

Bayesian generalized monotonic functional mixed models for the effects of radiation dose histograms on normal tissue complications

Matthew Schipper^{1,*},[†], Jeremy M. G. Taylor² and Xihong Lin³

¹*Department of Radiology, University of Michigan, Ann Arbor, MI 48109, U.S.A.*

²*Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, U.S.A.*

³*Department of Biostatistics, Harvard University, Boston, MA 02115, U.S.A.*

SUMMARY

When treating cancer patients with radiation therapy, the normal tissue in an organ close to the tumour usually receives some dose of radiation. The dose is not of the same intensity throughout the organ. This radiation can cause normal tissue complications, so for treatment planning purposes, it is important to understand the relationship between the distribution of dose intensities in the organ and the occurrence of complications. One general summary measure of the dose effect is obtained by integrating a weighting function ($w(d)$) over the dose distribution. The weighting function $w(d)$ should be monotone for biological reasons. Because the true shape of $w(d)$ is not known, we estimate it non-parametrically subject to the monotonicity constraint. In our approach $w(d)$ is written as a weighted sum of monotone basis functions. The weights in this sum are formulated as a mixture of point mass at zero and a Gamma random variable. A key feature of our method is that it allows for flat regions through the use of this mixture prior. The model is estimated using a Markov Chain Monte Carlo algorithm. We illustrate our method with data from a head and neck cancer study in which the irradiation of the parotid gland results in loss of saliva flow. Copyright © 2007 John Wiley & Sons, Ltd.

KEY WORDS: non-parametric regression; mixture prior; monotonicity; functional data; dose effect; normal tissue complications

1. INTRODUCTION

Cancer patients treated with radiation therapy can experience complications that are the consequence of the irradiation of normal tissue in an organ close to the tumour site. There is

*Correspondence to: Matthew Schipper, 200 Zina Pitcher Place, 3307A Kresge III, Ann Arbor, MI 48109, U.S.A.

[†]E-mail: mjschipp@med.umich.edu

Contract/grant sponsor: US National Institutes of Health; contract/grant numbers: CA-095096, CA-97248, CA-83654

Received 16 August 2006

Accepted 15 February 2007

Copyright © 2007 John Wiley & Sons, Ltd.

a large literature on normal tissue complication models (see, for example, [1, 2]). These models relate the dose received by the normal tissue to the probability of a complication. They are typically motivated by some radio-biological or bio-physical theory and sometimes take into account the physical structure of the normal tissue, e.g. whether it is a serial structure such as the spinal cord, or more homogenous such as the parotid gland [3]. The most commonly used model in treatment planning is the Lyman normal tissue complication probability model [4, 5]. This specific model applies to binary outcomes and specifies the probability of complication as a function of dose, the volume of normal tissue irradiated, and three parameters. In contrast to these models, which tend to be highly parametric, our approach is more non-parametric and not restricted to binary outcomes.

While the tumour site generally receives a fairly homogenous dose, the normal tissue in an organ close to the tumour usually receives a heterogeneous dose. It is possible to compute, for each subject and a given radiation treatment plan, the cumulative dose of radiation received at each location within the organ. This dose distribution is typically represented using the dose–volume histogram (DVH), a plot over dose of the proportion of normal tissue which received that dose. Thus, the DVH is a curve that integrates to 1; however, it does ignore the spatial aspect of the distribution of dose within the organ. In some organs, such as the parotid gland which is relatively homogenous, ignoring the spatial aspect seems to be reasonable. For less homogenous organs, such as the lung, this will be less reasonable but may still be used as an approximation or as a first step in a more complicated analysis taking the spatial component into account. See Figure 1 for an example of two DVH's. Denoting the DVH by $p(d)$, the goal is to find the relationship between the curve $p(d)$ and a scalar complication outcome. Specifically, we are interested in using $p(d)$ as a functional covariate to construct an appropriate summary measure of dose in a regression model for the complication outcome.

Functional data analysis is a branch of statistics concerned with curves, which can be used as a response or a predictor [6]. For the latter, one is interested in estimating the effect of a functional covariate $p(d)$ on a scalar outcome, and the functional form often takes $\int p(d)w(d) dd$, where $w(d)$ is an unknown weight function (Chapter 15 of [6]). We are interested in estimating $w(d)$ to construct a summary measure of $p(d)$ as a good explanatory variable of the outcome. This approach was discussed by Hastie and Mallows [7], and in the functional generalized linear model of James [8], where $\int p(d)w(d) dd$ was included as a covariate in a GLM and a regression spline was used to estimate $w(d)$, and by Zhang *et al.* [9] for a continuous outcome, where a smoothing spline is used to estimate $w(d)$. However, little work is done for such functional models with a monotonic constraint for $w(d)$ using mixed models. Schipper *et al.* [10] proposed a generalized monotone functional mixed model (GMFMM), which included $\int p(d)w(d) dd$ as a covariate in a generalized linear mixed model (GLMM) and estimated $w(d)$ monotonically and non-parametrically using a regression spline. Two special cases are worth noting. If $w(d) = d$ then the summary measure is the mean dose received by the organ. If $w(d) = I(d > d_0)$ then the summary measure is the proportion of the organ that receives a dose larger than d_0 .

In this paper, we propose a Bayesian version of the GMFMM in Schipper *et al.* [10]. The dose effect is summarized as $\int p(d)w(d) dd$. Any biologically meaningful estimate of $w(d)$ should be monotone, because increasing dose cannot lead to lower probability or severity of complications. Thus, we propose a Bayesian method for monotonic non-parametric estimation of $w(d)$ within this model framework.

