Generalized monotonic functional mixed models with application to modelling normal tissue complications

Matthew Schipper,

Innovative Analytics, Kalamazoo, USA

Jeremy M. G. Taylor

University of Michigan, Ann Arbor, USA

and Xihong Lin

Harvard School of Public Health, Boston, USA

Summary. Normal tissue complications are a common side effect of radiation therapy. They are the consequence of the dose of radiation that is received by the normal tissue surrounding the site of the tumour. Within a specified organ each voxel receives a certain dose of radiation, leading to a distribution of doses over the organ. It is often not known what aspect of the dose distribution drives the presence and severity of the complications. A summary measure of the dose distribution can be obtained by integrating a weighting function of dose ($w(d)$) over the density of dose. For biological reasons the weight function should be monotonic. We propose a generalized monotonic functional mixed model to study the dose effect on a clinical outcome by estimating this weight function non-parametrically by using splines and subject to the monotonicity constraint, while allowing for overdispersion and correlation of multiple obervations within the same subject. 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.

Keywords: Dose effect; Functional data; Monotonicity; Non-parametric regression; Normal tissue complications; Overdispersion; Splines

1. Introduction

Radiation therapy is commonly used to treat cancer. The goal is to deliver the highest possible dose to the site of a tumour and the lowest possible dose to the surrounding normal tissue. Higher doses to the tumour result in more damage to the cancer cells, whereas higher doses to the normal tissue cause damage that can lead to normal tissue complications. Pneumonitis is an example of a serious but rare normal tissue complication that is experienced by lung cancer patients. Other examples include rectal failure in colon cancer patients and xerostoma (loss of saliva production) in head and neck cancer patients. There are many potential treatment plans depending on the number, direction and intensity of the radiation beams. In choosing a treatment plan, the physician must trade off maximizing damage to the tumour with minimizing damage to the surrounding tissue. To do this efficiently, it is necessary to understand precisely

Address for correspondence: Matthew Schipper, Innovative Analysis, 161 East Michigan Avenue, Kalamazoo, MI 49007, USA.
E-mail: mjschipper@ianalytics.biz

how the dose of radiation to the normal tissue and the probability or the severity of normal tissue complications are related.

Modern treatment planning techniques allow the physician to compute, for a given treatment plan, the spatial distribution of dose within the tissue being irradiated. This dose distribution is typically represented by using the dose volume histogram (DVH) (Lichter, 1991). It is common for the region of the tumour to be given an approximately uniform dose. Such is not so for the surrounding normal tissue where the dose depends on its proximity to the tumour as well as on the treatment plan.

Our motivating example is the head and neck cancer data that were collected at the University of Michigan. This study involved 82 patients with a cancerous tumour in their head or neck who 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 two parotid glands, one in either cheek, are responsible for producing saliva. The doses of radiation to each voxel in the parotid ranged from 0 to 82 Gy and the DVH bin widths are 0.5 Gy. Fig. 1 is an example of the DVHs for the two glands (ipsilateral and contralateral) of a particular subject. For a more detailed description of the data, see Section 4.1. The question of major interest in such settings is how to relate the radiation dose distribution that is received by the normal tissue to the observed complication, e.g. saliva flow. Simple approaches include creating a summary measure of the dose distribution, e.g. the mean dose, and regressing the normal tissue complication on the summary dose measure.

It is of significant interest to estimate the effect of the dose distribution on a normal tissue complication non-parametrically. In other words, we are interested in relating a functional predictor to a scalar outcome. Ramsay and Silverman (1997), chapter 10, considered a functional model that relates a precisely measured functional covariate to a scalar outcome. A more general
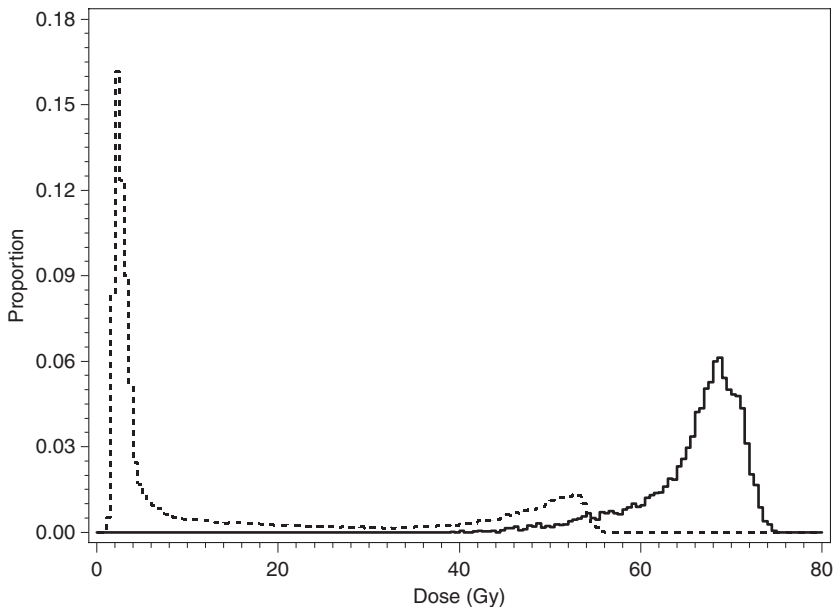

Fig. 1. Example of one subject's DVHs (the ipsilateral gland is the gland on the same side of the head as the tumour, whereas the contralateral gland is on the opposite side of the head): ———, ipsilateral parotid gland; ------, contralateral parotid gland

model is the functional generalized linear model (James, 2002; Zhang *et al.*, 2007), where the observed functional predictor is assumed to be measured over time and with error. Rather than time, our functional predictor is measured over dose. Denote by $p(d)$ the density of the dose distribution, i.e. the fraction of the organ receiving dose less than c is $\int_0^c p(d)\,\mathrm{d}d$. Then a general measure of the dose effect is given by $\int p(d)\,w(d)\,\mathrm{d}d$, where $w(d)$ is a weighting function to be estimated. We can easily see that commonly used dose summaries are special cases of this general summary measure. The mean dose is obtained with a linear weighting function ($w(d) = d$). The partial volume or proportion of tissue receiving dose greater than X Gy is obtained by an indicator function ($w(d) = I(d > X)$).

In this paper, we propose a generalized monotone functional mixed model which relates a multivariate complication outcome to a dose distribution. The dose effect is summarized by using the summary measure that was discussed above. Within this model framework we propose a new method for non-parametric estimation of $w(d)$ subject to two constraints. The first is that any biologically meaningful estimate of $w(d)$ should be monotone since increasing dose cannot lead to lower probability or severity of complications. The second is that $w(0) = 0$. The biological reason for this is clear since zero dose should correspond to no effect. The statistical reason for this constraint is model identifiability. We define $w(d)$ as the integral of a smooth positive function, where the smooth positive function is obtained as a positive transformation of an unconstrained regression spline. Our model differs from the functional generalized linear model in that we wish to estimate $w(d)$ monotonically, and we introduce random effects to accommodate correlation and overdispersion of multivariate complication outcomes. Maximum likelihood is used for model fitting.

