University of Michigan School of Public Health
The University of Michigan Department of Biostatistics Working Paper Series

Year 2006 *Paper* 61

Generalized Monotonic Functional Mixed Models with Application to Modeling Normal Tissue Complications

Matthew Schipper* Jeremy Taylor†
Xihong Lin‡

*University of Michigan, mjschipp@med.umich.edu
†University of Michigan, jmgt@umich.edu
‡Havard School of Public Health, xlin@hsph.harvard.edu

Generalized Monotonic Functional Mixed Models with Application to Modeling Normal Tissue Complications

Matthew Schipper, Jeremy Taylor, and Xihong Lin

Abstract

Normal tissue complications are a common side effect of radiation therapy. They are the consequence of the dose of radiation received by the normal tissue surrounding the tumor site. It is not known what function of the dose distribution to the normal tissue drives the presence and severity of the complications. Regarding the density of the dose distribution as a curve, a summary measure is obtained by integrating a weighting function of dose (w(d)) over the dose density. For biological reasons the weight function should be monotonic. We propose to study the dose effect on a clinical outcome using a nonparametric method by estimating this weight function smoothly and subject to the monotonicity constraint. In our approach w(d) is written as a integral of a smooth positive function of d. 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.

Generalized Monotonic Functional Mixed Models with Application to Modeling Normal Tissue Complications

Matthew Schipper[1], Jeremy Taylor[1] and Xihong Lin[2]

[1] Department of Biostatistics, University of Michigan

[2] Department of Biostatistics, Harvard University

August 14, 2006

Normal tissue complications are a common side effect of radiation therapy. They are the consequence of the dose of radiation received by the normal tissue surrounding the tumor site. It is not known what function of the dose distribution to the normal tissue drives the presence and severity of the complications. Regarding the density of the dose distribution as a curve, a summary measure is obtained by integrating a weighting function of dose ($w(d)$) over the dose density. For biological reasons the weight function should be monotonic. We propose to study the dose effect on a clinical outcome using a nonparametric method by estimating this weight function smoothly and subject to the monotonicity constraint. In our approach $w(d)$ is written as a integral of a smooth positive function of d. 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: Nonparametric regression, monotonicity, functional data, dose effect, normal tissue complications

1 Introduction

Radiation therapy is commonly used to treat cancer. The goal is to deliver the highest possible dose to the tumor site and the lowest possible dose to the surrounding normal tissue. Higher doses to the tumor result in more damage to the cancer cells while 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 experienced by lung cancer patients. Other examples include rectal failure in colon cancer patients and Xerostomia (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 tumor with minimizing damage to the surrounding tissue. To do this efficiently, it is necessary to understand precisely how the dose of radiation to the normal tissue and the probability or severity of normal tissue complications are related.

Modern treatment planning techniques allow the physician to compute, for a given treatment plan, the 3D distribution of dose within the tissue being irradiated (Lichter, 1991). It is common for the tumor region to be given an approximately uniform dose. Such is not the case for the surrounding normal tissue where the dose depends on its proximity to the tumor as well as on the treatment plan. There is clearly a spatial component to this 3D dose distribution, but in this paper we will be ignoring it. In some organs, such as the parotid gland which is relatively homogenous, this 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.

The analysis goal here is to relate the radiation dose distribution received by the normal tissue to the observed complication. The more general statistical problem is one of relating a functional predictor to a scalar outcome. This problem has been considered in the field of functional data analysis. See for example, chapter 10 of Ramsay and Silverman (1997). They propose a linear model which relates a functional predictor to a scalar response using a summary measure similar to the one discussed in this paper. Least squares, possibly with a roughness penalty term, is used as a fitting criterion. A more general model is the functional generalized linear model (FGLM) proposed by James (2002). In the FGLM, the observed functional predictor is assumed to be measured periodically over time and with error. Smoothing splines with common basis functions are used to parameterize each subject's observed predictor curve and a functional parameter curve. The integral of the product of these two curves is included in the mean structure of the model.

Rather than time, our functional predictor is measured over dose. Let F(d) denote the cumulative dose distribution function and p(d) denote the corresponding density. Then a general summary measure of the dose effect is given by $\int p(d)w(d)\,\mathrm{d}d$, where $w(d)$ is a weighting function we wish to estimate. 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 30 Gy is obtained by an indicator function ($w(d) = I(d > 30)$). Because the true form of $w(d)$ is not known, we wish to estimate it nonparametrically.

In this paper, we propose a generalized monotone functional mixed model which relates a complication outcome to a dose distribution. The dose effect is summarized using the summary measure discussed above. Within this model framework we propose a new method for nonparametric 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. There is also a statistical reason for this constraint which will be clear when we specify the model. 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 FGLM in that we wish to estimate $w(d)$ monotonically. In addition, our model allows for random effects which can be used to account for correlated and overdispersed data. Maximum likelihood is used for model fitting.

There are many approaches to monotone nonparametric regression (Friedman and Tibshirani (1984), Ramsay (1988), Kelly and Rice (1990), Ramsay (1998), Hall and Huang (2001), Gelfand and Kuo (1991), Holmes and Heard (2003) and Dunson (2005)). Attempting a summary here is beyond the scope of this paper. Instead we note why we are proposing a new approach. Existing techniques were 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 the head and neck cancer data in section 4, and use it to illustrate the proposed methods. In section 5 simulation results are presented. Section 6 concludes with some discussion and notes areas of open research.

2 The Model

In this section we propose a new model which we term a generalized monotone functional mixed model. Let Y_{ij} denote the complication outcome for the j^{th} observation of the i^{th} subject. We assume that conditional on random effects b_i, Y_{ij} follows an exponential family distribution with mean μ_{ij} given by

$$g(\mu_{ij}) = X_{ij}\alpha + \int_0^\infty p_{ij}(d)w(d)\mathrm{d}d + Z_{ij}b_i \tag{1}$$

where g 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 to the normal tissue of interest. We further assume that the random effects, b_i are independent and each follow a multivariate normal distribution with mean zero and covariance matrix Σ. Let $h(b_i)$ be the corresponding Gaussian density. Then the marginal likelihood is given by

$$L = \prod_i \int \prod_j f(Y_{ij}|b_i)h(b_i)\mathrm{d}b_i \tag{2}$$

where $f()$ is the conditional density of Y_{ij}. Although this is a cross-sectional model it does allow for multiple observations per subject at a single point in time. For example, in the head and neck cancer data, the outcome is saliva flow, which is measured separately from the subject's two parotid glands. The random effects are included to account for correlation between observations within subject as well as possible overdispersion. In principle, longitudinally measured complications could be handled by the model in (1). However, in (1), the dose effect, captured by $w(d)$ is assumed to be constant. In practice the dose effect will generally diminish over time so that (1) will not be realistic for longitudinally measured complications. The statistical reason we require $w(0) = 0$ can be seen from (1). If α includes an intercept term, it will be confounded with any

nonzero intercept $w(0)$, because this latter term can be pulled out from the integration in (1), since $p_{ij}(d)$ is a density and thus integrates to one.

