Epidemiol. Infect. (2008), **136**, 1599–1605. © 2008 Cambridge University Press
doi:10.1017/S0950268808000526 Printed in the United Kingdom

A statistical model investigating the prevalence of tuberculosis in New York City using counting processes with two change-points

J. A. ACHCAR[1], E. Z. MARTINEZ[1]*, A. RUFFINO-NETTO[1], C. D. PAULINO[2] AND P. SOARES[2]

[1] *Departamento de Medicina Social, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Brazil*
[2] *Departamento de Matemática, Universidade Técnica de Lisboa, Portugal*

(*Accepted 12 February 2008; first published online 17 March 2008*)

SUMMARY

We considered a Bayesian analysis for the prevalence of tuberculosis cases in New York City from 1970 to 2000. This counting dataset presented two change-points during this period. We modelled this counting dataset considering non-homogeneous Poisson processes in the presence of the two-change points. A Bayesian analysis for the data is considered using Markov chain Monte Carlo methods. Simulated Gibbs samples for the parameters of interest were obtained using WinBugs software.

INTRODUCTION

General overview

In 1993 the World Health Organization (WHO) declared tuberculosis (TB) a global public health emergency, being the only disease thus far to warrant that designation. Although hospitals have been established and chemotherapy has been developed to combat TB, bringing considerable reduction in incidence to developed nations, historical data calculated by the WHO indicate that there have not been great effects on the global problem since the time of Koch. Currently, TB is responsible for more human deaths than any other single infectious agent, representing 26% of all preventable deaths and 7% of all deaths [1].

TB resumption has been attributed to several factors, such as the increase in drug resistance, the HIV/AIDS pandemic (at the beginning of the 1980s), the increase of injecting drug users, changes in social structure, the increase of immigrants from high prevalence nations to developed ones, the ageing of the world's population, the active transmission in environments of human accumulation (e.g. prisons, hospitals, homeless shelters), and the dismantling of health-care systems [2]. Although TB became a re-emerging disease in European and North-American nations, TB is not an emergent nor re-emerging public health problem in developing countries such as Brazil, but rather a long lasting one [3].

In order to facilitate the comprehension of the various components involved in the interaction between these factors, Ruffino-Netto [4] proposed an expression which reflects the TB burden, represented by the components social inequality, prevalence of HIV-positive individuals, percentual default of treatment, prevalence of primary resistance plus acquired resistance, migration, age of the population, adequate health services, directly observed treatment short-course, educational level, nutrition level, human resources for TB control and degree of political participation of the population.

* Author for correspondence: Dr E. Z. Martinez, Departamento de Medicina Social, Faculdade de Medicina de Ribeirão Preto, USP, Avenida Bandeirantes 3900, 14049-900 Ribeirão Preto, SP, Brazil.
(Email: edson@fmrp.usp.br)

From the historical data of the observed numbers of TB cases, especially in developed countries, we observe a trend of declining incidence starting at the beginning of 1960 up to 1980, where there was a change in this trend. During the period between 1980 and 1990, we observe an increase in the incidence rates of TB cases; after 1990, we again observe a trend of declining incidence. That is, we have the presence of two change-points in the rates of TB, especially for developed countries.

The case of New York City

The incidence (notification cases) of TB disease in New York City (NYC) between 1970 and 2000 presents three trends (see Table 1): a first period (1970–1979) where the trend of declining incidence was probably associated with good control programmes; a second period (1979–1992) where there was an increase in incidence rates [5, 6], possibly associated with a systematic dismantling of public-health infrastructure of control programmes, social disruption (including homelessness, drug abuse, poverty and housing overcrowding), and mainly caused by the HIV epidemic; a third period (1992–2000) where again there is a decline in incidence rates. It is important to remember the many factors associated with this third period, i.e. implementation of directly observed therapy, broader chemotherapy regimens for patients with TB or suspected multidrug-resistant TB and improved therapeutics for the care of HIV-infected individuals [7].