There are many approaches to monotone non-parametric regression (Friedman and Tibshirani, 1984; Ramsay, 1988, 1998; Kelly and Rice, 1990; Hall and Huang, 2001; Gelfand and Kuo, 1991; Holmes and Heard, 2003; Dunson, 2005). Existing techniques were generally designed to estimate a monotone relationship between two scalar quantities and hence are not directly applicable to relating a scalar outcome to a functional predictor (dose distribution). Although some of the existing techniques could potentially be modified to this setting, they generally impose monotonicity through constraints on the parameter space. Such constraints complicate estimation and inference with the possibility of estimates on the boundary of the parameter space. In contrast, our proposed method requires no such constraints.

This paper is organized as follows. Section 2 describes the model as well as the formulation of $w(d)$. Section 3 discusses estimation of the model. We describe in detail the head and neck cancer saliva data in Section 4 and apply the proposed methods to analyse the saliva data. Section 5 presents simulation results, followed by discussion in Section 6.

2.   The generalized monotonic functional mixed model

Let Y_{ij} denote the complication outcome for the jth observation ($j = 1, \ldots, J$) of the ith subject ($i = 1, \ldots, n$). We assume that, conditionally on random effects b_i, Y_{ij} follows an exponential family distribution (McCullagh and Nelder, 1989) with mean μ_{ij} given by

$$g(\mu_{ij}) = X_{ij}^{\mathrm{T}}\alpha + \int_0^D p_{ij}(d)\,w(d)\,\mathrm{d}d + Z_{ij}^{\mathrm{T}}b_i \qquad (1)$$

where $g(\cdot)$ is a link function, α and b_i represent fixed and random effects with corresponding design matrices X_{ij} and Z_{ij}, and $p_{ij}(d)$ is the density of the dose distribution; $[0, D]$ is the range of dose. We further assume that random effects b_i are independent and each follows $N(0, \Sigma)$. The marginal likelihood is given by

$$L = \prod_{i=1}^{n} \int \prod_{j=1}^{J} f(Y_{ij}|b_i) f(b_i) \, db_i, \tag{2}$$

where $f(Y_{ij}|b_i)$ is the conditional density of $Y_{ij}|b_i$ and $f(b_i)$ is the normal likelihood density of b_i. For the head and neck cancer data, the outcome is saliva flow, which is measured separately from each subject's two parotid glands and hence $J = 2$. Since saliva flows are measured as rates, we assume that $Y_{ij}|b_i$ follows a Poisson distribution in our analysis. We could alternatively make a weaker assumption by using the quasi-likelihood with the conditional mean $E(Y_{ij}|b_i) = \mu_{ij}$ and the conditional variance $\mathrm{var}(Y_{ij}|b_i) = \mu_{ij}$ (McCullagh and Nelder, 1989; Breslow and Clayton, 1993). The random effects b_i account for correlation between bivariate observations within subject as well as possible overdispersion. Since α includes an intercept term and $p(d)$ is a density function satisfying $\int p(d) \, dd = 1$, we need to constrain $w(d)$ for identifiability. Specifically, if $w(d)$ is an arbitrary constant w, then

$$\int p(d) w(d) \, dd = w,$$

which would confound with the intercept α_0. We hence constrain $w(d)$ by $w(0) = 0$.

As discussed in Section 1, biologically, $w(d)$ needs to be monotone increasing. A convenient way to impose this constraint is to define $w(d)$ as the integral of a non-negative function. We consider the formulation

$$w(d) = \int_0^d s(c) \, dc \tag{3}$$

where

$$s(d) = h\{r(d)\} \tag{4}$$

and

$$r(d) = \sum_{l=1}^{L+4} \beta_l B_l(d). \tag{5}$$

Here $h(\cdot)$ is a known, non-negative continuous function defined on the real line, the $B_l(\cdot)$s are known basis functions which are defined by using L interior knots, e.g. B-spline bases, and the β_l are unknown parameters. In this formulation, equation (5) imposes smoothness whereas equations (3) and (4) impose monotonicity. No constraint is needed for the βs.

It follows that model (1) can be rewritten as

$$g(\mu_{ij}) = X_{ij}^{\mathrm{T}}\alpha + \int_0^D p_{ij}(d) \int_0^d h\left\{ \sum_{l=1}^{L+4} \beta_l B_l(c) \right\} dc \, dd + Z_{ij}^{\mathrm{T}} b_i. \tag{6}$$

In the absence of the functional effect $\int p(d) w(d) \, dd$, model (1) specifies a generalized linear mixed model (Breslow and Clayton, 1993). However, it can be seen from model (6) that, since $h(\cdot)$ can be non-linear, the second term need not be a linear function of β_l. Thus the dose summary measure in model (1) may also not be linear in β_l and model (6) is a functional extension of the generalized linear mixed model.