One way to constrain $w(d)$ to be monotone increasing is to define it as the integral of a positive function. We consider the following formulation.

$$w(d_c) = \int_0^{d_c} s(d)\ \mathrm{d}d \tag{3}$$

where

$$s(d) \propto \phi(r(d)) \tag{4}$$

and

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

where $\phi(.)$ is a positive function defined on the set of real numbers, the $B_l()'s$ are known basis functions and the β_l are parameters that need to be estimated. Note that no constraints are placed on β. In this formulation, (5) imposes smoothness while (3) and (4) impose monotonicity. Excluding the dose effect summary measure, (1) specifies a generalized linear mixed model (Breslow and Clayton, 1993). However, it can be seen from (4) that $w(d)$ need not be a linear function of β_l. Thus the dose effect summary measure in (1) may also not be linear in β_l.

Figure 1 illustrates a few of the possible shapes for $w(d)$ with this formulation, as well as the underlying $r(d)$ and $s(d)$ that result in these shapes. Any linear combination of basis functions could be used for $r(d)$. In the saliva flow example, we used a two knot cubic regression spline for $r(d)$ parameterized by six B-spline basis functions, but have left the notation general to allow other choices. The flexibility of $w(d)$ depends on the flexibility of $r(d)$. When using a regression spline formulation for $r(d)$, important choices to be made are the number and location of the knots.

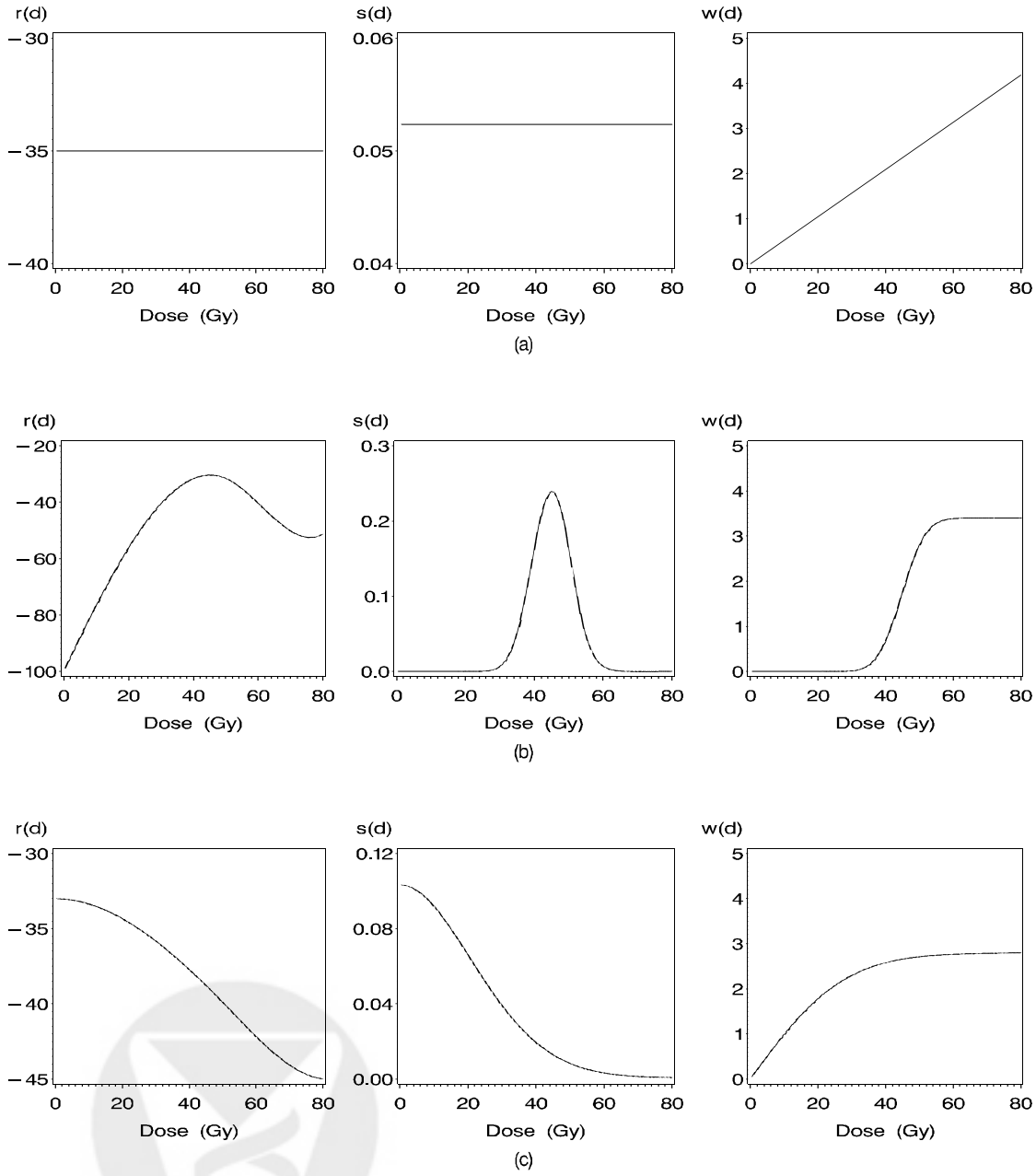

Figure 1: Three examples of w(d) with the corresponding r(d) and s(d): (a) Linear (b) Sigmoidal (c) Plateau

More knots will increase the flexibility of the estimate and better allow detection of local trends. However with no penalty or distributional assumption on the β_l, too many knots can result in overfitting the data. In simulations and the parotid example, we found relatively few (2 or 3) knots worked well. A standard choice for knot placement is to place the knots at the quantiles of the data, in our case at the quantiles of the observed dose range. Alternatively we could also place the knots in dose regions where we thought there might be more change.

