Appl. Statist. (2006)
55, *Part* 1, *pp.* 123–138

Assessing Poisson variation of intestinal tumour multiplicity in mice carrying a Robertsonian translocation

Michael A. Newton

University of Wisconsin, Madison, USA

and David I. Hastie

University of Bristol, UK

[Received May 2004. Final revision May 2005]

Summary. Tumour multiplicity is a frequently measured phenotype in animal studies of cancer biology. Poisson variation of this measurement represents a biological and statistical reference point that is usually violated, even in highly controlled experiments, owing to sources of variation in the stochastic process of tumour formation. A recent experiment on murine intestinal tumours presented conditions which seem to generate Poisson-distributed tumour counts. If valid, this would support a claim about mechanisms by which the adenomatous polyposis coli gene is inactivated during tumour initiation. In considering hypothesis testing strategies, model choice and Bayesian approaches, we quantify the positive evidence favouring Poisson variation in this experiment. Statistical techniques used include likelihood ratio testing, the Bayes and Akaike information criteria, negative binomial modelling, reversible jump Markov chain Monte Carlo methods and posterior predictive checking. The posterior approximation that is based on the Bayes information criterion is found to be quite accurate in this small n case-study.

Keywords: Cancer biology; Model averaging; Model choice; Multiple intestinal neoplasia; Negative binomial; Reversible jump Markov chain Monte Carlo methods; Tumour initiation

1. Introduction

The Poisson distribution naturally models the number of cancerous tumours that appear in a tissue during a fixed time period (e.g. Moolgavkar and Knudson (1981) and Kokoska (1987)): basically, there are many cells in the tissue and each has a small probability of becoming cancerous. However, extra-Poisson variation is widely observed in this sort of tumour count data (e.g. Drinkwater and Klotz (1981), Moser *et al.* (1990, 1992), Gould *et al.* (1996), Nagase *et al.* (1999), Chulada *et al.* (2000) and Ramachandran *et al.* (2002)). The overdispersion that is observed in many studies is not readily explained by environmental or genetic factors, since these factors are often well controlled; rather, the overdispersion is more likely attributable to sources of variation that are intrinsic to the basic process of tumour formation. For example, there may be different cellular or molecular mechanisms that mediate the same required set of genetic changes by which a normal intestinal epithelial cell becomes a tumour. Some of these mechanisms may be affected by factors that are localized within or near the cell, but others may be systemic,

Address for correspondence: Michael A. Newton, Department of Statistics, University of Wisconsin–Madison, 1300 University Avenue, Madison, WI 53706, USA.
E-mail: newton@stat.wisc.edu

 0035–9254/06/55123

inducing correlation between widely separated cells and producing extra-Poisson variation in the measured tumour count.

Those who carry a defective copy of the adenomatous polyposis coli gene (Apc in mice; APC in humans) are predisposed to intestinal cancer (e.g. Hardy *et al.* (2000)). The multiple intestinal neoplasia (MIN) strain of the laboratory mouse carries a nonsense Apc allele, and, consequently, develops tumours throughout the intestinal tract (Moser *et al.*, 1990; Su *et al.*, 1992) and thus provides an experimental system to study the biology and genetics of normal intestinal tissue and its neoplastic transformation (Dove *et al.*, 1998). As in humans, the murine intestinal tract comprises a large number of organized groups of proliferating cells called crypts, and it is from within aberrant crypts that tumours begin to form (Li *et al.*, 1994; Preston *et al.*, 2003).

A Poisson distribution of tumour counts would emerge if after the post-natal establishment of the crypt layer each crypt becomes aberrant with some small probability and does so independently of other crypts (Simon and Gordon, 1995). Considering Poisson approximation theory (Arratia *et al.*, 1990), we expect Poisson variation to hold even if there is a small degree of positive correlation between crypts, but if the dependence becomes substantial then the Poisson approximation breaks down and we would see extra-Poisson variation. Biological factors which act in a spatially localized way (e.g. within crypts or among only a few neighbouring crypts) are expected, therefore, to produce a Poisson tumour count distribution.

Haigis and Dove (2003) measured tumour multiplicities in four groups of genetically identical MIN mice: one control group and three groups which carried different forms of the Robertsonian translocation (Rb9). This is a chromosomal construct in which certain pairs of chromosomes are fused together. (The translocation alters the genome organization without changing its content.) The experimental mice, which carried a fusion of chromosome 7 with the Apc harbouring chromosome 18, were used primarily to study factors by which the wild-type Apc allele is inactivated within tumours; an important observation was that animals carrying an Rb9 translocation had significantly reduced tumour multiplicity, indicating that non-genetic chromosomal factors play a role in susceptibility to tumours. Haigis and Dove (2003) argued that the Rb9 translocation inhibits one mechanism of Apc inactivation, namely somatic recombination, but that it does not alter certain other mechanisms (chromosomal non-disjunction; gene silencing). Tumours formed in mice with the Rb9 translocation, but they must have done so through a reduced set of mechanistic pathways. A striking feature of the experimental data was that the tumour multiplicities in Rb9 animals appeared to exhibit purely Poisson variation. Never before in the extensive set of MIN mouse experiments had Poisson fluctuations been observed; such *random* variation would support the claim that these other mechanisms for Apc inactivation entail only spatially localized effects within the intestine, and further that the somatic recombination pathway entails non-localized effects. The Haigis–Dove experiment presented an intriguing interpretation of Apc inactivation: one based on positive evidence supporting the Poisson hypothesis. The purpose of the present paper is to examine this evidence statistically.

Table 1 reports the full set of tumour multiplicity data that were analysed by Haigis and Dove (2003). As in many mouse studies, the sample size here is modest; there are four groups of sizes 16, 17, 15 and 18. For notation we retain the group labels as in Haigis and Dove (2003) where the reader may find the precise definitions. Briefly, the '+/+' group is a control group in which animals do not carry any Rb9 translocation, and the other groups correspond to different ways that the fusion of chromosomes 7 and 18 can occur. Following a standard protocol, tumour counts were measured on the whole intestinal tract from animals that were sacrificed at 60 days of age. The significant reduction in tumour multiplicity in the Rb9 groups is evident from an inspection of Table 1 and this can be confirmed by a variety of statistical tests. (Haigis and Dove (2003) used the nonparametric Tukey–Kramer test.) The suitability of the Poisson hypothesis

Table 1. Data from the Haigis–Dove study†

Group	*Tumour multiplicities*	*Mean*	*Variance*
+/+	80, 103, 112, 121, 121, 121, 131, 140, 140, 150, 166, 169, 194, 199, 199, 262	150.5	2102.1
Rb9 trans	5, 7, 8, 8, 9, 9, 11, 12, 12, 13, 13, 13, 14, 15, 15, 16, 18	11.6	12.5
Rb9 cis	7, 7, 7, 8, 8, 8, 10, 10, 10, 10, 11, 11, 12, 12, 20	10.1	10.6
Rb9/Rb9	3, 4, 4, 5, 6, 6, 6, 6, 7, 7, 7, 9, 10, 10, 11, 11, 12, 15	7.7	10.3

†The rightmost columns are sample means and variances.

is suggested by the close agreement of sample means and sample variances in the Rb9 groups. We assess the evidence supporting this hypothesis.