Table 1 gives the yearly numbers and the accumulated numbers of TB cases in NYC. From Table 1, we see decreasing numbers of TB cases from 1970 to 1978, where there is a minimum. From 1978 to 1992, we observe increasing numbers of TB cases, where there is a maximum number of cases in 1992. From 1992 to 2000, we observe decreasing numbers of cases. That is, we have two change-points for the numbers of cases (see Fig. 1). It is interesting to note that the use of powerful antiviral drugs against HIV commenced around 1990.

To model the number of TB cases in NYC during the period 1970–2000, we consider the use of a point process to count the numbers of TB cases in each year starting in 1970. In this way, we considered a stratified sample of size $n=6721$ representing 10% of the total number of TB cases (259 cases in 1970, 257 cases in 1971, 227 cases in 1972 and so on), where for each year, we used an uniform distribution to have the times (in days) for the occurrence of each case since 1 January 1970 until 30 December 2000, i.e. with a total time of observation equal to $T=11\,323$ days.

Table 1. *Number of tuberculosis cases in NYC from 1970 to 2000*

Year	Year – 1970	Cases/year	Accumulated number of cases
1970	0	2590	2590
1971	1	2572	5162
1972	2	2275	7437
1973	3	2101	9538
1974	4	2022	11 560
1975	5	2151	13 711
1976	6	2151	15 862
1977	7	1605	17 467
1978	8	1307	18 774
1979	9	1566	20 340
1980	10	1478	21 818
1981	11	1582	23 400
1982	12	1597	24 997
1983	13	1648	26 645
1984	14	1629	28 274
1985	15	1843	30 117
1986	16	2223	32 340
1987	17	2197	34 537
1988	18	2317	36 854
1989	19	2545	39 399
1990	20	3520	42 919
1991	21	3673	46 592
1992	22	3811	50 403
1993	23	3235	53 638
1994	24	2999	56 637
1995	25	2445	59 082
1996	26	2053	61 135
1997	27	1730	62 865
1998	28	1558	64 423
1999	29	1460	65 883
2000	30	1332	67 215

Source: New York City Department of Health and Mental Hygiene, Bureau of Tuberculosis Control Information Survey.

For this dataset, we assume a non-homogeneous Poisson process (NHPP) in the presence of two change-points considering a Bayesian approach using Markov chain Monte Carlo (MCMC) methods (see e.g. Gelfand & Smith [8]). The use of Bayesian methods has been considered by many authors for analyses of homogeneous or non-homogeneous Poisson processes in the presence of change-points (see e.g. Raftery & Akman [9] considering the presence of a change-point in homogeneous Poisson processes, or

https://doi.org/10.1017/S0950268808000526 Published online by Cambridge University Press

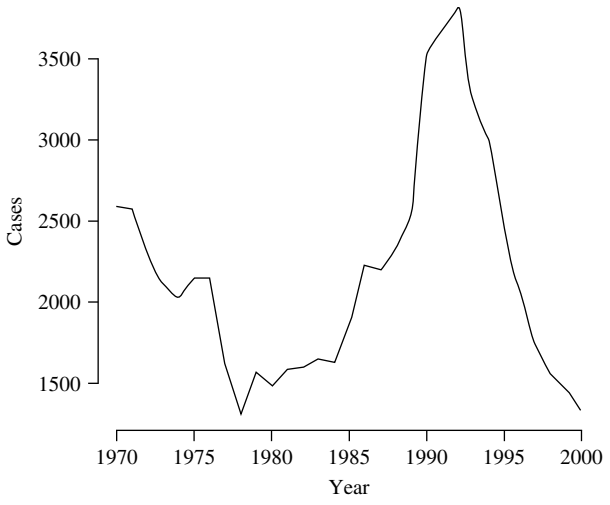

Fig. 1. Number of tuberculosis cases in New York City, 1970–2000.

Ruggeri & Sivaganesan [10] considering any number, random or fixed, of change-points in NHPP assuming power-law intensity functions).

The paper is organized as follows: in the Methods section, we introduce the likelihood function and a Bayesian analysis for the model; in the Results section, we introduce the analysis for the NYC data, and finally, in the Discussion, we present some concluding remarks.