Fig. 2 demonstrates the flexibility of the proposed functional dose summary measure $\int p(d) \times w(d) \, dd$ by illustrating a few of the possible shapes for $w(d)$ using this formulation, as well as the underlying $r(d)$ and $s(d)$ that result in these shapes. We here set $h(x) = 1000 \, \phi(x)$, where $\phi(\cdot)$ is the standard normal density. These results show that our model can cover a wide range of shapes of $w(d)$. The flexibility of $w(d)$ depends on the flexibility of $r(d)$. We approximate $r(d)$ by using a regression spline that is specified by using a small number of knots. In simulations
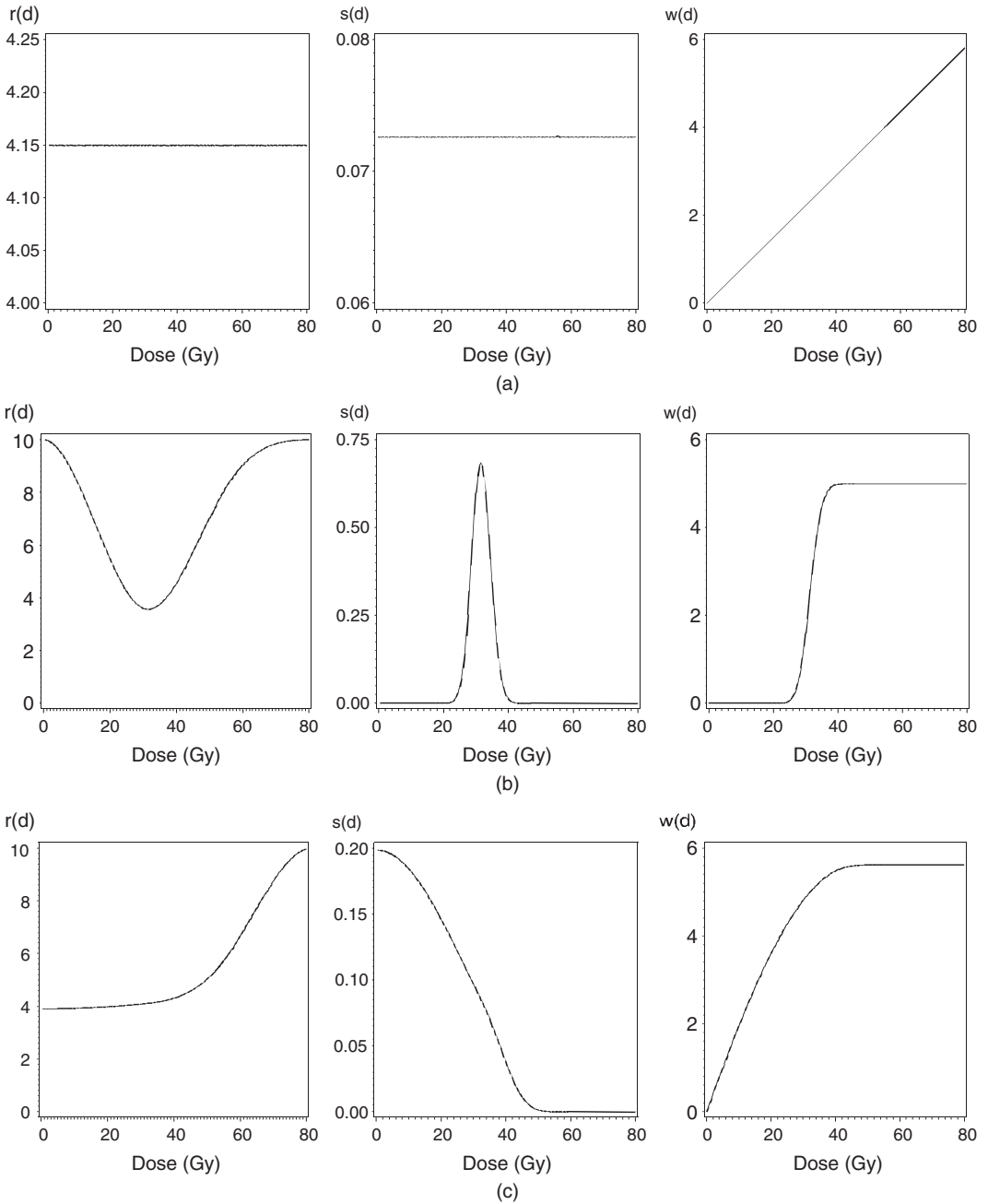
Fig. 2.   Three examples of $w(d)$ with the corresponding $r(d)$ and $s(d)$: (a) linear; (b) sigmoidal; (c) plateau

and the saliva flow example, we found that relatively few (two or three) knots worked well. A standard choice for knot placement is to place the knots by using percentiles of the data.

Monotonicity of $w(d)$ requires only that $s(d)$ and hence $h(\cdot)$ be non-negative. Possible choices for $h(z)$ include $\exp(z)$, z^2 or density functions, e.g. normal or logistic density functions. Note that if a density function is bounded, e.g. the normal density $\phi(\cdot)$, we need to multiply it by

a large constant so as not to restrict the range of $s(d)$. Since $r(d)$ is estimated non-parametrically, the estimates of $w(d)$ are robust to the choice of $h(\cdot)$. Our numerical results confirm this. However, if we wish to test the null hypothesis of no dose effect, it is desirable to choose $h(\cdot)$ so that $h(0) = 0$, e.g. $h(z) = z^2$ or $h(z) = a\{\phi(0) - \phi(z)\}$ if the normal density is used, where a is an arbitrary scale constant to make the range of $s(d)$ practically unconstrained. Then no dose effect occurs when $\beta_1 = \beta_2 = \ldots = \beta_{L+4} = 0$ and a likelihood ratio test can be used to test this null hypothesis. Another special case of our model is when $\beta_1 = \beta_2 = \ldots = \beta_{L+4}$, which corresponds to a mean dose model since $\Sigma_l^{L+4} B_l(t) = 1$ for all t. A likelihood ratio test can also be used to test whether mean dose is a suitable summary of the dose effect. Using a symmetric $h(\cdot)$ function technically results in non-identifiability of the signs of $r(d)$ since $h(z) = h(-z)$. In practice this will generally not be an issue and does not affect the estimator of $w(d)$.

3. Estimation

The specification of the generalized monotone functional mixed model (1) assumes that the dose distribution $p(d)$ is known and continuous. In practice only the empirical distribution of $p(d)$, which is called the DVH, is observed. The DVH discretizes dose into K bins and measures the proportion of the gland that received a dose within each bin. Fig. 1 shows two example DVHs of the saliva data of a representative subject.

Using the DVH as an empirical estimate of $p(d)$, we replace model (1) by

$$g(\mu_{ij}) = X_{ij}^{\mathrm{T}}\alpha + \sum_{k=1}^{K} p_{ij}(d_k)\, w(d_k) + Z_{ij}^{\mathrm{T}}b_i \qquad (7)$$

where K is the number of DVH bins, $d_k = (k-1)$ bin width, i.e. d_k is the left end point of the kth bin and $p(d_k)$ is the proportion of normal tissue receiving dose in bin k. We also approximate the integral in equation (3) as $w(d_1) = 0$ and

$$w(d_k) \approx \sum_{l=2}^{k} s(d_l)\, \Delta d$$

for $k = 2, \ldots, K$, where Δd is the bin width. Plugging this into model (7), we obtain

$$\begin{aligned} g(\mu_{ij}) &= X_{ij}^{\mathrm{T}}\alpha + \sum_{k=2}^{K} p_{ij}(d_k)\left\{ \sum_{m=2}^{k} s(d_m)\, \Delta d \right\} + Z_{ij}^{\mathrm{T}}b_i \\ &= X_{ij}^{\mathrm{T}}\alpha + \sum_{k=2}^{K} p_{ij}(d_k)\left[\sum_{m=2}^{k} h\left\{ \sum_{l=1}^{L+4} \beta_l\, B_l(d_m) \right\} \Delta d \right] + Z_{ij}^{T}b_i. \end{aligned} \qquad (8)$$

Estimation in this model proceeds with the maximum likelihood method which maximizes the likelihood function (2), where the conditional likelihood of $L(Y_{ij}|b_i)$ is specified by using the mean model (8) under the exponential family. The random effects b_i are assumed to be $N(0, \Sigma)$. In contrast with the generalized linear mixed model (Breslow and Clayton, 1993), where regression coefficients enter the linear predictor linearly, regression coefficients β_ls in model (8) enter $g(\mu_{ij})$ non-linearly. The marginal likelihood (2) generally does not have a closed form expression. We numerically approximate the integrated likelihood (2) by using adaptive quadrature (Pinheiro and Bates, 1995) and estimate the regression coefficients (α, β) and the variance components Σ by using the Newton–Raphson method. The maximum likelihood estimation can be implemented in SAS by using procedure NLMIXED, which allows the specification of an arbitrary mean function.