The proposed estimator of $w(d)$ in Schipper *et al.* [10] is strictly monotone. However, when modelling a dose effect, it is possible that there is no effect within a range of doses. This appears

to be the case for low doses in the saliva flow data we use to illustrate our model. This corresponds to flat regions in $w(d)$, and our Bayesian estimator allows for this, which is not allowed in the formulation of Schipper *et al.* [10]. We define $w(d)$ as the sum of monotone non-decreasing basis functions each multiplied by a non-negative coefficient. The non-negativity is imposed through priors in a Bayesian model. Similar to the approach of Dunson [11], we specify a mixture prior for the coefficients, which explicitly allow the function $w(d)$ to be flat, while at the same time imposing the monotonicity constraint. In addition to applying to functional covariates, the model proposed in this paper differs from that of Dunson [11] in that it applies to general distributions and correlated outcome measures.

This paper is organized as follows. Section 2 describes the model as well as the mixture prior. Section 3 discusses estimation of the model. In Section 4, the head and neck cancer data are described, and the proposed methods are illustrated. In Section 5, simulation results are presented, followed by discussions in Section 6.

2. THE MODEL

2.1. Generalized monotone functional mixed model

In this section we propose a Bayesian GMFMM, for relating a complication outcome to a dose distribution while accounting for other covariates as well as correlation between multiple non-Gaussian responses from the same subject. Let Y_{ij} denote the complication outcome for the j th observation of the i th subject, $j = 1, \dots, J$; $i = 1, \dots, I$. We note that there may be multiple measurements of the complication outcome for a particular subject. In the head and neck cancer example we consider later, the complication outcome is saliva flow, which is measured separately from the two parotid glands of each subject. We assume that conditional on random effects b_i , Y_{ij} follows an exponential family distribution with mean μ_{ij} given by

$$g(\mu_{ij}) = X_{ij}^T \alpha + \int p_{ij}(d)w(d) dd + Z_{ij}^T b_i \quad (1)$$

where g is a canonical link function, α and b_i represent fixed and random effects with corresponding design matrices X_{ij} and Z_{ij} , $p_{ij}(d)$ is the DVH in the organ of interest with $\int p_{ij}(d) dd = 1$, and $w(d)$ is an unknown weight. We assume that the random effects b_i are independent and follow a $N(0, \Sigma)$ distribution. The marginal likelihood is given by

$$\mathcal{L}\{\alpha, w(d), \Sigma\} = \prod_i \int \prod_j f\{Y_{ij}|b_i; \alpha, w(d)\} f(b_i; 0, \Sigma) db_i \quad (2)$$

where $f(Y_{ij}|b_i)$ is the conditional density of Y_{ij} and $f(b_i)$ is the normal density of b_i .

Overdispersion and correlation between observations from the same subject can be modelled using the random effects b_i . We further assume $w(d)$ is a smooth, non-decreasing function with $w(0) = 0$. The constraint $w(0) = 0$ is for the sake of identifiability. Specifically, from (1), if α includes an intercept term, it will be confounded with any intercept $w(0)$ because this latter term can be pulled out from the integration in (1), since $p_{ij}(d)$ integrates to one.

2.2. Formulation of $w(d)$

One way to constrain $w(d)$ to be non-decreasing is to define it as the sum of non-decreasing functions. We consider the following formulation:

$$w(d) = \sum_{l=1}^L s_l B_l(d) \quad (3)$$

Here $B_l(d)$ are known non-decreasing basis functions, L is the number of basis functions, and the s_l are non-negative parameter coefficients. With this definition of $w(d)$, the model discussed in the previous section can be written as a GLMM [12]. In particular,

$$\begin{aligned} g(\mu_{ij}) &= X_{ij}^T \alpha + \int p_{ij}(d) w(d) \, dd + Z_{ij}^T b_i \\ &= X_{ij}^T \alpha + \int p_{ij}(d) \left\{ \sum_{l=1}^L s_l B_l(d) \right\} \, dd + Z_{ij}^T b_i \\ &= X_{ij}^T \alpha + \sum_{l=1}^L s_l \int p_{ij}(d) B_l(d) \, dd + Z_{ij}^T b_i \end{aligned} \quad (4)$$

Many choices are possible for the basis functions, and we consider two here. The first is to use piecewise linear basis functions, considered for example by Neelon and Dunson [13]. Let $B_l(d) = \min(d, \gamma_l) - \gamma_{l-1}$ if $d > \gamma_{l-1}$ and $B_l(d) = 0$ otherwise, where $(0 = \gamma_0 < \gamma_1 < \dots < \gamma_L)$ are known knot locations. In words, $B_l(d)$ is zero till the beginning of the l th interval, then is linear (with slope 1) in the l th interval, and flat at its maximum value after the l th interval. Under this formulation, the slope of $w(d)$ within interval $(\gamma_{l-1}, \gamma_l]$ is given by s_l . The second set of basis functions we use are the I-spline functions [14]. These functions are obtained as the integral of B-spline basis functions, and are thus monotone since the B-splines are non-negative. While both basis functions give continuous estimates, a key feature of the I-Spline basis functions is that they result in a smoother estimate compared to the estimator based on the linear basis functions. Since $B_l(0) = 0$ for all l , $w(0) = 0$ from (3), for both sets of basis functions.

