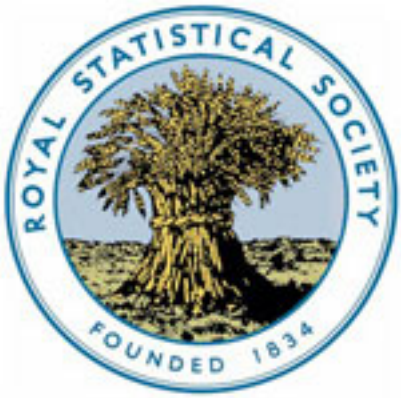

Tumour Incidence, Prevalence and Lethality Estimation in the Absence of Cause-of-Death Information
Author(s): T. Härkänen and E. Arjas
Source: *Applied Statistics*, Vol. 53, No. 4 (2004), pp. 601-617
Published by: Blackwell Publishing for the Royal Statistical Society
Stable URL: http://www.jstor.org/stable/3592577
Accessed: 14/05/2009 06:04

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=black.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit organization founded in 1995 to build trusted digital archives for scholarship. We work with the scholarly community to preserve their work and the materials they rely upon, and to build a common research platform that promotes the discovery and use of these resources. For more information about JSTOR, please contact support@jstor.org.

Royal Statistical Society and *Blackwell Publishing* are collaborating with JSTOR to digitize, preserve and extend access to *Applied Statistics.*

http://www.jstor.org

Appl. Statist. (2004)
53, *Part* 4, *pp.* 601–617

Tumour incidence, prevalence and lethality estimation in the absence of cause-of-death information

T. Härkänen and E. Arjas

University of Helsinki, Finland

[Received August 2001. Final revision October 2003]

Summary. A Bayesian intensity model is presented for studying a bioassay problem involving interval-censored tumour onset times, and without discretization of times of death. Both tumour lethality and base-line hazard rates are estimated in the absence of cause-of-death information. Markov chain Monte Carlo methods are used in the numerical estimation, and sophisticated group updating algorithms are applied to achieve reasonable convergence properties. This method was tried on the rat tumorigenicity data that have previously been analysed by Ahn, Moon and Kodell, and our results seem to be more realistic.

Keywords: Bayesian inference; Bioassay; Data augmentation; Intensity model; Markov chain Monte Carlo methods; Serial sacrifice

1. Introduction

This paper introduces a semiparametric Bayesian intensity model for estimating tumour incidence, prevalence and lethality, which is fitted to the rats data that were presented in Thurman *et al.* (1994) and analysed by Ahn *et al.* (2000). The various phases of the life of a rat are considered here in the form of the multistate model that is depicted in Fig. 1. A rat is considered to have a tumour when the tumour has grown to a size that is sufficiently large to be detectable in necropsy. A rat can die from two types of tumour (*mononuclear cell leukaemia* (MCL) and *pituitary adenoma or carcinoma*, denoted PIT), if present, or from competing risks.

In these data there is missing information on two levels of the data: first, the causes of death are unknown. The key to identifying the lethalities of different causes of death is that some of the rats died without tumours. If a rat had a tumour, the excess risk of death is assumed to be caused by the tumour. Here it turned out that the excess risk was small. Second, in this design there is severe interval censoring on the tumour onset times. To compensate for this, the tumour prevalences at certain ages were estimated by sacrificing some randomly chosen rats. This *serial sacrificing* scheme was compared with some other experimental designs in Borgan *et al.* (1984): it was found to be more efficient than a simple survival experiment, but less efficient than experiments in which rats were periodically examined to determine whether tumours were present, or complete observation in which the tumour onset times were observed exactly.

The model here is a synthesis of ideas that were presented by McKnight and Crowley (1984) and Dinse (1991). McKnight and Crowley (1984) introduced a continuous time intensity model for serial sacrificing experiments, and discussed its identifiability, but did not comment on meth-

Address for correspondence: T. Härkänen, Rolf Nevanlinna Institute, University of Helsinki, PO Box 4, FIN-00014 Helsinki, Finland.
E-mail: tommi.harkanen@rni.helsinki.fi

© 2004 Royal Statistical Society 0035–9254/04/53601

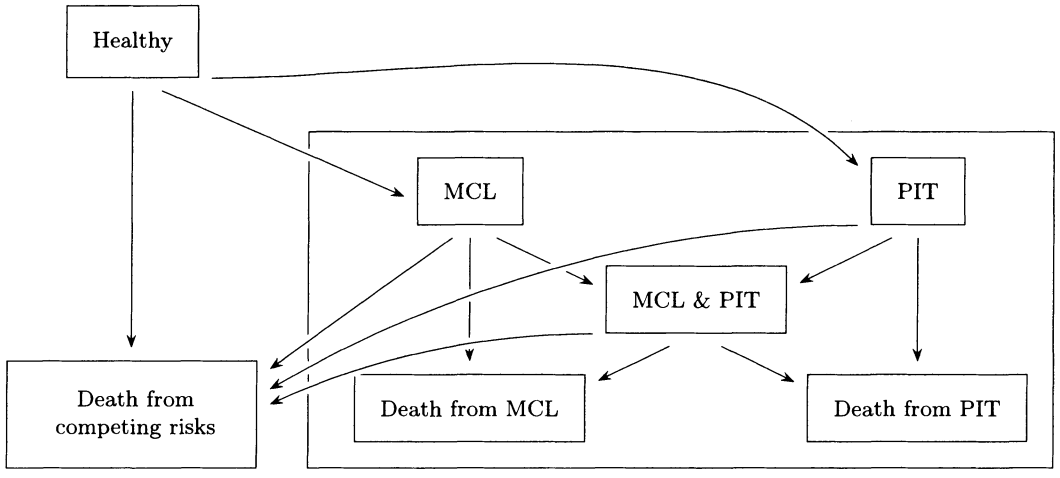

Fig. 1. Directed acyclic graph depicting the phases of a rat (MCL, mononuclear cell leukaemia; PIT, pituitary adenoma or carcinoma)

ods of estimation. Dinse (1991) presented a discrete time model with some constraints imposed on the hazard rates to make the model identifiable. For background on intensity models see Andersen *et al.* (1993) and Gelman *et al.* (1995) on hierarchical Bayesian methods and inference. Arjas and Gasbarra (1994) presented a nonparametric specification of intensity models which is applied here, and Härkänen (2003) has presented software for analysing such models.

The missing tumour onset times are augmented by generating synthetic values from the model, following the ideas of Tanner and Wong (1987). The values of the tumour onset times depend not only on the tumour incidence rate but also on the times of death of the rats and on the excess risk of death caused by the tumour: if the tumour is very lethal, the time from tumour onset to death is likely to be short and the prevalence of tumours low. In that case only few sacrificed rats could be expected to have a tumour, and a rat with one type of tumour would have very little time to develop a tumour of the other type.

The marginal expectations and distributions are here estimated numerically by using the Metropolis–Hastings algorithm. In this case finding a good proposal distribution is not easy because of strong dependences between the model parameters and the missing tumour onset times. Since the model is partly nonparametric, standard adaptive algorithms are not directly applicable, and therefore two sophisticated proposal distributions are used; see Appendix A for details. The software is available at `www.rni.helsinki.fi/~tth/bite.html`.

The structure of the paper is as follows. Section 2 introduces the intensity model, prior distributions, predictive survival functions and tumour lethality estimates. The results are presented in Section 3, and a comparison with the results of Ahn *et al.* (2000) is made in Section 4 where some earlier references in this field are also discussed.

2. Data and model

The data contain the lifetimes, tumours found and covariate information on 851 'Fischer 344' rats that have previously been analysed in Ahn *et al.* (2000). The covariates were gender (male or female) and nourishment (*ad libitum* (AL) or calorie restricted (CR)). The latter covariate is called diet in what follows. Together these covariates divide the rats into four groups.

Dead rats were subjected to necropsy, and the presence of tumours and the tumour types were recorded, except for 69 rats for which the PIT status was not recorded. The other rats gave

rise to 'complete' observations: 577 rats died from a natural cause; of these, 85 had no tumour of either kind, 194 had only PIT and 131 had both MCL and PIT. The remaining 205 rats were sacrificed: at the age of 369, 556, 755, 920 and 1097 days (denoted here by $t_1, t_2, \ldots, t_5$, and $t_0 := 0$) 12 randomly chosen rats in each group which were still alive were sacrificed, and finally the last four rats (all females in the CR group) who were still alive at the age of 1293 days ($=: t_6$) days were sacrificed. Very few of the rats under 1.5 years old, and approximately half of the older rats, had a tumour, suggesting that young rats have few tumours, and that the tumours are not very lethal.

