University of Michigan School of Public Health

The University of Michigan Department of Biostatistics Working Paper Series

Year 2004 *Paper* 50

A Bayesian Mixture Model Relating Dose to Critical Organs and Functional Complication in 3D Conformal Radiation Therapy

Tim Johnson* Jeremy Taylor†

Randall K. Ten Haken‡ Avraham Eisbruch**

*University of Michigan Biostatistics, tdjtdj@umich.edu

†University of Michigan, jmgt@umich.edu

‡University of Michigan Department of Radiation Oncology, rth@umich.edu

**University of Michigan Department of Radiation Oncology, eisbruch@umich.edu

A Bayesian Mixture Model Relating Dose to Critical Organs and Functional Complication in 3D Conformal Radiation Therapy

Tim Johnson, Jeremy Taylor, Randall K. Ten Haken, and Avraham Eisbruch

Abstract

A goal of radiation therapy is to deliver maximum dose to the target tumor while minimizing complications due to irradiation of critical organs. Technological advances in 3D conformal radiation therapy has allowed great strides in realizing this goal, however complications may still arise. Critical organs may be adjacent to tumors or in the path of the radiation beam. Several mathematical models have been proposed that describe a relationship between dose and observed functional complication, however only a few published studies have successfully fit these models to data using modern statistical methods which make efficient use of the data. One complication following radiation therapy of head and neck cancers is the patient's inability to produce saliva. Xerostomia (dry mouth) leads to high susceptibility to oral infection and dental caries and is, in general, unpleasant and an annoyance. We present a dose-damage-injury model that can accommodate any of the various mathematical models relating dose to damage. The model is a nonlinear, longitudinal mixed effects model where the outcome (saliva flow rate) is modeled as a mixture of a Dirac measure at zero and a gamma distribution whose mean is a function of time and dose. Bayesian methods are used to estimate the relationship between dose delivered to the parotid glands and the observational outcome – saliva flow rate. A summary measure of the dose-damage relationship is modeled and assessed by a Bayesian x2 test for goodness-of-fit.

A Bayesian Mixture Model Relating Dose to Critical Organs and Functional Complication in 3D Conformal Radiation Therapy

Timothy D. Johnson[1*] Jeremy M. G. Taylor[1]
Randall K. Ten Haken[2] Avraham Eisbruch[2]

[1]Department of Biostatistics
School of Public Health
[2]Department of Radiation Oncology
School of Medicine
University of Michigan
Ann Arbor, MI 48109

Abstract

A goal of radiation therapy is to deliver maximum dose to the target tumor while minimizing complications due to irradiation of critical organs. Technological advances in 3D conformal radiation therapy has allowed great strides in realizing this goal, however complications may still arise. Critical organs may be adjacent to tumors or in the path of the radiation beam. Several mathematical models have been proposed that describe the relationship between dose and observed functional complication, however only a few published studies have successfully fit these models to data using modern statistical methods which make efficient use of the data. One complication following radiation therapy of head and neck cancers is the patient's inability to produce saliva. Xerostomia (dry mouth) leads to high susceptibility to oral infection and dental caries and is, in general, unpleasant and an annoyance. We present a dose-damage-injury model that can accommodate any of the various mathematical models relating dose to damage. The model is a non-linear, longitudinal mixed effects model where the outcome (saliva flow rate) is modeled as a mixture of a Dirac measure at zero and a gamma distribution whose mean is a function of time and dose. Bayesian methods are used to estimate the relationship between dose delivered to the parotid glands and the observational outcome—saliva flow rate. A summary measure of the dose-damage relationship is modeled and assessed by a Bayesian χ^2 test for goodness-of-fit.

Key Words: Bayesian analysis, Dirac measure, MCMC, Mixture model, Goodness-of-fit, 3D conformal radiation therapy.

*Corresponding Author, email: tdjtdj@umich.edu, tel. (734) 936-1007

1 Introduction

3D conformal radiation therapy, (Lichter, 1991), is a general term used to describe the planning and delivery of radiation where the dose at every point in a three-dimensional target field is calculated and the alignment and position of the radiation source is chosen to optimize the dose distribution, within the target field, according to certain criteria. 3D conformal radiation therapy has enabled radiation oncologists to deliver higher, more accurate and homogeneous doses of radiation to a defined region of cancer. Despite these advances, adjacent normal tissue and organs may also receive high doses of radiation resulting in impaired functioning of that organ. Furthermore, the dose distribution to healthy tissue/organs can be highly heterogeneous. As is the case with advances in the technology of many medical devices, these advances have preceded the quantitative understanding and statistical evaluation of the relationship between the dose to healthy tissue/organs and the observed functional complications due to the damage caused by ionizing radiation.

Radiation therapy is given in a series of daily small fractionated doses (typical dose fraction sizes are 1.5-3 Gray (Gy), with the most common values begin 1.8 and 2.0 Gy) delivered to the target five days per week over approximately a six week period. The total dose delivered to the tumor is typically on the order of 60 to 75 Gray and is quite homogeneous. The fractionated dose distribution may differ from one day to the next due to designed field size changes and technique changes. However, in practice only one or two changes are made during the course of treatment. 3D treatment planning software calculates the dose to every voxel (volume element) at every fractionation in the path of the radiation beam based on volumetric data sets typically obtained by computed tomography (CT). The goal of the treatment plan is two-fold. First, to deliver the prescribed overall dose of radiation homogeneously to the target (typically a malignant tumor). Secondly, to minimize the dose delivered to the surrounding healthy tissues and organs and to those in the path of the radiation beam. The development of 3D conformal radiation therapy treatment has allowed radiation oncologists to meet these goals to a larger extent than previously possible.