4. Application to head and neck cancer data

4.1. Description of the data

In this section we fit the generalized monotone functional mixed model to the head and neck cancer data that were introduced in Section 1. 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 either cheek, are responsible for producing saliva. The doses of radiation to each voxel in the parotid ranged from 0 to 82 Gy and the DVH bin widths are 0.5 Gy; thus $K = 164$. Fig. 1 is an example of the DVHs for the two glands of a particular subject. Clearly one gland of this subject received a much larger dose than the other. This is common because in treatment planning the dose to the gland on the opposite side from the tumour (contralateral) is usually minimized, whereas no such attempt is made for the gland on the same side as the tumour (ipsilateral). Saliva flows are measured separately from both cheeks. We study in this paper the effect of the dose distribution on saliva flow at 6 months after the baseline. This analysis involved 157 saliva flow measurements (with corresponding DVHs) from 82 subjects.

4.2. Model description

The saliva flows are measured as rates (millilitres per minute), which suggests a Poisson distribution for the outcome distribution. Since the measured saliva flows are not integers, we multiplied the original flows by 150 and rounded to the nearest integer. We used 150 because it was sufficiently large to ensure that meaningful differences in saliva flow resulted in different transformed values, but sufficiently small to prevent a highly discrete distribution with gaps that are inconsistent with Poisson data. Continuous distributions such as the gamma distribution are not appropriate since roughly 30% of the saliva flows are exactly 0. Let Y_{ij} denote the rescaled saliva flow from gland j ($j = 1$ for the ipsilateral gland; $j = 2$ for the contralateral gland) of subject i ($i = 1, 2, \ldots, 82$). Let $b_i = (b_{i1} \quad b_{i2})^\tau$ denote the subject level random effects assumed to be independently distributed as $N(0, \Sigma)$, where Σ is an unstructured covariance matrix. Further, denote the log(baseline saliva flow) from gland j of subject i by base_{ij}. We assume that, conditionally on the random effects, the saliva flows follow a Poisson distribution with mean μ_{ij} given by

$$\log(\mu_{ij}) = \alpha_0 + \alpha_1 \, \text{base}_{ij} - \sum_{k=2}^{164} p_{ij}(d_k) \, w(d_k) + b_{ij} \tag{9}$$

where $p_{ij}(\cdot)$ is the DVH for the jth gland of the ith subject. We give a minus sign to the dose effect term because we wish to interpret this term as damage. With this parameterization, higher damage implies lower expected saliva flow. Exploratory analyses indicate that the marginal variance is approximately proportional to the square of the marginal mean. The gland-specific random effect b_{ij} approximates this marginal variance structure and also allows us to account for different degrees of overdispersion between the two sides. There was some thought by the physicians that there might be compensation between the two glands, so that, if one gland was heavily damaged, the other gland would compensate by producing more saliva. For this reason we allow the off-diagonal element in Σ, which represents the covariance between the two sides, to be estimated. Note that including a subject level random effect would only allow positive correlation between the sides and does not allow for different degrees of overdipersion of the two sides. Besides dose, we have only included baseline saliva flow as a covariate. Clinical covariates such as age, gender, chemotherapy, medications, comorbidity and surgery before

radiation therapy were previously found not to impact saliva flow (Eisbruch *et al.*, 1999). We used a regression spline with two interior knots at 30 and 60 Gy for $r(d)$. As discussed in Section 2, the two-knot cubic regression spline formulation for $r(d)$ is parameterized by six *B*-spline basis functions and we fit

$$\log(\mu_{ij}) = \alpha_0 + \alpha_1 \operatorname{base}_{ij} - \sum_{k=2}^{164} p_{ij}(d_k) \sum_{m=1}^{k} h\left\{ \sum_{l=1}^{6} \beta_l B_l(d_m) \right\} + b_{ij} \qquad (10)$$

where β_j are unconstrained parameters and $B_j(\cdot)$ are six *B*-spline basis functions defined on $(0, 82 \,\mathrm{Gy})$. Here we used $h(z) = 1000 \, \phi(z)$, where $\phi(\cdot)$ is the standard normal density. We fit model (10) by using the maximum likelihood method that was discussed in Section 3.

These data and earlier versions of them have previously been analysed by other investigators by assuming strong parametric models. Eisbruch *et al.* (1999) considered two models. The first was a generalized linear model in which the dose effect was summarized by a threshold mean dose effect and the outcome was the observed saliva flow rate. They also considered a normal tissue complication probability (NTCP) model (Lyman, 1985; Kutcher *et al.*, 1989) for binary outcomes. The NTCP model is a probit regression model which relates a DVH to the probability of complication by using three parameters. Patients whose post-baseline salivary flow was at or below 25% of their salivary flow at baseline were considered to have xerostoma. Subjects who did not meet this criterion were considered not to have xerostoma. Johnson *et al.* (2005) proposed a complex Bayesian model where the dose effect was assumed to be captured by a parametric model by using a percentile of the dose distribution (the DVH). Our approach differs from those above primarily in how it models the dose effect. Rather than assuming that the damage is given by the mean dose or a percentile of the dose distribution, we use the general functional summary measure $\int p_{ij}(d) \, w(d) \, \mathrm{d}d$ and estimate the dose effect $w(d)$ non-parametrically. Also, unlike the NTCP model, our model does not require a binary outcome and uses the observed saliva flow rates.