Monotonicity of $w(d)$ requires only that $s(d)$ be positive. Density functions such as the normal density could be used for $\phi()$. One potential disadvantage to using a bounded density function is that this also bounds the maximum possible slope of $w(d)$. However, since we can choose the proportionality constant in (4) to be arbitrarily large, this should not be a problem. Using a normal density function also technically results in non-identifiability since $s(r(d)) = s(-r(d))$. In practice this will generally not be an issue, because if we set the proportionality constant to a large value, the maximum of $s(d)$ will not be attained and $r(d)$ will be strictly positive or strictly negative. Another possible choice is $s(d) = e^{r(d)}$. This choice has the advantages that it is unbounded and is also a 1-1 function, so that $s(r(d)) \neq s(-r(d))$. However, we in general expect a function proportional to a normal density for $s(d)$ to lead to fewer numerical difficulties in estimation.

3 Estimation

The mean structure in (1) assumes that the known dose distribution $p(d)$ is continuous, but in practice only an estimate, called the dose volume histogram (DVH), is usually available. The DVH discretizes dose into K bins and measures what proportion of the gland received a dose

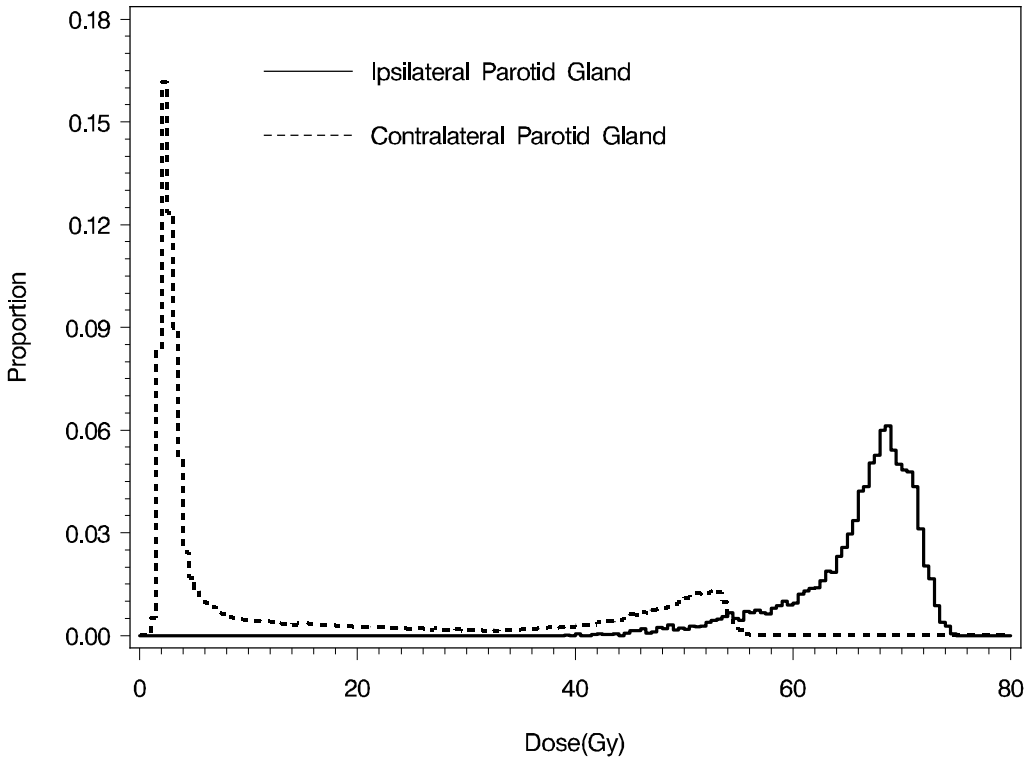

Figure 2: Examples of one subject's dose volume histograms. The ipsilateral gland is the gland on the same side of the head as the tumor, while the contralateral gland is on the opposite side of the head.

within each bin. Figure 2 shows two example DVH's.

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

$$g(\mu_{ij}) = X_{ij}\alpha + \sum_{k=1}^{K} p_{ij}(d_k)w(d_k) + Z_{ij}b_i \tag{6}$$

where d_k = $(k-1)$(bin width), that is, d_k is just the left endpoint of the k^{th} bin and $p(d_k)$ is the proportion of normal tissue receiving dose in bin k. We also approximate the integral in (3) as $w(d_1) = 0$ and $w(d_k) \approx \sum_{l=2}^{k} s(d_l)$ for k=2,3,... If the dose bins are not of width 1 then this sum is

only proportional to the integral, but the proportionality constant is absorbed into $s(d)$. Plugging this estimate of $w(d)$ into (6), we obtain

$$g(\mu_{ij}) = X_{ij}\alpha + \sum_{k=2}^{K} p_{ij}(d_k) \left[\sum_{l=2}^{k} s(d_l) \right] + Z_{ij} b_i \tag{7}$$

We propose use of maximum likelihood to fit this model. The mean structure given in (7) and the normal distributional assumption on the random effects b_i will fully specify the likelihood for the Poisson data that we consider in the next section. The scale parameter present in some exponential family distributions can also be estimated via maximum likelihood. As in the generalized linear mixed model setting, the marginal likelihood will generally not have a closed form solution and must be approximated. If the dimension of random effects is small (1 or 2), adaptive quadrature can be used (Pinheiro and Bates, 1995). We implemented our analysis in SAS (Proc NlMixed), which allows the specification of an arbitrary mean function.

4 Application to head and neck cancer data

4.1 Data Description

In this section we fit the above model to data from a head and neck cancer trial. Patients with a cancerous tumor 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 2, one in either cheek, are responsible for producing saliva. The radiation doses to each voxel in the parotid ranged from 0 to 82 Gy and the DVH bin widths are .5 Gy, thus K=164. Figure 2 is an example of the DVH's 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 as the tumor (Contralateral) is usually minimized, while no such attempt is made for the gland on the same side as the tumor (Ipsilateral). Saliva flows are obtained at baseline (before radiation therapy) and at 1, 3, 6, 12, 18 and 24 months after radiation therapy. At each timepoint, saliva flows are measured from both cheeks and under both stimulated and unstimulated conditions. Because our model is a cross-sectional model, we focus on modeling the relationship between the dose distribution to the parotid gland and the stimulated saliva flows at month 6. We chose month 6 because at earlier months the patients might still be recovering from the effects of radiation. Also relatively few patients dropped out by month 6. At month 6, there are 157 saliva flow measurements (with corresponding DVH's) from 82 subjects.