The dose distribution to a specified region is typically summarized by the dose volume histogram (DVH) based on the total delivered dose to that region. The DVH is analogous to the survivor function—it is one minus the distribution function of the total dose over the region of interest. Throughout this paper, the dose distribution function to the region of interest (in our case a critical, healthy organ) will be denoted by F.

The model we consider in this article can be broken down into two parts. The first part relates the DVH, or equivalently F, to the damage suffered by a specific healthy organ. The damage to the organ is unobserved. The second part of the model relates damage to injury incurred by the organ. Injury is the observed outcome. Several biomathematical models relating dose to damage incurred have been proposed, e.g. Chappell, Nondahl, and Fowler (1995), Dawson et al. (2002),Deasy et al. (2002),Jackson et al. (1995),Lyman (1985) and Seppenwoolde et al. (2003). However, few of these have been fit to data using modern statistical methodology as pointed out by Shultheiss, Orton, and Peck (1983). Some analyses of data from 3D conformal therapy have been published (Boersma et al., 1998; Chao et al., 2001; Jackson, Kutcher, and Yorke, 1993; Jackson et al., 1995; Kwa et al., 1998; Roesink et al., 2001; Scrimger et al., 2004; Theuws et al., 1998). In these papers the authors have attempted to estimate the relationship between the DVH and complications. However, the analyses, for the most part, have not been very sophisticated from a statistical point of view. Although the current usage of statistical methods in the field of 3D conformal radiation therapy is limited, there have been some improvements in recent years. Maximum likelihood estimation is used (Roberts and Hendry, 1993), profile likelihood confidence intervals have been advocated (Schilstra and Meertens, 2001), the use of simulation to assess goodness-of-fit has been suggested (Stavrev and Stavrev, 2000) and the Bootstrap has been used to obtain confidence intervals (Chao et al., 2001) and generalized linear models/generalized estimating equations (Eisbruch et al., 1999). Shultheiss (2001) lists a number of problem areas in modeling normal tissue radiation injury. The list includes (i) too few data for analysis; too few events, (ii) assigning values to parameters that should be estimated, (iii)

using inappropriate statistical methods, ad hoc methods, univariate analysis and (iv) ignoring model validation.

In this article we address the aforementioned issues. We use an empirical measure of damage and concentrate on modeling the relationship between damage and injury to critical organs, the parotid glands (the major salivary producing glands), that occurs in head and neck cancer patients. The data come from an ongoing study being conducted in the department of Radiation Oncology at the University of Michigan's Comprehensive Cancer Center. This article lays the foundation for future work where we will use the proposed model to compare, and develop, different biomathematical models that relate dose to damage.

The paper is organized as follows. The data set is described in Section 2, while our model is presented in section 3. Details of the MCMC estimation procedure are detailed in section 4. Results of the data analysis are given in section 5 along with an analysis of the goodness-of-fit of the model. We conclude with a discussion of future work in section 6.

2 The Parotid Data Set

The data set is comprised of 87 head and neck cancer patients. The outcome of interest is saliva flow rate from the parotid glands. Saliva flow rate (ml/min) was obtained by placing a suction cup over the secretory duct of the parotid gland (Heft and Baum, 1984). Saliva was collected for 2 minutes for each of the two parotid glands under two stimulatory conditions: normal and stimulated. Saliva flow rate was then estimated by taking the volume collected (ml) and dividing by time (120 sec). The parotid gland was stimulated by citric acid (lemon juice) place on the tongue. Subjects began the study prior to radiation therapy, typically 1 to 3 weeks prior to therapy (month 0). The subjects were also scheduled to return for saliva flow rate measurements at 1, 3, 6, 12, 18 and 24 months after termination of radiation therapy. We note here that many of the saliva measurements at each post radiation follow-up are zero. Computed tomography (CT) treatment planning software was used to develop the treatment plan for each subject. Each parotid gland from each subject was outlined

on the CT by a radiation oncologist and the treatment planning software computed the cumulative DVH for each of the parotid glands. Examples of 8 DVHs from 4 patients are shown in Figure 1. Sixty-three of the 87 subjects have DVHs for both parotids. The other 24 subjects had one of their parotid glands surgically removed due to disease involvement and thus have a DVH for only one parotid (the contralateral side). As is the case with most, if not all, longitudinal studies, there are missing observations. In this study, 23% of the data are missing. We assume that the data are missing completely at random (MCAR), (Little and Rubin, 2002). We will have more to say about this assumption in Section 4. Percentages of missing data at each time point are given in Table 1

Table 1: Percentages of missing data by month.

Month	0	1	3	6	12	18	24
% Missing	0	13.33	15.33	16.00	24.67	48.00	45.33

We found from the goodness-of-fit summary (section 5.1) that a transformation of the data was necessary. The GOF summary suggested that there were too many larger than expected observed values. Since all but two observations were less than two, we transformed the data by squaring the observed values divided by two. That is, if y is an observation, we took $y^* = (y/2)^2$.