4.3. Results

The maximum likelihood estimate and 95% pointwise confidence intervals of $w(d)$ are shown in Fig. 3. Very similar results are obtained with different choices for $r(d)$ (two knots at 25 and 50 Gy or three knots at 20, 40 and 60 Gy) and $h(z)$ ($\exp(z)$ or z^2). The pointwise confidence intervals are calculated by using a delta method estimate of the standard error. To interpret the estimated $w(d)$, it is helpful to note that regression coefficients have both subject-specific and population-average interpretations in Poisson–normal random-intercept models. In other words, the marginal means still have the same exponential form except that the intercept is changed (Breslow and Clayton, 1993). Consider two parotid glands, say gland A and gland B, which are of the same type (ipsilateral or contralateral) but differ only in the dose of radiation received. Assume that gland A receives a uniform dose d_c and gland B receives no dose of radiation. Let μ_A and μ_B be the expected saliva flows for glands A and B respectively. Then, from equation (9), some calculations show that $\log(\mu_A) - \log(\mu_B) = -w(d_c)$. Thus $w(d_c)$ could be interpreted as the difference in expected saliva flow (on the log-scale) between a gland that received uniform dose d_c and a gland that received no dose of radiation. Similarly, $\exp\{-w(d_c)\}$ can be interpreted as the ratio μ_A/μ_B.

Our results show that the estimated $w(d)$ is flat for the first 40 Gy. The clinical implication of this finding is that the parotid gland could be uniformly irradiated at 40 Gy with minimal effect on saliva flow. It is also interesting that the estimated $w(d)$ is not linear, implying that mean dose does not seem to be a suitable summary of the DVH. We can test this by using a likelihood
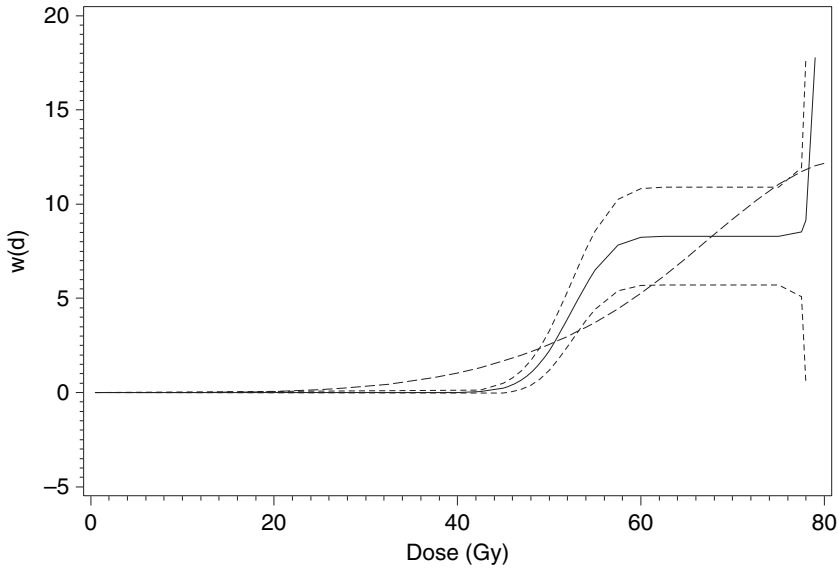
Fig. 3. Maximum likelihood estimate and pointwise 95% confidence intervals of $w(d)$: ———, generalized monotone functional mixed model estimate; - - - - - -, 95% confidence interval; − − −, estimate by using *I*-splines

Table 1. Parameter estimates and standard errors for the saliva flow data

Parameter	Estimate	Standard error
α_0	1.23	0.09
α_1	0.72	0.02
Σ_{11}	11.77	3.01
Σ_{12}	0.005	0.17
Σ_{22}	0.72	0.12

ratio test by testing $H_0 : \beta_1 = \beta_2 = \ldots = \beta_6$. The χ^2-statistic with p-value in parentheses for the likelihood ratio test is 29.9 ($p < 0.0001$). Thus we strongly reject the mean dose model in favour of our non-parametric model.

Estimates of the baseline saliva effect and the variance components are shown in Table 1. The results show that baseline saliva flow is a significant factor for saliva flow at month 6. Overdispersion for the ipsilateral and contralateral glands is captured by $\Sigma_{1,1}$ and $\Sigma_{2,2}$ respectively. Overdispersion is significant for both but much larger for the ipsilateral glands. The correlation between glands could be measured by using the covariance parameter estimates $\Sigma_{12}/\sqrt{(\Sigma_{11}\Sigma_{22})} = 0.002$, which is not significantly different from 0. A significantly negative Σ_{12} would have given support to the notion of compensation whereas a significantly positive correlation would indicate that the two glands within a subject tend to behave similarly. We thus find no support for the notion of compensation between glands.

Fig. 4 plots, for each gland, the log-proportion change in saliva flow against the estimated summary measure of the dose effect $\Sigma_{k=2}^{k=164} p_{ij}(d_k)\,\hat{w}(d_k)$. Because some of the saliva flows are 0, we added 1 before taking logarithms. This plot does not exactly correspond to the model
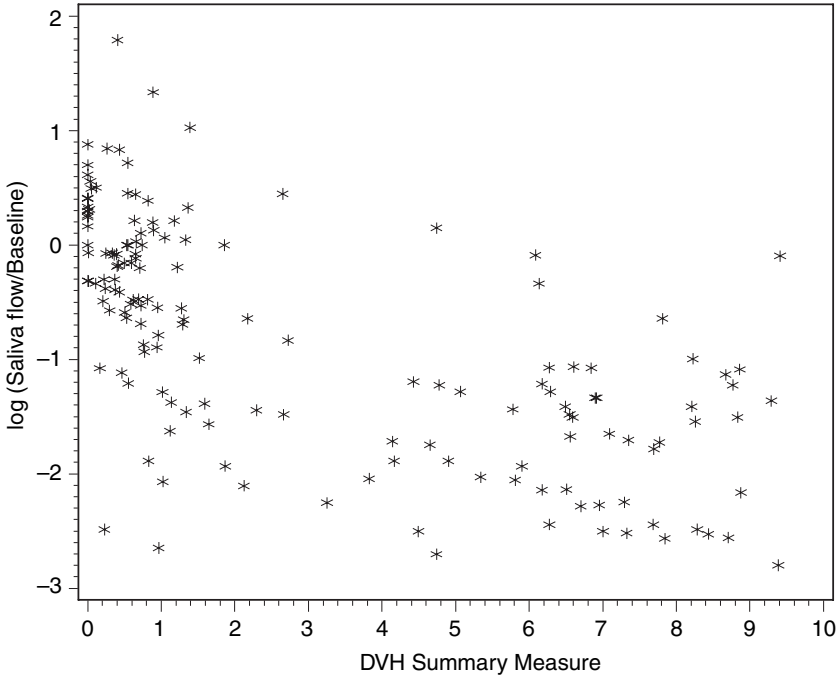
Fig. 4. Change in log-saliva-flow from baseline *versus* estimated dose effect summary measure

given by equation (9) since $\hat{\alpha}_1 \neq 1$. However, the plot changes very little if instead we plot $\log\{(Y_{ij} + 1)/(\hat{\alpha}_1 \, \text{base}_{ij} + 1)\}$. There is clearly a decreasing and approximately linear trend. It is also apparent that there is substantial variation in the change in saliva flows from baseline, even at a fixed dose effect level. Some of this variation is due to variability between glands and subjects.