4.2 Model Description

The saliva flows are measured as rates (milliliters per minute), so the Poisson distribution is a natural choice for 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 ($j = 1$ for ipsilateral gland, 2 for contralateral gland) of subject i ($i = 1, 2, ...82$). Let $b_i = [b_{i1} \quad b_{i2}]'$ 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, conditional on the random effects, the

saliva flows follow a Poisson distribution with mean μ_{ij} given by

$$log(\mu_{ij}) = a_0 + a_1 base_{ij} - \sum_{k=2}^{k=164} p_{ij}(d_k) w(d_k) + b_{ij} \tag{8}$$

where $p_{ij}()$ is the DVH for the j^{th} gland of the i^{th} subject. We are subtracting 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 level random effect included as part of the linear predictor will approximate this variance structure and allow us to account for the overdispersion present in the data. 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 positive or negative. Including a subject level random effect would only allow positive correlation between the sides. Besides dose, we have only included baseline saliva flow as a covariate. Clinical covariates such as age, gender, chemotherapy, medications, comorbidity and pre-RT surgery were previously found not to impact saliva flow (Eisbruch et al, 1999). We use a two knot cubic regression spline formulation for $r(d)$ parameterized by 6 B-spline basis functions.

$$r(d) = \sum_{j=1}^{6} \beta_j B_j(d) \tag{9}$$

where $\beta_1, ..., \beta_6$ are unconstrained parameters and $B_1(), ..., B_6()$ are B-spline basis functions defined on (0,82) with knots at 30 and 60 Gy.

These data have been previously analyzed by other investigators. Eisbruch et al (1999) considered two models. The first was a GLM in which the dose effect was summarized by a threshold mean

dose effect and the outcome was the observed saliva flow rate. Longitudinal saliva flows were included, and generalized estimating equations were used to estimate the model parameters. They also considered a normal tissue complication probability (NTCP) model (Lyman, 1985). Assume a fractional volume v of a normal tissue receives a uniform dose d. Then the Lyman NTCP model gives the probability of complication as

$$NTCP = \Phi((d - D_0 v^{-n})/mD_0 v^{-n}) \tag{10}$$

where D_0, n and m are parameters. It is seldom the case that the normal tissue receives a uniform dose. Thus to use this model it is necessary to convert the observed dose distributions into ones where a partial volume receives a uniform dose. How to do this is described in detail in Kutcher et al, (1989). This model applies only to binary outcomes, so patients whose salivary flow at 12 months was at or below 25% of their salivary flow at baseline, were considered to have Xerostomia. Subjects who did not meet this criteria were considered not to have Xerostomia. Johnson et al, (2005) proposed a complex Bayesian model which incorporated the longitudinally measured saliva flows. In their model the dose effect was assumed to be captured by a percentile of the dose distribution (DVH). Our model differs from these models principally in how it includes the dose effect. Rather than assuming the damage is given by the mean dose or a percentile of the dose distribution, we have used the general summary measure $\int p_{ij}(d)w(d)\,\mathrm{d}d$. Also unlike the NTCP model, our model does not require a binary outcome and we use the observed saliva flow rates.

4.3 Results

The maximum likelihood estimate and 95% pointwise confidence intervals of $w(d)$ are shown in Figure 3. To interpret the estimated $w(d)$, it is helpful to imagine 2 parotid glands, say gland A

and gland B, that differ only in the dose of radiation received. Assume 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 8,

$$\begin{aligned} log(\mu_A) - log(\mu_B) &= -\sum_{k=2}^{k=164} p_A(d_k)w(d_k) + \sum_{k=2}^{k=164} p_B(d_k)w(d_k) \\ &= -w(d_c) + w(0) \\ &= -w(d_c) \end{aligned} \tag{11}$$

Thus $w(d_c)$ may be interpreted as the difference in expected saliva flow (on log scale) between a gland that received uniform dose d_c and a gland that received no dose of radiation. Similarly, $e^{-w(d_c)}$ can be interpreted as the ratio μ_A/μ_B. The estimates for the largest dose bins are not plotted because the extremely large values obscure the shape of $w(d)$. What is striking is that the estimated $w(d)$ is nearly flat for the first 30 Gy. The clinical implication of this finding is that the parotid gland could be uniformly irradiated at 30 Gy with almost no effect on saliva flow. It is also interesting that the estimated $w(d)$ is not linear, implying that mean dose is not the best way to summarize the DVH. We are able to test this conclusion using a likelihood ratio test since the mean dose model is nested within our more complicated model (1). If $\beta_1 = ... = \beta_6$ then $w(d)$ is linear because $\sum_l B_l(d) = 1$ by construction. When $w(d)$ is linear, the resulting model is equivalent to a mean dose model. The chi-square statistic (p-value) for the LRT is $34.2 (< .0001)$. Thus we strongly reject the mean dose model in favor of our more complicated model. The pointwise confidence intervals are calculated using a delta method estimate of the standard error. We also note here that very similar estimates of $w(d)$ are obtained by using 3 equally spaced knots as well as by varying the placement of the 2 knots.

In Figure 4 we have plotted, for each gland, the log proportion change in saliva flow versus the