2. Methods

2.1. Testing strategies

Experimentalists often adopt a hypothesis testing posture when evaluating data. A simple testing approach is to assume that the data are Poisson distributed, with group-specific means perhaps, and then to assess the evidence against this hypothesis. Conditioning on group sums to eliminate nuisance parameters, the Poisson counts become multinomials. We can consider the conditional distribution of some test statistic—for example each group provides a sample coefficient of dispersion (sample variance divided by sample mean)—and calibrate this test statistic by using Monte Carlo sampling (Barnard, 1963). This testing approach is straightforward but it does not quantify the positive evidence supporting the Poisson hypothesis. Non-rejection implies consistency of the hypothesis with the data and a paucity of contradictory evidence. The usual interpretation is that we should retain the null hypothesis, but we are faced with the fact that all historical data in the MIN system exhibit extra-Poisson variation. Taking the Poisson hypothesis as the null hypothesis seems to be an undue contrivance.

Professor Drinkwater (personal communication) suggested a different hypothesis testing strategy for the Haigis–Dove problem based on his experience that the negative binomial distribution often fits tumour count data (e.g. Drinkwater and Klotz (1981)). Indeed the negative binomial distribution is widely used to model count data (e.g. McCullagh and Nelder (1989)). Recall that this two-parameter family has probability mass function, for $x \geqslant 0$,

$$p(x|\kappa, \lambda) = \frac{1}{\kappa^{1/\kappa}\,\Gamma(1/\kappa)} \frac{\Gamma(x+1/\kappa)}{\Gamma(x+1)} \frac{\lambda^x}{(\lambda+1/\kappa)^{x+1/\kappa}}. \tag{1}$$

In this parameterization, the expected tumour count is λ and the variance is $\lambda(1+\kappa\lambda)$ for $\kappa \geqslant 0$. The limit $\kappa \to 0$ corresponds to the Poisson distribution. Recall also that, when viewing this distribution as a mixture of Poisson components, $1/\kappa$ is the shape parameter of the gamma mixing distribution (Greenwood and Yule, 1920). Drinkwater's proposal was to allow group-specific λ-values and to perform a likelihood ratio test for a common shape parameter (i.e. common κ). The rationale is that the shape parameter characterizes the degree of overdispersion and so testing for equal shapes would assess the significance of variation in the observed coefficients of dispersion. An advantage in using Drinkwater's approach is that we are not assuming the Poisson null hypothesis. However, the inference does not measure the positive evidence supporting the Poisson hypothesis. An alternative approach, which is not obvious to many experimentalists, is to eschew hypothesis testing and to frame the problem as one of model choice or model averaging.

2.2. Choice of model

We adopt the negative binomial model as a general description of tumour multiplicity. This is supported by historical data in the MIN mouse and also by goodness-of-fit diagnostic tests (Section 3.3). Various submodels are obtained by restricting the parameters; for instance setting $\kappa = 0$ corresponds to the Poisson case. To deal properly with the multiple groups, we introduce group labels $i = 1, \ldots, 4$ and allow the possibility that each group can have its own mean parameter λ_i and dispersion parameter κ_i. Therefore, the full model contains eight free parameters. Drinkwater's likelihood ratio test considered above compares this full model with the five-parameter null model in which $\kappa_1 = \kappa_2 = \kappa_3 = \kappa_4 \geqslant 0$ and the four means are unrestricted. The simple Monte Carlo tests work on the Poisson null hypothesis that $\kappa_i = 0$.

A host of different submodels emerges by considering equality constraints among $\{\lambda_i\}$, equality constraints among $\{\kappa_i\}$ and the possibility that each κ_i may equal 0. Table 2 notes the 15 possible equality patterns among the mean parameters; these correspond to the Bell number of set partitions of four objects (Bell, 1934). We do not tabulate them all, but we observe that likewise there are 52 distinguishable constraint patterns for the dispersion parameters. To see this, note that

$$52 = \sum_{j=0}^{4} \binom{4}{j} a_j \tag{2}$$

where the a_j are also Bell numbers: $(a_0, a_1, a_2, a_3, a_4) = (1, 1, 2, 5, 15)$. There are $\binom{4}{j}$ ways to choose j groups that will have non-zero κ_i-values; for each such choice there are a_j patterns of equality among the j non-zero values. Taken together, we have $780 = 15 \times 52$ distinct submodels of the general negative binomial model.

Table 2. Possible patterns of equality among components of $\boldsymbol{\lambda}$ across the four groups†

Index l	*l-pattern*				*Dimension* q_l
	$i=1$	$i=2$	$i=3$	$i=4$	
1	1	1	1	1	1
2	1	1	1	2	2
3	1	1	2	1	2
4	1	2	1	1	2
5	1	2	2	2	2
6	1	1	2	2	2
7	1	2	1	2	2
8	1	2	2	1	2
9	1	1	2	3	3
10	1	2	1	3	3
11	1	2	3	1	3
12	1	2	2	3	3
13	1	2	3	2	3
14	1	2	3	3	3
15	1	2	3	4	4

†Within each row, two entries are equal if and only if the corresponding group mean values are equal. The first column indexes the pattern and the last column gives the number of free parameters.

We label each submodel by $m = (l, k)$ where $l \in \{1, \ldots, 15\}$ is an index of the equality pattern among components of $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_4)$, and likewise $k \in \{1, \ldots, 52\}$ is an index for the equality and zero pattern for $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_4)$. By our coding, for example, $m = (12, 20)$ entails three free λ_is and two free non-zero κ_is. More specifically, the means satisfy $\lambda_2 = \lambda_3$, $\lambda_1 \neq \lambda_3$, $\lambda_1 \neq \lambda_4$ and $\lambda_3 \neq \lambda_4$ (this is $l = 12$); the dispersion parameters satisfy $\kappa_2 = \kappa_3 = 0$, $\kappa_1 > 0$, $\kappa_4 > 0$ and $\kappa_1 \neq \kappa_4$ (this is $k = 20$).

We can readily maximize the log-likelihood within each submodel and note the number of free parameters. These quantities can be combined in various ways to score a balance between model fit and model complexity: the Akaike information criterion AIC (Akaike, 1983) and the Bayes information criteria BIC (Schwarz, 1978; Smith and Spiegelhalter, 1980). Compared with testing strategies, the model choice approach more directly measures support for the Poisson hypothesis. But beyond ranking it does not quantify the extent of the favourable evidence.

2.3. Bayesian approaches

Going further, we pursue posterior analysis by using the negative binomial likelihood and prior distributions over the multitude of submodels. Our first calculation transforms the BIC-values into approximate submodel posterior probabilities (Kass and Wasserman, 1995). This does not change the ranking of submodels, because the approximate posterior mass for submodel m is

$$p(m|\text{data}) \propto \exp(\text{BIC}/2),$$

but it does allow us to calibrate the ranking and it enables the calculation of marginal posterior summaries such as the posterior probability of the Poisson hypothesis for each group.

The BIC-based posterior computations rely on large sample theory and ought to be viewed as a first approximation to a more explicit Bayesian analysis. Unsure of the degree of this approximation, we implement a fully specified Bayesian analysis via reversible jump (i.e. trans-dimensional) Markov chain Monte Carlo (MCMC) sampling. An advantage of this approach is that the issues of model choice and parameter estimation are dealt with simultaneously and there are no large sample approximations; a challenge is that care is required in setting the prior and in implementing the posterior sampling.