3 The Dose-Damage-Injury Model

A radiobiologically based model that is central to the field of radiation therapy and that is amenable to statistical data analysis is the dose-damage-injury model (Jackson, Kutcher, and Yorke, 1993; Jackson et al., 1995; Niemierko and Goitein, 1992; York et al., 1993). Conceptually, the model is based on the premise that radiation causes damage to tissue/organs at the cellular level which, for the most part, is unobservable. A consequence of this damage is injury, which is observable. The radiation dose distribution is known and well characterized by 3D treatment planning software. Damage can be conceptualized as the surviving

fraction of critical cells, or functional subunits. More generally, damage can be thought of as a continuous single number summary that is solely responsible for determining the clinically observed outcome. We assume that damage is monotonically related to dose in the sense that a uniform increase in dose results in an increase in damage.

We present a dose-damage-injury model and focus on the relationship between damage and injury. We take a Bayesian approach to estimating model parameters. This part of the model assumes that the injury incurred by the parotid gland is measured by saliva flow rate. Saliva flow rate is assumed to be a smooth, monotone decreasing function of damage. Random effects for model parameters are used to account for overdispersion. Before introducing this part of the model, we first discuss the dose-damage relationship.

3.1 Dose-Damage Relation

The first part of our model relates dose (observed) to damage (unobserved). In general damage will be a functional $g(F)$ of the dose distribution F to the parotid. As mentioned in the introduction, several biomathematical models that relate dose to damage have been proposed. However, in this article our focus is on the relationship between damage and injury and not on validating the previously proposed dose-damage models. To this end, we take an empirical approach to estimating the damage from the dose distribution. In particular, we use the inverse cumulative distribution function (inverse cdf),

$$d = F^{-1}(p),\ p \in [0, 1],$$

to specify the relationship between dose and damage. Formally d is the dose at which $100p$ percent of the parotid received dose d or less . We take d to be a measure of damage, thus d can take values from 0 to the maximum possible dose in this study, 81.5 Gray. We estimate the posterior distribution of p from the data. A relatively large value of p suggests that the dose to a majority of the parotid gland is important in assessing overall damage. Whereas a relatively small value of p suggests that the dose to only a small portion of the parotid gland is necessary. This could have implications as to whether an attempt should be made

to reduce the overall dose distribution to the parotid gland, or to attempt to spare a portion of the gland and not worry about the rest. We assign the quantile p, which we assume to be the same for all subjects, a $U[0,1]$ (uniform) prior distribution. To estimate the inverse cdf at any quantile $p \in [0,1]$ we first estimate the dose density with the *density* function in R[1] at 101 equally spaced points from 0 to 81.5. From this we have estimates of damage at intervals of 0.01 from 0 to 1. Then for any $p \in [0,1]$ not a multiple of 0.01, we linearly interpolate to estimate the inverse cdf.

3.2 Damage-Injury Relation

The second part of our model relates damage to the injury incurred by the parotid gland. Specifically, the (transformed) expected saliva flow rate is modeled as a smooth, monotone decreasing parametric function of damage. In our particular application saliva flow rate is used to determine the extent of injury to the parotid glands. This part of the model is a non-linear mixture model with both fixed and random effects. As mentioned in section 2, there are numerous outcome measures of zero. To account for these zero saliva flow rates, we model saliva flow rate at a given level of damage and at each time, as a mixture of a Dirac measure (point mass) at zero and a gamma distribution. This type of mixture model (a point mass at zero mixed with a standard distribution function) is reminiscent of Lambert's (Lambert, 1992) zero-inflated Poisson regression model. The parotid glands are indexed by k (= 0 for ipsilateral and = 1 for contralateral), time is indexed by j and the stimulation condition is indexed by ℓ (= 0 for normal and = 1 for stimulated). Subjects are indexed by i. d_{ik} will denote the damage incurred by parotid k of subject i.

Let $Z_{ijk\ell}$ be a latent allocation variable taking values in $\{0,1\}$ such that $p(Z_{ijk\ell} = 1) = \pi_{j\ell}(d_{ik})$. Let, $N(\mu, \sigma^2)$ denote the normal distribution with mean μ and variance σ^2 and $G(\mu, \nu)$ denote the gamma distribution with mean μ and shape parameter ν. Thus, if $X \sim G(\mu, \nu)$ the density of X is given by $(\nu/\mu)^\nu x^{\nu-1} e^{-\nu x/\mu}/\Gamma(\nu)$. Then saliva flow rate $Y_{ijk\ell}$

[1]The R Project for Statistical Computing. `http://www.r-project.org/`

conditional on the allocation variable has the following distribution:

$$\begin{aligned} Y_{ijk\ell}|Z_{ijk\ell} = 0 &\sim N(0,0) \quad \text{(Dirac measure at zero)} \\ Y_{ijk\ell}|Z_{ijk\ell} = 1 &\sim G(\mu_{j\ell}(d_{ik}), \nu_{j\ell}). \end{aligned}$$

Marginalizing over $Z_{ijk\ell}$, $Y_{ijk\ell}$ is a mixture of a point mass at zero and a gamma random variable:

$$Y_{ijk\ell} \sim [1 - \pi_{j\ell}(d_{ik})]N(0,0) + \pi_{j\ell}(d_{ik})G(\mu_{j\ell}(d_{ik}), \nu_{j\ell}).$$