2.1. *Intensity model*

Since there is very little information on the times of tumour onset, and even on their order of onset, the following simplifying independence assumptions are made here.

(a) The risk of onset of tumour $y \in \{\text{MCL, PIT}\}$ does not depend on the presence or absence of the other tumour type $y' \neq y$.
(b) The risk of death from tumour y does not depend on the presence or absence of the other tumour type $y' \neq y$, nor on the age of the rat, nor of the tumour (model 1).

In reality, the presence of a tumour might increase the probabilities of onset of other types of tumour, and if a rat has both types of tumour the risk of tumour-related death might be higher than simply the sum of the two tumour-specific risks. Moreover, the risk of death from a tumour is likely to depend on time, possibly effectively only through the age of the tumour (although, in a study of mice, Dinse (1991) concluded that the effect of the age of the mouse was not statistically significant). Estimation of time dependences seems to be very difficult, essentially because the present study design provides so little information about the individual tumour onset times. Therefore the possible dependence on tumour age is incorporated in the models (model 2) by making a strong prior assumption about its functional form. This can be partly compensated by carrying out suitable sensitivity analyses afterwards.

Table 1 describes the notation that is used for the event times and the corresponding base-line hazards. If rat i was sacrificed, then we can view the death time $T_i^{C \vee D}$ from tumours or from competing risks as being right censored.

The hazard rates for tumour onset can now be specified for each of the four possible covariate values: for rat i, $h^T_{\text{MCL},i}(t) := f^T_{\text{MCL}}(t|X_i)$ (where X_i stands for the covariates), and similarly $h^T_{\text{PIT},i}(t) := f^T_{\text{PIT}}(t|X_i)$.

Table 1. Notation for rat i†

Event type	*Base-line hazard rate*	*Event time*
MCL incidence rate	$f^T_{\text{MCL}}(t\|X_i)$	$T^T_{\text{MCL},i}$
PIT incidence rate	$f^T_{\text{PIT}}(t\|X_i)$	$T^T_{\text{PIT},i}$
Sacrifice		T_i^S
Death from MCL	$f^D_{\text{MCL}}(X_i)$	
Death from PIT	$f^D_{\text{PIT}}(X_i)$	
Death from competing risks	$f^C(t\|X_i)$	
Natural death	See model 1	$T_i^{C \vee D}$

†Covariates are denoted by X_i, and the age of rat i by t.

Obviously, rat i can die from tumour y only after the tumour onset at age $T^T_{y,i}$, and before that only death from 'other causes' is possible. By the independence properties (a) and (b) above, the stochastic hazard rate for death given the (past) tumour onset times will in model 1 have the form

$$h_i^{C\vee D}(t) := \overbrace{f^C(t|X_i)}^{h_i^C(t)} + \overbrace{f^D_{\mathrm{MCL}}(X_i)\,\mathbb{1}_{[T^T_{\mathrm{MCL},i},\infty]}(t)}^{h^D_{\mathrm{MCL},i}(t)} + \overbrace{f^D_{\mathrm{PIT}}(X_i)\,\mathbb{1}_{[T^T_{\mathrm{PIT},i},\infty]}(t)}^{h^D_{\mathrm{PIT},i}(t)}.$$

An alternative model, in which (b) is relaxed slightly and it is assumed that the risk of dying from a tumour depends on the age of that tumour via a Weibull-type hazard rate, is the following (model 2):

$$h_i^{C\vee D}(t) := \overbrace{f^C(t|X_i)}^{h_i^C(t)} + \overbrace{f^{D*}_{\mathrm{MCL}}(X_i)(t - T^T_{\mathrm{MCL},i} + 1)^{\gamma_{\mathrm{MCL}}}\,\mathbb{1}_{[T^T_{\mathrm{MCL},i},\infty]}(t)}^{h^D_{\mathrm{MCL},i}(t)} + \underbrace{f^{D*}_{\mathrm{PIT}}(X_i)(t - T^T_{\mathrm{PIT},i} + 1)^{\gamma_{\mathrm{PIT}}}\,\mathbb{1}_{[T^T_{\mathrm{PIT},i},\infty]}(t)}_{h^D_{\mathrm{PIT},i}(t)}.$$

Owing to the severe interval censoring in the data, reliable estimation of the powers γ_yMCL and γ_{PIT} is virtually impossible, and therefore a model with $\gamma_{\mathrm{MCL}} := \gamma_{\mathrm{PIT}} := 2$ was used here. Note that model 2 reduces to model 1 if $\gamma_{\mathrm{MCL}} = \gamma_{\mathrm{PIT}} = 0$ and $f^{D*}_y = f^D_y$.

The (stochastic) intensity functions for tumour onset are now given by

$$\lambda^T_{\mathrm{MCL},i}(t) := h^T_{\mathrm{MCL},i}(t)\,\mathbb{1}_{[0,T^T_{\mathrm{MCL},i}]}(t),$$

$$\lambda^T_{\mathrm{PIT},i}(t) := h^T_{\mathrm{PIT},i}(t)\,\mathbb{1}_{[0,T^T_{\mathrm{PIT},i}]}(t)$$

and for death by

$$\lambda_i^{C\vee D}(t) := h_i^{C\vee D}(t)\,\mathbb{1}_{[0,T_i^{C\vee D}]}(t).$$

2.2. *Prior distributions for base-line hazards*

As an approximation facilitating the numerical integration of the posterior distribution, we assume that the nonparametric random base-line hazard functions are piecewise constant, as in Arjas and Gasbarra (1994). Using the generic notation f for such functions supported by an interval $[0, \mathfrak{T}_{\max}]$, we can write

$$f(t) := \sum_{n=0}^{n(\mathfrak{T})} a_n\,\mathbb{1}_{(\mathfrak{T}_n,\,\mathfrak{T}_{n+1}]\cap[0,\,\mathfrak{T}_{\max}]}(t),$$

where $n(\mathfrak{T})$ is the number of jump points on $(0, \mathfrak{T}_{\max})$. The prior distribution of the jump points $0 =: \mathfrak{T}_0 < \mathfrak{T}_1 < \ldots < \mathfrak{T}_{n(\mathfrak{T})+1} := \mathfrak{T}_{\max}$ is assumed to be a Poisson process on $(0, \mathfrak{T}_{\max})$ with (hyper)parameter μ. The prior distribution for the levels $(a_n)_{n\geqslant 0}$ is defined by

$$a_n \sim \begin{cases} \mathrm{gamma}(\cdot|\alpha_0, \beta_0), & n = 0, \\ \mathrm{gamma}(\cdot|\alpha, \alpha/a_{n-1}), & n > 0. \end{cases}$$

Nonparametric estimation of the base-line hazards f^T_{MCL} and f^T_{PIT} did not succeed because the estimation algorithm turned out to be numerically unstable, and therefore their jump points were fixed: $\mathfrak{T}_n := t_n$ for $n = 1, 2, \ldots, 5$. The base-line hazard for death from competing risks f^C is

estimated nonparametrically by setting $\mu := 1$. See Härkänen *et al.* (2000) for more details concerning the estimation. The hyperparameter values for the levels of f^T_{MCL}, f^T_{PIT} and f^C are fixed here at values $(\alpha_0, \beta_0, \alpha) := (0.1, 1, 1)$. These values do not reflect particularly vague prior knowledge but should be sufficiently flexible to let the observed information in the data control effectively the posterior distribution. The hyperparameter values for f^D_y are $(\alpha_0, \beta_0) := (0.1, 0.01)$. The prior expectation of f^D_y is 10, but the prior variance is 1000, and in this sense the prior can be considered as being rather vague. In model 2, the parameters f^{D*}_y are given the same prior. The relative risk that is associated with tumour age, $(t - T^T_{y,i})^{\gamma_y}$, is also approximated by a piecewise constant function.