We consider a range of priors which for convenience share the factorization

$$p(m, \boldsymbol{\lambda}, \boldsymbol{\kappa}) = p(m)\ p(\boldsymbol{\lambda}|l)\ p(\boldsymbol{\kappa}|k) \tag{3}$$

but which differ in how they assign mass $p(m)$ to the submodels. One prior, denoted A, entails

$$\begin{aligned} p(m) &= p(l)\ p(k) \\ &= 1/15 \times 1/52 = 1/780 \end{aligned}$$

for all submodels m. This has the basic appeal of non-informativeness, but also it aims to mimic the previous BIC-based computations. The combinatorics of this uniform submodel prior imply $P(\kappa_i = 0) = 15/52 \approx 0.29$ for each i, and thus global uniformity does not confer a balance in this marginal prior probability. A second prior B is set up specifically to address the marginal Poisson hypothesis; it entails $\omega = P(\kappa_i = 0) = \frac{1}{2}$ for all i. To achieve this balance we retain independence of the pattern indices l and k and uniformity of l, but we modify $p(k)$ so that it boosts mass on the Poisson submodels. If the pattern k has j non-zero κ_i-values, then we take

$$p(k) = \omega^j (1 - \omega)^{4-j} / a_j$$

where a_j is the jth Bell number, as in equation (2). The thinking is that each group i tosses an ω-coin to decide whether or not it will be Poisson distributed; conditionally on the four coin

tosses there are j non-zero κ_i-values, and their internal pattern of equality constraints is selected uniformly from the a_j possibilities. Prior B corresponds to one case, $\omega = \frac{1}{2}$, in a continuum of hyperparameter settings. We also compute posterior summaries under the full range of ω-values. To complete the prior settings, we take vague independent gamma priors for the free parameters within each model. Specifically, free components of $\boldsymbol{\lambda}$ are regulated by a gamma(2,0.1) distribution and those of $\boldsymbol{\kappa}$ are regulated by a gamma(1,2) distribution. Posterior analysis indicates little sensitivity to these choices.

Posterior computation is done by MCMC sampling. Standard Metropolis–Hastings updates are available to sample the free parameters in $\boldsymbol{\lambda}$ and $\boldsymbol{\kappa}$ conditionally on the data and the submodel m (e.g. Robert and Casella (2002)); however, it is evident that different submodels can have different dimensions, and thus these basic updates are not suitable for updating m. To accommodate the varying dimensionality, we invoke a transdimensional MCMC sampler that allows jumping between submodels of different dimensions (Green, 1995, 2003). In contrast with many other applications of transdimensional MCMC sampling, this case is interesting because there is only a partial nesting of submodels. Furthermore, the presence of two pattern variables l and k in m means that the situation is slightly more general than the typical problem involving only a single model index. In updating m, we accommodate the partial nesting and cycle through two move classes: one proposes a new mean pattern l given the dispersion pattern k, and the other proposes a new k given l. The close linkage between m and $(\boldsymbol{\lambda}, \boldsymbol{\kappa})$ means that free-parameter values need to be created and destroyed during these proposals. Further details are in Appendix A.

The output from the MCMC algorithm provides direct access to properties of marginal posterior distributions. We can look not only at the probabilities of the pattern indices l and k and

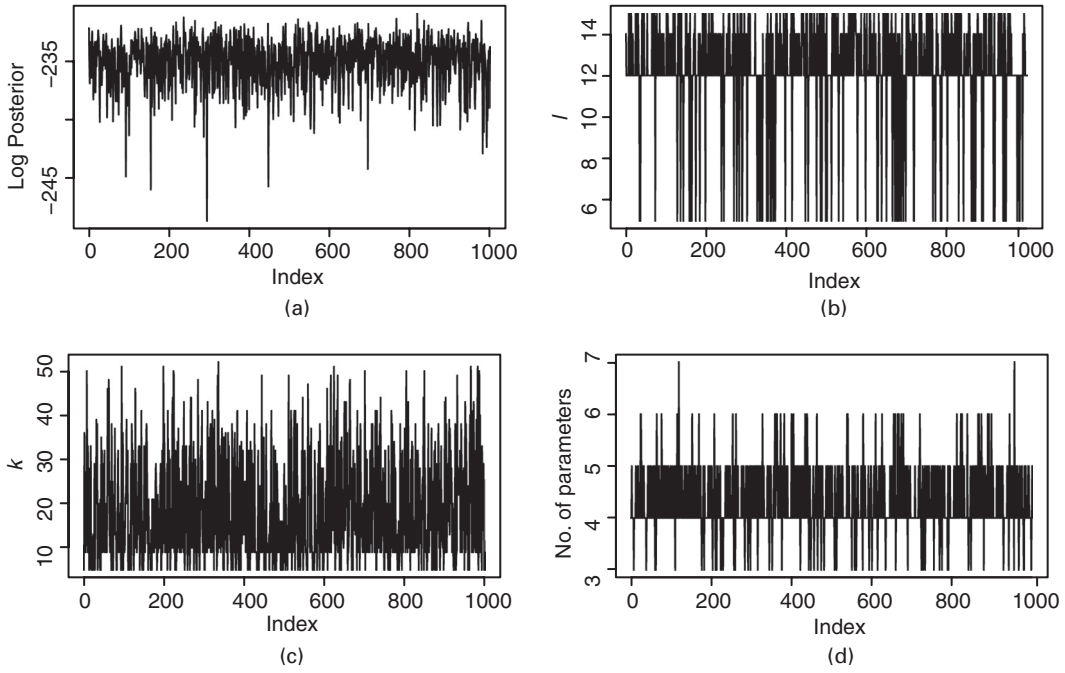

Fig. 1. Trace plots from MCMC runs (chains thinned to 1000 observations): (a) log-posterior value (prior A); (b) mean pattern index l (prior B); (c) dispersion pattern index k (prior A); (d) number of free parameters, $q_l + q_k$ (prior B)

various parameters, but we can also evaluate the marginal posterior probabilities of the Poisson hypothesis for each treatment group. Validity of these numerical summaries rests on properties of the chain and the quality of our implementation. Part of the MCMC output analysis is to study trace plots such as in Fig. 1, which help to confirm that our MCMC sampler is performing sufficiently. Our analysis indicates that the reported posterior probabilities are accurate to two decimal places.

Although the general negative binomial model fits well in historical data (e.g. Drinkwater and Klotz (1981)), its validity ought to be checked in the present case. Posterior predictive checks are suitable for this purpose (Gelman *et al.*, 2003). For each four-group data set that is simulated from the posterior predictive distribution, we find the best fitting negative binomial distribution in each group, and then we measure the maximum distance between each fitted distribution function and the corresponding empirical distribution.

3. Results

3.1. Testing strategies

We find via Monte Carlo testing that each of the Rb9 groups is consistent with the Poisson hypothesis but the +/+ group is not. At the same time, we find that the data are consistent with a common shape negative binomial hypothesis allowing group-specific means (Table 3). The testing approach is informative but it remains inconclusive as a way to quantify the evidence supporting the Poisson hypothesis.

3.2. Choice of model

Table 4 indicates the position of several submodels by the criteria BIC and AIC. Both criteria concur that, in the context of the general negative binomial model, the best explanation of the

Table 3. Testing results†

Group	*p-value*	*Results for the following hypotheses:*			
		Alternative		*Null*	
		$\hat{\kappa}$	L	$\hat{\kappa}$	L_0
+/+	0.001	0.076	−82.52	0.054	−82.94
Rb9 trans	0.37	0.001	−45.21	0.054	−45.99
Rb9 cis	0.38	0.000	−37.43	0.054	−38.21
Rb9/Rb9	0.15	0.033	−45.34	0.054	−45.41
			−210.50		−212.55

†The second column shows p-values from the Monte Carlo test of the Poisson null hypothesis. Subsequent columns give results of the likelihood ratio test for common κ. In each case the maximum likelihood estimate of the group-specific λ is the sample mean (see Table 1). Shown here are maximum likelihood estimates of the group-specific κ-values on the alternative hypothesis and the common estimate on the null hypothesis. The contribution to the log-likelihood is L for the unrestricted case and L_0 for the null hypothesis. The generalized likelihood ratio statistic is $4.1 = 2(212.55 - 210.50)$, which is not significant by standard assessments.