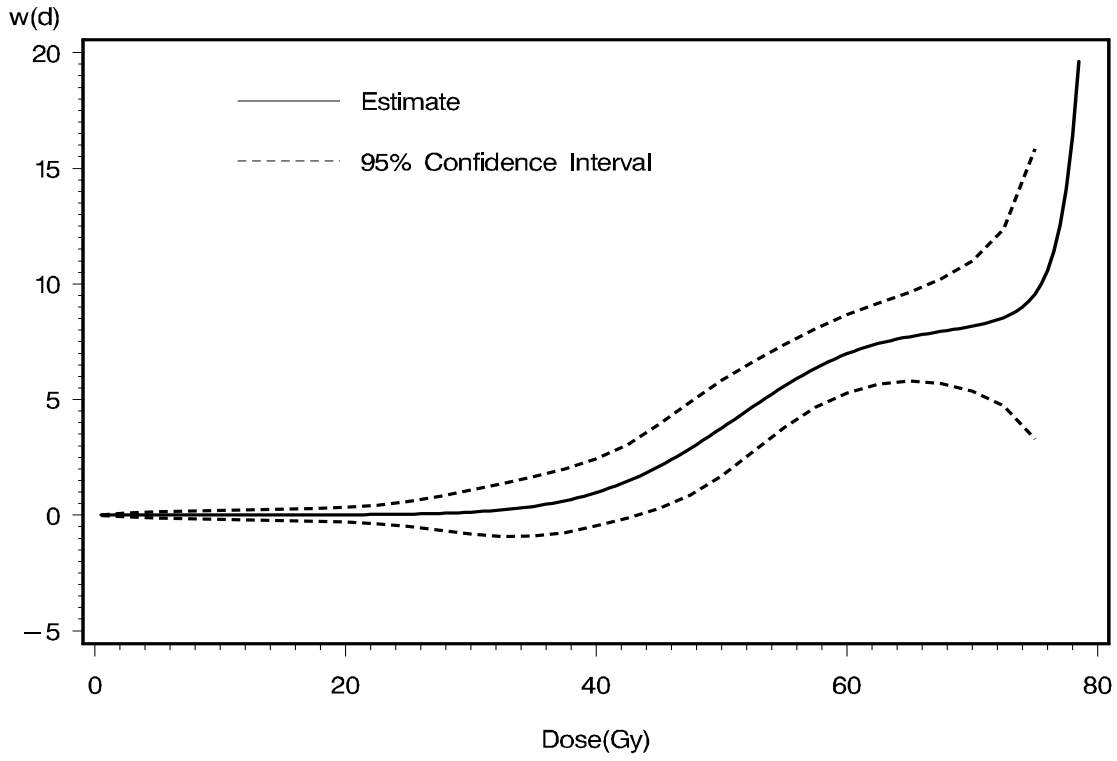

Figure 3: Maximum likelihood estimate and pointwise 95% confidence intervals of $w(d)$

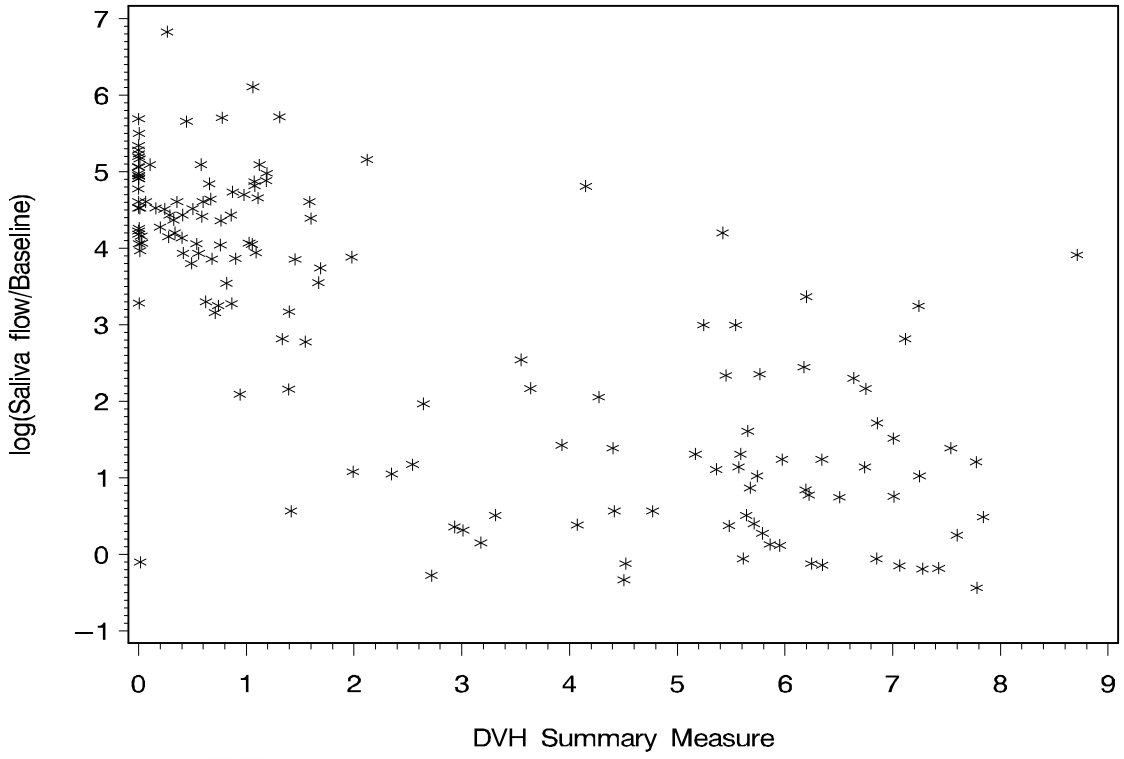

Figure 4: Change in log saliva flow from baseline versus estimated dose effect summary measure

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. This plot does not exactly correspond to the model given by (8) since $\hat{a_1} \neq 1$. However, the plot changes very little if instead we plot $log((Y_{ij}+1)/(\hat{a_1}base_{ij}+1))$. 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. Estimates of the random effect variances and other parameters are shown in table 1. It is apparent that baseline saliva flow is an important factor since a_1 is significantly different from zero. Since $\hat{a_1} \neq 1$ and $\hat{a_0} \neq 0$, the predicted saliva flow at month 6, for a gland that received zero dose of radiation, is not equal to that gland's baseline saliva flow. However, in our data all glands received some dose of radiation. 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 estimated correlation between glands may be calculated from the covariance parameter estimates as $\Sigma_{1,2}/\sqrt{\Sigma_{1,1}\Sigma_{2,2}} = .12$,and is not significantly different from zero. A significantly negative correlation coefficient would have given support to the notion of compensation while 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.

An alternative approach to estimating $w(d)$ is to define it directly as a regression spline using I-spline basis functions (Ramsay, 1988). The I-spline basis functions are defined at each dose as the integral of the corresponding B-spline basis function up to that dose and are thus monotone increasing. If the parameter coefficients are constrained to be positive, then the resulting regression spline will be monotone increasing. We used this approach to estimate $w(d)$ for the month 6 saliva

Table 1: Parameter estimates and standard errors

Parameter	Estimate	SE
a_0	1.38	0.51
a_1	0.74	0.12
$\Sigma_{1,1}$	9.53	3.21
$\Sigma_{1,2}$	0.30	0.41
$\Sigma_{2,2}$	0.62	0.12

flow data. Of the six parameters, two were estimated to be at the boundary value of zero. The resulting estimate of $w(d)$ is shown in Figure 5 and is somewhat different from that in Figure 3. Rather than being flat for doses up to 30 Gy, the I-spline estimator starts to increase after 10 Gy and is more linear thereafter than our estimate. The associated negative 2 log-likelihoods are 1094.8, 1062.1 and 1060.6 for models based on meandose, I-spline, and our method, respectively. The I-spline model and our model have the same number of parameters and likelihood values favor our model, although we are not able to test this, since the two models are not nested.

