0.86	0.0251	0.0175	(0.0041, 0.0672)	-0.54
29	0.0247	0.0210	(0.0031, 0.0796)	-0.89	0.0239	0.0181	(0.0032, 0.0683)	-0.50
30	0.0204	0.0189	(0.0020, 0.0717)	-0.90	0.0221	0.0184	(0.0020, 0.0687)	-0.47
31	0.0182	0.0188	(0.0013, 0.0694)	-0.91	0.0206	0.0191	(0.0011, 0.0692)	-0.39
32	0.0175	0.0196	(0.0008, 0.0712)	-1.12	0.0200	0.0214	(0.0004, 0.0747)	-0.22
33	0.0182	0.0230	(0.0006, 0.0813)	-1.39	0.0206	0.0259	(0.0001, 0.0872)	-0.04
34	0.0200	0.0283	(0.0004, 0.0973)	-1.37	0.0228	0.0337	(0.0000, 0.1112)	0.12
35	0.0226	0.0367	(0.0003, 0.1199)	-1.25	0.0277	0.0492	(0.0000, 0.1607)	0.27

t denotes time to colonization, $\hat{h}(t)$ is the estimated hazard function,

SD is the standard deviation of $\hat{h}(t)$, and Z score is the test statistic.

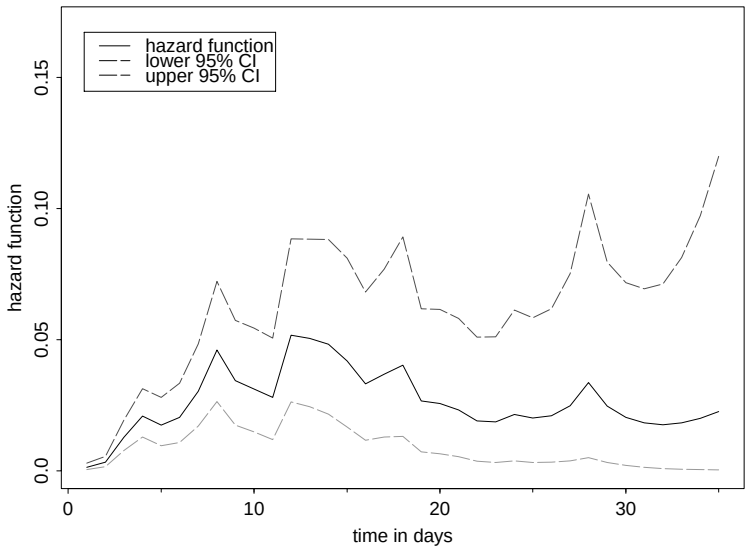

Figure 2: Estimates of colonization hazard rate for the MRSA data using the first difference prior model.

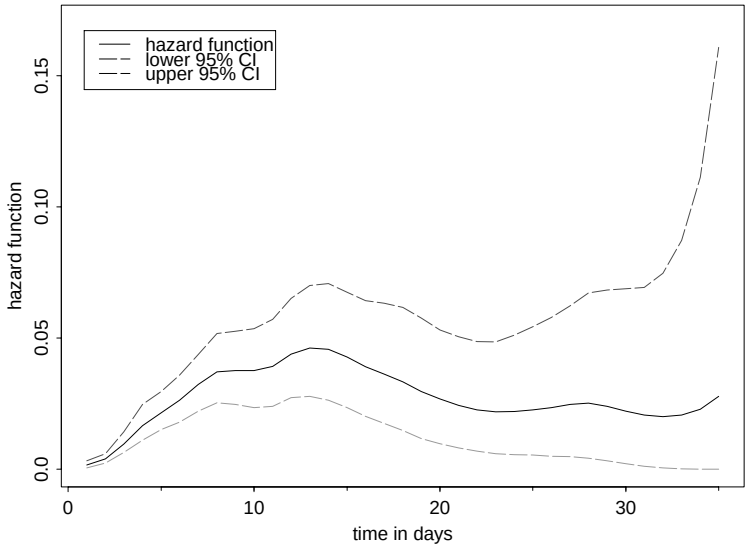

Figure 3: Estimates of colonization hazard rate for the MRSA data using the second difference prior model.

hazard function for days > 13 are not reliable due to the small number of patients.