2.3. Prior specification

The coefficients s_l are restricted to be non-negative through the prior distributions. Adapting the ideas of Dunson [11], we propose a mixture prior for s_l which imposes the monotonicity constraint on $w(d)$ while at the same time allowing for flat regions. Let δ_l be a latent Bernoulli variable indicating zero or non-zero coefficient and β_l be a positive continuous random variable. Then define s_l , as

$$s_l = \delta_l \beta_l \quad (5)$$

Thus, β_l is the value of s_l , given s_l is non-zero. Plugging (5) into (4) we obtain

$$g(\mu_{ij}) = X_{ij}^T \alpha + \sum_{l=1}^L \delta_l \beta_l \int p_{ij}(d) B_l(d) \, dd + Z_{ij}^T b_i \quad (6)$$

Next we define the priors for δ_l and β_l as

$$\delta_l \sim \text{Bernoulli}(\pi) \quad (7)$$

$$\beta_l \sim \text{Gamma}(c_1, \lambda) \quad (8)$$

where $\text{Gamma}(c_1, \lambda)$ denotes a Gamma distribution with mean c_1/λ and variance c_1/λ^2 . Rather than choosing fixed values for π and λ , we specify hyperpriors to reduce the sensitivity of our conclusions to the chosen values. We choose to set c_1 to a fixed value. If we choose $c_1 > 1$ then the prior for β_l in (8) will have a non-zero mode. This should prevent non-identifiability arising from very small β_l with positive δ_l and zero δ_l , both leading to similar values for s_l . Choosing $c_1 > 1$ also ensures that the prior for β_l is log-concave. This allows efficient sampling in the Gibbs sampler, as discussed in Section 3. For the hyperpriors, we assume

$$\pi \sim \text{Uniform}(0, 1) \quad (9)$$

$$\lambda \sim \text{Gamma}(c_2, c_3) \quad (10)$$

We note here that $w(d)$ depends on δ_l , in addition to β_l . This gives some robustness in the estimated $w(d)$ to the choice of the prior on β_l . To complete the model specification, we place a vague Normal prior on α and an inverse Wishart prior on the covariance matrix Σ .

3. ESTIMATION

The DVH, or $p(d)$, is obtained by discretizing dose into K bins and measuring what proportion of the normal tissue received a dose within each bin. Figure 1 shows two example DVH's from

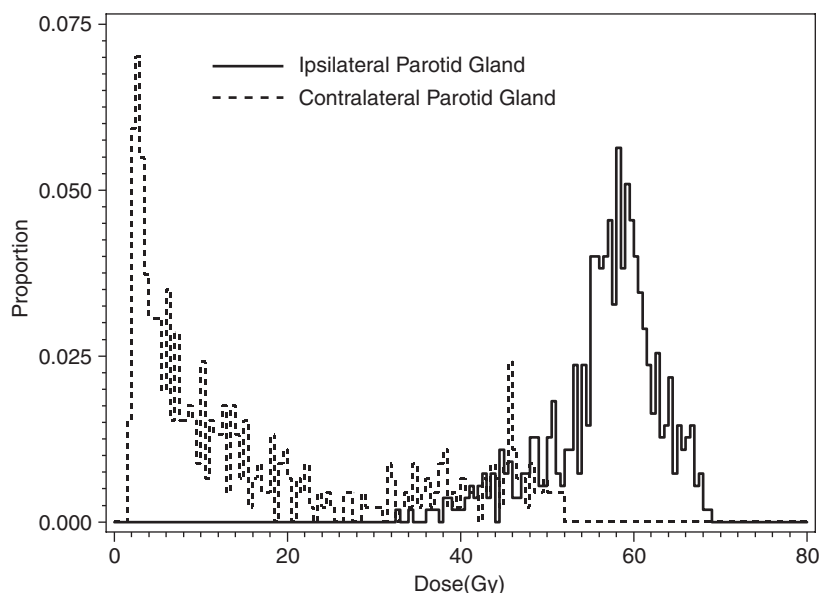


Figure 1. Examples of one subject's dose volume histograms. The ipsilateral gland is the gland on the same side of the head as the tumour, while the contralateral gland is on the opposite side of the head.

the saliva flow data we discuss in the next section. Thus, we can rewrite (1) by replacing the integration with summation over dose bins as

$$g(\mu_{ij}) = X_{ij}^T \alpha + \sum_{k=1}^K p_{ij}(d_k) w(d_k) + Z_{ij}^T b_i \quad (11)$$

where $d_k = (k - 1)$ (bin width). That is, d_k is the left endpoint of the k th bin and $p(d_k)$ is the proportion of normal tissue receiving dose in bin k . Similarly we can also rewrite (6) as

$$g(\mu_{ij}) = X_{ij}^T \alpha + \sum_{l=1}^L \delta_l \beta_l \left[\sum_{k=1}^K p_{ij}(d_k) B_l(d_k) \right] + Z_{ij}^T b_i \quad (12)$$

Gibbs sampling can be used to estimate the parameters in this model by iteratively sampling from each of the full conditional distributions. Conditional on δ and β , (12) specifies a GLMM with an offset term of $\sum_{l=1}^L \delta_l \beta_l \sum_{k=1}^K p_{ij}(d_k) B_l(d_k)$. Gibbs sampling for GLMMs has been discussed by many authors, e.g. Zeger and Karim [15] and Dellaportas and Smith [16], among others. These methods could be used to sample from the full conditional distributions for α , b_i , and Σ . The only additional computational work is in computing the value of the offset term at each iteration of the Gibbs sampler.