2.3. *Survival functions, and tumour prevalence and lethality estimates*

The posterior expectations and credibility intervals of the random variables are presented as the outcomes of the analysis. To compare our results with those of Ahn *et al.* (2000) we define the (random) function

$$S^T_y(t) := \exp\left\{-\int_0^t \hbar^T_y(s)\,\mathrm{d}s\right\}, \qquad y \in \{\mathrm{MCL}, \mathrm{PIT}\}. \tag{1}$$

Even when we view the function $\hbar^T_y$ as given, expression (1) does not have a direct probabilistic interpretation in terms of 'real rat lives' because it refers to a model from which the possibility of dying from other causes or from the other tumour type $y' \neq y$ have been removed. In spite of this, we use in the following the term (*tumour onset*) *survival function* (SF) for this function. The SF for death is defined similarly, by

$$S^{C\vee D}(t) := \exp\left\{-\int_0^t \hbar^{C\vee D}(s)\,\mathrm{d}s\right\}. \tag{2}$$

In practice $\hbar^T_y$ and $\hbar^{C\vee D}$ are unknown, and we can only deal with the posterior distribution of expressions (1) and (2). We could therefore report, for each t, the (pointwise) posterior credibility intervals for $S^T_y(t)$ and $S^{C\vee D}(t)$, combined for example with their posterior means. This is actually done in Figs 3 and 4 later. Viewed as a function of t, the latter has an interpretation of a predictive SF, corresponding to the conditional probability given the data, that a generic individual (rat) survives at least until age t. Since the tumour onset times T^T_y appearing in the definition of $\hbar^{C\vee D}$ (model 1 and model 2) are random, the posterior expectation of expression (2) needs to be taken over all tumour onset times T^T_y. When applying Markov chain Monte Carlo (MCMC) methods, this is done by generating, at every iteration k, the tumour onset times from the corresponding tumour onset distributions (based on $\hbar^T_y[k](\cdot)$) and then by averaging the values of $\exp\{-\int_0^t \hbar^{C\vee D}[k](s)\,\mathrm{d}s\}$.

Tumour prevalence, the proportion of living rats carrying tumour y at age t, is estimated by the posterior predictive probability that a generic rat that is alive at age t would have a tumour of type y, i.e.

$$\mathbb{P}\{T^T_y < t | T^{C\vee D} t, \mathrm{data}\} = \frac{\mathbb{P}\{T^T_y < t, T^{C\vee D} t | \mathrm{data}\}}{\mathbb{P}\{T^{C\vee D} > t | \mathrm{data}\}}. \tag{3}$$

The proportion of deaths from a tumour to all deaths is estimated by using the concept of the *attributable fraction*

$$\eta_{y,i} := \frac{\hbar^D_{y,i}(T^{C\vee D}_i)}{\hbar^{C\vee D}_i(T^{C\vee D}_i)} \tag{4}$$

which is often given a causal interpretation, as the probability that 'the death of a rat i was caused by tumour of type y'; see Greenland and Robins (1988). Note that $\eta_{y,i}$ has value 0 if tumour y was not found in necropsy. Here we consider the posterior expectation

$$\tilde{\eta}_{y,i} := \mathbb{E}[\eta_{y,i}|\text{data}], \qquad y \in \{\text{MCL, PIT}\}. \tag{5}$$

In our model tumours cannot be beneficial to health, i.e. $h^D_{y,i}(t) \geqslant 0$ and $\eta_{y,i} \in [0, 1]$ for all t, i and y. The number of rats $r_{y,k}$ which died from tumour y during the sacrifice interval $(t_{k-1}, t_k]$ is estimated simply by the sum of expression (5):

$$\tilde{r}_{y,k} := \sum_{i:T_i^{C\vee D}\in(t_{k-1},t_k]} \tilde{\eta}_{y,i}. \tag{6}$$

3. Results

An MCMC simulation of 100000 iterations (containing 10000 iterations of burn-in) on an 800 MHz Pentium III personal computer took about 17 h. Several simulations were run, and the diagnostic tests provided in the CODA package (see Best *et al.* (1996)) were executed. All Geweke's Z-score (Geweke, 1992) values were in the range $(-2, 2)$; the 2.5% quantiles were estimated with accuracy 0.02 with probability 0.95; all chains were stationary according to the Heidelberger and Welch (1983) test, and the accuracy of the parameters was for most parameters better than 0.1; autocorrelations of the chains were close to zero with lags greater than 450. Therefore the mixing of f^D_y was judged to be fairly good. The following results are for model 1. A comparison with the results from model 2 is made in Section 4.

Although the posterior expectations of the hazard rates $f^D_y(g, d)$ of death from tumour y, gender $g \in \{\text{male, female}\}$ and diet $d \in \{\text{AL, CR}\}$ in Table 2 show some differences, the posterior probabilities $\mathbb{P}\{f^D_y(g, d) > f^D_y(g', d'|\text{data}\}$ where either $g \neq g'$ and $d = d'$ or $g = g'$ and $d \neq d'$ show no clear differences between gender and diet. This is apparent when considering the lower 2.5% quantiles which are all very close to 0 in Table 2.

For the competing risk hazard rates $f^C(\cdot)$ the situation is different and the covariates are clearly influential, as can be seen in Fig. 2. Diet seemed to have more influence than gender: in the AL groups rats had a higher risk of dying from the competing risks than in the CR groups. The difference seemed clear after the age of 1.5 years in males and 2 years in females. The influence of gender was weaker, although males in the AL groups seemed to have a slightly higher risk of dying from competing risks than females. Nonparametric modelling of the competing risks hazard rate $f^C(\cdot)$ revealed details of time dependence, which the fixed jump points of $f^C(\cdot)$ might hide.

Table 2. Posterior expectations of f^D_{MCL} and f^D_{PIT} with 95% credibility intervals for model 1

Gender	*Diet*	*Results for MCL*		*Results for PIT*	
		$\mathbb{E}[f^D_{\text{MCL}}\|data]$	*(2.5%, 97.5%)*	$\mathbb{E}[f^D_{\text{PIT}}\|data]$	*(2.5%, 97.5%)*
Male	AL	0.27	(0, 1.30)	0.03	(0, 0.44)
Male	CR	0.46	(0, 1.14)	0.48	(0, 1.54)
Female	AL	1.59	(0, 4.07)	0.02	(0, 0.25)
Female	CR	0.60	(0, 1.74)	0.59	(0, 2.21)