We model the mixing parameters, $\pi_{j\ell}$, as a smooth, monotone decreasing functions of damage allowing for the possibility that the mixing proportion at a damage of zero is less than one ($\pi_{j\ell}(0) < 1$). This allows subject to have zero saliva flow at baseline, prior to radiation therapy, due to xerostomia (dry mouth)—a condition that is not uncommon among the elderly. $\pi_{j\ell}(d_{ik})$ can also be interpreted as the probability of non-zero saliva flow rate for a subject that has incurred damage level d_{ik} to parotid k under stimulation condition ℓ at time j. In particular we model the logit of $\pi_{j\ell}$ as a linear function of damage:

$$\text{logit}[\pi_{j\ell}(d_{ik})] = \theta_\ell - \phi_j d_{ik}. \tag{1}$$

The intercept, $\theta_\ell > 0$, is indexed by stimulation condition and the slope, $\phi_j > 0$, is indexed by time. Indexing the intercept by stimulation condition allows for differing baseline proportions of subjects with no saliva flow between the stimulated and normal conditions. Indexing the slope by time allows for the probability of non-zero saliva flow to change over time at a given level of damage, thus allowing for the possibility that the parotid recovers from damage over time. θ_ℓ and ϕ_j are assigned vague, proper exponential priors with rates 0.01 ($\mathcal{E}(0.01)$). At baseline, prior to radiation therapy, we set $\phi_0 \equiv 0$ since no radiation has been given, no damage has occurred due to radiation exposure. Therefore we estimate the posterior distribution of ϕ_j only at the six post-treatment times $j = 1, 2, 3, 4, 5, 6$ corresponding to 1, 3, 6, 12, 18 and 24 months post radiation therapy, respectively.

When $Z = 1$ we require that the conditional expected outcome, $\mu_{j\ell}(d_{ik})$, to be a (parametrized) smooth monotone decreasing function of the damage to the parotid: $\mu_{j\ell}(d_{ik})$:

$[0,\infty] \to [0,\infty]$. Let $\alpha_{ik\ell}, \beta_{ij\ell}, \gamma_{ik\ell} > 0$. We use a scaled version (scaled by $\gamma_{ik\ell}$) of the incomplete gamma function

$$\mu_{j\ell}(d_{ik}) = \gamma_{ik\ell} \underbrace{\frac{\beta_{ij\ell}^{\alpha_{ijk\ell}}}{\Gamma(\alpha_{ijk\ell})} \int_{d_{ik}}^{\infty} t^{\alpha_{ijk\ell}-1} e^{-t\beta_{ij\ell}} dt}_{\text{incomplete gamma function}} \tag{2}$$

(Here we have taken a bit of notational liberty. Technically we should have shown the conditional dependence of $\mu_{j\ell}$ on the parameters $\alpha_{ijk\ell}$, $\beta_{ij\ell}$ and $\gamma_{ik\ell}$. Henceforth, this conditional dependence will be implicit.) $\mu_{j\ell}$ achieves its conditional maximum, $\gamma_{ik\ell}$, at $d_{ik} = 0$ and has an asymptote of zero as $d_{ik} \uparrow \infty$. We chose this function because of it flexibility in shape. It takes three parameters, γ, which scales the range and α, β which together control location and shape. It has a sigmoid shape and inflection point at $\alpha(\alpha-1)/\beta$ for $\alpha > 1$, is exponential when $\alpha = 1$ and is supra-exponential for $\alpha < 1$. One could also choose the survivor function of any distribution with non-negative support in lieu of the incomplete gamma function, for example, the Weibull distribution.

We model the priors for the parameters of the mean function, $\mu_{j\ell}$, in a hierarchical fashion, thus accounting for some of the correlation within subjects across parotid glands, stimulation condition and time. In particular, let $\boldsymbol{\gamma}_i = (\gamma_{i00}, \gamma_{i01}, \gamma_{i10}, \gamma_{i11})^{\mathrm{T}}$, all i. Further, let $W_d(S)$ denote the Wishart distribution with d degrees of freedom and symmetric, positive definite scale matrix S for which the dimension will be obvious. To satisfy the positivity constraint on $\boldsymbol{\gamma}_i$ we use a lognormal prior:

$$\begin{aligned} \ln(\boldsymbol{\gamma}_i) &\sim N(\ln(\boldsymbol{\gamma}), \Phi) \\ \ln(\boldsymbol{\gamma}) &\sim N(\mathbf{g}, V) \\ \Phi^{-1} &\sim W_5(R) \end{aligned}$$

where $\mathbf{g} = \ln(0.1, 0.1, 0.1, 0.1)^{\mathrm{T}}$, $V = 100\,\mathrm{I}_4$, $R = .01\,\mathrm{I}_4$ and I_d is the $d \times d$ identity matrix. This gives a rather vague, but proper prior on both the random effects and their mean. To account for overdispersion, random parotid and stimulation effects within each subject, $\boldsymbol{\gamma}_i$, are centered about the population mean, $\boldsymbol{\gamma}$, with covariance Φ. Φ accounts for correlation

within subjects between and across parotids and stimulation conditions at baseline. Apriori, we assume the components of the random effects vectors are uncorrelated as well as the components of the population mean vector.