Table 4. Choice of model†

Submodel $m=(l,k)$	*l-pattern*				*k-pattern*				L	q	$BIC(r)$	$AIC(r)$
	$i=1$	$i=2$	$i=3$	$i=4$	$i=1$	$i=2$	$i=3$	$i=4$				
(12,9)	1	2	2	3	1	0	0	1	−211.62	4	1	1
(12,5)	1	2	2	3	1	0	0	0	−211.68	4	2	2
(12,13)	1	2	2	3	1	0	1	1	−212.43	4	3	6
(12,14)	1	2	2	3	1	1	0	1	−212.46	4	4	7
(15,9)	1	2	3	4	1	0	0	1	−210.71	5	8	3
(15,5)	1	2	3	4	1	0	0	0	−210.76	5	9	4
(12,20)	1	2	2	3	1	0	0	2	−211.42	5	14	5
(14,9)	1	2	3	3	1	0	0	1	−212.80	4	7	31
(14,5)	1	2	3	3	1	0	0	0	−213.31	4	11	35
(15,16)	1	2	3	4	1	1	1	1	−212.55	5	40	52
(15,52)	1	2	3	4	1	2	3	4	−210.50	8	160	91

†For a collection of 11 interesting submodels (out of 780), shown are the submodel index m and its components l, the mean pattern index, and k, the dispersion pattern index. The l-pattern is as in Table 2. The k-pattern is similar except that 0 indicates $\kappa_i = 0$ (Poisson variation). L is the maximized log-likelihood, q is number of free parameters and BIC(r) and AIC(r) are respectively the rank of the submodel in terms of BIC and AIC.

Haigis–Dove data entails Poisson variation and a common mean in Rb9 cis and Rb9 trans groups, and common shape but different means in the other two groups. Poisson variation in all Rb9 groups is a close second-place explanation for both criteria, though they balance the fit and complexity differently for other submodels.

3.3. Bayesian approaches

In the first calculation we transform the BIC-values into submodel posterior probabilities. The top two submodels from Table 4 each carry 10% of the posterior mass, the best 11 submodels account for half the posterior mass and 80% probability concentrates on the best 36 submodels. Again we can conclude that evidence favours Poisson variation in the Rb9 cis and Rb9 trans groups.

The full Bayesian analysis via MCMC sampling allows a range of inferences and also measures the accuracy of the BIC-based computation. Of interest to us is the marginal posterior distribution over the discrete set of submodels. Table 5 records the (non-zero) marginal posterior

Table 5. Full Bayesian approach†

Index l	*l-pattern*				*Posterior probability*	
	$i=1$	$i=2$	$i=3$	$i=4$	*Prior A*	*Prior B*
5	1	2	2	2	0.09	0.07
12	1	2	2	3	0.57	0.59
13	1	2	3	2	0.01	0.01
14	1	2	3	3	0.19	0.18
15	1	2	3	4	0.15	0.15

†Shown are mean patterns l with non-zero marginal posterior probabilities (see Table 2).

probabilities of the mean pattern index l. The modal pattern $l = 12$, which accounts for just over half the posterior mass under either prior, entails three different rates of occurrence of tumours: a common rate for Rb9 trans and Rb9 cis, and distinct rates for the other groups. Significantly, there is no posterior support for any of the values of l in which the control group +/+ shares a mean parameter with any of the other groups. This is consistent with earlier calculations indicating the strong effect of the Rb9 translocation on expected tumour count.

A similar analysis of the dispersion pattern k indicates that two patterns contain most of the posterior mass. They are $k = 5$, in which only the +/+ group is not Poisson, and $k = 9$, in which +/+ and Rb9/Rb9 share a non-zero κ-value and the other groups are Poisson. Other dispersion patterns are considerably less probable, although there is more mass outside the modal two values using prior A than using prior B.

Table 6 shows 11 submodels $m = (l, k)$ which include the top six submodels for priors A and B. Ranking is by marginal posterior probability. The top two submodels, $m = (12, 5)$ and $m = (12, 9)$, are the same as those selected by BIC and AIC (Table 4), although the order is reversed; further differences emerge as we look down the ranking. These top two submodels

Table 6. Full Bayesian approach†

Submodel index $m = (l, k)$	*l-pattern*				*k-pattern*				*Posterior probability*		*Ranking*	
	$i=1$	$i=2$	$i=3$	$i=4$	$i=1$	$i=2$	$i=3$	$i=4$	*Prior A*	*Prior B*	*Prior A*	*Prior B*
(12,5)	1	2	2	3	1	0	0	0	0.09	0.23	1	1
(12,9)	1	2	2	3	1	0	0	1	0.09	0.11	2	2
(14,9)	1	2	3	3	1	0	0	1	0.04	0.05	3	5
(15,5)	1	2	3	4	1	0	0	0	0.03	0.06	8	3
(12,20)	1	2	2	3	1	0	0	2	0.04	0.04	5	6
(14,5)	1	2	3	3	1	0	0	0	0.02	0.05	11	4
(12,13)	1	2	2	3	1	0	1	1	0.04	0.02	4	12
(15,9)	1	2	3	4	1	0	0	1	0.03	0.03	9	9
(12,14)	1	2	2	3	1	1	0	1	0.03	0.02	6	15
(15,16)	1	2	3	4	1	1	1	1	0.00	0.00	65	90
(15,52)	1	2	3	4	1	2	3	4	0.00	0.00	162	—

†Shown are posterior probabilities and rankings for 11 submodels containing the six most highly ranked submodels for both priors.

Table 7. Bayesian approach†

Method	+/+	$P(\kappa_i = 0 \mid data)$		
		Rb9 trans	*Rb9 cis*	*Rb9/Rb9*
BIC-approximation	0.00	0.58	0.57	0.37
Full, prior A	0.00	0.61	0.60	0.32
Full, prior B	0.00	0.80	0.80	0.53

†Shown are posterior probabilities that tumour multiplicities follow a Poisson distribution for each group. Recall that prior B has $P(\kappa_i = 0) = \frac{1}{2}$ and in other cases $P(\kappa_i = 0) = 15/52$.

account for 18% of the total posterior mass under the flat prior A and 34% under the Poisson-balanced prior B.

Of primary interest is the posterior probability of the Poisson hypothesis, i.e. $P(\kappa_i = 0|\text{data})$ for each group i. Table 7 shows these marginal probabilities for the full analysis via MCMC sampling using priors A and B, and for the BIC-approximation. The final evidence is relatively