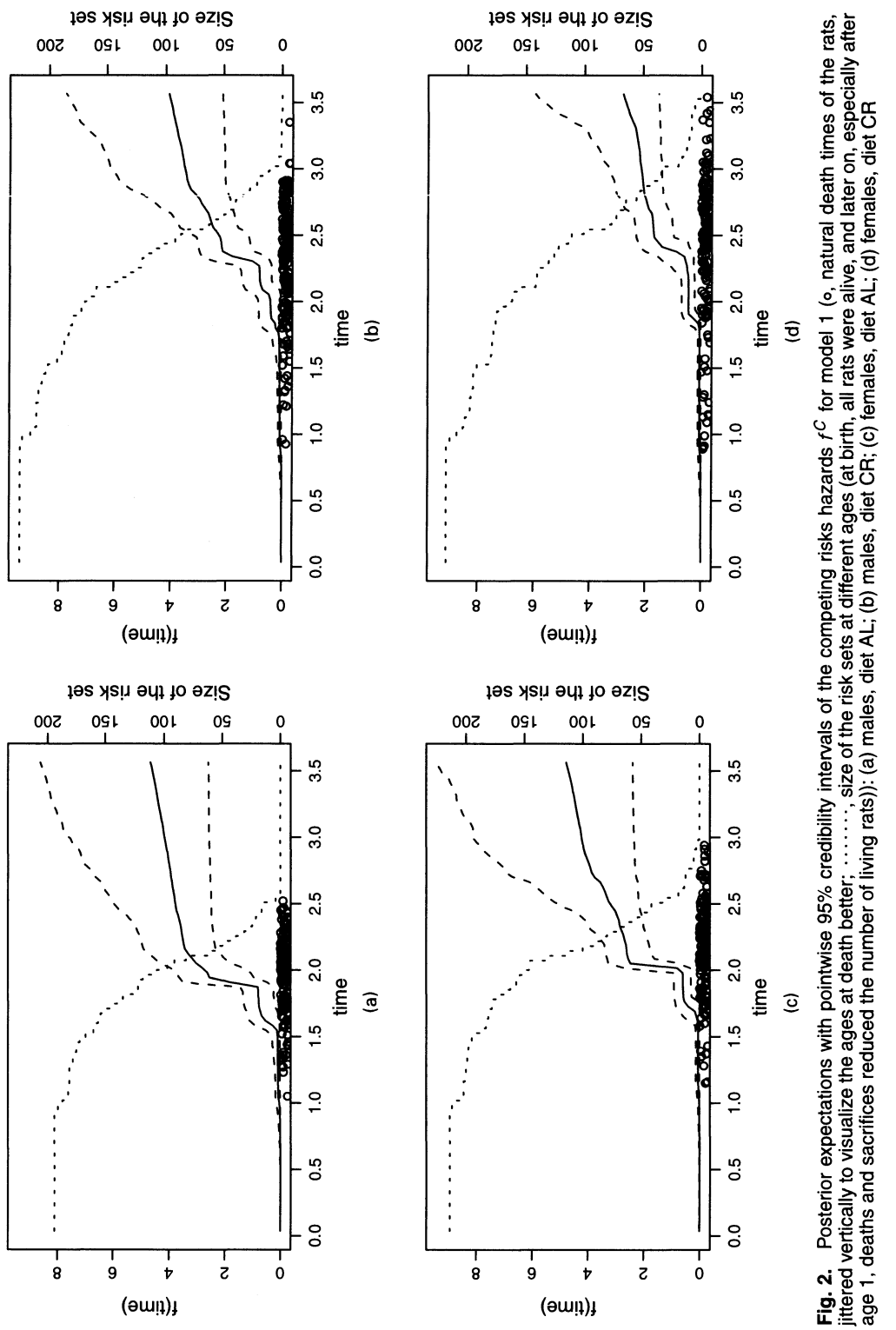

Fig. 2. Posterior expectations with pointwise 95% credibility intervals of the competing risks hazards f^C for model 1 (o, natural death times of the rats, jittered vertically to visualize the ages at death better; ·····, size of the risk sets at different ages (at birth, all rats were alive, and later on, especially after age 1, deaths and sacrifices reduced the number of living rats)): (a) males, diet AL; (b) males, diet CR; (c) females, diet AL; (d) females, diet CR

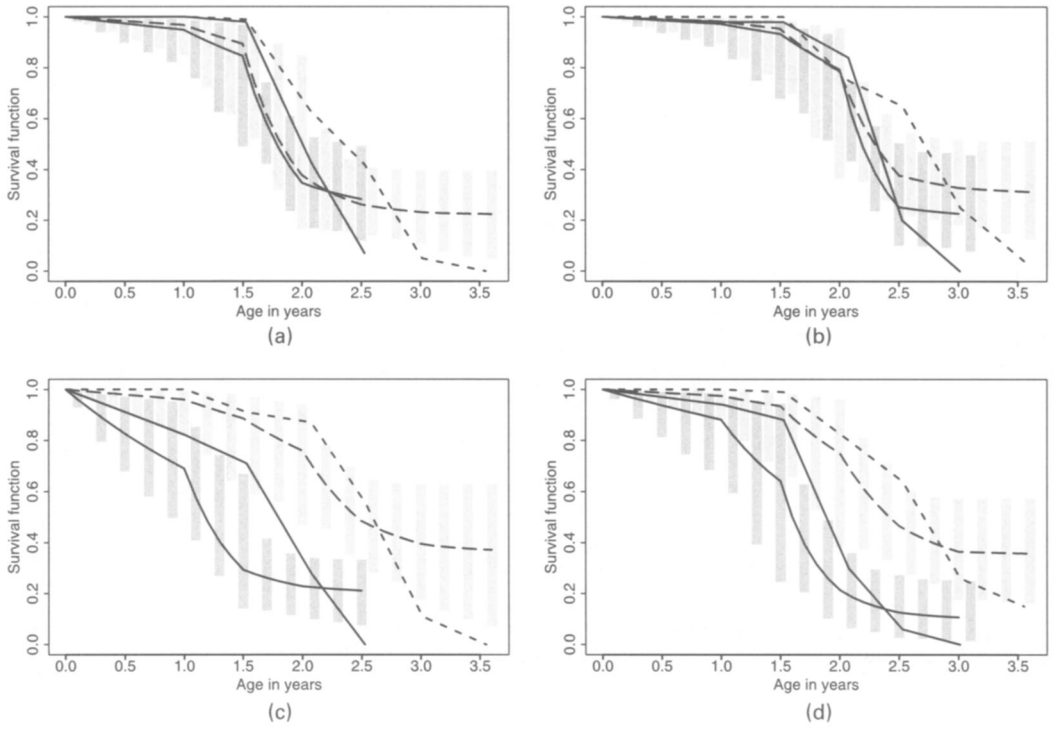

Fig. 3. Posterior expectations of $S_\gamma^T(t)$ for tumour onset times with 95% pointwise credibility intervals (———, AL by model 1; - - - - - - -, CR by model 1; ———, AL by Ahn *et al.* (2000); · · · · · · ·, CR by Ahn *et al.* (2000); ■, AL credibility interval; ■, CR credibility interval): (a) MCL tumour, males; (b) MCL tumour, females; (c) PIT tumour, males; (d) PIT tumour, females

In Fig. 3 the posterior distributions of the tumour onset SF (1) for tumour onset are compared with the results of Ahn *et al.* (2000). The tumour onset SF estimates derived from them seem to have some peculiarities, especially at early and at late ages: if the rats that were sacrificed at the sacrificing ages t_1 and t_2 did not have tumours, then the estimate for tumour onset SF in Ahn *et al.* (2000) was 1 before t_2, which is certainly an underestimate of the tumour onset risk during $(0, t_2)$. The posterior expectations of expression (1) avoid these problems. There is a difference in the PIT onset probabilities between the diets: in the AL groups the rats develop tumours earlier than in the CR groups. In MCL no such difference was found. These findings agree with those of Ahn *et al.* (2000).

A traditional estimator of an SF is the Kaplan–Meier estimator (Andersen *et al.* (1993), pages 255–286). Fig. 4 shows that the posterior expectations of SFs for death (2) correspond closely to the Kaplan–Meier estimates, and the latter are almost everywhere contained in the narrow 50% credibility regions that are defined by the corresponding pointwise posterior distributions.

As Dinse (1991) noted, if tumours have no effect on the risk of death, the proportion of rats with a tumour in some age interval should be approximately the same among the sacrificed rats and those that died naturally. Looking at Fig. 5, this appeared to be so for the rats on the AL diet and having a PIT tumour. In the other cases the rats that died naturally appeared to have a slightly higher tumour prevalence than those which were sacrificed, suggesting higher tumour lethality. These findings are in good agreement with the posterior expectations of the hazard rates that are shown in Table 2. The estimates of MCL prevalence (3) may be slightly