The NTCP model is equivalent to a probit regression model with a power law mean dose term for the linear predictor:

$$\text{NTCP}_{ij} = \Phi\left[\beta_0 + \beta_1\left\{\sum_{k=2}^{164} d_k^\lambda \, p_{ij}(d_k)\right\}^{1/\lambda}\right] \tag{11}$$

where NTCP_{ij} is the probability of complication for the jth gland of the ith subject. This model cannot be directly compared with the generalized monotonic functional mixed model that we used to analyse the saliva data because it applies to binary outcomes. It is, however, possible to compare how each model summarizes the dose effect. To do so we replace the summary measure of dose effect in model (10) with the summary measure of dose effect in the NTCP model and fit the model with mean given by

$$\log(\mu_{ij}) = \alpha_0 + \alpha_1 \, \text{base}_{ij} + \alpha_2\left\{\sum_{k=2}^{164} d_k^\lambda \, p_{ij}(d_k)\right\}^{1/\lambda} + b_{ij}. \tag{12}$$

The resulting $-2\log$-likelihood value is 1081.8, which is larger than 1075.5, the value that was obtained from our generalized monotonic functional mixed model. Because the two models are not nested a likelihood ratio test is not possible.

An alternative approach to estimating $w(d)$ monotonically is to approximate $w(d)$ directly by using a regression spline with I-spline basis functions (Ramsay, 1988) as $w(d) = \sum_{l=1}^{L+4} \beta_l I_l(d)$, where the β_ls are constrained to be positive and $I_l(d) = \int_0^d B_l(c) \, \mathrm{d}c$ are the I-spline basis functions defined as the integral of the B-spline basis functions and are thus monotone increasing. We used this approach to estimate $w(d)$ for the month 6 saliva flow data. Of the six parameters, two were estimated to be at the boundary value of 0. The resulting estimate of $w(d)$ is shown in Fig. 3 and is somewhat different from the generalized monotonic functional mixed model estimate. Rather than being flat for doses up to 50 Gy, the I-spline estimator starts to increase after 30 Gy and is more linear thereafter than our estimate. We show in the simulation study in Section 5 that I-splines do not work well for estimating flat regions. The associated $-2\log$-likelihoods are 1105.4, 1079.5 and 1075.5 for models that are based on the mean dose, I-spline and our method respectively. The I-spline model and our model have the same number of parameters and likelihood values favour our model, although we cannot test this, since the two models are not nested.

4.4. Goodness of fit

Goodness-of-fit statistics are used to assess whether the model fits the data reasonably well. One commonly used such statistic is the Pearson X^2-statistic and is computed as

$$\sum_i \sum_j (y_{ij} - \mu_{ij})^2 / \mu_{ij} \tag{13}$$

where $\mu_{ij} = E[y_{ij}|b_i]$ is the conditional mean and is obtained from equation (9) by plugging in estimates of the fixed effect parameters α and β and of the random-effect estimates b_i. Conditional on the random effect b_i, Y_{ij} follows a Poisson distribution and so the variance is simply equal to the (conditional) mean and thus we divide by μ_{ij}. There are two variants on statistic (13) that we also consider. The first is obtained with the same numerator but a different standardization. Rather than dividing by μ_{ij}, we divide by the conditional mean-squared error of prediction (Booth and Hobert, 1998). Doing so accounts for the variability in the estimates of the random effects. The second is to subtract the marginal mean and to divide by the marginal variance of Y_{ij}. For a detailed discussion of goodness of fit in generalized non-linear mixed models, and in particular how residuals that are based on marginal and conditional means are sensitive to different model assumptions, see Vonesh *et al.* (1996). The null distribution of these statistics is not known, so we use a parametric bootstrap approach to obtain empirical estimates. Specifically we simulate 100 data sets of the same size as our observed data set, using the parameter estimates from the observed data. For each data set we fit the model and compute the goodness-of-fit statistic. Of the 100 simulated data sets, 21 had goodness-of-fit statistics (based on the marginal mean and variance) that were larger than that of our observed data. Results are similar (*p*-values 0.55 and 0.15) for the other two goodness-of-fit statistics. Thus, on the basis of these metrics, our proposed model appears to be consistent with the data.

5. Simulations

To evaluate our method, we simulate Poisson data with a mean that is dependent on a variety of shapes for $w(d)$. We then estimate $w(d)$ with our method where $r(d)$ is a cubic regression spline with knots at 30 and 60 Gy, and $h(x) = 1000\,\phi(x)$, where $\phi(\cdot)$ is the standard normal density. Specifically, for each of the three shapes for $w(d)$ in Fig. 1, we simulate 100 data sets each with 157 independent Poisson variables with means μ_i given by

$$\log(\mu_i) = \alpha_0 - \sum_{k=2}^{k=164} w(d_k)\, p_i(d_k) \tag{14}$$

where we set $\alpha_0 = 5$, which is roughly the mean of the sum of the estimated intercept and baseline flow terms in the observed data. To mimic the saliva data, the DVHs were simulated as normal densities (truncated to lie in (0,82)) with means drawn uniformly in (5,76.5) and standard deviations drawn uniformly on (1,10). Fig. 5 shows the mean and median of the estimated values of $w(d) - \alpha_0$ for each shape along with the underlying true $w(d) - \alpha_0$. We have used $w(d) - \alpha_0$ because estimates of this quantity are more stable than estimates of $w(d)$ alone.

We next conduct a simulation study to assess the performance of this method for overdispersed Poisson data such as the saliva flow data. We simulate data by using the model

$$\log(\mu_i) = \alpha_0 - \sum_{k=2}^{k=164} w(d_k)\, p_i(d_k) + b_i$$

where we set $\alpha_0 = 5$ and the random effects $b_i \sim N(0, 1)$. We simulate 100 data sets of size 157 for each of the three different shapes of $w(d)$. The results are shown in Fig. 5. For both the Poisson and the overdispersed Poisson data, there is little bias except in the tails. There are relatively little data in these regions for some of the simulated data sets so the estimates tend to be more variable with a few extreme estimates having a large effect on the mean.

We wished to assess the robustness of our proposed estimator to the choice of $h(\cdot)$. To do so we analysed the Poisson data sets that were discussed above, which were generated by using $h(z) = 1000\,\phi(z)$, with $h(z) = z^2$. The results are in Fig. 5, where we see no evidence of bias from using a different $h(\cdot)$ function in the analysis. We also wished to assess the robustness of our proposed estimator to the choice of the number of knots that are used to define $r(d)$. To do so we generated 100 simulated data sets as discussed above by using a three-knot regression spline formulation for $r(d)$ with knots at 20, 40 and 60 Gy. We then analysed these data sets by using a two-knot regression spline with knots at 30 and 55 Gy. The results (which are not shown) indicate a good fit with no evidence of bias.