Lognormal priors are also place on the other two parameters of the mean function. Specifically, let $\boldsymbol{\alpha}_{ij} = (\alpha_{ij00}, \alpha_{ij01}, \alpha_{ij10}, \alpha_{ij11})^{\mathrm{T}}$, all i and j. Let $\boldsymbol{\beta}_{i\ell} = (\beta_{i1\ell} \ldots \beta_{i6\ell})^{\mathrm{T}}$ for all i, ℓ. Then

$$\begin{aligned} \ln(\boldsymbol{\alpha}_{ij}) &\sim N(\ln(\boldsymbol{\alpha}), \Theta) \\ \ln(\boldsymbol{\alpha}) &\sim N(\mathbf{a}, U) \\ \Theta^{-1} &\sim W_5(T) \end{aligned} \tag{3}$$

where $\mathbf{a} = \ln(3, 3, 3, 3)^{\mathrm{T}}$, $U = \mathrm{I}_4$ and $T = 2\mathrm{I}_4$; and

$$\begin{aligned} \ln(\boldsymbol{\beta}_{i\ell}) &\sim N(\ln(\boldsymbol{\beta}), \Psi) \\ \ln(\boldsymbol{\beta}) &\sim N(\mathbf{h}, W) \\ \Psi^{-1} &\sim W_7(S) \end{aligned} \tag{4}$$

where $\mathbf{h} = (0.3, \ldots, 0.3)^{\mathrm{T}}$, $W = \mathrm{I}_6$ and $S = 2\mathrm{I}_6$.

Θ accounts for post-treatment correlation between stimulation conditions and parotid glands. We assume this correlation is constant over time. Ψ accounts for correlation over time within a subject. We assume that this correlation is the same for both normal and stimulated conditions. We note here that the prior specifications for the mean of the random effects, the prior distributions in (3) and (4) are somewhat informative. These priors were chosen, however, in combination to give a broad range of shapes (Figure 2) for the prior mean function $\mu_{j\ell}$ which is of particular interest to this study. Thus, we believe that this prior specification does not overly influence the posterior distribution of the mean function. Finally, the shape parameters, $\nu_{j\ell}$, are given an exponential prior distribution with rate 0.1. This completes our model specification.

Accounting for correlation over time within a parotid gland , between parotid gland and across stimulation conditions, along with numerous outcome measures of zero and any po-

tential overdispersion requires one of two solutions: 1) Either make simplifying assumptions and ignore some or all potential correlation within the data or 2) fit the data with a necessarily complex model that accounts for the various forms of correlation, numerous outcome measures of zero and included random effects to help account for overdispersion. The latter choice is preferred from a statistical point of view.

4 MCMC Estimation

Because of the complexity of the model, the posterior distribution of the model parameters is not analytically tractable and so we estimate the posterior distribution through Markov Chain Monte Carlo (MCMC) simulation. The MCMC simulation is a hybrid Metropolis-Hastings and Gibbs algorithm. For those parameters whose full conditional posterior distributions are analytically tractable $(\Psi, \Theta, \Phi, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\beta})$ we perform a Gibbs update and perform a Metropolis-Hastings step for all other parameters. Let $p(\theta \mid \cdot)$ denote the full conditional posterior distribution of a generic random variate θ. We update parameters in the following order: θ_ℓ, $\ell = 0, 1$; ϕ_j, $j = 1, \ldots, 6$; $\boldsymbol{\gamma}_{ik}, i = 1, \ldots, N$, $k = 0, 1$; Φ; $\boldsymbol{\gamma}$; $\boldsymbol{\beta}_{i\ell}, i = 1, \ldots, N$, $\ell = 0, 1$; Ψ; $\boldsymbol{\beta}$; $\boldsymbol{\alpha}_i, i = 1, \ldots, N$; Θ; $\boldsymbol{\alpha}$; $\nu_{j\ell}$, $j = 0, 1, \ldots, 6$, $\ell = 0, 1$; and p.

A complication arises in estimating the mean function $\mu_{j\ell}(d_{ik})$ for a particular parotid gland from a specific subject under a distinct stimulation condition at a given time in that the observation is obtained at only one level of damage, d_{ik}. However, we would like to obtain a smooth, monotone decreasing estimate of saliva flow rate as a function of damage incurred. And we would like to have this estimate for each parotid under both conditions. To circumvent this complication, we use the baseline observation as a second point, which corresponds to the observed saliva flow rate at $d_{ik} = 0$. This assumes that saliva flow rate is not a function of time when no radiation has been delivered, which we believe is a reasonable assumption. A third observation can be obtained as a latent variable evaluated at the estimated damage incurred by the opposing parotid gland within the particular patient of interest. This is accomplished by drawing the latent observation from its posterior predictive

density. These draws are easily obtained at each step of the MCMC simulation. First draw $Z^*_{ij\ell k}$ from a Bernoulli distribution with parameter $\pi_{j\ell}(d_{ik^*})$, where $k^* = |k-1|$ (i.e. at the damage received by the other parotid gland). Then if $Z^*_{ij\ell k} = 0$, set $Y^*_{ij\ell k} = 0$ and if $Z^*_{ij\ell k} = 1$ draw $Y^*_{ij\ell k}$ from a gamma distribution with mean $\mu_{j\ell}(d_{ik^*})$ and shape parameter $\nu_{j\ell}$. Here, $\pi_{j\ell}(d_{ik^*})$, $\mu_{j\ell}(d_{ik^*})$ and $\nu_{j\ell}$ are the current draws from the posterior distribution of the parameters. By the hierarchical structure placed on the parameters of the mean function and assuming that the covariance matrix Θ is constant over time and Ψ is the same regardless of stimulation condition, we borrow strength over time, across stimulation conditions and between parotid glands when drawing these latent variables from the aforementioned gamma distribution.