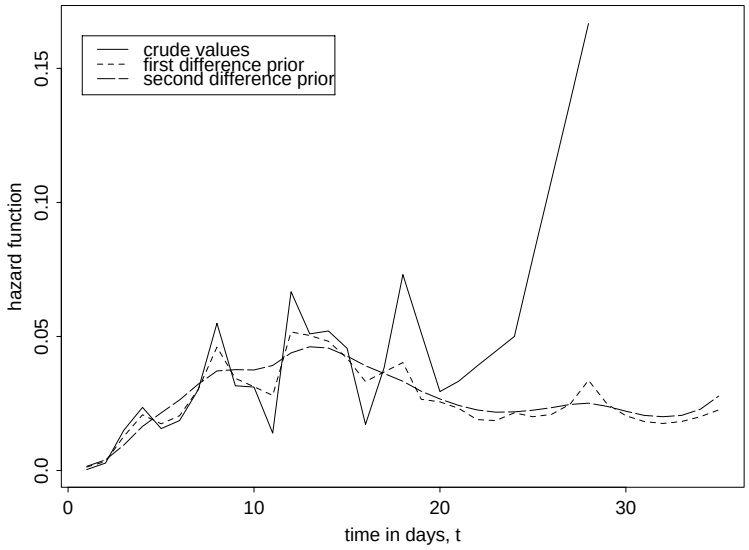

Figure 4: Estimates of colonization hazard rate for the MRSA data using crude data, the first difference prior model and the second difference prior model.

Using the relationship between the hazard and survival functions in equations (1) and (2), the posterior distribution of the survival functions can be calculated. These are shown in Table 3 with their standard errors, 95% credible intervals and the Z scores. We observe that the standard deviations increase with time in days until colonization, resulting in wider confidence intervals. However, the ranges are reasonable for both Bayesian smoothing models. The fifth and the ninth columns of Table 3 show the Z scores obtained from the first and second difference model, respectively. All the Z scores are within the acceptable range.

In the next section, we discuss how the results obtained from nonparametric autoregressive smoothing are used to predict MRSA colonization and to set up control limits of the control chart. We also limit the number of days stay to 13 since the analysis or estimates for days greater than 13 are not reliable and were not monotone increasing.

Table 3: Summary of the posterior inference of the survival functions, their standard deviations, 95% credible intervals and the Z scores. The results were shown for both inferences obtained from the first and second difference prior models.

	First Difference				Second Difference			
Day t	$\hat{S}(t+1)$	SD	95 % Credible Interval	Z score	$\hat{S}(t+1)$	SD	95% Credible Interval	Z score
1	0.9986	0.0006	(0.9971, 0.9996)	0.45	0.9984	0.0007	(0.9968, 0.9995)	-1.38
2	0.9953	0.0014	(0.9923, 0.9976)	0.21	0.9944	0.0015	(0.9913, 0.9970)	-1.31
3	0.9827	0.0033	(0.9755, 0.9884)	-1.32	0.9850	0.0026	(0.9794, 0.9897)	-0.21
4	0.9622	0.0060	(0.9496, 0.9728)	-0.83	0.9687	0.0052	(0.9572, 0.9778)	1.01
5	0.9454	0.0078	(0.9289, 0.9596)	-0.51	0.9479	0.0076	(0.9316, 0.9617)	0.89
6	0.9261	0.0100	(0.9051, 0.9443)	0.29	0.9230	0.0101	(0.9019, 0.9417)	0.48
7	0.8980	0.0130	(0.8709, 0.9217)	0.28	0.8933	0.0129	(0.8668, 0.9177)	0.31
8	0.8567	0.0172	(0.8209, 0.8884)	0.44	0.8602	0.0159	(0.8281, 0.8905)	0.33
9	0.8273	0.0198	(0.7863, 0.8640)	-0.34	0.8279	0.0186	(0.7902, 0.8637)	0.12
10	0.8015	0.0220	(0.7561, 0.8423)	-0.50	0.7969	0.0212	(0.7547, 0.8379)	-0.12
11	0.7790	0.0238	(0.7300, 0.8229)	-0.45	0.7657	0.0238	(0.7184, 0.8109)	-0.16
12	0.7388	0.0273	(0.6831, 0.7892)	-0.62	0.7323	0.0263	(0.6794, 0.7821)	0.00
13	0.7017	0.0303	(0.6390, 0.7582)	-0.27	0.6986	0.0289	(0.6398, 0.7533)	0.22
14	0.6680	0.0329	(0.6006, 0.7300)	-0.07	0.6669	0.0315	(0.6023, 0.7260)	0.34
15	0.6401	0.0351	(0.5685, 0.7059)	-0.03	0.6385	0.0337	(0.5696, 0.7020)	0.38
16	0.6190	0.0366	(0.5444, 0.6875)	0.03	0.6138	0.0357	(0.5416, 0.6809)	0.35
17	0.5963	0.0385	(0.5184, 0.6684)	0.15	0.5917	0.0377	(0.5161, 0.6630)	0.26
18	0.5725	0.0409	(0.4896, 0.6493)	0.17	0.5722	0.0397	(0.4931, 0.6485)	0.13
19	0.5573	0.0423	(0.4716, 0.6372)	0.24	0.5554	0.0417	(0.4721, 0.6368)	-0.02
20	0.5432	0.0435	(0.4551, 0.6254)	0.26	0.5408	0.0437	(0.4538, 0.6262)	-0.14
21	0.5307	0.0448	(0.4405, 0.6156)	0.23	0.5279	0.0456	(0.4380, 0.6170)	-0.22
22	0.5208	0.0458	(0.4288, 0.6073)	0.24	0.5162	0.0473	(0.4232, 0.6083)	-0.28
23	0.5112	0.0468	(0.4170, 0.6000)	0.26	0.5051	0.0490	(0.4093, 0.5999)	-0.30
24	0.5004	0.0481	(0.4036, 0.5920)	0.35	0.4943	0.0505	(0.3954, 0.5917)	-0.28
25	0.4905	0.0493	(0.3912, 0.5846)	0.49	0.4834	0.0520	(0.3813, 0.5839)	-0.24
26	0.4805	0.0508	(0.3784, 0.5771)	0.62	0.4724	0.0536	(0.3671, 0.5759)	-0.18
27	0.4688	0.0527	(0.3633, 0.5673)	0.76	0.4611	0.0554	(0.3522, 0.5681)	-0.11
28	0.4535	0.0556	(0.3398, 0.5562)	0.84	0.4498	0.0572	(0.3373, 0.5613)	-0.04
29	0.4427	0.0579	(0.3242, 0.5488)	0.90	0.4395	0.0592	(0.3223, 0.5551)	0.01
30	0.4340	0.0598	(0.3111, 0.5438)	0.94	0.4302	0.0613	(0.3097, 0.5497)	0.04
31	0.4265	0.0618	(0.2986, 0.5399)	0.98	0.4218	0.0633	(0.2975, 0.5439)	0.07
32	0.4195	0.0638	(0.2879, 0.5364)	1.04	0.4139	0.0656	(0.2828, 0.5389)	0.08
33	0.4124	0.0664	(0.2724, 0.5329)	1.12	0.4060	0.0686	(0.2643, 0.5349)	0.08
34	0.4049	0.0697	(0.2556, 0.5293)	1.19	0.3977	0.0726	(0.2407, 0.5311)	0.07
35	0.3968	0.0741	(0.2330, 0.5264)	1.24	0.3883	0.0785	(0.2127, 0.5289)	0.03