We now focus on the full conditional distributions for the dose effect parameters δ_l , β_l , π , and λ . Some of these distributions involve the likelihood

$$\mathcal{L} = \prod_{i,j} f(Y_{ij}|b_i) \quad (13)$$

where $f(Y_{ij}|b_i)$ is the density of y_{ij} conditional on b_i . Let θ_{-v} denote the set of all unknown parameters except v , and $[a_1|a_2]$ denote the distribution of a_1 given a_2 . We consider the δ_l first. The full conditional distribution for δ_l is given by

$$[\delta_l|\theta_{-\delta_l}] \propto [\delta_l|\pi] \mathcal{L} \quad (14)$$

where the proportionality constant can be easily calculated since the δ_l are binary.

Next we consider the full conditional for the π :

$$\begin{aligned} [\pi|\theta_{-\pi}] &\propto [\pi] \prod_l [\delta_l|\pi] = \pi^{\sum_l \delta_l} (1 - \pi)^{\sum_l (1 - \delta_l)} \\ &\propto \text{Beta} \left\{ 1 + \sum_l \delta_l, 1 + \sum_l (1 - \delta_l) \right\} \end{aligned} \quad (15)$$

Thus, the full conditional for π is a Beta distribution for which standard sampling methods are available.

The full conditional for the β_l is given by

$$[\beta_l|\theta_{-\beta_l}] \propto [\beta_l|\lambda] \mathcal{L} \quad (16)$$

When $\delta_l = 0$, it is clear from (12) that μ_{ij} does not contain any terms for β_l . Thus, the likelihood \mathcal{L} is independent of β_l , and the full conditional is proportional to the prior. Conditional on $\delta_l = 1$, the right-hand side of (12) is a linear function of α , β_l , and b_i and thus specifies a GLMM. If we condition on $\delta_l = 1$ and on b_i , (12) specifies a GLM with offset term $Z_{ij} b_i$. Thus, the likelihood is a

log-concave function of β_l [16]. Since we have also specified a log-concave prior (Gamma(c_1, λ)) for β_l , the full conditional (16) is log-concave because the product of two log-concave functions is also log-concave. Adaptive rejection sampling may be used to efficiently sample from this distribution [17].

The full conditional for λ is given by

$$\begin{aligned} [\lambda|\theta_{-\lambda}] &\propto [\lambda] \prod_l [\beta_l|\lambda] \\ &\propto \text{Gamma}\left(c_1 \mathcal{L} + c_2, c_3 + \sum_l \beta_l\right) \end{aligned} \quad (17)$$

Thus, the full conditional for λ is a Gamma distribution, for which straightforward sampling methods are available. The Gibbs sampling discussed here can be conveniently implemented in the program WinBUGS. Convergence of the Markov chain can be assessed using standard methods.

4. APPLICATION TO HEAD AND NECK CANCER DATA

4.1. Data description

The data come from a head and neck cancer study conducted at the University of Michigan. Patients with a cancerous tumour in their head or neck were treated with external beam radiation therapy. A common side effect in these patients is loss of saliva flow due to the irradiation of the parotid glands. The parotid glands, of which there are two, one in each cheek, are responsible for producing the majority of a person's saliva. The radiation doses ranged from 0 to 82 and the DVH bin widths are 0.5 Gy, thus $K = 164$. Figure 1 is an example of the DVH's for the two glands of a particular subject. Saliva flows are obtained at baseline (before radiation therapy) and at 6 months after radiation therapy. At each timepoint, saliva flows are measured from both parotids. The ipsilateral glands (gland on the same side of head as tumour) generally received a much higher dose than the contralateral glands (glands on the opposite side of head as tumour). This can be seen in the example DVH's plotted in Figure 1. As a consequence, the ipsilateral glands typically had much lower saliva flows. In total, there are 157 saliva flow measurements (with corresponding DVH's) from 82 subjects.

4.2. Model description

The units of saliva flow are millilitres per minute, so we use the Poisson distribution as outcome distribution. Since the measured saliva flows are not integers, we multiplied the original flows by 150 and rounded to the nearest integer. Let Y_{ij} denote the (transformed) saliva flow from gland j of subject i . Let $b'_i = [b_{i1} \ b_{i2}]$ denote the unobserved random effects for subject i , assumed to be independently distributed as $N(0, \Sigma)$. Further, denote the log of the baseline saliva flow from gland j of subject i by base_{ij} . We assume that, conditional on the random effects b_i , the saliva flows follow a Poisson distribution with mean μ_{ij} given by

$$\log(\mu_{ij}) = a_0 + a_1 \text{base}_{ij} - \sum_{k=1}^{k=164} p_{ij}(d_k) w(d_k) + b_{ij} \quad (18)$$

where $p_{ij}(\cdot)$ is the DVH for the j th gland of the i th subject.