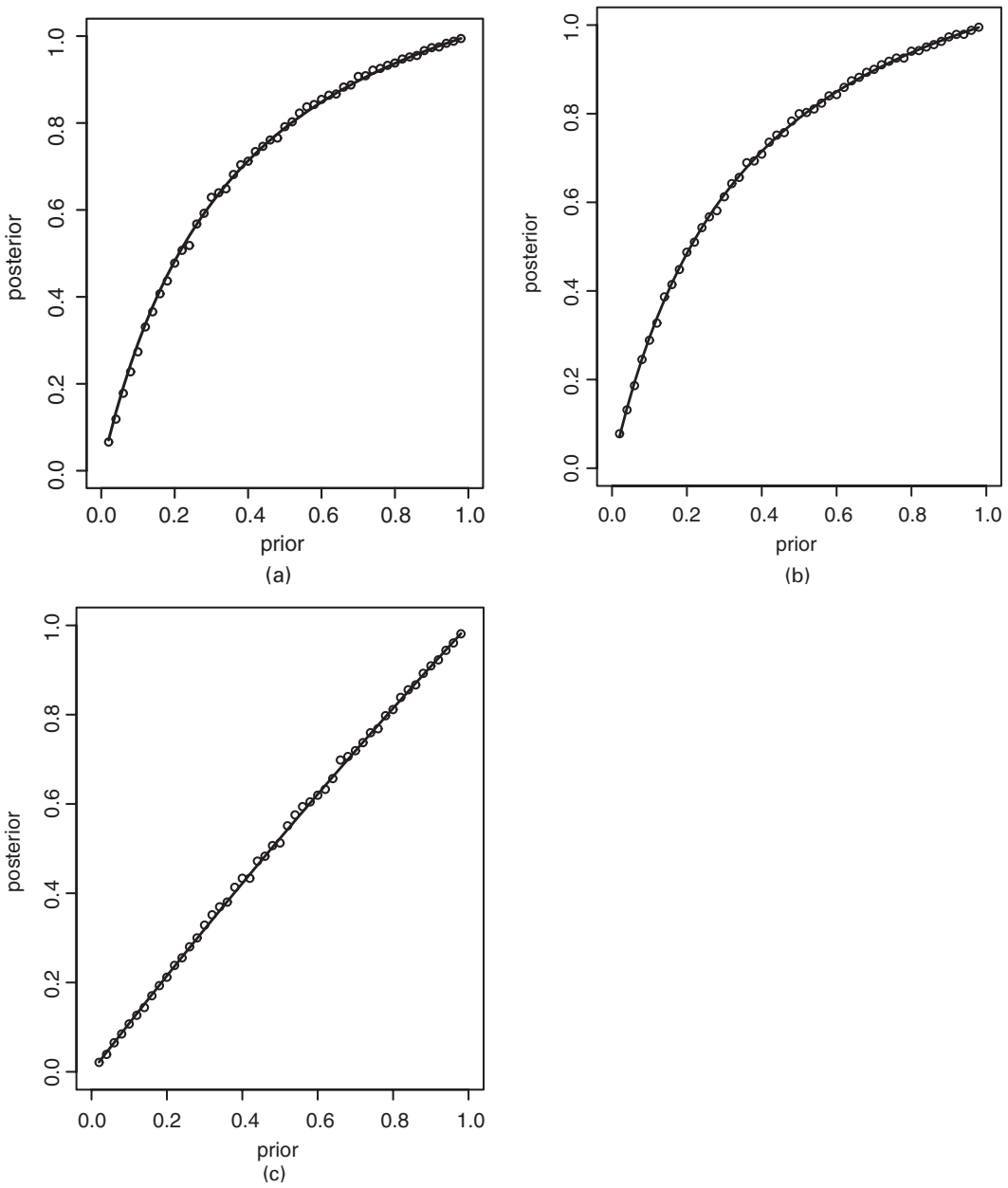

Fig. 2. Marginal posterior probability (∘) of the Poisson hypothesis *versus* prior probability on a grid of 48 prior values (———, constant Bayes factor BF): (a) Rb9 trans group, BF = 3.7; (b) Rb9 cis group, BF = 3.8; (c) Rb9/Rb9, BF = 1.1

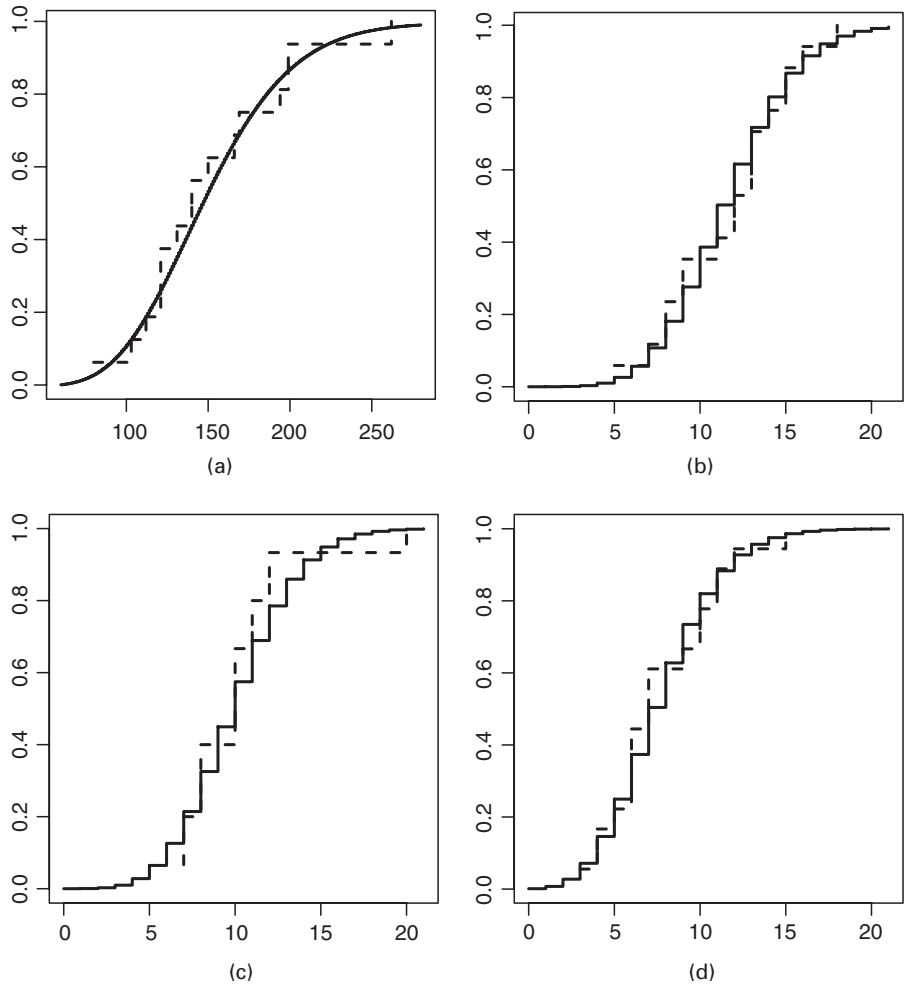

Fig. 3. Empirical distributions (⌴⌝) of tumour counts from Table 1 and best fitting negative binomial distributions (⌟⌜) (fitting was done by maximum likelihood separately in each group): (a) +/+ group; (b) Rb9 trans group; (c) Rb9 cis group; (d) Rb9/Rb9

weak, but all calculations support the Poisson model for Rb9 trans and Rb9 cis. The support is positive and in favour of the Poisson hypothesis, with posterior probability 0.6 and 0.8 for priors A and B respectively. We find that the Bayes factor

$$\mathrm{BF}_i = \frac{P(\kappa_i = 0|\mathrm{data})/P(\kappa_i > 0|\mathrm{data})}{P(\kappa_i = 0)/P(\kappa_i > 0)}$$

is insensitive to the prior setting; $\mathrm{BF}_i \approx 4$ for Rb9 trans and Rb9 cis, and $\mathrm{BF}_i \approx 1$ for Rb9/Rb9. Interestingly, the stability of the Bayes factor remains evident when we rerun the MCMC calculations over a grid of 49 different prior settings for $P(\kappa_i = 0)$ in (0.02, 0.98) (Fig. 2).