t denotes time to colonization, $\hat{S}(t+1)$ is the survival function,

SD is the standard deviation of $\hat{S}(t)$, and Z score is the test statistic.

4 Prediction

The analysis in the previous sections allow us to make prediction on the number of colonizations which might occur in the next few days. We can then set up control charts, say, for the next s days, given information about the ICU length of stay of individual patients. The probability that a patient is free from colonization in the next few days given that he or she is free from it now can be calculated using the estimates of the hazard and survival functions given in Section 3.2.

In the PA hospital, the ICU has a total of twelve beds and hence it can only accommodate a maximum twelve patients per day. However, the recorded number of total patients can be more than 12 per day because of the possibility of one or more patients leaving and others entering the ICU on the same day. Let us say that at the present time there are N patients and the corresponding lengths of stay in the ICU up to the present time are denoted by $t_1, \ldots, t_N$. The length of stay for patient j, t_j can take values greater or equal to one. At this stage, the colonization status for each patient is unknown, but it could take values of either one or zero. In short, at the end of day s, colonization can be written as $Y_1, \ldots, Y_N$ where

$$\begin{aligned} Y_j &= 0 \quad \text{if patient } j \text{ not colonized with MRSA} \\ &= 1 \quad \text{if patient } j \text{ is colonized.} \end{aligned}$$

s can take values greater than 0. Assume that the Y_js are independently Bernoulli distributed with parameter p_j. In other words, we are assuming that colonization of a patient is independent of the other, although it might not be true in real life. However, in the study of nosocomial infection surveillance, the same assumption was used [26, 27, 28].

Let p_j equal $p_s(t_j)$, the probability of a patient colonized in the next s days given he or she is not colonized now. Since we have a Bernoulli variable, the expected value for a patient colonized in the next 7 days is equal to p_j, that is $E(Y_j) = p_j$. It follows that if $S = Y_1 + Y_2 + \ldots + Y_N$

denotes the number of colonized patients in the next s days, and $\hat{N}$ denotes the expected number of colonized patients in the next s days, then $\hat{N}$ is given by $\sum_{j=1}^{N} p_j$.