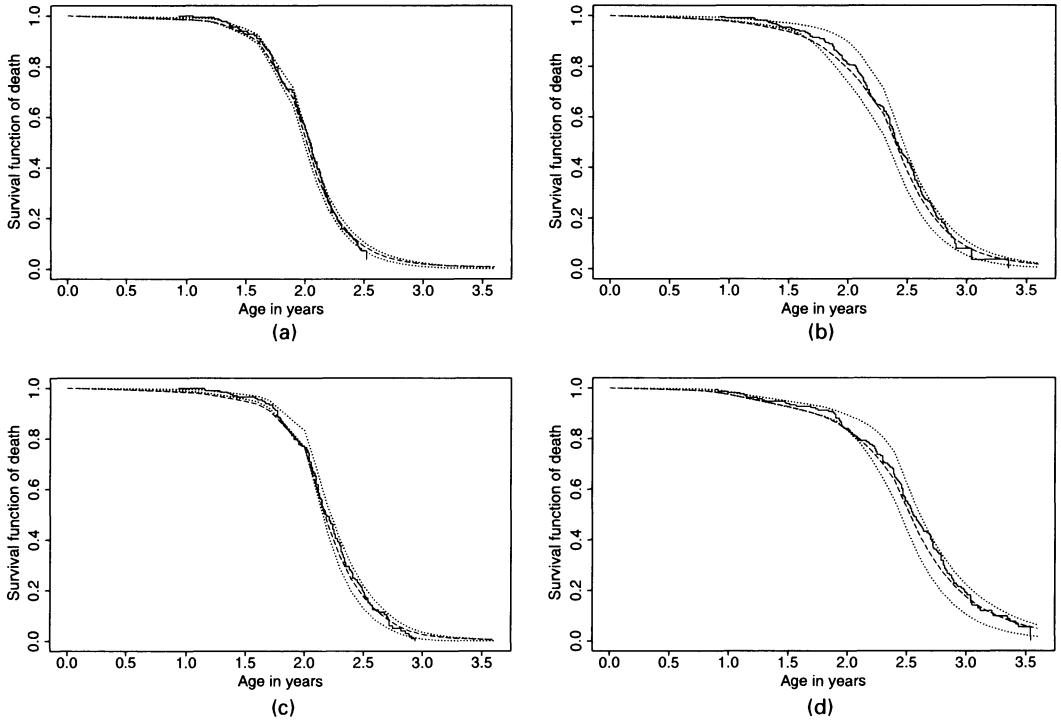

Fig. 4. Posterior expectations (-------) and 50% pointwise credibility intervals (· · · · · · ·) of the survival function for death, and the corresponding Kaplan–Meier estimates (———): (a) males, diet AL; (b) males, diet CR; (c) females, diet AL; (d) females, diet CR

too high, but PIT prevalences match quite well with the observed prevalences in the sacrificed rats.

The numbers of rats which died from a tumour was estimated in each case by using expression (6). Table 3 shows that the results which are presented here are not very different from those in Ahn *et al.* (2000) at early ages but suggests that later far fewer deaths were caused by tumours. This can be understood by considering the simulation study in Ahn *et al.* (1999) which shows that there is some bias in their estimator, especially at later ages when tumour lethality is low, and comparing the posterior expectations of death from tumour hazard rates in Table 2 with those of death from competing risks in Fig. 2: the hazard rates corresponding to death from competing risks are low in young animals, and therefore the rats which died young and were carrying a tumour probably died from the tumour. Old rats die relatively quickly whether they carry a tumour or not because the hazard of death from competing risks has much higher values than the hazard of dying from a tumour. The most dramatic difference between this study and that of Ahn *et al.* (2000) is in the PIT lethality of the AL groups. They noted that

> 'pathologists often claim that accurate determinations of the cause of death are impossible, and classification errors can produce biases'.

Otherwise, and disregarding the pathologists' assessments, our result seems more realistic.

4. Discussion

The study of tumour incidence and lethality estimation has been a popular research topic during

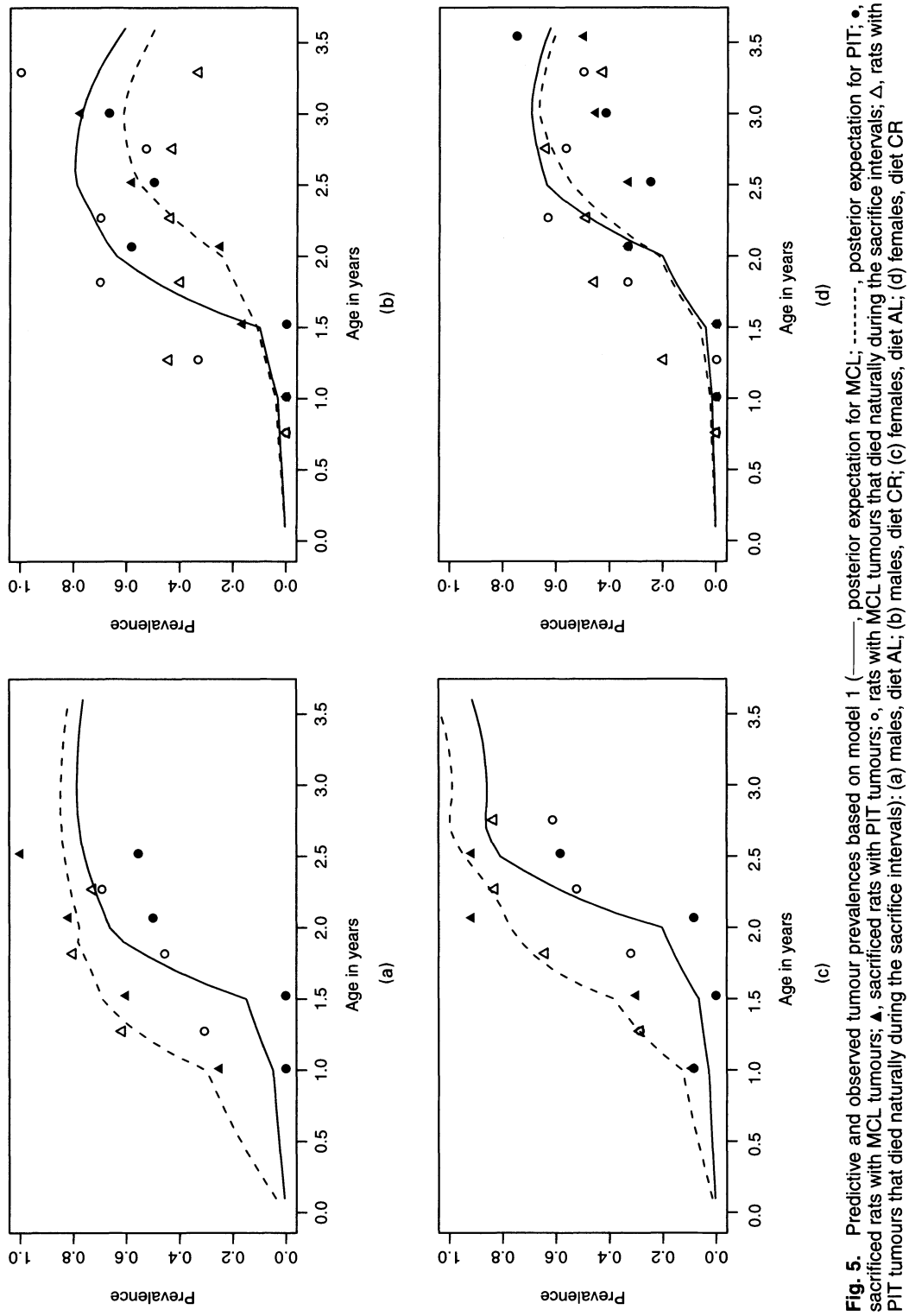

Fig. 5. Predictive and observed tumour prevalences based on model 1 (——, posterior expectation for MCL; - - - - - -, posterior expectation for PIT; ●, sacrificed rats with MCL tumours; ▲, sacrificed rats with PIT tumours; ○, rats with MCL tumours that died naturally during the sacrifice intervals; △, rats with PIT tumours that died naturally during the sacrifice intervals): (a) males, diet AL; (b) males, diet CR; (c) females, diet AL; (d) females, diet CR

Table 3. Assigned *versus* estimated (model 1) tumour lethality†

Gender	*Diet*	*Numbers in the following age intervals (days):*					
		0–368	*369–555*	*556–754*	*755–919*	*920–1096*	*1097–1293*
MCL							
Male	AL	*0*	*4*	*35*	*36*	*0*	*0*
		0	3.8	16.8	18.9	0	0
		0	**1.7**	**4.4**	**2.4**	**0**	**0**
Male	CR	*0*	*2*	*21*	*50*	*27*	*3*
		0	1.9	16.6	33.2	12.2	1.5
		0	**1.8**	**17**	**17**	**4.5**	**0.47**
Female	AL	*0*	*2*	*16*	*36*	*6*	*0*
		0	1.9	12.8	22.3	5.5	0
		0	**1.9**	**12**	**19**	**4.2**	**0**
Female	CR	*0*	*0*	*8*	*30*	*26*	*6*
		0	0	3.8	22.8	16.5	2.8
		0	**0**	**5.2**	**13**	**7.9**	**1.9**
PIT							
Male	AL	*0*	*3*	*44*	*12*	*0*	*0*
		0	2.1	32.9	13.4	0	0
		0	**0.62**	**2.6**	**0.6**	**0**	**0**
Male	CR	*0*	*3*	*5*	*14*	*12*	*1*
		0	2.6	5	13.1	7.3	0.5
		0	**3.2**	**6.8**	**13**	**5.8**	**0.28**
Female	AL	*0*	*2*	*25*	*56*	*14*	*0*
		0	1.3	10	32.7	8	0
		0	**0.28**	**0.85**	**0.44**	**0.08**	**0**
Female	CR	*0*	*1*	*7*	*11*	*22*	*1*
		0	1	6.3	10	17	0.3
		0	**1.1**	**8**	**12**	**13**	**1.9**

†The three numbers in each cell represent the numbers of rats which died from a tumour (in italics) as assigned by pathologists, the corresponding estimates in Ahn *et al.* (2000) (in Roman) and the posterior expectations (6) (in bold).

the past 25 years. The aim here is not to give a systematic review of this literature, but only to establish some connections to work that has been done in the past.