METHODS

The likelihood function

Let $N(t)$ be the cumulative number of TB cases that are observed during the interval $(0, t)$ and assume that $N(t)$ is modelled by a NHPP with intensity function $\lambda(t) = \mathrm{d}m(t)/\mathrm{d}t = \mathrm{d}E[N(t)]/\mathrm{d}t$, where $m(t)$ is the mean value function (see e.g. Cox & Lewis [11]). Different parametrical forms could be assumed for the intensity function $\lambda(t)$ (increasing, decreasing, bathtub shape, unimodal, among many others, see e.g. Musa & Okumoto [12] or Muldholkar *et al.* [13]). We assume power-law processes (PLP) in the presence of two change-points with intensity function for the overall process given by

$$\lambda(t|\boldsymbol{\theta}) = \begin{cases} \lambda_1 = \dfrac{\beta_1}{\alpha_1}\left(\dfrac{t}{\alpha_1}\right)^{\beta_1 - 1} & \text{if } 0 < t < \zeta_1 \\ \lambda_2 = \dfrac{\beta_2}{\alpha_2}\left(\dfrac{t}{\alpha_2}\right)^{\beta_2 - 1} & \text{if } \zeta_1 \leqslant t < \zeta_2 \\ \lambda_3 = \dfrac{\beta_3}{\alpha_3}\left(\dfrac{t}{\alpha_3}\right)^{\beta_3 - 1} & \text{if } t \geqslant \zeta_2 \end{cases} \tag{1}$$

where $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3, \zeta_1, \zeta_2)$.

Equivalently, letting $m_j(t) = m(t|\boldsymbol{\theta}j)$, the corresponding mean value function is given by

$$m(t|\boldsymbol{\theta}) = \begin{cases} m_1(t) & \text{if } 0 < t < \zeta_1 \\ m_2(t) - m_2(\zeta_1) + m_1(\zeta_1) & \text{if } \zeta_1 \leqslant t < \zeta_2 \\ m_3(t) - m_3(\zeta_2) + m_2(\zeta_2) - m_2(\zeta_1) + m_1(\zeta_1) & \\ & \text{if } \zeta_2 \leqslant t < T, \end{cases} \tag{2}$$

where $m_1(t) = (t/\alpha_1)^{\beta_1}$, $m_2(t) = (t/\alpha_2)^{\beta_2}$ and $m_3(t) = (t/\alpha_3)^{\beta_3}$.

Observe that the intensity function $\lambda_j(t)$ in equation (1) is constant for $\beta_j = 1$, decreases for $\beta_j < 1$ and increases for $\beta_j > 1$, $j = 1, 2, 3$. This process is related to the Weibull probability model [14] (α_j, β_j), $j = 1, 2, 3$.

Assuming that the data are observed up to a total time T, where the epochs of occurrence of cases are denoted by t_i, $i = 1, \ldots, n$, $0 < t_1 < t_2 < \ldots < t_{N(\zeta_1)} < t_{N(\zeta_1)+1} < \ldots < t_{N(\zeta_2)} < t_{N(\zeta_2)+1} < \ldots < t_n < T$, the likelihood function for $\boldsymbol{\theta}$ in the presence of two change-points ζ_1 and ζ_2 is given by

$$L(\theta) = \prod_{i=1}^{N(\zeta_1)} \lambda_1(t_i) \mathrm{e}^{-m_1(\zeta_1)} \prod_{i=N(\zeta_1)+1}^{N(\zeta_2)} \lambda_2(t_i) \mathrm{e}^{-m_2(\zeta_2)+m_2(\zeta_1)} \times \prod_{i=N(\zeta_2)+1}^{N(T)} \lambda_3(t_i) \mathrm{e}^{-m_3(T)+m_3(\zeta_2)} \tag{3}$$