Using the information obtained in Section 3.2, we calculate the probability, denoted by $p_s(t_j)$, of a patient being colonized with MRSA in the next s days given that he or she is not colonized before day t for each patient.

$$\begin{aligned} p_s(t) &= P(\text{colonized in the next } s \text{ days given not colonized before } t) \\ &= \sum_{j=0}^{s-1} P(\text{colonized on day } t+j | \text{not colonized before } t) \\ &= \sum_{j=0}^{s-1} \frac{P(\text{colonized on day } t+j \text{ and not colonized before } t)}{\text{P(not colonized before } t)} \\ &= \sum_{j=0}^{s-1} \frac{P(\text{colonized on day } t+j)}{\text{P(not colonized before } t)} \\ &= \frac{1}{S(t)} \sum_{j=0}^{s-1} p(t+j) \end{aligned} \tag{5}$$

where $p(t)$ can be obtained by using the following equation $p(t) = S(t) - S(t+1)$ for $t = 1, 2, \ldots$. Table 4 gives the estimates of $p_7(t_j)$, the chances of a patient being colonized in the next 7 days given the patient's length of stay in the ICU. These estimates are calculated using equation (5) with $s = 7$ and BUGS. Besides giving estimates of these probabilities, BUGS automatically calculates the corresponding standard deviations and 95% credible intervals for $p_7(t)$. Table 4 also shows that the standard deviations increase with time, and as a result, wider confidence intervals are produced as the length of stay in the ICU increases. The Z scores are within the acceptable range.

Figure 5 and Figure 6 give the estimates of $p_7(t)$ together with the 95% credible intervals obtained using the first difference prior model and the second difference prior model, respectively. In both figures, we can see that the credible intervals are larger for longer ICU length of stay.

Figure 7 give plots of the estimated $p_7(t)$ for both models. The estimates obtained using the

Table 4: Probability of colonization in the next 7 days given the length of stay in the ICU, standard deviations, 95% credible intervals and the Z scores. The estimates are obtained using the first and second difference prior models.

	First Difference				Second Difference			
Day t	$p_7(t)$	SD	95 % Credible Interval	Z score	$p_7(t)$	SD	95% Credible Interval	Z score
1	0.1019	0.0130	(0.0782, 0.1290)	-0.28	0.1066	0.0129	(0.0822, 0.1331)	-0.31
2	0.1420	0.0173	(0.1103, 0.1781)	-0.41	0.1383	0.0160	(0.1079, 0.1705)	-0.42
3	0.1688	0.0199	(0.1319, 0.2098)	0.36	0.1674	0.0187	(0.1313, 0.2055)	-0.25
4	0.1843	0.0221	(0.1434, 0.2300)	0.40	0.1909	0.0212	(0.1499, 0.2334)	0.10
5	0.1903	0.0241	(0.1461, 0.2398)	0.32	0.2095	0.0241	(0.1635, 0.2574)	0.29
6	0.2185	0.0278	(0.1667, 0.2759)	0.57	0.2274	0.0269	(0.1766, 0.2820)	0.11
7	0.2422	0.0314	(0.1839, 0.3073)	0.34	0.2431	0.0299	(0.1867, 0.3046)	-0.14
8	0.2561	0.0345	(0.1920, 0.3275)	0.13	0.2535	0.0331	(0.1908, 0.3216)	-0.29
9	0.2528	0.0372	(0.1844, 0.3301)	0.14	0.2576	0.0362	(0.1900, 0.3308)	-0.32
10	0.2518	0.0395	(0.1796, 0.3337)	-0.14	0.2586	0.0391	(0.1852, 0.3371)	-0.34
11	0.2560	0.0424	(0.1787, 0.3445)	-0.33	0.2574	0.0420	(0.1777, 0.3419)	-0.33
12	0.2652	0.0464	(0.1811, 0.3627)	-0.35	0.2528	0.0449	(0.1685, 0.3438)	-0.22
13	0.2457	0.0484	(0.1587, 0.3466)	-0.55	0.2415	0.0476	(0.1538, 0.3398)	0.03

$p_7(t)$ is the expected number of colonization in the next 7 days given t as the length of stay in the ICU, SD is the standard deviations of $p_7(t)$ and Z score is a test statistic.

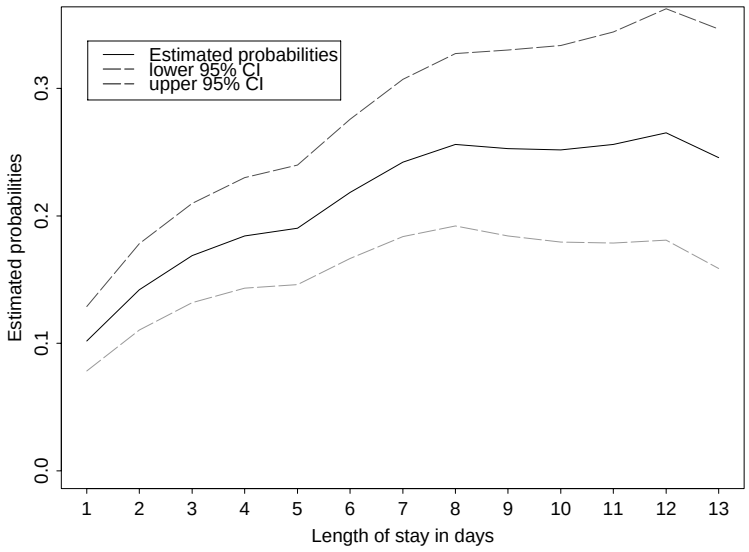

Figure 5: Probability of colonizations in the next 7 days given length of stay obtained using the first difference prior model.