To use the prior and hyperprior distributions specified in (7) and (10), we must specify values for c_1 , c_2 , and c_3 . We wish to choose these values so that reasonable prior mass is placed on both linear $w(d)$ (i.e. meandose models) and step function $w(d)$ which correspond to a threshold effect. We use the link between model (18) and the simpler meandose model to aid in setting these values. Specifically, if $s_1 = s_2 = \dots = s_{20} = s$, then some algebra shows that (18) can be rewritten as

$$\log(\mu_{ij}) = a_0 + a_1 \text{base}_{ij} - s \times \text{meandose}_{ij} + b_{ij} \quad (19)$$

We fit this mean dose model to the saliva flow data and obtain $\hat{s} = 0.1$. The implied linear $w(d)$ can be computed by replacing each s_l with \hat{s} in equation (3). Using the implied $w(d)$ from the fitted meandose model, we compute $\hat{w}(82) \approx 8$. If the true $w(d)$ was a step function such that $w(82) = 8$, we would require $s_l \approx 2$ for a particular l , since the maximal value of $B_l(d)$ is 4 for both sets of basis functions. Given these calculations, we set $c_1 = 1.5$, $c_2 = 1.5$, and $c_3 = 0.75$. The resulting prior and hyperprior in (8) and (10) assign large prior probability to a range of plausible values for β_l including 0.1 and 2.

The marginal variance of the saliva flows is larger than we would expect from the Poisson assumption and is approximately proportional to the square of the marginal mean. The random effects included as part of the linear predictor approximate this variance structure and account for the overdispersion. The two parameters of Σ on the main diagonal capture overdispersion for the ipsilateral and contralateral glands. We do not constrain them to have the same value. Most of the subjects have two measurements, one from each parotid. The off diagonal element of Σ is the covariance between the two sides within a subject. To fit the model, we use Gibbs sampling implemented in WinBUGS.

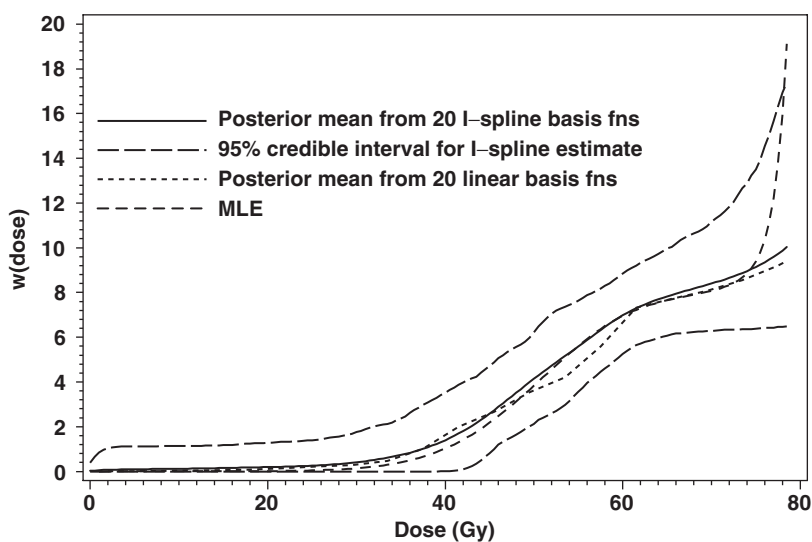
4.3. Results

We fit the model separately using the piecewise linear and I-spline basis functions to define $w(d)$. The reported results are based on using 20 basis functions ($L = 20$), but similar results were obtained using 10 or 40 functions. Using only five knots resulted in noticeably different estimates of $w(d)$. For each of the I-spline and linear basis function models, we ran two chains for 200 000 iterations discarding the first 50 000 as burn-in and retaining every 10th iteration thereafter. Thus estimates are based on 30 000 iterations. Trace plots of the parameters were compared between chains and within each chain over iteration to visually assess convergence. The chains mixed relatively rapidly for most parameters but were slowest to mix for δ_l . WinBUGS provides an estimate of the Monte Carlo error of the posterior mean estimate, and at 200 000 iterations the estimated Monte Carlo error is less than 10 per cent of the standard error for each parameter. It took about 4 h to run one chain for 200 000 iterations on a PC. Table I gives posterior mean estimates and 95 per cent credible intervals for the parameters based on the I-spline model. Results were very similar for the linear basis function model. The lower 95 per cent credible bounds for $\Sigma_{1,1}$ and $\Sigma_{2,2}$ are not close to 0, suggesting significant overdispersion for both the ipsilateral and contralateral glands, although it is much larger for the former. The estimated covariance between the two sides ($\Sigma_{1,2}$) does not appear to be significantly different from zero.

The posterior mean estimates of $w(d)$ and their 95 per cent pointwise credible intervals are shown in Figure 2. The estimated $w(d)$ obtained using linear and I-spline basis functions are similar, although not surprisingly, the linear basis functions lead to a less smooth estimate. For comparison we included an alternative estimate of $w(d)$ obtained from the method of

Table I. Parameter estimates and standard errors for model (18).

Parameter	Estimate	95 per cent credible intervals
a_0	1.54	(0.36, 2.95)
a_1	0.74	(0.45, 1.04)
p	0.30	(0.06, 0.69)
β	3.20	(0.65, 8.20)
$\Sigma_{1,1}$	10.35	(4.85, 20.20)
$\Sigma_{1,2}$	0.27	(-0.62, 1.24)
$\Sigma_{2,2}$	0.78	(0.54, 1.11)

Figure 2. Posterior mean estimate of $w(d)$ with pointwise 95 per cent credible intervals.

Schipper *et al.* [10]. We briefly describe the method here. Starting with an unconstrained regression spline $r(d)$, $w(d)$ is defined as the integral of a positive transformation of $r(d)$ as

$$w(d_c) = \int_0^{d_c} \phi(r(d)) dd \quad (20)$$

where ϕ is chosen to be proportional to a Normal density. This formulation of $w(d)$ is used in the same model given in (1), but we fit it using maximum likelihood. It is reassuring that the estimates are similar despite coming from different models and different estimation procedures. The Bayesian estimates of $w(d)$ do not exhibit the poor tail behaviour for the largest doses that is present in the Frequentist estimator.