To assess the goodness of fit of the general negative binomial model we consider posterior predictive checks. Fig. 3 compares the empirical distribution of tumour counts within each group to a best fitting negative binomial model for that group. To calibrate these fits, we compare Fig. 3 with similar results for 1000 data sets simulated from the posterior predictive distribution. The 1000 parameter settings were taken by thinning one of the MCMC runs under prior A. The

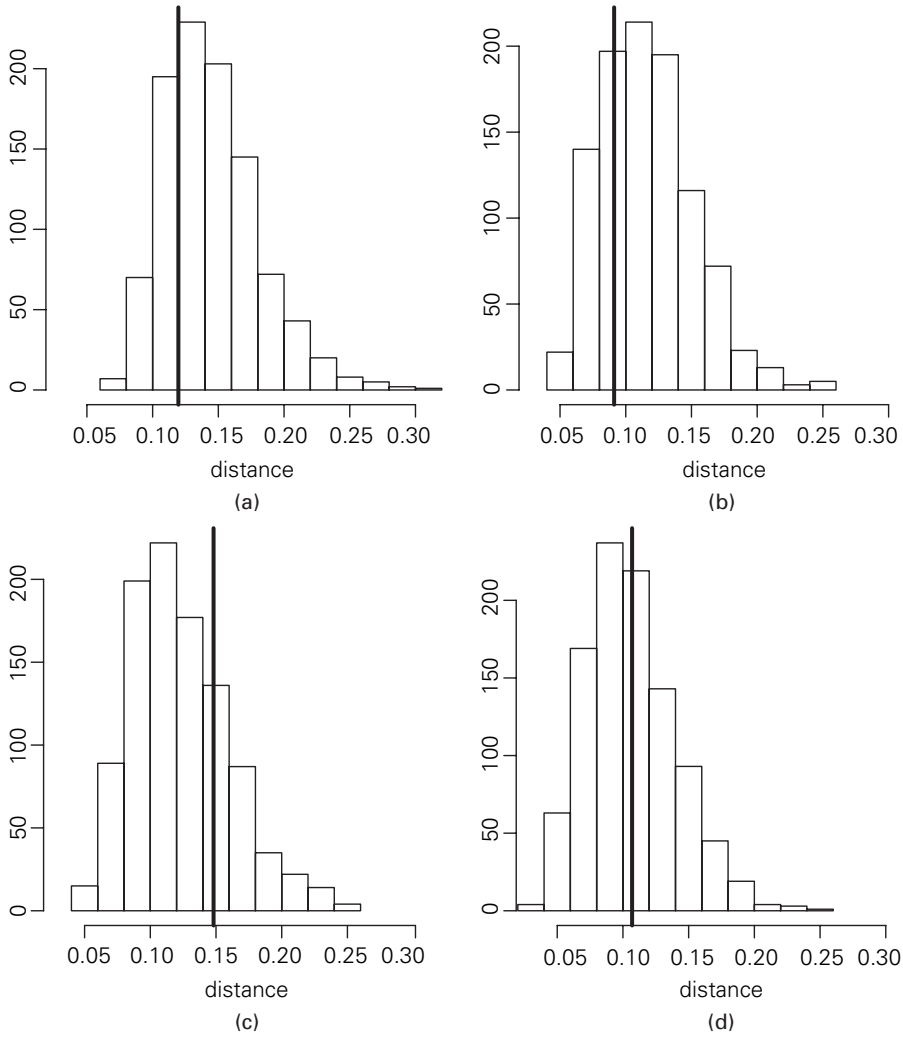

Fig. 4. Posterior predictive checks using as test statistic the maximum difference between predicted sample empirical and best fitting negative binomial models, as in Fig. 3 (, values for observed data, and thus we have no evidence against the general negative binomial model): (a) +/+ group; (b) Rb9 trans group; (c) Rb9 cis group; (d) Rb9/Rb9

results in Fig. 4 indicate an adequate fit of this general model and thus support the model-based computations.

4. Discussion

The biological mechanisms that regulate the intestinal epithelium must co-ordinate the continual proliferation of tens of millions of cells. We know that tumours form when this regulation fails, and yet much remains unknown about the early stages of tumour formation. A necessary early genetic step in the process is loss of the APC or Apc gene product. One pathway to Apc inactivation (somatic recombination) was closed in the Rb9 animals of the Haigis–Dove experiment, and so their tumours were obliged to lose Apc by some other mechanism (e.g. chro-

mosomal non-disjunction or gene silencing). By our statistical analysis we have measured the distribution of the resulting tumour counts and have found positive evidence supporting Poisson variation. Poisson variation is predicted if the Apc inactivation mechanisms are spatially localized in the intestine, rather than involving systemic factors which would induce extensive spatial correlation and extra-Poisson variation. Indeed it is difficult to see how Poisson variation could occur in the absence of spatial independence (or at least weak spatial dependence), considering various characterizations of the Poisson distribution in terms of weakly dependent Bernoulli or count variables (Arratia *et al.* (1990), theorem 1, and Daley and Vere-Jones (1988), theorem 2.2).

The statistical tools that were employed in this case-study include hypothesis testing, model selection, model averaging and posterior predictive checking. Though they vary in terms of computational demands and the nature of their conclusions, all the approaches are either consistent with or directly support the hypothesis that Rb9 data are Poisson distributed. This case-study provides a context for comparing different statistical approaches. We find limitations of classical hypothesis testing. We find that different model choice schemes can agree on the best explanation of data and can provide positive support for certain hypotheses. The most extensive inferences are provided by a full model-based Bayesian analysis. As expected by theory (e.g. O'Hagan (1994)), the flat prior calculations are in closer agreement with the BIC-approximation than those for the prior which distributes mass unequally to control the marginal Poisson probability. Noting the small sample size, a striking feature in this case-study is the quality of the asymptotic BIC-approximation. We gain insights into the use and effectiveness of fully specified Bayesian analysis fitted with transdimensional MCMC methods. Though it is more computationally demanding than the other approaches and requires more detailed specification, it also provides the richest and most quantitative conclusions. This approach deals simultaneously with submodel constraints and free-parameter values, and thus avoids problems that are inherent in other approaches. The posterior distributions provide detailed information beyond what we require to address the primary scientific question concerning Poisson fluctuations, such as information on differences between the Rb9 classes. Future experiments illuminating the mechanisms of intestinal tumour growth may provide a context for this information.

Acknowledgements

The authors are grateful to Professor Dove, Dr Haigis and Professor Drinkwater for useful discussions and to two referees for helpful comments on the first submission. An earlier and much less extensive version of this manuscript appeared as an unrefereed Web note supporting Haigis and Dove (2003). See `http://www.nature.com/ng/journal/v33/n1/suppinfo/ng1055_S1.html`. This research was supported by grants R01 CA64364 and R37 CA63677 from the National Cancer Institute of the USA.

Appendix A: Markov chain Monte Carlo implementation

Our sampler is designed to take advantage of the partial nesting of the $\boldsymbol{\lambda}$ and $\boldsymbol{\kappa}$ constraint patterns. In particular, suppose that the current mean pattern is indexed by l. All moves that propose a new state l' of different dimension (i.e. $q_{l'} \neq q_l$) are such that either the pattern of $\boldsymbol{\lambda}$ corresponding to l is nested within the pattern corresponding to l' or vice versa. At most a move that alters the dimension by 1 is attempted. Table 8 details the permissible models l' which can be proposed given that the current model is l. We emphasize again that for such moves k remains constant between the current and new model. A similar table (which is not shown) can be created for the second move class which updates k while keeping l fixed.