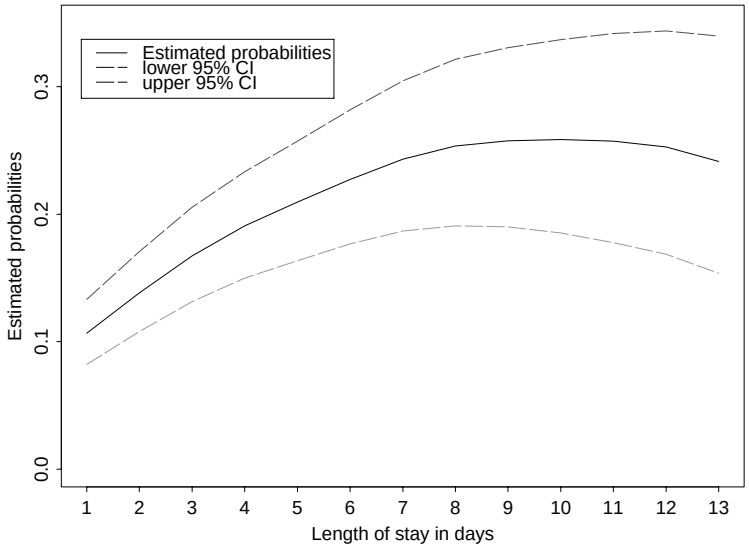

Figure 6: Probability of colonizations in the next 7 days given length of stay obtained using the second difference prior model.

second difference model give a smoother plot than the first difference model. For both models, the probability, $p_7(t)$ of colonization in the next 7 days increases with the length of patient stay in the ICU. For example, in the second difference model, a patient who stays in the ICU for 3 days has a 17% chance of being colonized in the next 7 days as compared to 25% for a patient who stays for 8 days. The estimates decrease at $t = 12$ due to the unreliable estimates of the hazard functions for day $t > 13$ for both plots. Note that to estimate the number of colonizations in the next 7 days given the patient's length of stay is 13 days, we need to use information up to day $t + s - 1 = 13 + 7 - 1 = 19$.

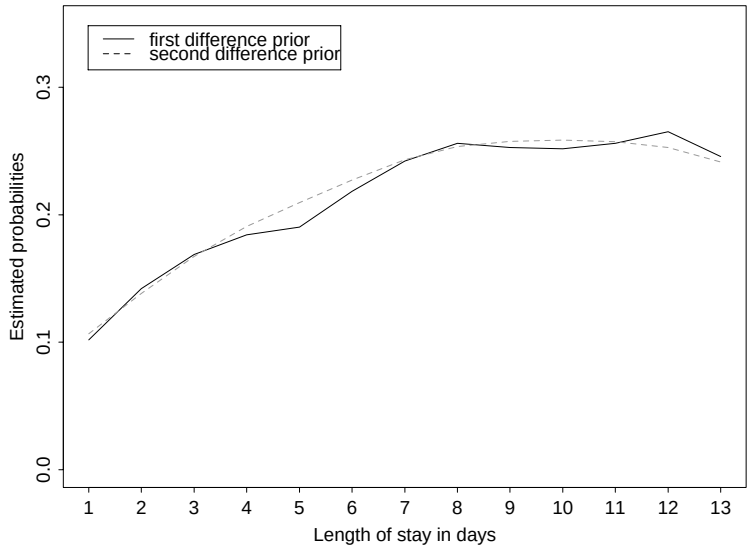

Figure 7: Probability of colonization in the next 7 days given length of stay obtained using the first difference prior model and the second difference prior model.

The estimates of colonizations for the next 7 days might not be useful since the average length of patient stay in the ICU is about 4 days with median equal to 2. More useful estimates would be estimates of $p_s(t_j)$ where $s \leq 7$. To calculate $p_s(t_j)$, the same procedure is employed as for calculating $p_7(t)$ except that now the Gibbs samplers are run for different values of s. We

Table 5: Probability of colonization in the next s days given length of stay in the ICU where $s = 1, \ldots, 7$. Estimates are obtained using the second difference prior.