Figure 3 plots for each gland the log proportion change in saliva flow *versus* the estimated summary measure of the dose effect $\sum_{k=2}^{k=164} p_{ij}(d_k) \hat{w}(d_k)$. Because some of the saliva flows are zero, we added one before taking logs. The plot shows proportion change in saliva flow from baseline (on log scale) decreasing linearly as a function of the estimated dose effect term.

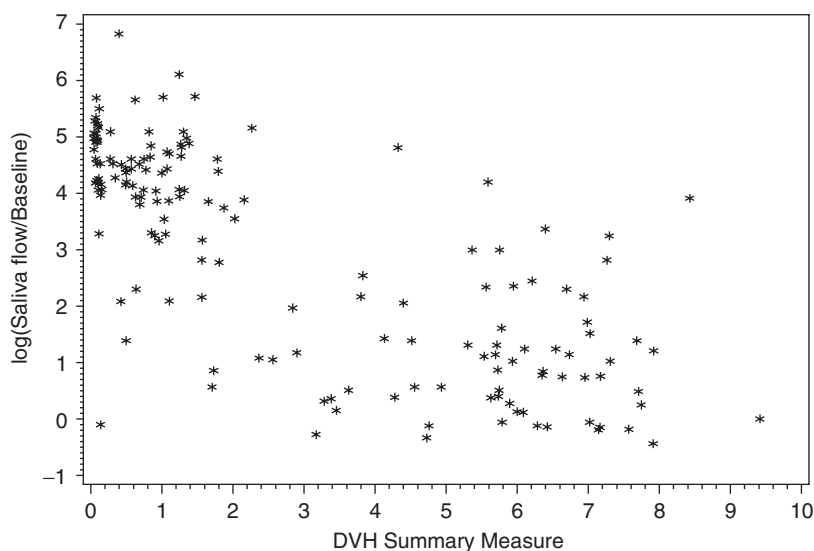


Figure 3. Change in log saliva flow from baseline *versus* estimated dose effect summary measure.

It is notable in Figure 2 that the estimated $w(d)$ is flat for doses less than 30 Gy. This implies that the parotid gland could be given a uniform dose of 30 Gy with minimal effect on saliva flow. In addition it implies that meandose is not the best way to summarize the DVH. When $w(d)$ is defined using linear basis functions it will be flat in a region $(\gamma_{j-1}, \gamma_{j+k})$ only if $\delta_j = \delta_{j+1} = \dots = \delta_{j+k} = 0$. The posterior probability of this event can easily be calculated as the proportion of draws (after discarding burn in) for which it holds. The posterior probabilities of $w(d)$ being flat in the four quartiles (0, 20), (20, 40), (40, 60), and (60, 81) are 0.72, 0.19, 0.02, and 0.27, respectively. A weakness of these calculations is that they ignore the size of the non-zero increases in $w(d)$. In particular they make no distinction between large coefficients and small coefficients close to zero.

The same calculations can be done for the I-spline basis functions but are slightly more involved. Rather than one, there are five I-spline basis functions with non-zero slope at any dose and thus we must check whether all five coefficients are zero. As a result, the corresponding probabilities are smaller than those based on the linear basis functions. Using the analysis based on the I-spline basis functions, the posterior probabilities of $w(d)$ being flat in the four quartiles of dose are 0.54, 0.004, 0 and 0.06, respectively. These probabilities are noticeably smaller than those from the analysis using linear basis functions. The reason for the difference is that these calculations dichotomize the basis functions as flat or not flat. At a given dose, however, some of the non-flat I-spline basis functions have large slope while some have very small slopes. This is not an issue for the linear basis functions since they have constant slope in their increasing regions. It is an undesirable property of these calculations for the I-spline functions that, although the estimated function $w(d)$ is similar to that using the linear basis functions, the estimated probabilities of flat $w(d)$ are much different. For this reason if one wishes to perform these calculations, the linear basis functions are preferable to the I-spline basis functions. Instead of relying solely on δ_l , the 'flatness' of $w(d)$ can be assessed visually using $\hat{w}(d)$ and pointwise credible intervals presented in Figure 2.

4.4. Goodness of fit

We assess goodness of fit of our model using posterior predictive simulation [18]. In this method we simulate new data y^{rep} from the posterior predictive distribution $P(y^{\text{rep}}|y)$. To compare the simulated and observed data we use a discrepancy variable, $T(y, \theta)$, which will take on large values when the data do not fit the model well. For each draw of θ , we compute $T(y^{\text{obs}}, \theta)$ and $T(y^{\text{rep}}, \theta)$. A P -value can be calculated as the proportion of times that $T(y^{\text{rep}}, \theta)$ exceeds $T(y^{\text{obs}}, \theta)$.

We use the deviance as a discrepancy variable, where the deviance is defined as

$$D = \sum_{i,j} y_{ij} \log(y_{ij}/\mu_{ij}) - (y_{ij} - \mu_{ij}) \quad (21)$$

The mean $\mu_{ij} = E[Y_{ij}|b_{ij}]$ is obtained from (18) by plugging in the parameter draws from the posterior distribution. For each of the 15 000 post-burn in samples (of one chain), we used the parameter draws to simulate a vector of Poisson data. We then computed the deviance for the observed data and the simulated data using the posterior parameter draws. In 49 per cent of the samples the deviance from the simulated data was greater than that of the observed data. Another commonly used discrepancy variable is the Pearson Chi-square GOF statistic, calculated as $\sum_{i,j} (y_{ij} - \mu_{ij})^2 / \mu_{ij}$. In 36 per cent of the samples the Pearson statistic from the simulated data was larger than that of the observed data. Thus for both of these discrepancy variables, we find no lack of fit.