where $\lambda_j(t)$ is given in equation (1) and $m_j(t)$ is given in equation (2) for $j = 1, 2, 3$. To justify the likelihood function [equation (3)], observe that $N(s+t) - N(s)$ given $\boldsymbol{\theta}$ has a Poisson distribution $P(m(s+t|\boldsymbol{\theta}) - m(s|\boldsymbol{\theta}))$ for $t > 0$ and independent increments [9]. Thus, the sampling distribution for the between occurrence times, say U_i, has density $f_{U_1|\theta}(t) = \lambda(t|\boldsymbol{\theta}) \exp[-m(t|\boldsymbol{\theta})]$, $f_{U_2|U_1=s}(t) = \lambda(s+t|\boldsymbol{\theta}) \exp[-m(s+t|\boldsymbol{\theta}) + m(s|\boldsymbol{\theta})]$, and so on. In this way, we obtain the likelihood of the data $D_T = \{n; t_1, \ldots, t_{N(\zeta_1)}, t_{N(\zeta_1)+1}, \ldots, t_{N(\zeta_2)}, t_{N(\zeta_2)+1}, \ldots, t_n, T\}$ in the presence of two change-points. Moreover, observe that homogeneous Poisson processes in the presence of one change-point is a special case of equation (3) [9].

A Bayesian analysis

For a Bayesian analysis of the PLP with intensity function given in equation (1) in the presence of two

https://doi.org/10.1017/S0950268808000526 Published online by Cambridge University Press

change-points ζ_1 and ζ_2, we assume uniform prior distributions for α_j and β_j given by

$$\alpha_j \sim U(0, a_j) \quad \text{and} \quad \beta_j \sim U(b_{1_j}, b_{2_j}), \tag{4}$$

for $j=1, 2, 3$, where a_j, $b_{1_j}$and b_{2_j} are known hyperparameters, b_{11} and b_{13} are assumed to be equal to 0 and b_{21} and b_{23} are assumed to be equal to 1 in order to have decreasing intensity functions in the intervals $0<t<\zeta_1$ and $\zeta_2<t<T$; b_{12} is assumed to be equal to 1 to have increasing intensity function in the interval $\zeta_1<t<\zeta_2$, and a_j and b_{22} are assumed to have large values (non-informative prior distributions for α_j, $j=1, 2, 3$). We also assume uniform prior distributions for the change-points ζ_1 and ζ_2, given by

$$\zeta_\ell \sim U(c_\ell, d_\ell), \tag{5}$$

where c_ℓ and d_ℓ are known hyperparameters, $\ell=1, 2$ is assumed to have $\zeta_1<\zeta_2$. We further consider prior independence among the parameters.

The joint posterior distribution for $\boldsymbol{\theta}$ is given by [15]

$$\Pi(\boldsymbol{\theta}|D_T) \propto \left(\frac{\beta_1}{\alpha_1}\right)^{N(\zeta_1)} \left(\frac{\beta_2}{\alpha_2}\right)^{N(\zeta_2)-N(\zeta_1)} \left(\frac{\beta_3}{\alpha_3}\right)^{N(T)-N(\zeta_2)} \times \left[\prod_{i=1}^{N(\zeta_1)} \left(\frac{t_i}{\alpha_1}\right)^{\beta_1-1}\right] \left[\prod_{i=N(\zeta_1)+1}^{N(\zeta_2)} \left(\frac{t_i}{\alpha_2}\right)^{\beta_2-1}\right] \left[\prod_{i=N(\zeta_2)+1}^{N(T)} \left(\frac{t_i}{\alpha_3}\right)^{\beta_3-1}\right]$$
$$\times \exp\left\{-\left(\frac{\zeta_1}{\alpha_1}\right)^{\beta_1-1} - \left[\left(\frac{\zeta_2}{\alpha_2}\right)^{\beta_2} - \left(\frac{\zeta_1}{\alpha_2}\right)^{\beta_2}\right] - \left[\left(\frac{T}{\alpha_3}\right)^{\beta_3} - \left(\frac{\zeta_2}{\alpha_3}\right)^{\beta_3}\right]\right\}, \tag{6}$$

where $D_T=\{n;\ t_1, \ldots,\ t_n;\ T\}$, $0<\alpha_j<a_j$, $0<\beta_j<b_j$, $c_\ell<\zeta_l<d_\ell$, $j=1, 2, 3$ and $\ell=1, 2$.