Length of stay	Probability of colonization in the next s days						
in days, t	1	2	3	4	5	6	7
1	0.0016	0.0055	0.0150	0.0313	0.0521	0.0770	0.1066
2	0.0039	0.0134	0.0297	0.0506	0.0755	0.1052	0.1383
3	0.0095	0.0259	0.0468	0.0719	0.1016	0.1349	0.1674
4	0.0165	0.0377	0.0630	0.0931	0.1266	0.1594	0.1909
5	0.0215	0.0472	0.0778	0.1120	0.1453	0.1773	0.2095
6	0.0263	0.0576	0.0925	0.1265	0.1593	0.1921	0.2274
7	0.0322	0.0680	0.1030	0.1366	0.1703	0.2066	0.2431
8	0.0371	0.0733	0.1080	0.1428	0.1803	0.2180	0.2535
9	0.0376	0.0737	0.1098	0.1488	0.1879	0.2248	0.2576
10	0.0375	0.0752	0.1156	0.1563	0.1946	0.2287	0.2586
11	0.0392	0.0812	0.1235	0.1632	0.1987	0.2298	0.2574
12	0.0438	0.0879	0.1292	0.1662	0.1985	0.2273	0.2528
13	0.0461	0.0895	0.1281	0.1619	0.1920	0.2187	0.2415

only do this analysis using the second difference prior model. Table 5 shows the estimated values of $p_s(t_j)$ for $s = 1, \ldots, 7$ and the plots of $p_s(t_j)$ against s are given in Figure 8 for the length of stay t equal to $1, \ldots, 5$. For each t, the values of $p_s(t_j)$ increases with s. This result supports the statement that the risk of MRSA increases with the number of days stay in the ICU. Note that the values in the first column (where $s = 1$) are equivalent to the hazard functions in Table 2.

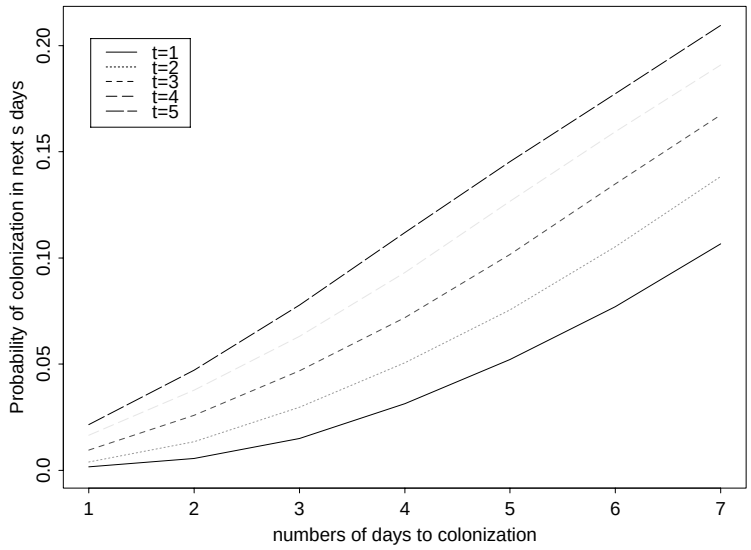

Figure 8: Probability of colonization in the next s days given length of stay is t days and there is no colonization before day t. Results given for $t = 1, \ldots, 5$ days, obtained using the second difference prior model.

Once $p_s(t_j)$ is calculated, we can then calculate $\hat{N} = \sum_{t=1}^{N} p_s(t)$, the expected number of patients to be colonized in the next s days. Then $\hat{N}$ can be used as a control limit for the control chart. We need to determine which distribution is suitable for our problem. Since S is defined as $\sum_{j=1}^{N} Y_j$ and the Y_j are independent Bernoulli variables, naturally the sum of these independent variables has a binomial distribution with the same parameter p_j, as long as p_js are all the same. On the other hand, in our situation, not all the patients enter the ICU at the same

time, resulting in different values of the t_js and thus the p_js vary from one patient to another where $p_j = p_s(t_j)$. This situation is known as Bernoulli trials with variable probabilities.

On the other hand, if all the p_js are equal and if the Y_js are independent, then S is a random variable whose distribution can be approximated by the Poisson distribution. Feller [29] showed that this conclusion remains valid under the assumption that the probabilities p_j are not equal for N large and for moderate values of S. Another approximation can be made by using binomial$(N, \bar{p})$ where $\bar{p} = \dfrac{\sum_{j=1}^{N} p_j}{N}$ is the average value of $p_s(t_j)$.

Let us denote p_{bk}, p_{pk} and $p_{b_i k}$ be the probabilities of $S = k$ obtained using Bernoulli trials with variable probabilities, Poisson approximation and binomial approximation with $\bar{p}$, respectively. These probabililies are calculated using probability generating functions, $E(Z^s)$ which is defined by $\prod_{j=1}^{N}(1 - p_j + p_j z)$. To see how large the difference is between p_{bk} and p_{pk} and between p_{bk} and $p_{b_i k}$, we look at the maximum absolute errors for the two probabilities which can be calculated as $|p_{pk} - p_{bk}|$ and $|p_{b_i k} - p_{bk}|$, respectively.