4.4 Goodness of Fit

Goodness of fit (GOF) statistics are used to assess whether the model fits the data reasonably well. One commonly used such statistic is the Pearson Chi-Square Statistic and is computed as,

$$\sum_i \sum_j (y_{ij} - \mu_{ij})^2 / \mu_{ij} \tag{12}$$

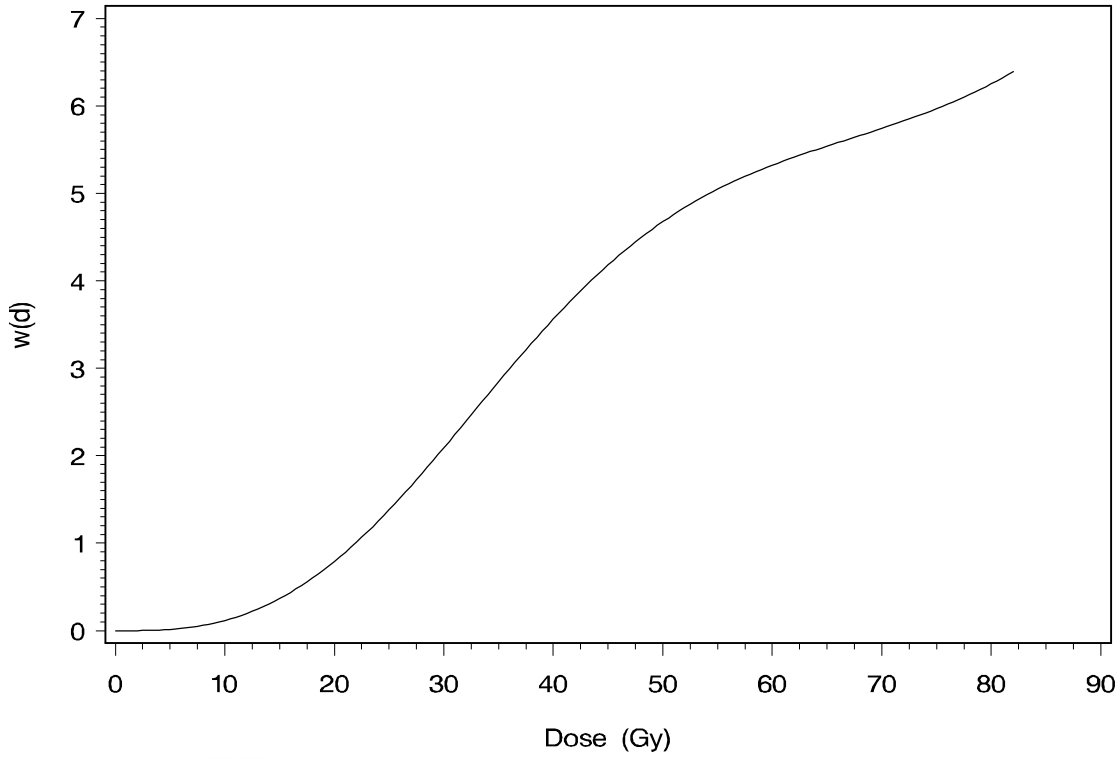

Figure 5: Maximum likelihood estimate of w(d) using I-spline method of Ramsay(1988)

where $\mu_{ij} = E[y_{ij}|b_i]$ is the conditional mean and is obtained from (8) by plugging in estimates of the fixed effect parameters α and β and of the random effect parameters 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 (12) 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 divide by the marginal variance of Y_{ij}. For a detailed discussion of goodness of fit in generalized nonlinear mixed models, and in particular how residuals 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 datasets of the same size as our observed dataset, using the parameter estimates from the observed data. For each dataset we fit the model and compute the GOF statistic. A histogram of the statistics (based on the marginal mean and variance) from the simulated datasets is plotted in Figure 6. Of the 100 simulated datasets, 18 had GOF statistics larger than that of our observed data. Results are similar for the other two GOF statistics. Thus, based on these metrics, our proposed model appears to be consistent with the data.

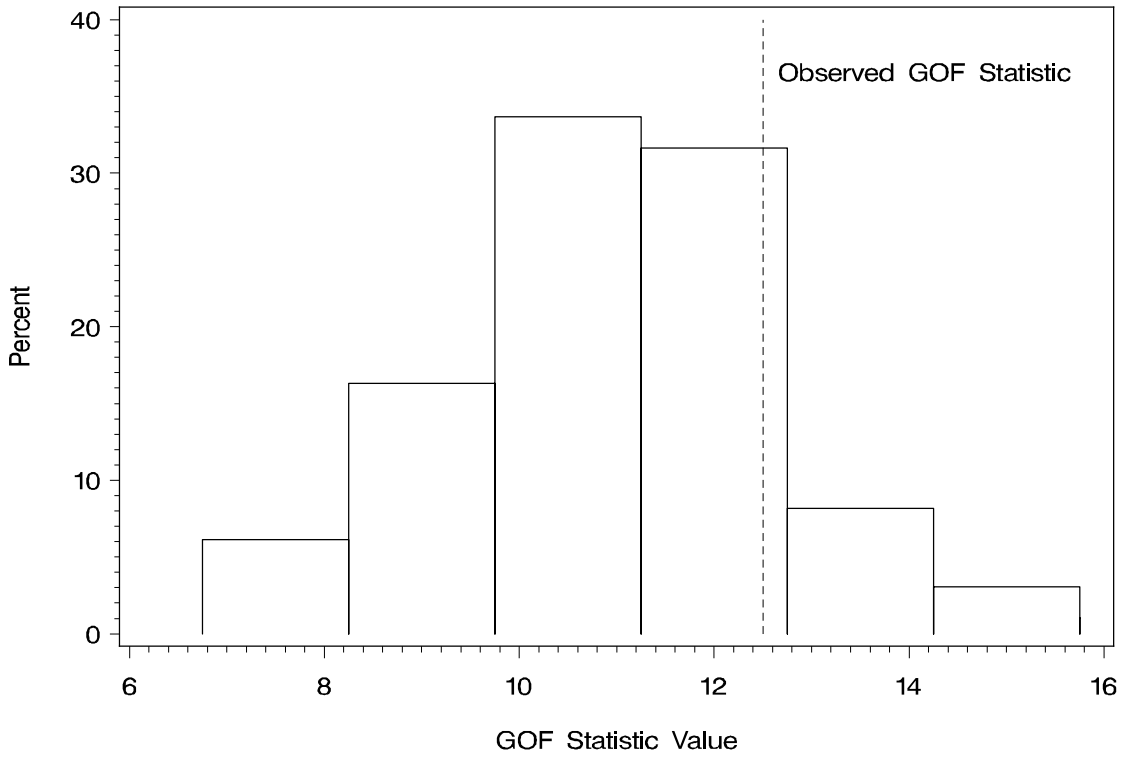

Figure 6: Histogram of GOF statistics from simulated data. The dashed line represents the GOF statistic value for the observed data

5 Simulations

To evaluate our method we simulate Poisson data with a mean dependent on a variety of shapes for $w(d)$ and then estimate $w(d)$ with our method. Specifically, for each of 3 shapes for $w(d)$, we simulate 100 datasets each with 157 independent Poisson variables with means μ_i given by

$$log(\mu_i) = a - \sum_{k=2}^{k=164} w(d_k)p_i(d_k) \tag{13}$$

where we set $a = 5$, which is roughly the mean of the sum of the estimated intercept and baseline flow terms in the observed data. The DVH's 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). Figure 7 shows the mean and median of the estimated values of $(w(d) - a)$ for each shape along with the underlying true $(w(d) - a)$. We have used $(w(d) - a)$ because estimates of this quantity are more stable than estimates of $w(d)$ alone.