To simulate samples for the joint posterior distribution [equation (6)], we could consider standard MCMC methods such as the Gibbs sampling algorithm [8] or the Metropolis–Hastings algorithm [16]. In this case, we need all full conditional posterior distributions $\Pi(\theta_j|\boldsymbol{\theta}_{(j)},\ D_T)$, $j=1,\ 2, \ldots,\ K$ and $\boldsymbol{\theta}_{(j)}=(\theta_1, \ldots,\ \theta_{j-1},\ \theta_{j+1}, \ldots,\ \theta_K)$. A great computational simplification is given by WinBugs software [17], where we only need to specify the joint distribution for the data and the prior distributions for the parameters.

RESULTS

For a Bayesian analysis of the NYC TB data, we assumed the uniform prior distributions [equation (4)] for α_j and β_j with $a_j=100$, $j=1, 2, 3$, $b_{21}=b_{23}=1$, $b_{11}=b_{13}=0$ (related to decreasing functions), $b_{12}=1$ and $b_{22}=10$ (related to an increasing function between the first and second change-points). We also assumed prior distributions [equation (6)] for the change-points ζ_1 and ζ_2 with $c_1=2558$, $d_1=4383$, $c_2=7671$ and $d_2=9131$ (number of days since 1 January 1970). This choice of prior distributions, especially for the change-points ζ_1 and ζ_2 are based on medical knowledge of the epidemic, or if it is known that the first change-point is between 1977 and 1982 and the second change point is between 1991 and 1995. We are assuming in the prior distribution [equation (5)] that the two intervals do not overlap. Using WinBugs software and considering a burn-in sample of size 40 000, we simulated a Gibbs sample of size 100 000 choosing every 50th sample for each parameter to have approximately uncorrelated samples, i.e. obtaining a final Gibbs sample of size 1200 to get the posterior summaries for each parameter. The WinBugs code is given in the Appendix.

Table 2. *Posterior summaries for the parameters*

Parameter	Median	Standard deviation	95% credible interval
α_1	0·8262	0·1625	0·5599–1·184
α_2	166·1	16·48	132·2–195·4
α_3	0·1968	0·4842	0·003–1·771
β_1	0·9127	0·0217	0·8701–0·9546
β_2	2·087	0·0507	1·978–2·175
β_3	0·8033	0·1047	0·5967–0·9843
ζ_1	3002	88·02	2909–3080
ζ_2	9123	4·639	9112–9130

Table 2 gives the posterior summaries for each parameter. Convergence of the Gibbs sampling algorithm was monitored by checking the plots of the simulated samples for each parameter to verify if a stationary distribution was obtained by the 1200 simulated Gibbs samples and also using other existing methods to check the convergence of the Gibbs sampling algorithm [18].

https://doi.org/10.1017/S0950268808000526 Published online by Cambridge University Press