Consider the dynamics of one particular transdimensional type of move. Suppose that we are in submodel $m = (l, k)$. We propose a move to submodel $m' = (l', k)$ as follows. With probability b_l we propose a

Table 8. MCMC implementation†

Current mean structure l	*Permitted birth proposals l′*	*Permitted death proposals l′*	*Permitted switch proposals l′*
1	2, 3, 4, 5, 6, 7, 8	—	—
2	9, 10, 12	1	3, 4, 5, 6, 7, 8
3	9, 11, 13	1	2, 4, 5, 6, 7, 8
4	10, 11, 14	1	2, 3, 5, 6, 7, 8
5	12, 13, 14	1	2, 3, 4, 6, 7, 8
6	9, 14	1	2, 3, 4, 5, 7, 8
7	10, 13	1	2, 3, 4, 5, 6, 8
8	11, 12	1	2, 3, 4, 5, 6, 7
9	15	2, 3, 6	10, 11, 12, 13, 14
10	15	2, 4, 7	9, 11, 12, 13, 14
11	15	3, 4, 8	9, 10, 12, 13, 14
12	15	2, 5, 8	9, 10, 11, 13, 14
13	15	3, 5, 7	9, 10, 11, 12, 14
14	15	4, 5, 6	9, 10, 11, 12, 13
15	—	9, 10, 11, 12, 13, 14	—

†Shown are the values of the mean pattern index l' that can be proposed when the current mean pattern index is l (the labels are as in Table 2).

move to l' such that $q_{l'} = q_l + 1$ (a *birth* move). Alternatively, with probability d_l we propose a move to l' such that $q_{l'} = q_l - 1$ (a *death* move). We set $b_0 = 1$, $d_0 = 0$, $b_{15} = 0$ and $d_{15} = 1$ and for all other values of l we take $b_l = d_l = \frac{1}{2}$. Having chosen between a birth and death move, we choose the particular value of l at random from the $r_{l,q_{l'}}$ particular candidates. With the model proposed in place, we need to propose new parameter values $\boldsymbol{\lambda}'$. Importantly, we note that we are actually proposing a move from the q_l unique parameter values $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_{q_l}$ to the $q_{l'}$ unique parameters $\tilde{\lambda}'_1, \ldots, \tilde{\lambda}'_{q_{l'}}$. For the birth move, this is achieved by generating a lognormal$(0, \sigma_\lambda)$ random variable u and then setting

$$\tilde{\lambda}'_{i_1} = \tilde{\lambda}_{i_1} u,$$
$$\tilde{\lambda}'_{i_2} = \tilde{\lambda}_{i_1} / u,$$
$$\tilde{\lambda}'_j = \begin{cases} \tilde{\lambda}_j & j \neq i_1,\ j < i_2, \\ \tilde{\lambda}_{j+1} & j \neq i_1,\ j > i_2. \end{cases}$$

The values of indices i_1 and i_2 are obvious from the current and proposed mean patterns l and l' as shown in the example below. The proposed parameter vector $\boldsymbol{\lambda}'$ is then immediate from $\tilde{\lambda}'_1 \ldots, \tilde{\lambda}'_{q_{l'}}$ and $\boldsymbol{\kappa}$ remains unchanged.

To clarify this move, consider the example of moving from model $l = 8$ to model $l' = 12$. This is a birth move since $q_l = 2$ and $q_{l'} = 3$ and is allowed since $l' = 12$ is tabulated as an allowable candidate (Table 8) for $l = 8$. Suppose that the value of $\boldsymbol{\lambda}$ is $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_4)$. The two unique parameters are $\tilde{\lambda}_1 = \lambda_1 = \lambda_4$ and $\tilde{\lambda}_2 = \lambda_2 = \lambda_3$. Here $i_1 = 1$ and $i_2 = 3$, which means that we propose a move to $\tilde{\lambda}'_1, \ldots, \tilde{\lambda}'_3$, by setting $\tilde{\lambda}'_1 = \tilde{\lambda}_1 u$, $\tilde{\lambda}'_2 = \tilde{\lambda}_2$ and $\tilde{\lambda}'_3 = \tilde{\lambda}_1 / u$. This gives $\boldsymbol{\lambda}' = (\lambda'_1, \ldots, \lambda'_4) = (\tilde{\lambda}'_1, \tilde{\lambda}'_2, \tilde{\lambda}'_2, \tilde{\lambda}'_3)$.

Calculating the acceptance probability of a birth move is straightforward once the dynamics of the reverse death move are considered. To do so, we suppose now that the chain is in model $m' = (l', k)$, with parameters $\boldsymbol{\lambda}'$ and $\boldsymbol{\kappa}$. Having chosen a death move (with probability $d_{l'}$) we propose a new mean structure l, at random from the r_{l',q_l} possible candidates. The proposed mean parameter $\boldsymbol{\lambda}$ is then given by

$$\tilde{\lambda}_{i_1} = \sqrt{(\tilde{\lambda}'_{i_1} \tilde{\lambda}'_{i_2})},$$
$$\tilde{\lambda}_j = \begin{cases} \tilde{\lambda}'_j & j \neq i_1,\ j < i_2, \\ \tilde{\lambda}'_{j-1} & j \neq i_1,\ j \geqslant i_2, \end{cases}$$

and the dummy random variable u is given by

$$u = \sqrt{(\tilde{\lambda}'_{i_1} / \tilde{\lambda}'_{i_2})},$$

where, as above, i_1 and i_2 are apparent from the new and existing mean structures l and l'.

Following the methods that were prescribed in Green (2003) it follows that the acceptance probabilities for the birth move and reverse death moves are given by $\min\{1, A\}$ and $\min\{1, A^{-1}\}$ respectively, where

$$A = \frac{\mathcal{L}(\mathbf{x}|m', \boldsymbol{\lambda}', \boldsymbol{\kappa})\ p(m', \boldsymbol{\lambda}', \boldsymbol{\kappa})}{\mathcal{L}(\mathbf{x}|m, \boldsymbol{\lambda}, \boldsymbol{\kappa})\ p(m, \boldsymbol{\lambda}, \boldsymbol{\kappa})} \frac{d_{l'}}{b_l} \frac{r_{l,q'_l}}{r_{l',q_l}} \frac{1}{g_{\sigma_\lambda}(u)} |J|. \tag{4}$$

Here the first factor contains the ratio of posteriors, comprising the ratio of likelihoods $\mathcal{L}$ and the ratio of priors which depends on whether prior A or prior B is used. The second factor is the ratio of the proposal distributions, where g_{σ_λ} is the probability density function of the lognormal$(0, \sigma_\lambda)$ distribution. The final factor is the absolute value of the Jacobian J, which simplifies to

$$\begin{aligned} J &= \begin{vmatrix} \dfrac{\partial \tilde{\lambda}'_{i_1}}{\partial \tilde{\lambda}_{i_1}} & \dfrac{\partial \tilde{\lambda}'_{i_2}}{\partial \tilde{\lambda}_{i_1}} \\ \dfrac{\partial \tilde{\lambda}'_{i_1}}{\partial u} & \dfrac{\partial \tilde{\lambda}'_{i_2}}{\partial u} \end{vmatrix} \\ &= \begin{vmatrix} u & \dfrac{1}{u} \\ \tilde{\lambda}_{i_1} & -\dfrac{\tilde{\lambda}_{i_1}}{u^2} \end{vmatrix} \\ &= -\frac{2\tilde{\lambda}_{i_1}}{u}. \end{aligned}$$

In addition to the move that is detailed above, at each sweep of our sampler we include a birth–death move pair to update the dispersion structure k while keeping l fixed. The dynamics of this move are similar to the above and we omit the details. To improve the mixing properties of our sampler, we also include a second model changing move type. Again, this move type, which we call a switch move, either proposes a new mean structure l while retaining the current dispersion index k or vice versa. At each sweep, an attempt is made to update both l and k in this manner. Unlike the birth–death type of move, rather than proposing a model with one more or one fewer free parameters, the switch move proposes a model with the same dimension. The new proposed values $\tilde{\boldsymbol{\lambda}}'$ (or $\tilde{\boldsymbol{\kappa}}'$) for the underlying parameters are then drawn from independent log-normal distributions, independently of the current values $\tilde{\boldsymbol{\lambda}}$ and $\tilde{\boldsymbol{\kappa}}$. The acceptance probability is then easily derived following standard methods. Further details of the reversible jump moves along with all aspects of the MCMC sampler can be found in chapter 3 of Hastie (2004).