Missing data are treated as missing completely at random (Little and Rubin, 2002). Since this is an ongoing study, missing data can occur because patients have not completed the study, missed appointments, loss to follow-up and even death. However, complications to the parotid glands are most likely not the cause of death and so we feel that this assumption is valid. Further, in a related analysis on a subset of this data, Eisbruch et al. (2001) found no relationship between xerostomia questionnaire scores and patterns of missing data. Missing observations are also treated as latent variables and the values are drawn from the posterior predictive distribution as above. Drawing the missing data gives us a complete balanced data set which simplifies computation for some of the model parameters. In particular it allows us to easily draw the precision matrices Φ^{-1}, Θ^{-1} and Ψ^{-1} from their conjugate full conditional distributions.

We ran the simulation for 250,000 iterations with a burn-in of 50,000 iterations. We assessed convergence graphically and ran the chain from several different initial states. The posterior distribution was estimated by saving every 10th iteration after burn-in for a total of 20,000 draws from the posterior.

5 Data Analysis Results

Damage, as we have modeled it, is a function of a random parameter p. As such, damage has a posterior distribution and can cause difficulty in interpreting results of the model, especially for radiation oncologists who are interested, among other things, the probability of non-zero saliva flow at a specified level of damage. That is, at a point estimate of damage. The estimated marginal posterior density of p is shown in Figure 3 and has a median of 0.341. The estimated 95% Bayesian credible interval (equal tail area) for p is (0.295, 0.381). We can see that this distribution is fairly tight around its median value and roughly symmetric. Therefore, we reran our simulation, conditional on $p = 0.341$. Results of the unconditional simulation and the conditional simulation are virtually indistinguishable. Hence, we report results conditional on $p = 0.341$.

Two summary measures of interest are the probability of non-zero saliva flow at each time point under both conditions, $\pi_{j\ell}(d)$, and the marginal population mean saliva flow rate as a function of damage for each condition and parotid gland, $\mu_{j\ell}(d)$. The first of these is obtained from equation (1). The second can be obtained by equation (2) evaluated at the median values of $\alpha_{ik\ell}$, $\beta_{ij\ell}$ and $\gamma_{ik\ell}$. We concentrate on the 12 month results.

The median probability of non-zero saliva flow rate is displayed in Figure 4 for both stimulation conditions along with the Bayesian 95% point-wise credible intervals (Wahba, 1983). For normal saliva flow rate the damage with a fifty percent chance of a zero saliva flow rate at 1 year is estimated to be $\text{LD}_{50} = 26.8$ Gy. Let $\tilde{\theta}_\ell$ and $\tilde{\phi}_j$ denote the median values of the estimated marginal posterior distributions of θ_ℓ and ϕ_j, respectively. An estimate of the LD_{50} at time j for stimulation condition ℓ is $\tilde{\theta}_\ell/\tilde{\phi}_j$. Under stimulated conditions, the estimated $\text{LD}_{50} = 48.3$ Gy at 1 year. The estimated LD_{50} at each of the post-radiation therapy follow-up times are tabulated in Table 2. We see that there is an increase in the LD_{50} up to 18 months and then a slight drop-off. The increase indicates that the parotid glands recover from radiation induced damage, to some degree. The slight drop-off may be due to chance variation or may be due to the pattern of missing observations shown in Table 1.

Table 2: Estimated LD_{50} for each follow-up time with the 95% Bayesian credible intervals.

	Stimulation Condition			
	Normal		Stimulated	
Month	LD_{50}	95% Credible Int.	LD_{50}	95% Credible Int.
1	19.08	(16.42, 22.11)	35.05	(30.43, 39.95)
3	19.31	(16.54, 22.44)	35.45	(30.75, 40.47)
6	22.03	(19.15, 25.29)	40.42	(35.79, 45.36)
12	26.39	(23.04, 30.07)	48.33	(43.44, 53.62)
18	29.91	(25.95, 34.61)	54.85	(48.90, 61.77)
24	28.09	(24.38, 32.54)	51.57	(45.69, 58.21)

The median of the estimated marginal population mean saliva flow rate posterior distribution for each condition by parotid gland is shown in Figure 5 along with the Bayesian 95% credible bands. In this figure, the gray shaded area represents the range of values of damage estimated from the observed data for that particular parotid gland. The solid black dots represent sample means from the observed data at 12 months. The observed data were first grouped into the lower, middle and upper third quantiles of their estimated damage evaluated at the median value of p. Then for each grouped set of data, the mean saliva flow rate was plotted against the mean damage. The sample baseline mean saliva rate is also depicted in the graphs as a solid dot at a damage of 0. The 'Xs' represent the mean saliva flow rate of the means of the latent observations evaluated by grouping the latent observations according to the grouping performed on the observed observations from the opposing parotid gland. From Figure 5 is appears that our model does a decent job at modeling the marginal mean saliva flow rate. Note the similarity in shapes of the predicted marginal mean saliva flow rates in the four figures. Especially the similarity in shape between the ipsilateral and contralateral predicted mean saliva flow rates under normal conditions and between the two parotid glands under stimulated conditions. Also note the positions of the means of the latent variables relative to the positions of the observed means from the opposing parotid gland under each stimulation condition. The latent means track the relative positions of the observed means fairly closely. This is a consequence of borrowed strength across the glands,

stimulation condition and time that was alluded to in section 4.