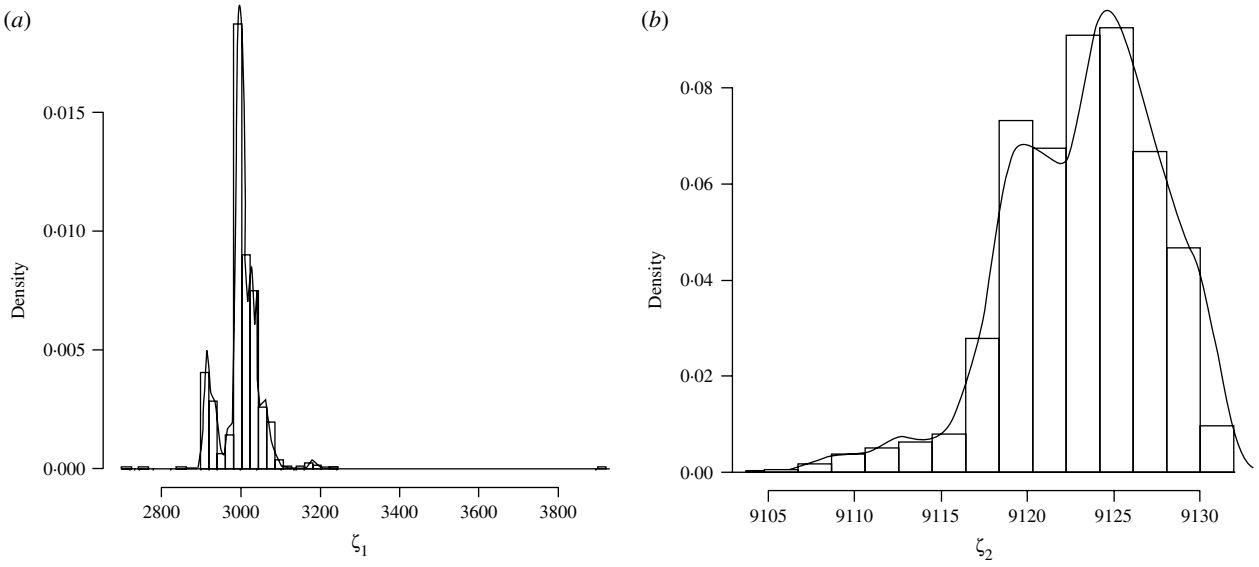

Fig. 2. Marginal posterior distributions for the change-points. (*a*) ζ_1; (*b*) ζ_2.

Figure 2 shows the plots of the marginal posterior distributions for the change-points ζ_1 and ζ_2 approximated by the simulated Gibbs samples. Inferences for the change-points are of great interest to epidemiologists.

Considering the Monte Carlo estimators for the posterior means of α_1, α_2, α_3, β_1, β_2, β_3, ζ_1 and ζ_2 given in Table 2, we obtain Bayesian estimators for the mean value function $m(t)$ given by equation (2), i.e.

$$m(t|\boldsymbol{\theta}) = \begin{cases} \left(\dfrac{t}{0{\cdot}8262}\right)^{0{\cdot}9127} & \text{if } 0 < t < 3002, \\ \left(\dfrac{t}{166{\cdot}1}\right)^{2{\cdot}087} - \left(\dfrac{3002}{166{\cdot}1}\right)^{2{\cdot}087} + \left(\dfrac{3002}{0{\cdot}8262}\right)^{0{\cdot}9127} & \text{if } 3002 \leqslant t < 9123, \\ \left(\dfrac{t}{0{\cdot}1968}\right)^{0{\cdot}8033} - \left(\dfrac{9123}{0{\cdot}1968}\right)^{0{\cdot}8033} + \left(\dfrac{9123}{166{\cdot}1}\right)^{2{\cdot}087} - \left(\dfrac{3002}{166{\cdot}1}\right)^{2{\cdot}087} + \left(\dfrac{3002}{0{\cdot}8262}\right)^{0{\cdot}9127} & \text{if } 9123 \leqslant t < 11\,323. \end{cases}$$

Table 3 gives Monte Carlo Bayesian estimators for $m(t)$ based on the 1200 simulated Gibbs samples and the observed accumulated numbers of TB cases for each year. Figure 3 shows the plot of the estimated mean vale function and the observed accumulated number of TB cases against the years (in days). We observe a good fit for the PLP in the presence of two change-points.

DISCUSSION

There was a great increase in TB prevalence in NYC during the 1980s and at the start of the 1990s, with a peak of 3811 cases in 1992. In 1978 a very low number of TB cases in NYC (1307 cases) was observed, following a long period of decreasing numbers in the prevalence of the disease. We observe that the proposed model was well fitted to the data of TB cases in NYC, and it is straightforward to implement this model in WinBugs.