A common factor in many of the applied papers is discretization of time, with the effect that the likelihood has a simple form. But there are also drawbacks: discrete time models lose information in the data by grouping survival times into a few intervals, although the times of death are usually recorded more accurately. Borgan *et al.* (1984) compared the efficiencies of some such observational plans.

McKnight and Crowley (1984) proposed a hazard model that is similar to our model 1. They discussed the identifiability of the hazards in the case of simple survival and serial sacrifice experiments. They also noted that a model in which the risk of death from a tumour depends on the tumour onset time is generally not identifiable from data where tumour onset times are not observed. The death hazards from competing risks before and after the tumour onset were not assumed equal as here, but their estimation without additional information such as the cause of death, or further restrictions on the model, is not possible. Their model cannot easily be estimated in a general situation where the number of tumours can be greater than 1 and the tumour onset times are unknown: in the case of two tumour types the number of hazard rates to be estimated would be 8 (which is the number of possible moves in Fig. 1 when aggregating the

three causes of death). Here, by assumptions (a) and (b) in Section 2.1, the number of hazard rates is 5. Parise *et al.* (2001) used the ideas of McKnight and Crowley (1984) in a discrete time model which was parameterized by prevalence and tumour lethality $l(t)$ functions of subject's age t; see Section 2.3. The main difference between this work and that of Parise *et al.* (2001) is that here the tumour onset and hazard-of-death rates were estimated directly. In this sense we follow McKnight and Crowley (1984) who proposed that the tumour onset hazards should provide the basis for inference.

These identifiability and estimation problems can be overcome by making some restrictive assumptions in the model. For example, Ahn *et al.* (2000) imposed independence assumptions that were similar to our assumptions (a) and (b) in Section 2.1 by analysing the occurrence of the two tumour types separately (i.e. by including the other type of tumour as a part of the competing risks). Dinse (1991) proposed that the presence of a tumour increases the risk of death by multiplying the death without a tumour hazard by a (possibly rat–age-dependent) parameter, or by adding such a parameter to it. The additive model in Dinse (1993) was found to perform very well in mice, and no significant mouse–age dependence was found.

Bayesian inference, latent variables and MCMC methods have also been used in incidence analysis before: Dunson and Dinse (2001) studied a parametric discrete time model in which the dependence of tumour incidence and mortality on rat age were modelled by a linear term (although their model could handle non-linearities also, e.g. by using polynomials). They augmented a latent individual indicator variable for the cause of death and used the Gibbs sampler, whereas here the cause of death is estimated by expression (5). If they had applied a more flexible model like that in Parise *et al.* (2001), sampling from the posterior distribution by using a single-component Gibbs sampler might give poor mixing for reasons that are discussed in Appendix A. The tumour onset time intervals were augmented, and in this sense the approach of Dunson and Dinse (2001) is similar to ours.

Ahn *et al.* (2000), which is the main reference in this work, studied the risk of dying from a tumour *versus* dying from other causes, by using a discrete time model. If a dead rat had a tumour, the cause of death was considered unknown, and the proportion of deaths was estimated by the model and afterwards compared with the fatality or incidence information that was provided by pathologists. The time to tumour onset was unknown; thus they introduced a model for that as well. Since a rat can die from a tumour only if it has one, Ahn *et al.* (2000) assumed that the values of the 'SF of tumour onsets' are smaller than those of the 'SF of deaths from tumour'. Their model has some weaknesses.

(a) The time is discretized and so the drawbacks that were mentioned above apply also to the method of Ahn *et al.* (2000).
(b) The simulation study in Ahn *et al.* (1999) indicates that there is some bias in their estimator for tumour lethalities, especially if the lethality is low. Their method also seemed to have difficulties in distinguishing low and intermediate levels of tumour lethality.
(c) The concept of latent cause-specific failure times can be criticized in several ways (e.g. Kalbfleisch and Prentice (1980), pages 172–175).
(d) The SF of tumour onsets does not have a probabilistic interpretation, as noted in Section 2.3.

This study aimed to overcome such restrictions by using Bayesian intensity models which provide a unified framework for analysing various multistate models. The data augmentation method that was used in this work would have allowed us to consider various censoring protocols. With reference to criticisms (a)–(d) above, we note the following.

(a) The exact times of death were used in modelling, and therefore no observed information was lost. Also, if there were more accurate observations on some rats, e.g. in the form of cause-of-death information, then the results could be refined by including this information in the analysis.
(b) Although the posterior expectations of the death from tumour hazard rates did not match exactly with the true hazards (as demonstrated by the simulation study in Appendix B), there was no difficulty in distinguishing low, intermediate and high tumour lethalities. The results here appeared to provide reasonably accurate results according to the simulation study, although interval censoring of tumour onset times seems to erode the quality of the data, as well as the estimates of unknown causes of death.
(c) The difficulty is solved here by the intensity model which does not need cause-specific times of death.
(d) The hazard rate describing the distribution of tumour onset times has a clear probabilistic interpretation unlike the SF of tumour onsets: if a rat is alive at age t and has no tumour, the hazard rate multiplied by a small positive constant dt is approximately the probability that the rat develops a tumour before age $t + dt$.

Model 2 was constructed using a strong prior assumption about the form of dependence of the risk of death from a tumour on the age of the tumour. When the tumour prevalences and tumour lethalities based on model 2 were compared with those obtained from model 1, the differences were quite small. Moreover, the results were not particularly sensitive to the choice of γ_y, which corresponds to the poor identifiability of death from tumour risk when considered as a function of tumour age.

The intensity models that are presented in this paper combined ideas of McKnight and Crowley (1984) and Dinse (1991) for handling the various problems that are posed by serially sacrificed data in an assessment of tumour incidences, prevalences and lethalities. All the observed information was used in the analysis, and no discretization of times of death was necessary. Using a continuous time model in an analysis is conceptually attractive because discretization of the death (and tumour onset) times is always based on some artificially chosen and often small set of time points. The predictive survival functions and tumour prevalences fit reasonably well to the observations that were presented by the Kaplan–Meier estimates and the observed prevalences in the sacrificed rats. Models 1 and 2 suggested that PIT tumours are less fatal at later ages than the results in Ahn *et al.* (2000) indicated: our result was found to be in good agreement with the prevalence data.

Acknowledgements

The prompt provision of the data by Hongshik Ahn, Thomas Bucci and Ralph Kodell is gratefully acknowledged. We also thank Timo Hakulinen, Marja Mutanen and two referees for their comments on our manuscript.

Appendix A: On the proposal distributions of the Markov chain Monte Carlo sampler

The interval-censored tumour onset times were augmented following the principles of Tanner and Wong (1987). In their method the model parameters and the missing data values were updated separately at every iteration of the MCMC sampler. In the posterior step the model parameters were assigned new values by sampling from the full conditional distribution given the current values of the missing data and the observed data. Then, in the imputation step the missing data values were augmented by sampling from the likelihood given the current values of the parameters.

Single-component updating which is often used in MCMC algorithms did not suffice here: the imputed tumour onset times and the tumour onset hazards $f_y^T(\cdot)$ (where y was MCL or PIT) were strongly dependent. Therefore integration of the imputation and posterior steps of Tanner and Wong (1987) was necessary. The Metropolis–Hastings algorithm was used for group updating both the tumour onset hazard rates and the missing tumour onset times because sampling directly from their joint distribution seemed virtually impossible.