Figure 6 shows the median of the marginal population mean saliva flow rate posterior distribution superimposed on the 12 month, stimulated data. This figure also displays the baseline observations as well as the means of the latent and missing values that were drawn from the estimated posterior predictive distribution. Again, the latent/missing predicted observations mix nicely throughout the observed data, due to the strength borrowed from the observed data. At damage levels larger than about 25, the latent contralateral predicted observations are a bit higher than the observed ipsilateral observations. This is a consequence of the larger predicted baseline population mean (0.51 for the contralateral side, 0.36 for the ipsilateral side, both under stimulated conditions).

5.1 Goodness-of-Fit

We assessed model adequacy by a Bayesian χ^2 goodness of fit statistic (Johnson, 2004). This goodness of fit test is similar to the classical χ^2 goodness of fit test except in the Bayesian version, the bin probabilities p_b are held fixed at their null values and the bin counts are considered random.

Let $\tilde{\theta}$ denote a sample from the posterior distribution and F the marginal sampling distribution after integrating out the missing data and latent variables. For this goodness of fit test we used 10 equally spaced bins such that the bin counts $m_b(\tilde{\theta})$ are determined by counting the number of $F(y_{ijk\ell} \mid \tilde{\theta}) \in ((b-1)/10, b/10]$ for $b = 1, \ldots, 10$. When $Y_{ijk\ell} \mid \tilde{\theta} = 0$, we use a randomization procedure to determine bin assignment. That is, we draw $z \sim U(0, 1 - \pi_{j\ell}(d_{ik}))$, then if $F(z) \in ((b-1)/10, b/10]$ we increment m_b by one. Johnson's Bayesian χ^2 statistic is then

$$R^B(\tilde{\theta}) = \sum_{b=1}^{10} \left[\frac{m_b(\tilde{\theta}) - np_b}{\sqrt{np_b}} \right]^2,$$

where n is the sample size and p_b is the theoretical probability of bin assignment, in our case with 10 bins, $p_b = 0.1$. Johnson suggests two summary measures of the distribution of $R^B(\tilde{\theta})$. First is the proportion of times it exceeds the critical value of a χ^2_9 distribution, 16.919. That is, the .95 quantile of a χ^2 distribution with 9 degrees of freedom. Values much larger than the nominal 0.05 suggest lack of fit. The second suggestion is the probability that $R^B(\tilde{\theta})$ is greater than a χ^2_9 random deviate. The nominal value of this probability is 0.5 and values much greater than this nominal value suggest lack of fit. For our model, the proportion of times $R^B(\tilde{\theta})$ exceeds the critical value of a χ^2_9 distribution is 0.052. The probability that $R^B(\tilde{\theta})$ is greater than a χ^2_9 random deviate is 0.51. Both of these summaries suggest that our model fits the data adequately. Figure 7 depicts the quantile-quantile plot of R^B values calculated from 20,000 samples from the posterior distribution.

As Johnson points out, the Bayesian χ^2 goodness-of-fit statistic is also a useful tool for code verification. The distribution of R^B tends to deviate remarkably from its null distribution when the model is misspecified or there are coding errors. It is evident from Figure 7 that R^B does not deviate from its null distribution and thus we are confident that we have coded our model correctly and that this is an adequate model for the data.

6 Discussion

We have developed and implemented a realistic, albeit complicated, model to analyze the parotid data set outlined in section 2. The need to account for the numerous outcome measures of zero and to account for within parotid gland correlation over time, correlation between stimulation conditions and correlation between the two parotid glands within a subject requires such a complex model. In this model we used an empirical measure of damage, namely the inverse cdf of the dose distribution and concentrated on modeling the damage-injury relationship in the dose-damage-injury model. Further, our analysis shows that the parotid gland recovers with time. A similar result was found by Scrimger et al. (2004). Given that we now have a statistical model that fits the damage-injury relationship and

takes into account the correlation inherent in the data (over time, across parotid glands and stimulation conditions), our goal is to concentrate on the dose-damage relationship. Several theoretical models have been proposed (see the Introduction) and we will pursue the use of these as well as modified versions of each. Other empirical relationships will also be explored, such as the generalized mean of the dose distribution. The challenge will be in comparing models. Bayes factors are the traditional method of comparing models, however, except for the simplest of models, they are computationally burdensome. Johnson's goodness-of-fit statistic, R^B could be used to compare models. The model whose R^B distribution is closest to it null (in some objective sense) would be chosen as the better model for the data at hand. Other summary measures of model adequacy will also be explored, such as the sum of the logs of the conditional predictive ordinates (Gelfand, Dey, and Chang, 1992).