We also wished to assess how well this method works in more variable data such as the overdispersed Poisson data that we observed in the saliva flow data. Here we simulate from the following model:

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

where we set $a = 5$ and the random effects $b_i \sim N(0, 1)$. As above we simulate 100 datasets of size 157 for each of the three different shapes of $w(d)$. The results are shown in Figure 8. Overall, the method seems to work well for both the Poisson and the overdispersed Poisson data. There is little bias except in the tails. There is relatively little data in these regions for some of the simulated datasets so the estimates tend to be more variable with a few extreme estimates having a large effect on the mean.

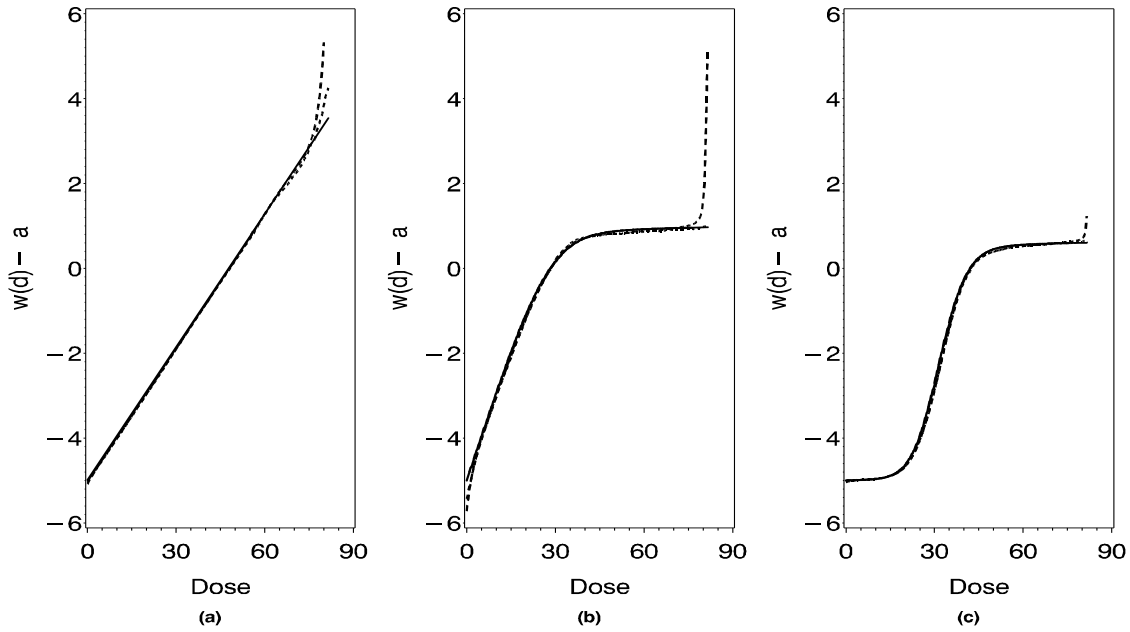

Figure 7: Simulation results for Poisson data: (a) Linear (b) Plateau (c) Sigmoidal $w(d)$. The long and short dashed lines are the mean and median, respectively of the 100 estimates of $w(d)$ at each dose.

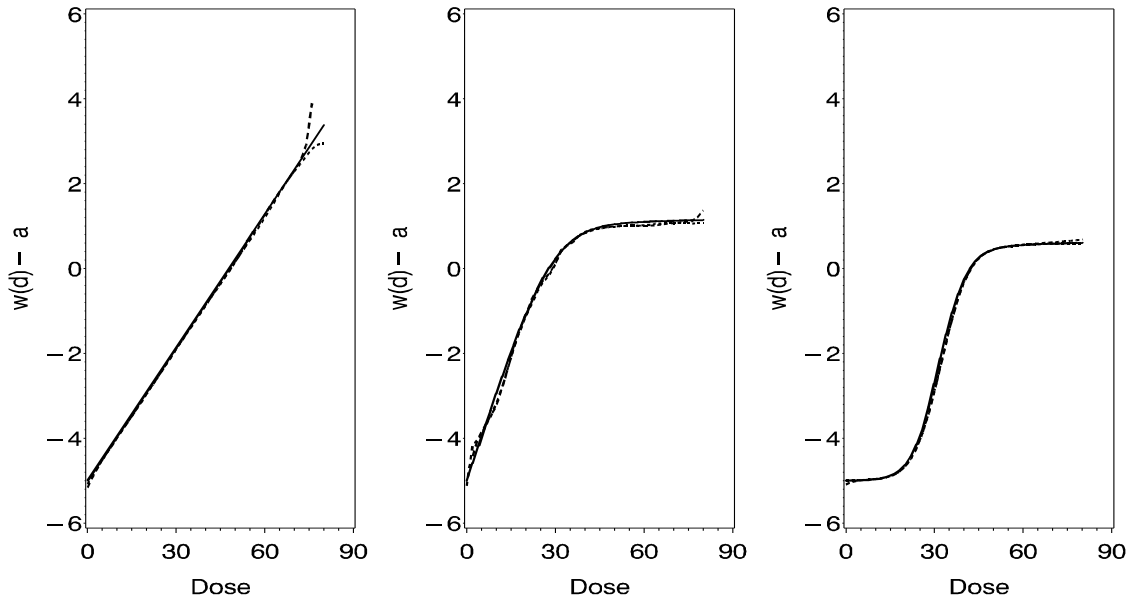

Figure 8: Simulation results for Poisson mixed data: (a) Linear (b) Plateau (c) Sigmoidal $w(d)$. The long and short dashed lines are the mean and median, respectively of the 100 estimates of $w(d)$ at each dose.

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 dataset based on this $w(d)$ from (13). We then fit both our model and the I-spline based model to this data. The resulting estimates are shown in Figure 9. The I-spline method severely oversmooths this step function. The reason for this is that the slope and maximum value of the regression spline are linked. To get a steeper function you need a larger parameter coefficient which also means a larger maximum value. It is thus not possible to get 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, our proposed method is able to indicate whether the true function being estimated is a step function.