Two group updating proposals in the Metropolis–Hastings algorithm were applied at every iteration: in the first proposal scheme, with probability 0.5 (scheme A) the onset times $(T_{y,i}^T)_{i\in\mathcal{I}_x}$ were moved closer to the times of death and the base-line hazard $f_y^T(t|X_i=x)$ was made to increase more steeply. In the reverse move, (scheme B) the onset times were moved away from the times of death (closer to the times of birth) and the base-line hazard was made to decrease more steeply. To be precise, let k denote the kth iteration of the MCMC algorithm, and let $l^{[k]} \sim \text{uniform}\{1, 2, \ldots, 5\}$. In scheme A the proposal onset time was

$$T_{y,i}^{T*} \sim \text{uniform}(\mathfrak{T}_{l^{[k]}}[k], T_i^{C\vee D}) \qquad \text{for } i\in\mathcal{I}_x \text{ and } T_{y,i}^T[k] < \mathfrak{T}_{l^{[k]}}[k] < T_i^{C\vee D},$$

where $\mathfrak{T}_{l^{[k]}}[k]$ was the lth $[k]$ jump point of $f_y^T(t|X_i=x)[k]$. The proposal levels a_j^* of $f_y^T(t|X_i=x)^{[k]}$ were deterministic:

$$a_j^* := \begin{cases} \frac{1}{5}a_j[k] & \text{if } j < l^{[k]}, \\ \frac{3}{2}a_j[k] & \text{if } j \geqslant l^{[k]}. \end{cases}$$

In scheme B the proposal was almost the opposite:

$$T_{y,i}^{T*} \sim \begin{cases} \text{uniform}(0, \mathfrak{T}_{l^{[k]}}[k]) & \text{with probability 0.1,} \\ \mathbb{1}_{\{T_{y,i}^T[k]\}}(\cdot) & \text{with probability 0.9} \end{cases}$$

for $i\in\mathcal{I}_x$,

$$a_j^* := \begin{cases} 5a_j[k] & \text{if } j < l^{[k]}, \\ \frac{2}{3}a_j[k] & \text{if } j \geqslant l^{[k]}. \end{cases}$$

The second proposal scheme was more random. The proposals for the levels $(a_j)_j$ of f_y^T were generated from gamma distributions: $a_0^* \sim \text{gamma}(1, 10)$ and $a_j^* \sim \text{gamma}(1, 1)$ for $j > 0$. The tumour onset times were then updated by generating the proposals $T_{y,i}^{T*}$ from the truncated distribution

$$\mathbb{P}(T_{y,i}^{T*} \mid f_y^{T*}, \{T_{y,i}^T < T_i^{C\vee D}\}).$$

The corresponding proposal for the death from tumour was

$$f_y^{D*} \sim \text{uniform}\{\cdot \mid (1-\varepsilon)\hat{f}_y^D, (1+\varepsilon)\hat{f}_y^D\}$$

where $\hat{f}_y^D$ was the maximum likelihood estimate given $(T_{y,i}^{T*})$, $f^C(\cdot)[k]$ and $f_{y'}^D[k](y' \neq y)$, and $\varepsilon > 0$. As this scheme generates values more evenly on the parameter space, the acceptance probabilities were found to be smaller than when following the first proposal scheme.

Appendix B: A simulation study

To evaluate the performance of our estimation method in a situation where the true parameter values are known, 851 simulated life histories divided into four groups were generated from model 1 with a similar sacrificing protocol to that in the real data: 12 'rats' were sacrificed at ages $t_1, \ldots, t_5$ in each group.

The true hazard rates were chosen so that they would demonstrate the functionality of the model in different cases: in the first group (male, AL) the risks of death from a tumour were assumed to be low. In the second (male, CR) and the third (female, AL) groups the risks were more intermediate, but in the third group the MCL tumour onset hazard rate had larger values between ages 1 and 2 years. In the fourth group (female, CR) the death from the MCL hazard rate had very large values.

Some frequentist ideas were applied in summarizing these results. This seemed natural because in a computer simulation a large number of random samples are actually drawn from a known model with known parameter values. 100 data sets were generated ($l = 1, \ldots, 100$), and the posterior expectations of the parameters were calculated in each case by running 9000 iterations of the MCMC sampler. The convergence of the risk of death from tumours was found to be satisfactory using CODA; see Best *et al.* (1996). In the case of one group, Fig. 6 shows the true hazard rates, the means of the posterior expectations from those 100 samples and the 90% pointwise confidence intervals, using directly the 5% and 95% quantiles

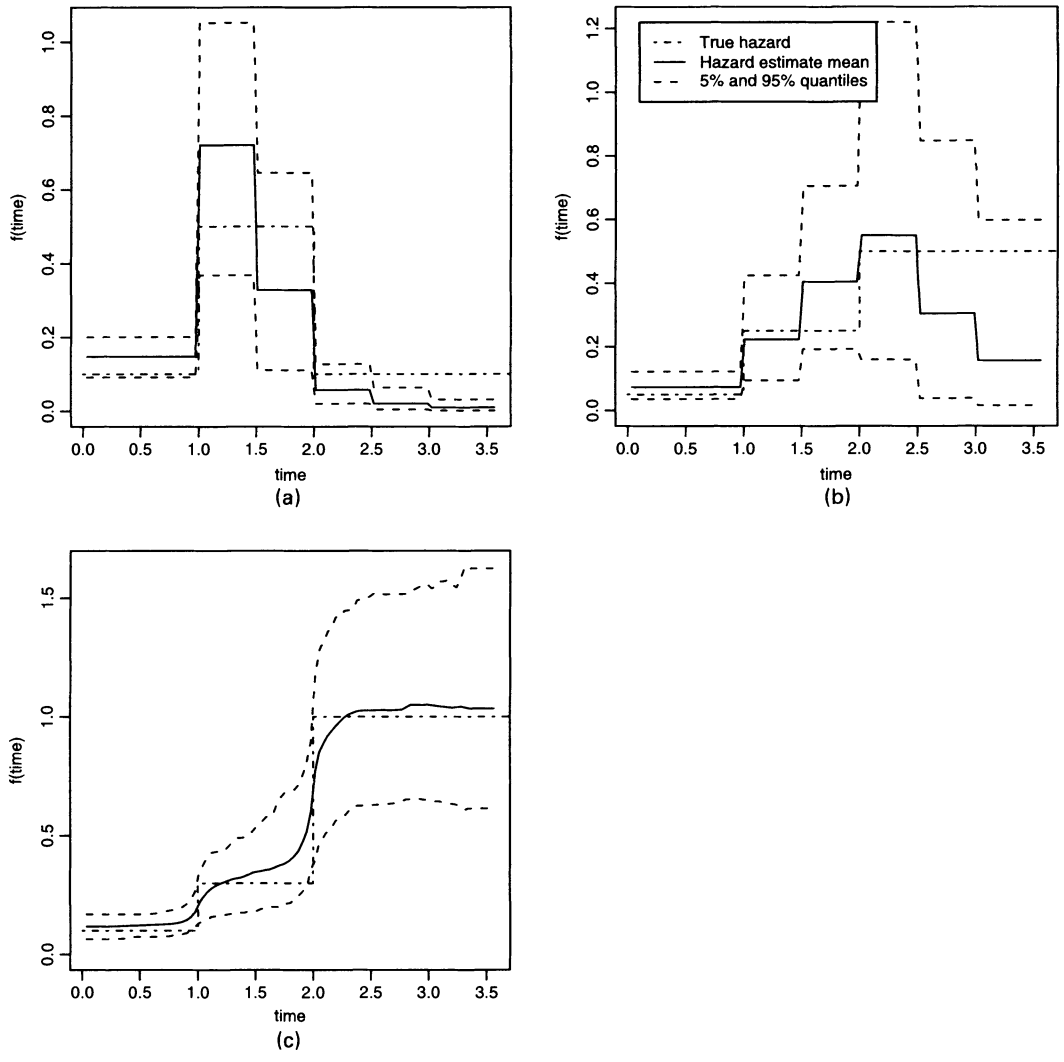

Fig. 6. Pointwise means (———) and 90% confidence intervals (-------) of the posterior expectations, and the true values of the base-line hazard rates (·····) in the simulation study in the group of female rats on the AL diet (model 1): (a) f^{MCL}; (b) f^{PIT}; (c) f^{C}

of the empirical distribution of the 100 computed posterior expectations. This was a typical result here: the posterior expectations of the tumour onset hazards matched quite well with the true values, although the estimated time dependences were not quite correct. The posterior expectations of the competing risks hazards fitted better to the true values. Table 4 shows that in group 1 the posterior expectations of death from tumour hazard rates seemed to be fairly accurate, but the tumour lethalities were slightly overestimated. In groups 2 and 3 no systematic error was found, but in group 4 the high MCL tumour lethalities were underestimated and PIT lethalities slightly overestimated.