To illustrate the increased flexibility in our estimator compared with the I-spline estimator, we consider the extreme case where $w(d)$ is a step function. We simulate a single Poisson data set based on this $w(d)$ from equation (14). We then fit both our model and the I-spline-based model to these data. The resulting estimates are shown in Fig. 6. The I-spline method severely oversmooths this step function. The reason for this is that the slope and maximum value of the I-regression spline are linked. To obtain a steeper function, we need a larger parameter coefficient, which also means a larger maximum value. It is thus not possible to obtain a 'small steep' increase. To be fair, our method is also not ideally suited to estimating step functions. Extremely large parameter values are required to do so. Our method also suffers from a lack of identifiability when $w(d)$ is a step function, because there are many sets of parameters that yield nearly identical estimates. This complicates tasks such as computing confidence intervals for $w(d)$. Nevertheless, the method can indicate whether the true function being estimated is a step function.

6. Discussion

We present a flexible generalized monotone functional mixed model for modelling the effect of a radiation dose distribution on a normal tissue complication outcome variable. This model allows us to estimate a general summary measure of the dose distribution by non-parametrically
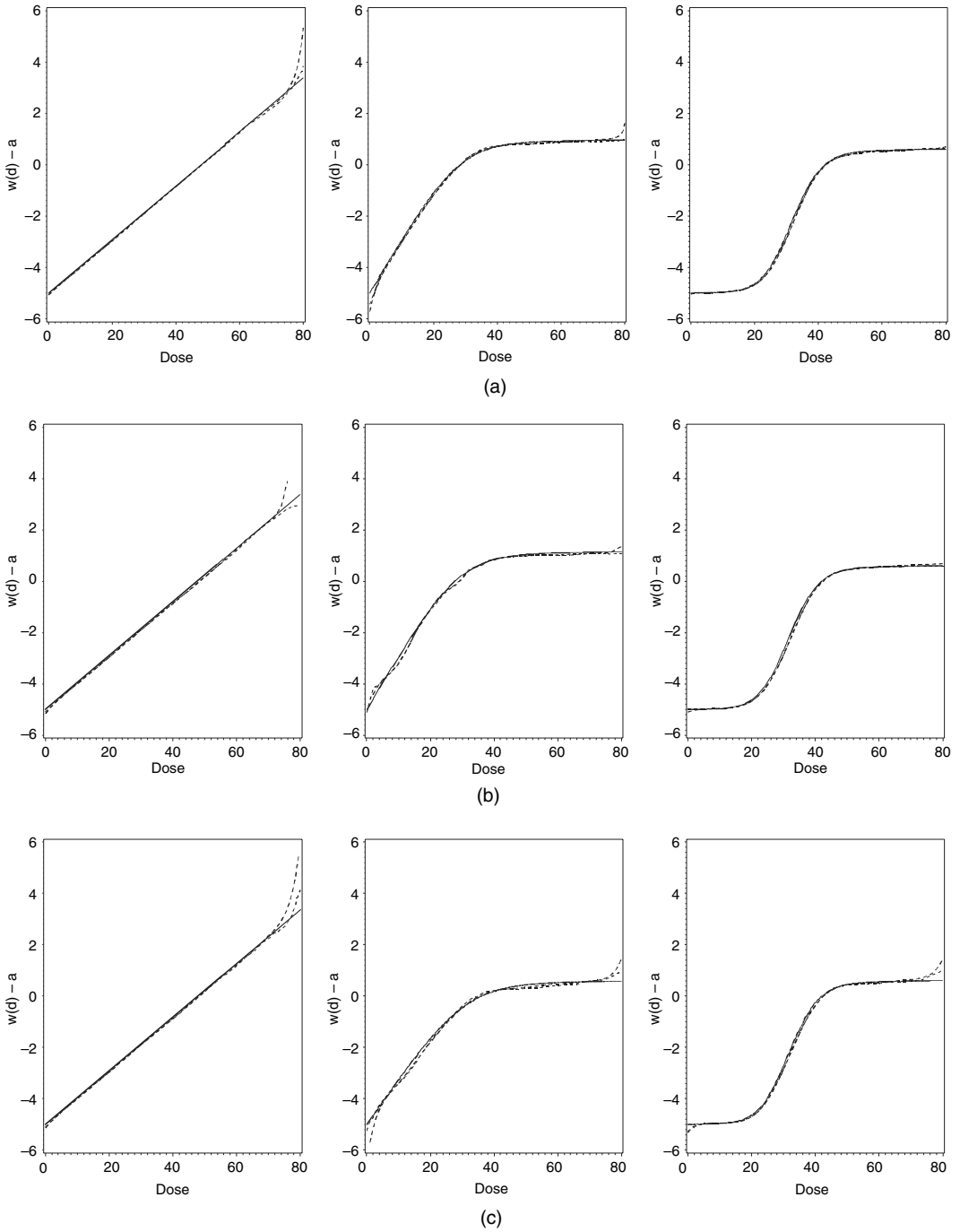
Fig. 5. Simulation results for Poisson data (------, mean and median of the 100 estimates of $w(d)$ at each dose): (a) Poisson data; (b) overdispersed Poisson data; (c) same data sets as in (a) analysed by using $h(z) = z^2$
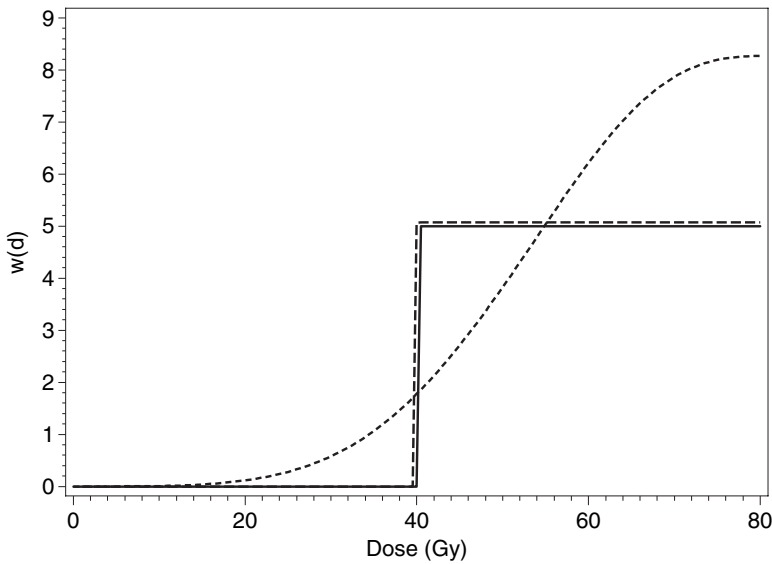
Fig. 6.   Comparison of our method and the *I*-spline method for estimating a step function: ———, true function; -------, *I*-spline estimator; – – –, our estimator

estimating a monotone weight function $w(d)$, which includes several commonly used summaries such as the mean dose as special cases. This non-parametric method is in contrast with typical radiobiological models which tend to be highly parametric. Simulations for both Poisson and overdispersed Poisson data show that the method performs well in estimating various shapes of the weight function $w(d)$.