The presence of more than one change-point is common in many applications of medical counting data. Considering the NYC TB data of Table 1, the results obtained in the present study could be easily extended for other epidemiological datasets, where we could have the presence of a finite number of change-points. In this way, the likelihood function [equation (6)] could be easily generalized to accommodate more than two change-points. We usually have great difficulty in obtaining classical inference results for the parameters of NHPP in the presence of change-points and the use of MCMC methods is a suitable way of obtaining Bayesian inferences for this family of models. Using WinBugs software greatly simplifies obtaining the posterior summaries of interest.

https://doi.org/10.1017/S0950268808000526 Published online by Cambridge University Press

Table 3. *Estimators for the mean value function and observed accumulated numbers*

	Time	Estimated number	Observed number
1	365	2619	2590
2	730	4919	5162
3	1096	7118	7437
4	1461	9244	9538
5	1826	11 325	11 560
6	2191	13 366	13 711
7	2557	15 386	15 862
8	2922	17 363	17 467
9	3287	18 678	18 774
10	3652	19 925	20 340
11	4018	21 332	21 818
12	4383	22 871	23 400
13	4748	24 572	24 997
14	5113	26 414	26 645
15	5479	28 403	28 274
16	5844	30 541	30 117
17	6209	32 830	32 340
18	6574	35 279	34 537
19	6940	37 876	36 854
20	7305	40 624	39 399
21	7670	43 526	42 919
22	8035	46 572	46 592
23	8401	49 790	50 403
24	8766	53 163	53 638
25	9131	56 647	56 637
26	9496	58 422	59 082
27	9862	60 200	61 135
28	10 227	61 961	62 865
29	10 592	63 717	64 423
30	10 957	65 457	65 883
31	11 323	67 181	67 215

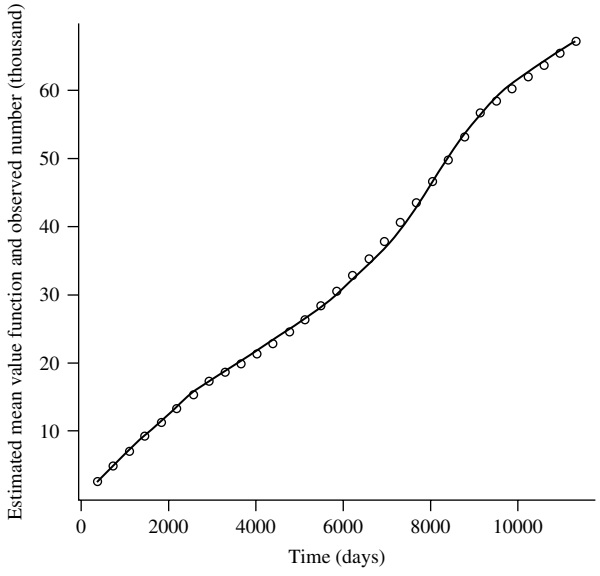

Fig. 3. Mean value function. —, Estimated; ○, observed.

Other parametrical forms for the intensity functions [equation (1)] could be considered in place of PLP. In this case we could consider other usual intensity functions commonly used in software reliability studies, e.g. Gompertz growth, logistic growth, etc. [19].

DECLARATION OF INTEREST

None.