A final comparison was made between the simulated 'true' number $r_{yjk}^{[l]}$ of deaths from a tumour, and the corresponding estimated numbers were obtained by the method of Ahn *et al.* (2000) and our method using expression (6), denoted respectively by $\hat{r}_{yjk}^{[l]}$ and $\tilde{r}_{yjk}^{[l]}$:

$$\hat{\Delta}_{yjk}^{[l]} := |r_{yjk}^{[l]} - \hat{r}_{yjk}^{[l]}|,$$
$$\tilde{\Delta}_{yjk}^{[l]} := |r_{yjk}^{[l]} - \tilde{r}_{yjk}^{[l]}|$$

Table 4. Means, and 90% confidence intervals of the posterior expectations of death from tumour hazard rates f_y^D of the study of 100 simulated data sets (model 1)

Group	*Gender*	*Diet*	*Results for MCL tumours*			*Results for PIT tumours*		
			Mean	*(Lower, Upper)*	*True value*	*Mean*	*(Lower, Upper)*	*True value*
1	Males	AL	0.098	(0.0127, 0.273)	0.05	0.121	(0.0246, 0.302)	0.05
2	Males	CR	0.332	(0.0569, 0.789)	0.25	0.345	(0.0464, 0.928)	0.25
3	Females	AL	1.02	(0.598, 1.54)	1.5	0.327	(0.0477, 0.965)	0.25
4	Females	CR	2.05	(0.741, 3.34)	5	0.875	(0.0686, 2.25)	0.5

Table 5. Average pairwise differences $\Sigma_l \tilde{\Delta}^{[l]}_{yjk}/100$ of the true and expected numbers of deaths from tumours (and with respect to the posterior model 1, in bold), and the corresponding averages $\Sigma_l \hat{\Delta}^{[l]}_{yjk}/100$ based on the estimator of Ahn *et al.* (2000) in the simulation study (model 1)

Gender	*Diet*	*Differences for the following age intervals (days):*					
		0–368	*369–555*	*556–754*	*755–919*	*920–1096*	*1097–1293*
MCL							
Male	AL	**0.29**	**0.70**	**1.12**	**1.41**	**1.45**	**1.26**
		0.41	0.99	2.56	3.03	4.37	4.80
Male	CR	**0.09**	**0.58**	**1.36**	**1.84**	**1.71**	**1.39**
		0.04	0.55	1.74	3.21	2.94	1.84
Female	AL	**0.64**	**1.72**	**2.78**	**1.92**	**1.04**	**0.54**
		0.97	2.82	2.34	1.66	0.91	0.57
Female	CR	**0.03**	**1.01**	**1.71**	**4.25**	**2.58**	**0.94**
		0.03	0.90	1.46	3.41	2.23	0.91
PIT							
Male	AL	**0.22**	**0.48**	**0.72**	**1.12**	**1.10**	**1.19**
		0.22	0.65	1.59	2.36	2.71	2.65
Male	CR	**0.12**	**0.48**	**0.97**	**1.23**	**1.20**	**1.02**
		0.03	0.34	1.16	2.45	2.10	1.40
Female	AL	**0.53**	**0.87**	**1.54**	**1.40**	**1.15**	**0.89**
		0.55	1.39	2.26	2.25	1.78	1.11
Female	CR	**0.10**	**0.63**	**0.97**	**1.34**	**0.96**	**0.60**
		0.03	0.67	1.20	2.09	1.66	0.78

where the tumour type is $y \in \{\text{MCL, PIT}\}$, group $j = 1, \ldots, 4$ and sacrifice interval $k = 1, \ldots, 6$. Table 5 presents the average pairwise differences $\Sigma_l \tilde{\Delta}^{[l]}_{yjk}/100$ and $\Sigma_l \hat{\Delta}^{[l]}_{yjk}/100$. Our method seems to work better in most cases, and even in the cases of high tumour lethalities (MCL tumour in groups 3 and 4) the method of Ahn *et al.* (2000) does not work much better than ours. For an overall comparison, we computed the pairwise differences $\Sigma_{yjk}(\hat{\Delta}^{[l]}_{yjk} - \tilde{\Delta}^{[l]}_{yjk})$ from the 100 simulations. Their mean was 24.9 and range (1.6, 46), corresponding to an approximate p-value of 10^{-15} in a t-test with the null hypothesis of no difference in the expected performance.

References

Ahn, H., Kodell, R. L. and Moon, H. (1999) Attribution of tumor lethality in the absence of cause-of-death information. *Technical Report SUNYSB-AMS-99-02*. Department of Applied Mathematics and Statistics, State University of New York at Stony Brook, Stony Brook.

Ahn, H., Moon, H. and Kodell, R. L. (2000) Attribution of tumour lethality and estimation of the time to onset of occult tumours in the absense of cause-of-death information. *Appl. Statist.*, **49**, 157–169.

Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1993) *Statistical Models based on Counting Processes*. New York: Springer.

Arjas, E. and Gasbarra, D. (1994) Nonparametric Bayesian inference from right censored survival data, using the Gibbs sampler. *Statist. Sin.*, **4**, 505–524.

Best, N. G., Cowles, M. K. and Vines, S. K. (1996) *CODA, 0.30 Version*. Cambridge: Medical Research Council Biostatistics Unit.

Borgan, Ø., Liestøl, K. and Ebbesen, P. (1984) Efficiencies of experimental designs for an illness-death model. *Biometrics*, **40**, 627–638.

Dinse, G. E. (1991) Constant risk differences in the analysis of animal tumorigenicity data. *Biometrics*, **47**, 681–700.

Dinse, G. E. (1993) Evaluating constraints that allow survival-adjusted incidence analyses in single-sacrifice studies. *Biometrics*, **49**, 399–407.

Dunson, D. B. and Dinse, G. E. (2001) Bayesian incidence analysis of animal tumorigenicity data. *Appl. Statist.*, **50**, 125–141.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (1995) *Bayesian Data Analysis*. London: Chapman and Hall.

Geweke, J. (1992) Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In *Bayesian Statistics 4* (eds J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith), pp. 169–193. Oxford: Clarendon.

Greenland, S. and Robins, J. M. (1988) Conceptual problems in the definition and interpretation of attributable fractions. *Am. J. Epidem.*, **128**, 1185–1197.

Härkänen, T. (2003) BITE: Bayesian intensity estimator. *Comput. Statist.*, **18**, 565–583.

Härkänen, T., Virtanen, J. I. and Arjas, E. (2000) Caries on permanent teeth: a nonparametric Bayesian analysis. *Scand. J. Statist.*, **27**, 577–588.

Heidelberger, P. and Welch, P. D. (1983) Simulation run length control in the presence of an initial transient. *Oper. Res.*, **31**, 1109–1144.

Kalbfleisch, J. D. and Prentice, R. L. (1980) *The Statistical Analysis of Failure Time Data*. New York: Wiley.

McKnight, B. and Crowley, J. (1984) Tests for differences in tumor incidence based on animal carcinogenicity. *J. Am. Statist. Ass.*, **79**, 639–648.

Parise, H., Dinse, G. E. and Ryan, L. M. (2001) Flexible estimates of tumour incidence for intermediately lethal tumours in a typical long-term animal bioassay. *Appl. Statist.*, **50**, 171–185.

Tanner, M. A. and Wong, W. H. (1987) The calculation of posterior distributions by data augmentation. *J. Am. Statist. Ass.*, **82**, 528–550.

Thurman, J. D., Bucci, T. J., Hart, R. W. and Turturro, A. (1994) Survival, body-weight, and spontaneous neoplasms in ad libitum-fed and food-restricted Fischer-344 rats. *Toxicol. Pathol.*, **22**, 1–9.