5. SIMULATIONS

To evaluate our method, we simulate overdispersed Poisson data with a mean dependent on a sigmoidal shape for $w(d)$. We then estimate $w(d)$ with our method using both the I-spline and piecewise linear basis functions. Specifically, we simulate 50 data sets each with 157 (the number of glands for which we have both dose data and saliva flows) independent Poisson variables with means μ_i given by

$$\log(\mu_i) = a - \sum_{k=1}^{k=164} w(d_k) p_i(d_k) + b_i \quad (22)$$

where $a = 5$ and $b_i \sim N(0, 1)$. The DVHs were simulated as truncated normal densities with means drawn uniformly in (5, 76.5) and standard deviations drawn uniformly on (1, 10). For each of the 50 simulated data sets we estimated $w(d) - a$. Figure 4 shows the mean and the 5th and 95th percentiles of these 50 estimates using I-spline basis functions along with the underlying true value. Similar results were obtained from piecewise linear basis functions. Overall, the method seems to work well with little bias except in the tails. In addition, the percentiles are close together and indicate that the estimated $w(d)$ is reasonably stable.

We also fit the model from Schipper *et al.* [10] to these same 50 data sets in order to compare the efficiency of the two estimators of $w(d)$. Table II contains the standard deviations of the 50 estimated values of $w(d)$ at various doses for each of the estimators. Despite using a different model and estimation procedure, the standard deviations for the method in Schipper *et al.* [10] are generally similar to those of the Bayesian models. As expected, the variability of each estimator is highest at 40 Gy where $w(d)$ is changing most rapidly. The variability of the two Bayesian estimators is also generally similar although the linear basis functions lead to more variability near

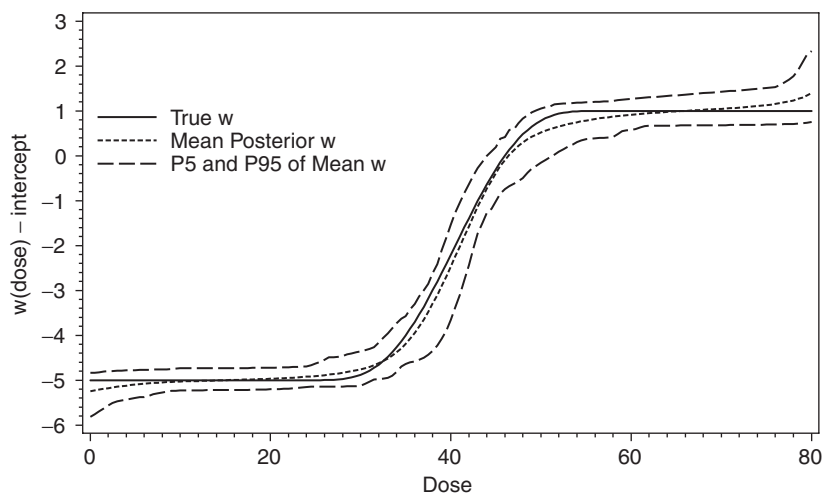


Figure 4. Simulation results for Poisson mixed data. The short dashed line is the mean of the 50 estimates of $w(d)$. The long dashed lines are the 5th and 95th percentiles of the 50 estimates.

Table II. Standard deviations of estimated $w(d)$ from 50 simulated data sets for the maximum likelihood (ML) estimator in Schipper *et al.* [10] and Bayesian posterior mean estimators.

Dose (Gy)	ML STD	Bayesian STD	
		I-spline basis functions	Linear basis functions
10	0.10	0.16	0.17
20	0.18	0.17	0.19
30	0.24	0.22	0.29
40	0.68	0.67	0.82
50	0.32	0.35	0.32
60	0.33	0.22	0.34
70	0.34	0.25	0.37
80	0.36	0.44	0.44

40 Gy. This is not surprising as the I-spline method has the advantage that the ‘true’ $w(d)$ used to simulate the data sets was created using 20 I-spline basis functions. We found that all of the methods can give poor estimates at 80 Gy, which is at the edge of the design space.

6. DISCUSSION

We have presented a flexible method for non-parametric and monotone estimation of a dose effect within a Bayesian GMFMM. Estimation of the model parameters can be accomplished via Gibbs sampling and is implemented in a standard program (WinBUGS). Our general summary measure of the dose effect allows us to estimate what aspect of the dose distribution best predicts

complications. Because the doctor has some control over the dose distribution $p(d)$, having an estimate of $w(d)$ would allow a dose distribution to be chosen to minimize $\int p(d)w(d) dd$, and thus keep the expected saliva flow rate at a high level. Of course the desire to minimize $\int p(d)w(d) dd$ must be balanced with other considerations, including the need to keep the dose to the tumour high.