Beyond model changing moves, our sampler also includes standard Metropolis–Hastings moves to update each of the free parameters in $\boldsymbol{\lambda}$ and $\boldsymbol{\kappa}$ in turn. For the component being updated, the proposed new value is the product of the current value and a random log-normal increment and is accepted with probability derived in the usual fashion.

For both priors we run the MCMC algorithm for 1 100 000 sweeps and discard the first 100 000 as burn-in. We subsample every 100 observations, leaving 10 000 observations. Our sampler typically gives transdimensional acceptance rates of between 6% and 12% for prior A and 5% and 8% for prior B, depending on the type of move. In both cases the fixed dimensional acceptance rates vary between 10% and 47%. On the sensitivity analysis in Fig. 2, we ran the same codes but subsampled every 10 observations.

Although some tuning has been carried out, the sampler has not been optimized and more desirable acceptance rates could no doubt be achieved. However, we think that with the current parameter settings the mixing is adequate. This is supported by the trace plots in Fig. 1 and the fact that multiple MCMC runs started from random initial states produce similar numerical results. Reported posterior probabilities are averages across two independent MCMC runs. In all instances we estimate the Monte Carlo standard error to be less than 0.005.

References

Akaike, H. (1983) Information measures and model selection. *Bull. Int. Statist. Inst.*, **50**, 277–290.

Arratia, R., Goldstein, L. and Gordon, L. (1990) Poisson approximation and the Chen-Stein method. *Statist. Sci.*, **5**, 403–424.

Barnard, G. A. (1963) Discussion on 'The spectral analysis of point processes' (by M. S. Bartlett). *J. R. Statist. Soc.* B, **25**, 294.

Bell, E. T. (1934) Exponential polynomials. *Ann. Math.*, **35**, 258–277.
Chulada, P. C., Thompson, M. B., Mahler, J. F., Doyle, C. M., Gaul, B. W., Lee, C., Tiano, H. F., Morham, S. G., Smithies, O. and Langenbach, R. (2000) Genetic disruption of *Ptgs-1*, as well as of *Ptgs-2*, reduces intestinal tumorigenesis in *Min* mice. *Cancer Res.*, **60**, 4705–4708.
Daley, D. J. and Vere-Jones, D. (1988) *An Introduction to the Theory of Point Processes*. Berlin: Springer.
Dove, W. F., Cormier, R. T., Gould, K. A., Halberg, R. B., Merritt, A. J., Newton, M. A. and Shoemaker, A. R. (1998) The intestinal epithelium and its neoplasms: genetic, cellular, and tissue interactions. *Phil. Trans. R. Soc. Lond.* B, **353**, 915–923.
Drinkwater, N. R. and Klotz, J. H. (1981) Statistical methods for the analysis of tumor multiplicity data. *Cancer Res.*, **41**, 113–119.
Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2003) *Bayesian Data Analysis*. New York: Chapman and Hall.
Gould, K. A., Luongo, C., Moser, A. R., McNeley, M. K., Borenstein, N., Shedlovsky, A., Dove, W. F., Hong, K., Dietrich, W. F. and Lander, E. S. (1996) Genetic evaluation of candidate genes for the *Mom1* modifier of intestinal neoplasia in mice. *Genetics*, **144**, 1777–1785.
Green, P. J. (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**, 711–732.
Green, P. J. (2003) Trans-dimensional Markov chain Monte Carlo. In *Highly Structured Stochastic Systems*, ch. 6, pp. 179–196. Oxford: Oxford University Press.
Greenwood, Major, and Yule, G. U. (1920) An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. *J. R. Statist. Soc.*, **83**, 255–279.
Haigis, K. M. and Dove, W. F. (2003) A Robertsonian translocation suppresses a somatic recombination pathway to loss of heterozygosity. *Nature Genet.*, **33**, 33–39.
Hardy, R. G., Meltzer, S. J. and Jankowski, J. A. (2000) ABC of colorectal cancer: molecular basis for risk factors. *Br. Med. J.*, **321**, 886–889.
Hastie, D. I. (2004) Towards automatic reversible jump MCMC. *PhD Thesis*. Department of Mathematics, University of Bristol, Bristol.
Kass, R. E. and Wasserman, L. (1995) A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criteria. *J. Am. Statist. Ass.*, **90**, 928–934.
Kokoska, S. M. (1987) The analysis of cancer chemoprevention experiments. *Biometrics*, **43**, 525–534.
Li, Y. Q., Roberts, S. A., Paulus, U., Loeffler, M. and Potten, C. S. (1994) The crypt cycle in mouse small intestinal epithelium. *J. Cell Sci.*, **107**, 3271–3279.
McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models*, 2nd edn. London: Chapman and Hall.
Moolgavkar, S. H. and Knudson, A. G. (1981) Mutation and cancer: a model for human carcinogenesis. *J. Natn. Cancer Inst.*, **66**, 1037–1052.
Moser, A. R., Dove, W. F., Roth, K. A. and Gordon, J. I. (1992) The *Min* (multiple intestinal neoplasia) mutation: its effect on gut epithelial cell differentiation and interaction with a modifier system. *J. Cell Biol.*, **116**, 1517–1526.
Moser, A. R., Pitot, H. C. and Dove, W. F. (1990) A dominant mutation that predisposes to multiple intestinal neoplasia in the mouse. *Science*, **247**, 322–324.
Nagase, H., Mao, J. H. and Balmain, A. (1999) A subset of skin tumor modifier loci determines survival time of tumor-bearing mice. *Proc. Natn. Acad. Sci. USA*, **96**, 15032–15037.
O'Hagan, A. (1994) *Kendall's Advanced Theory of Statistics*, vol. 2B, *Bayesian Inference*. London: Arnold.
Preston, S. L., Wong, W. M., Chan, A. O., Poulsom, R., Jeffery, R., Goodlad, R. A., Mandir, N., Elia, G., Novelli, M., Bodmer, W. F., Tomlinson, I. P. and Wright, N. A. (2003) Bottom-up histogenesis of colorectal adenomas: origin in the monocryptal adenoma and initial expansion by crypt fission. *Cancer Res.*, **63**, 3819–3825.
Ramachandran, S., Fryer, A. A., Lovatt, T., Smith, A., Lear, J., Jones, P. W. and Strange, R. C. (2002) The rate of increase in the numbers of primary sporadic basal cell carcinomas during follow up is associated with age at first presentation. *Carcinogenesis*, **23**, 2051–2054.
Robert, C. P. and Casella, G. (2002) *Monte Carlo Statistical Methods*. New York: Springer.
Schwarz, G. (1978) Estimating the dimension of a model. *Ann. Statist.*, **6**, 461–464.
Simon, T. G. and Gordon, J. I. (1995) Intestinal epithelial cell differentiation: new insights from mice, flies and nematodes. *Curr. Opinion Genet. Devlpmnt*, **5**, 577–586.
Smith, A. F. M. and Spiegelhalter, D. J. (1980) Bayes factors and choice criteria for linear models. *J. R. Statist. Soc.* B, **42**, 213–220.
Su, L. K., Kinzler, K. W., Vogelstein, B., Preisinger, A. C., Moser, A. R., Luongo, C., Gould, K. A. and Dove, W. F. (1992) A germline mutation of the murine homolog of the APC gene causes multiple intestinal neoplasia. *Science*, **256**, 668–670.