6 Discussion

One way to summarize a radiation dose distribution is to integrate a weighting function $w(d)$ over the density of the dose distribution. We have presented a flexible method for nonparametric and monotone estimation of this weighting function within a generalized monotone functional mixed model. Simulations for both Poisson and overdispersed Poisson data showed little bias in the estimated $w(d)$ over most of the dose range. There was some bias in the tails which is not surprising given that our formulation of $w(d)$ is based on a regression spline. In comparing our estimator with the estimator of Ramsay (1988) with the same number of knots, we found ours to

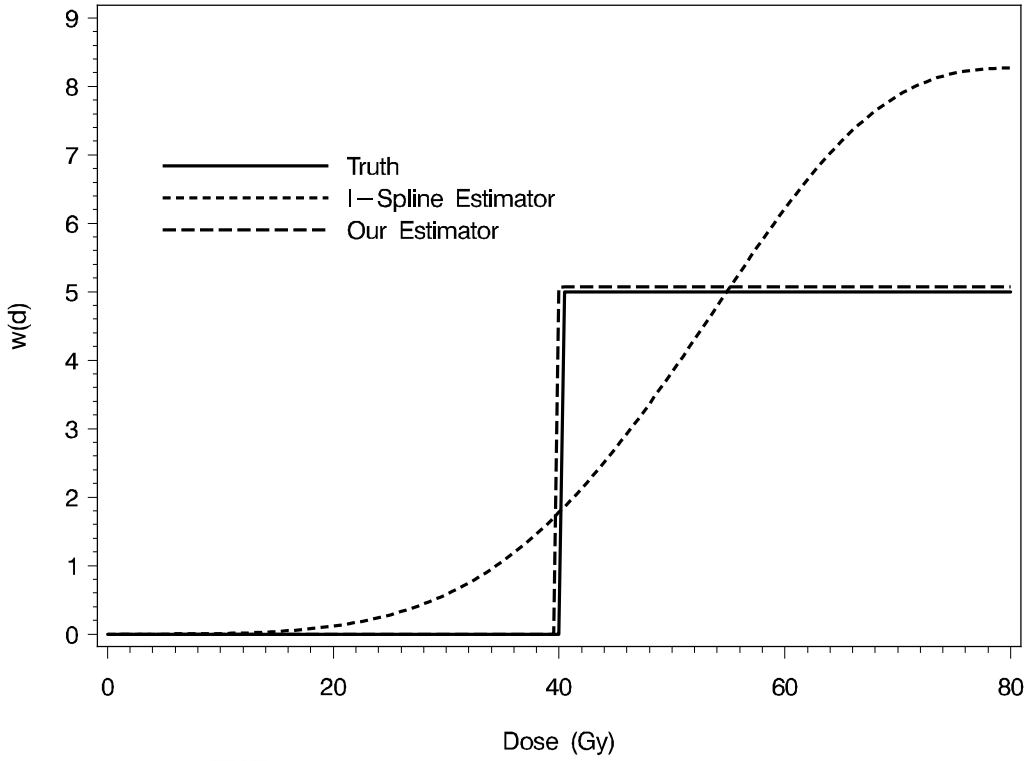

Figure 9: A comparison of our method and the I-spline method for estimating a step function

be more flexible. In contrast to many monotone smoothing methods, ours requires no restrictions on the parameter space.

Our proposed model makes two important contributions to modeling normal tissue complications. First, it utilizes a general summary measure of the dose effect which includes several commonly used summaries as special cases. Estimation of the dose effect does not assume a particular form and is nonparametric in this regard. This is in contrast to typical radiobiological models which tend to be highly parametric. Second, our model does not require the complication to be binary, and continuous outcomes are allowed.

Although we have focused on the case of a functional predictor, the formulation of $w(d)$ given in (3)-(5) can easily be used for relating scalar outcomes with scalar covariates when the relationship is known to be monotone. In this situation, there are many other existing methods that could also be applied.

In equation (5) we have written $r(d)$ as a linear combination of basis functions. We used a 2 knot cubic regression spline in the example and simulations. While this appeared sufficiently flexible in estimating the shapes we chose for the simulations, regression splines have been criticised for undesirable behaviour near the endpoints as well as the need to choose the number and placement of knots. Smoothing splines address these concerns. It would thus be useful to define $r(d)$ as a smoothing spline with number of knots corresponding to the number of bins in the DVH.

In some cases like the step function example, as well as the parotid analysis, a method that explicitly 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 proposed in this paper.

7 Bibliography

Booth, J.G. and Hobert, J.P. (1998). Standard errors of prediction in generalized linear mixed models. *J. of the American Statistical Association* **93**, 262-272.

Breslow, N.E. and Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *J. of the American Statistical Association* **88**, 9-25.

Dunson, D.B. (2005). Bayesian semiparametric isotonic regression for count data. *J. of the American Statistical Association* **100**, 618-627.

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

Eisbruch, A., TenHaken, R.K., Kim, H.M. et al. (1999). Dose, volume, and function relationships in parotid salivary glands following conformal and intensity-modulated irradiation of head and neck cancer. *Int J Radiat Oncology Biol Phys*, **, 45**, 577-587.

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

Hall, P. and Huang, L. (2001). Nonparametric kernel regression subject to monotonicity constraints. *The Annals of Statistics*, **29**, 624-647.

Holmes, C.C, and Heard, N.A. (2003). Generalised monotonic regression using random change points. *Statistics in Medicine* **22**, 623-638.

James, G.M. (2002). Generalized linear models with functional predictors. *J. Royal Statistical Society Series 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.

Lichter, A.S. (1991). Three-dimensional conformal radiation therapy: a testable hypothesis. *Int. J. of Radiation Oncology Biology Physics* **21**, 853-855.

Pinheiro, J.C. and Bates, D.M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *J. of Computational and Graphical Statistics* **4**, 12 -35.

Ramsay, J.O. (1988). Monotone regression splines in action, *Statistical Science* **3**, 425-461.

Ramsay, J.O. and Silverman, B.W. (1997). Functional data analysis. New York, *Springer-Verlag*, 157-177.

Ramsay, J.O. (1998) Estimating smooth monotone functions. *J. Royal Statistical Society Series B*, **60**, 365-375.

Vonesh, E.F., Chinchilli, V.M. and Pu, K. (1996). Goodness-of-fit in generalized nonlinear mixed models. *Biometrics* **52(2)**, 572-587.