Of the two basis functions we considered, the I-splines result in a smoother, and thus more visually appealing and biologically plausible, estimate of $w(d)$. The drawback of using the I-spline basis functions is that computing the probability of flat $w(d)$ is not as straightforward. Thus if primary interest is in estimating the posterior probabilities of flat $w(d)$, the linear basis functions are preferable. It is the overlapping nature of the I-Spline basis functions that both results in smoothness and makes computing these probabilities problematic. However, estimating the probability of flat $w(d)$ by $P(\delta=0)$ using either linear or I-spline basis functions ignores information in β . Thus, we prefer to assess flatness visually using the estimated $w(d)$ and pointwise CI's.

The model given in (1) is an example of a dose-damage-injury model [19, 20]. In these models, the damage to the tissue is considered to be a deterministic function of the DVH. This damage need not be directly interpretable in terms of the anatomical structure of the organ. Rather it can be viewed as a single number which summarizes the effect of the dose of radiation received. The damage is not observable but is responsible for causing the injury, loss of saliva flow in our case, which is observable. Denote the damage by D and the injury by Y . Then (1) can be seen to be a special dose-damage-injury model relating μ , the expected injury level, to the damage $D = \int p(d)w(d) dd$. Rather than including D linearly in (1), we might more generally include $H(D)$ where $H(\cdot)$ is a function to be estimated. Future work could be directed at investigating conditions under which $H(\cdot)$ and $w(\cdot)$ may both be non-parametrically estimated without loss of identifiability. These conditions will depend in part on the observed DVHs. An alternative is to fix $w(\cdot)$, possibly to a linear form (equivalent to meandose), and then only $H(\cdot)$ need be estimated. Generalized additive mixed models [21] could be used to estimate $H(\cdot)$ non-parametrically. To estimate $H(\cdot)$ both non-parametrically and monotonically, the model of Neelon and Dunson [13] could be directly applied.

ACKNOWLEDGEMENTS

The authors wish to thank the associate editor and two reviewers for helpful and constructive comments. This work was partially supported by the following grants from the US National Institutes of Health, CA-095096, CA-97248 and CA-83654.

REFERENCES

1. Martel M, Sahijdak W, TenHaken R. Fraction size and dose parameters related to the incidence of pericardial effusions. *International Journal of Radiation Oncology Biology Physics* 1998; **40**:155–161.
2. Price RE, Ang KK, Stephens LC, Peters LJ. Effects of continuous hyperfractionated accelerated and conventionally fractionated radiotherapy on the parotid and submandibular salivary glands of rhesus monkeys. *Radiotherapy Oncology* 1995; **36**(2):156–157.
3. Marks L. The impact of organ structure on radiation response. *International Journal of Radiation Oncology Biology Physics* 1996; **34**:1165.
4. Kutcher GJ, Burman C, Brewster L. Calculation of complication probability factors for non-uniform tissue irradiation: the effective volume method. *International Journal of Radiation Oncology Biology Physics* 1989; **16**:1623–1630.

5. Lyman JT. Complication probability as assessed from dose–volume-histograms. *Radiation Research* 1985; **104**: s13–s19.
6. Ramsay JO, Silverman BW. *Functional Data Analysis* (2nd edn). Springer: Berlin, 2005.
7. Hastie T, Mallows C. Discussion of A statistical view of some chemometrics regression tools. *Technometrics* 1993; **35**:140–143.
8. James GM. Generalized linear models with functional predictors. *Journal of the Royal Statistical Society, Series B* 2002; **64**:411–432.
9. Zhang D, Lin X, Sowers MF. Two-stage functional mixed models for evaluating the effect of longitudinal covariate profiles on a scalar outcome. *Biometrics* 2006, in press. doi: 10.1111/j.1541-0420.2006.00713.x
10. Schipper MJ, Taylor JMG, Lin X. Generalized monotone functional mixed models with application to modeling normal tissue complications. *Journal of the Royal Statistical Society, Series C*, under revision.
11. Dunson DB. Bayesian semiparametric isotonic regression for count data. *Journal of the American Statistical Association* 2005; **100**:618–627.
12. Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* 1993; **88**:9–25.
13. Neelon B, Dunson DB. Bayesian isotonic regression and trend analysis. *Biometrics* 2004; **60**:398–406.
14. Ramsay JO. Monotone regression splines in action. *Statistical Science* 1988; **3**:425–461.
15. Zeger SL, Karim MR. Generalized linear models with random effects: a Gibbs sampling approach. *Journal of the American Statistical Association* 1991; **86**:79–86.
16. Dellaportas P, Smith AF. Bayesian inference for generalized linear and proportional hazards models via Gibbs sampling. *Journal of the Royal Statistical Society, Series C* 1993; **42**:443–459.
17. Gilks WR, Wild P. Adaptive rejection sampling for Gibbs sampling. *Journal of the Royal Statistical Society, Series C* 1992; **41**:337–348.
18. Gelman A, Meng XL. Model checking and model improvement. In *Markov Chain Monte Carlo in Practice*, Gilks WR, Richardson S, Spiegelhalter DJ (eds). Chapman & Hall: London, 1998; 189–200.
19. Jackson A, Kutcher GJ, Yorke ED. Probability of radiation induced complications for normal tissues with parallel architecture subject to non-uniform radiation. *Medical Physics* 1993; **20**:613–625.
20. Yorke ED, Kutcher GJ, Jackson A, Ling CC. Probability of radiation-induced complications in normal tissue with parallel architecture under conditions of uniform whole or partial organ irradiation. *Radiotherapy and Oncology* 1993; **26**:226–237.
21. Lin X, Zhang D. Inference in generalized additive mixed models by using smoothing splines. *Journal of the Royal Statistical Society, Series B* 1999; **61**:381–400.