As illustrations, we examine 4 situations, each involving 12 patients with different lengths of stays. Estimations of $p_s(t_j)$ from the second difference prior are used and our aims are to find the expected number of patients colonized with MRSA in the next 7 days and also to find the maximum absolute error of the two approximations, $|p_{pk} - p_{bk}|$ and $|p_{b_i k} - p_{bk}|$, for the situations described. If $|p_{pk} - p_{bk}|$ is smaller than $|p_{b_i k} - p_{bk}|$, then the Poisson approximation is the better one, otherwise the binomial approximation is the more acceptable one.

In the first situation, out of 12 patients, 6 had already stayed in the ICU for one day and the rest already stayed for 9 days without being colonized. In the second, the lengths of stays are 1, 2, 4 and 9 and each length of stay has three patients. Then in the third situation, the lengths of stays are 1 and 2 days, each having only one patient, two patients stayed for 3 days, 4 stayed for 4 days, 2 stayed for 5 days and another 2 stayed for 6 days. Lastly, all the 12 patients in the ICU

have stayed for 3 days in the fourth situation. The expected number of colonized patients at the end of 7 days are 2.19, 2.08, 2.22 and 2.01 for situation 1, 2, 3 and 4, respectively. From the maximum absolute errors calculated, we find that binomial(12,$\bar{p}$) is a better approximation than Poisson ($\sum_{j=1}^{N} p_j$). If all the patients have the same probability (situation 4), then the probability calculated using binomial(12,$\bar{p}$) is the same as the exact probability.

5 Conclusion

In this paper, we developed a Bayesian smooth estimate of the hazard function in Section 3, which is substantially more appealing than the simple estimate. The Bayesian smooth estimates also automatically overcome the missing day problem. We also found that we obtained an increasing hazard function up to day 13 (that is, a hazard function which is not constant and not decreasing), which agrees with the studies by Boyce [6], Morita [10] and Layton et al [11] that found that the risk of colonization increases with the length of stay in the ICU.

We then used these estimates, the probability of colonization in the next s days given no colonization before day t. We noted the strong dependence on length of stay, t. The prediction rule we have developed is reliable only for patients with short ICU length of stay. However, this is considered sufficient because only a small percentage of the ICU patients are not covered by this rule. The estimated number of colonizations can be used as a control limit in a control chart and we show that the binomial distribution can be used to discover the probability of departures in the number of colonizations from the state of statistical control.

We also noted at the beginning of this paper that the times until colonization were in fact censored in the sense that the true value was less than the recorded value. Consequently, our conculsion must be interpeted cautiously in the sense that our times to colonization are larger than the true value and consequently hazard estimates may be larger systematically than our

estimates. Certainly, the resulting survival function is a conservative estimate of that which would result if the true times to colonization were available.

In general, there will be a time lapse between the collection of the reference database and application. By using a hospital's own database to calculate the probability of the outcome, we can overcome some of the time lapse between the reference database and the new data. It also would permit a more responsive and flexible approach to data management with the impact of any local changes in areas such as medical practices or socioeconomic status being assessed rapidly.

Invasive medical devices were found to contribute substantially to the increased risk of infection in the ICU patients [30]. Since there can be no infection without colonization, a further analysis should distinguish between ventilated and non-ventilated patients, as we expect the hazard to be greater for a ventilated patient.

ACKNOWLEDGEMENT

The authors wish to express their appreciation to all the staff at the Infectious Disease Department and the Intensive Care Unit, Princess Alexandra Hospital, Brisbane, Australia for providing the data and for their interest and willing cooperation. Also they wish to express their appreciation to Dr. A. P. Morton and M. Whitby for their help and support.

References

[1] Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, Sirio CA, Murphy DJ, Lotring T, Damiano A, and Harrell FE. The apache ii prognostic system: Risk

prediction of hospital mortality for critically ill hospitalized adults. *Chest*, 100(6):1619–1636, 1991.

[2] APACHE MEDICAL SYSTEMS, INC. *APACHE III Management System: Training Manual*, March 1994. Version 1.3.

[3] APACHE MEDICAL SYSTEMS, INC. *A Guide to APACHE Reports*, March 1994.

[4] Noor Azina Ismail. *Statistical Methods for the Improvement of Health Care*. PhD thesis, School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia, 2000.

[5] Bradley JS and Connor JD. Ceftriaxone failure in meningitis caused by streptococcus pneumoniae with reduced susceptibility to β-lactam antibiotics. *Pediatr. Infect. Dis.*, 10:871–873, 1991.