REFERENCES

1. **Pio A, Chaulet P.** *Tuberculosis Handbook* (*WHO/TB/98.253*). Geneva: World Health Organization, 1998.
2. **Ducati RG, *et al.*** The resumption of consumption – a review on tuberculosis. *Memórias do Instituto Oswaldo Cruz* 2006; **101**: 697–714.
3. **Ruffino-Netto A.** Tuberculosis: the neglected calamity. *Revista da Sociedade Brasileira de Medicina Tropical* 2002; **35**: 51–58.
4. **Ruffino-Netto A.** Tuberculosis load: reflections on a theme. *Jornal Brasileiro de Pneumologia* 2004; **30**: 307–309.
5. **Coker R.** Lessons from New York's tuberculosis epidemic. *British Medical Journal* 1998; **317**: 616.
6. **Wallace DN.** Discriminatory public policies in the New York City tuberculosis epidemic, 1975–1993. *Microbes and Infection* 2001; **3**: 515–524.
7. **Paolo Jr. WF, Nosanchuk JD.** Tuberculosis in New York city: recent lessons and a look ahead. *Lancet Infection Diseases* 2004; **4**: 287–293.
8. **Gelfand AE, Smith AFM.** Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* 1990; **85**: 398–409.
9. **Raftery AE, Akman VE.** Bayesian analysis of a Poisson process with a change-point. *Biometrika* 1986; **73**: 85–89.
10. **Ruggeri F, Sivaganesan S.** On modeling change points in non-homogeneous Poisson processes. *Statistical Inference for Stochastic Processes* 2005; **8**: 311–329.
11. **Cox DR, Lewis PA.** *Statistical Analysis of Series of Events*. London: Methuen, 1966.
12. **Musa JD, Okumoto K.** A logarithm Poisson execution time model for software reliability measurement. *Proceedings of Seventh International Conference on Software Engineering*. Orlando, 1984, pp. 230–238.
13. **Mudholkar GS, Srivastava DK, Friemer M.** The exponentiated Weibull family: a reanalysis of the bus-motor failure data. *Technometrics* 1995; **37**: 436–445.
14. **Kuo L, Yang TY.** Bayesian computation for non-homogeneous Poisson process in software reliability. *Journal of the American Statistical Association* 1996; **91**: 763–773.
15. **Box GEP, Tiao GC.** *Bayesian Inference in Statistical Analysis*. New York: Addison-Wesley, 1973.

https://doi.org/10.1017/S0950268808000526 Published online by Cambridge University Press

16. **Smith AFM, Roberts GO.** Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo methods. *Journal of the Royal Statistical Society B* 1993: **55**: 3–23.
17. **Spiegelhalter D, Thomas A, Best N.** *WinBUGS Version 1.3 User Manual*. Cambridge: Medical Research Council Biostatistics Unit, 2000.
18. **Gelman A, Rubin DB.** Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science* 1992; **7**: 457–511.
19. **Musa JD, Iannino A, Okumoto K.** *Software Reliability: Measurement, Prediction, Application*. New York: McGraw Hill, 1987.

APPENDIX

The WinBUGS code used to fit the Bayesian model is given below:

```
model {
c<- 1000
for (i in 1:N) {
zeros[i] <- 0
phi[i] <- -log(L[i])+c
zeros[i] ~dpois(phi[i])
log(lambda[i]) <- log(beta[J[i]]) - log(alpha[J[i]])
+ (beta[J[i]]-1)*(log(t[i])-log(alpha[J[i]]))
L[i] <- lambda[i]*m
J[i] <- 1+step(t[i]-tau1-.5) + step(t[i]-tau2-.5)
}
m <- exp (-(pow((tau1/alpha[1]),beta[1])
       + pow((tau2/alpha[2]),beta[2])
       + pow((T/alpha[3]),beta[3])-pow((tau1/alpha[2]),beta[2])
       - pow((tau2/alpha[3]),beta[3]))/N)
tau1 ~dunif(c1,d1)
tau2 ~dunif(c2,d2)
alpha[1] ~dunif(0,a1)
alpha[2] ~dunif(0,a2)
alpha[3] ~dunif(0,a3)
beta[1] ~dunif(0,b21)
beta[2] ~dunif(1,b22)
beta[3] ~dunif(0,b23)
}
list(N=6721,T=11323)
list(a1= , a2= , a3= , b21= , b22= , b23= , c1= , d1= , c2= , d2= )
```

In this WinBUGS code, a_1, a_2, a_3 are the hyperparameters of the uniform prior distribution of α_1, α_2 and α_3, respectively; b_1, b_2, b_3 are the hyperparameters of the uniform prior distribution of β_1, β_2 and β_3, respectively [see equation (4)]; and c_1, d_1, c_2 and d_2 are the hyperparameters in equation (5). These hyperparameters are declared in the last line of the WinBugs code.

https://doi.org/10.1017/S0950268808000526 Published online by Cambridge University Press