There are two principal uses of the model that is presented in this paper. The first is to aid in the choice of treatment plan. To choose between multiple potential treatments plans for a new patient, the radiation oncologist could choose the plan with the lowest value of $\Sigma_{k=1}^{164} p_{ij}(d_k) \, w(d_k)$. Alternatively the non-parametric estimate can be used to motivate a parametric summary measure. For instance the non-parametric estimate of $w(d)$ in Fig. 3 is similar to a step function with a step at 50 Gy. Thus, rather than using $\hat{w}(d)$, the radiation oncologist could instead use V50, the proportion of the gland receiving a dose more than 50 Gy, which is easier to compute, to choose between plans. The second use of our proposed model is in prediction. Predicted saliva flow rates 6 months after treatment with a given plan can be obtained from equation (10) by plugging in parameter estimates. To obtain expected rates for an average patient, the random effect b_{ij} could be replaced by 0.

In comparing our estimator with the estimator that is based on *I*-spline functions with the same number of knots, we found our method to be more flexible especially for flat regions. It also does not require constraints on the regression parametric space. We consider in this paper a regression spline method. A key advantage of the regression spline approach is that it is computationally easy. Drawbacks include unstable behaviour near the end points as well as the need to choose the number and placement of knots. The data example and simulation study, however, show that the regression spline method works well in modelling our normal tissue complication data except for the boundary. An alternative approach is to use smoothing splines (Green and Silverman, 1994) and *P*-splines (Ruppert *et al.*, 2003). A major disadvantage of such approaches is computational burden, especially for non-linear mixed models for non-Gaussian outcomes.

In some cases like the step function example, as well as the parotid analysis, a method that

explicictly allows for flat regions is desirable. Dunson (2005) proposed such a method within a Bayesian framework, for relating a scalar dose value with a scalar outcome. This approach could be adapted to estimate $w(d)$ within the generalized monotone functional mixed model that is proposed in this paper. The results are in Schipper *et al.* (2007).

References

Booth, J. G. and Hobert, J. P. (1998) Standard errors of prediction in generalized linear mixed models. *J. Am. Statist. Ass.*, **93**, 262–272.

Breslow, N. E. and Clayton, D. G. (1993) Approximate inference in generalized linear mixed models. *J. Am. Statist. Ass.*, **88**, 9–25.

Dunson, D. B. (2005) Bayesian semiparametric isotonic regression for count data. *J. Am. Statist. Ass.*, **100**, 618–627.

Eisbruch, A., TenHaken, R. K., Kim, H. M., Marsh, L. H. and Ship, O. (1999) Dose, volume, and function relationships in parotid salivary glands following conformal and intensity-modulated irradiation of head and neck cancer. *Int. J. Radian. Oncol. Biol. Phys.*, **45**, 577–587.

Friedman, J. and Tibshirani, R. (1984) The monotone smoothing of scatterplots. *Technometrics*, **26**, 243–250.

Gelfand, A. E. and Kuo, L. (1991) Nonparametric Bayesian bioassay including ordered polytomous response. *Biometrika*, **78**, 657–666.

Green, P. J. and Silverman, B. W. (1994) *Nonparametric Regression and Generalized Linear Models*. London: Chapman and Hall.

Hall, P. and Huang, L. (2001) Nonparametric kernel regression subject to monotonicity constraints. *Ann. Statist.*, **29**, 624–647.

Holmes, C. C. and Heard, N. A. (2003) Generalised monotonic regression using random change points. *Statist. Med.*, **22**, 623–638.

James, G. M. (2002) Generalized linear models with functional predictors. *J. R. Statist. Soc.* B, **64**, 411–432.

Johnson, T. D., Taylor, J. M. G., TenHaken, R. K. and Eisbruch, A. (2005) A Bayesian mixture model relating dose to critical organs and functional complication in 3D conformal radiation therapy. *Biostatistics*, **6**, 615–632.

Kelly, C. and Rice, J. (1990) Monotone smoothing with application to dose-response curves and the assessment of synergism. *Biometrics*, **46**, 1071–1085.

Kutcher, G. J., Burman, C. and Brewster, L. (1989) Calculation of complication probability factors from non-uniform tissue irradiation: the effective volume method. *Int. J. Radian Oncol. Biol. Phys.*, **16**, 1623–1630.

Lichter, A. S. (1991) Three-dimensional conformal radiation therapy: a testable hypothesis. *Int. J. Radian Oncol. Biol. Phys.*, **21**, 853–855.

Lyman, J. T. (1985) Complication probability as assessed from dose-volume-histograms. *Radian Res.*, **104**, 513–519.

McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models*, 2nd edn. London: Chapman and Hall.

Pinheiro, J. C. and Bates, D. M. (1995) Approximations to the log-likelihood function in the non-linear mixed-effects model. *J. Computnl Graph. Statist.*, **4**, 12–35.

Ramsay, J. O. (1988) Monotone regression splines in action. *Statist. Sci.*, **3**, 425–461.

Ramsay, J. O. (1998) Estimating smooth monotone functions. *J. R. Statist. Soc.* B, **60**, 365–375.

Ramsay, J. O. and Silverman, B. W. (1997) *Functional Data Analysis*, pp. 157–177. New York. Springer.

Ruppert, D., Wand, M. P. and Carroll, R. J. (2003) *Semiparametric Regression*. Cambridge: Cambridge University Press.

Schipper, M. J., Taylor, J. M. G. and Lin, X. (2007) Bayesian generalized monotonic functional mixed models for the effects of radiation dose histograms on normal tissue complications. *Statist. Med.*, **26**, 4643–4656.

Vonesh, E. F., Chinchilli, V. M. and Pu, K. (1996) Goodness-of-fit in generalized nonlinear mixed models. *Biometrics*, **52**, 572–587.

Zhang, D., Lin, X. and Sowers, M. F. (2007) Two-stage functional mixed models for evaluating the effect of longitudinal covariate profiles on a scalar outcome. *Biometrics*, **63**, 351–362.