[6] Boyce JM. Methicillin-resistant staphylococcus aureus in hospitals and long-term care facilities: Microbiology, epidemiology and preventive measures. *Infection Control and Hospital Epidemiology*, 13(12):725–737, 1992.

[7] Jernigan JA, Clemence MA, Stott GA, Titus MG, Alexander CH, Palumbo CM, and Farr BM. Control of methicillin-resistant *staphylococcus aureus* at a university hospital: One decade later. *Infection Control and Hospital Epidemiology*, 16(12):686–696, 1995.

[8] Crowcroft N, Maguire H, Fleming M, Peacock J, and Thomas J. Methicillin-resistant *staphylococcus aureus*: Investigation of a hospital outbreak using a case-control study. *Journal of Hospital Infection*, 34:301–309, 1996.

[9] Mulligan ME, Murray-Leisure KA, Ribner BS, Standiford HS, John JF, Korvick JA, Kauffman CA, and Yu VL. Methicillin-resistant *staphylococcus aureus*: A concensus review of

the microbiology, phatogenesis and epidemiology with implementation for prevention and management. *The American Journal of Medicine*, 94(3):313–326, 1993.

[10] Morita MM. Methicillin-resistant *staphylococcus aureus*. *Nursing Clinics of North America*, 28(3):625–637, 1993.

[11] Layton MC, Hierholzer WJ, and Patterson JE. The evolving epidemiology of methicillin-resistant *staphylococcus aureus* at a university hospital. *Infection Control and Hospital Epidemiology*, 16(1):12–17, 1995.

[12] Breslow NE and Clayton DG. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88(421):9–25, 1993.

[13] Spiegelhalter D, Gilks ATW, Best N, and Gilks W. *An Introduction to Bayesian Computation Using BUGS: BUGS Examples*. MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 2SR, November 1995. Version 0.5, Volume 2.

[14] Kalbfleisch JD and Prentice RL. *The Statistical Analysis of Failure Time Data*. Wiley, New York, 1980.

[15] Clayton DG. *Generalized Linear Mixed Models*. Chapman and Hall, 1996. In Markov Chain Monte Carlo in Practice.

[16] Best NG, Spiegelhalter DJ, Thomas A, and Brayne CEG. Bayesian analysis of realistically complex models. *Journal of the Royal Statistical Society A*, 159(Part 2):323–342, 1996.

[17] Jiang J. Consistent estimators in generalized linear mixed models. *Journal of the American Statistical Association*, 93(442):720–729, 1998.

[18] Henderson CR, Kempthorne O, Searle SR, and VonKrosigk CN. Estimation of environmental and genetic trends from records subject to culling. *Biometrics*, 30:583–588, 1959.

[19] Lee Y and Nelder JA. Hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society B*, 58:619–678, 1996.

[20] Harrison PJ and Stevens CF. Bayesian forecasting (with discussion). *Journal of the Royal Statistical Society B*, 38:205–247, 1976.

[21] West M and Harrison J. *Bayesian Forecasting and Dynamic Models*. Springer-Verlag, New York, 1989. Springer Series in Statistic.

[22] Whittaker J. *Graphical Models in Applied Multivariate Statistics*. Wiley, New York, 1990. Wiley Series in Probability and Mathematical Statistics.

[23] Spiegelhalter D, Thomas A, Best N, and Gilks W. *BUGs 0.5: Bayesian Inference Using Gibbs Sampling Manual (Version ii)*. MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 2SR, May 1996.

[24] Spiegelhalter DJ, Best NG, Gilks WR, and Inskip H. *Hepatitis B: A Case Study in MCMC methods*. Chapman and Hall, London, 1996. In Markov Chain Monte Carlo in Practice.

[25] Geweke J. *Evaluating the Accuracy of Sampling Based Approaches to Calculating Posterior Moments*. Claredon Press, Oxford, UK, 1992. In Bayesian Statistic 4.

[26] Birnbaum D. Analysis of hospital infection surveillance data. *Infection Control*, 5:332–338, 1984.

[27] Birnbaum D. Nosocomial infection surveillance programs. *Infection Control*, 8:474–479, 1987.

[28] Morrison AJ, Kaiser DL, and Wenzel RP. A measurement of the efficacy of nosocomial infection control using the 95 per cent confidence interval fpr infection rates. *American Journal of Epidemiology*, 126(2):292–297, 1987.

[29] Feller W. *An Introduction to Probability Theory and Its Applications.* Wiley, third edition, 1968. Volume I, New York.

[30] Flaherty JP and Weinstein RA. Nosocomial infection caused by antibiotic-resistant organisms in the intensive-care unit. *Infection Control and Hospital Epidemiology*, 17(4):236–248, 1996.