References

Boersma, L. J., et al. (1998). Estimation of the incidence of late bladder and rectum complications after high-dose (70-85 GY) conformal radiotherapy for prostate cancer, using dose-volume histograms. *Int. J. Radiation Oncology, Biology, Physics* **41**, 83–92.

Chao, K., et al. (2001). A prospective study of salivary function sparing in patients with head-and-neck cancers receiving intensity-modulated or three-dimensional radiation therapy: initial results. *Int. J. Radiation Oncology, Biology, Physics* **49**, 907–916.

Chappell, R., Nondahl, D. M., and Fowler, J. (1995). Modelling does and local control in radiotherapy. *Journal of the American Statistical Association* **90**, 829–938.

Dawson, L. A., et al. (2002). Analysis of radiation-induced liver disease using the lyman ntcp model. *Int J Radiat Oncol Biol Phys* **53**, 810–82.

Deasy, J. O., et al. (2002). Methodological issues in radiation dose-volume outcome analyses: Summary of a joint aapm/nih workshop. *Medical Physics* **29**, 2109–2127.

Eisbruch, A., 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. Radiation Oncology Biology Physics* **45**, 577–587.

Eisbruch, A., et al. (2001). Xerostomia and its predictors following parotid-sparing irradiation of head-and-neck cancer. *Int. J. Radiation Oncology Biology Physics* **50**, 695–704.

Gelfand, A. E., Dey, D. K., and Chang, H. (1992). *Bayesian Statistics 4*, chapter Model Determination using predictive distributions with implementation via sampling-based methods (with Discussion). Oxford: Oxford University Press.

Heft, M. W. and Baum, B. J. (1984). Unstimulated and stimulated parotid salivary flow rate in individuals of different ages. *J Dent Res* **63**, 1182–1185.

Jackson, A., Kutcher, G. J., and Yorke, E. D. (1993). Probability of radiation induced complications for normal tissues with parallel architecture subject to non-uniform radiation. *Medical Physics* **20**, 613–625.

Jackson, A., et al. (1995). Analysis of clinical complication data for radiation hepatitic using a parallel architecture model. *Int. J. Radiation Oncology, Biology, Physics* **31**, 883–891.

Johnson, V. E. (2004). A Bayesian chi-squared test for goodness of fit. *Annals of Statistics, to appear* **32**.

Kwa, S. L. S., et al. (1998). Radiation pneumonitis as a function of mean dose lung dose: an analysis of pooled data of 540 patients. *Int. J. Radiation Oncology, Biology, Physics* **42**, 1–9.

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* **34**, 1–14.

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

Little, R. J. A. and Rubin, D. B. (2002). *Bayesian Data Analysis*. New York: Wiley-Interscience.

Lyman, J. T. (1985). Complication probability as assessed from dose-volume-histograms. *Radiation Research* **104**, s13–19.

Niemierko, A. and Goitein, M. (1992). Modeling of normal tissue response to radiation: the critical volume model. *Int. J. Radiation Oncology, Biology, Physics* **25**, 135–145.

Roberts, S. and Hendry, J. (1993). A direct maximum-likelihood analysis of a collection of worldwide tumour-control data. *Radiotherapy and Oncology* **29**, 69–74.

Roesink, J., et al. (2001). Quantitative dose-volume response analysis of changes in parotid gland function after radiotherapy in the head-and-neck region. *Int. J. Radiation Oncology, Biology, Physics* **51**, 938–946.

Schilstra, C. and Meertens, H. (2001). Calculation of the uncertainty in complication probability for various does-response models. *Int. J. Radiation Oncology, Biology, Physics* **50**, 147–158.

Scrimger, R. A., et al. (2004). Phenomenologic model describing flow reduction for parotid gland irradiation with intensity-modulated radiotherapy: evidence of significant recovery effect. *Int J Radiation Oncology, Biology, Physics* pages 178–185.

Seppenwoolde, Y., et al. (2003). Comparing different ntcp models that predict the incidence of radiation pneumonitis. *Int J Radiat Oncol Biol Phys* **55**, 724–735.

Shultheiss, T. E. (2001). The controversies and pitfalls in modeling normal tissue radiation injury/damage. *Seminars in Radiation Oncology* **11**, 210–214.

Shultheiss, T. E., Orton, C. G., and Peck, R. A. (1983). Models in radiotherapy: volume effects. *Medical Physics* **10**, 410–415.

Stavrev, P. and Stavrev, N. (2000). Fraction size and dose parameters related to the incidence of pericardial effusions: regarding Martel et al. *Int. J. Radiation Oncology,*

Biology, Physics **48**, 609–617.

Theuws, J. C., et al. (1998). Prediction of overall pulmonary function loss in relation to the 3-d dose distribution for patients with breast cancer and malignant lymphoma. *Radiotherapy and Oncology* **49**, 233–243.

Wahba, G. (1983). Bayesian "confidence intervals" for the cross-validated smoothing spline. *Journal of the Royal Statistical Society, Series B, Methodological* **45**, 133–150.

York, E. D., et al. (1993). Probability of radiation-induced complications in normal tissue with parallel architecture under conditions of uniform whole or partial organ irradiation. *Radiotherapy and Oncology* **26**, 226–237.

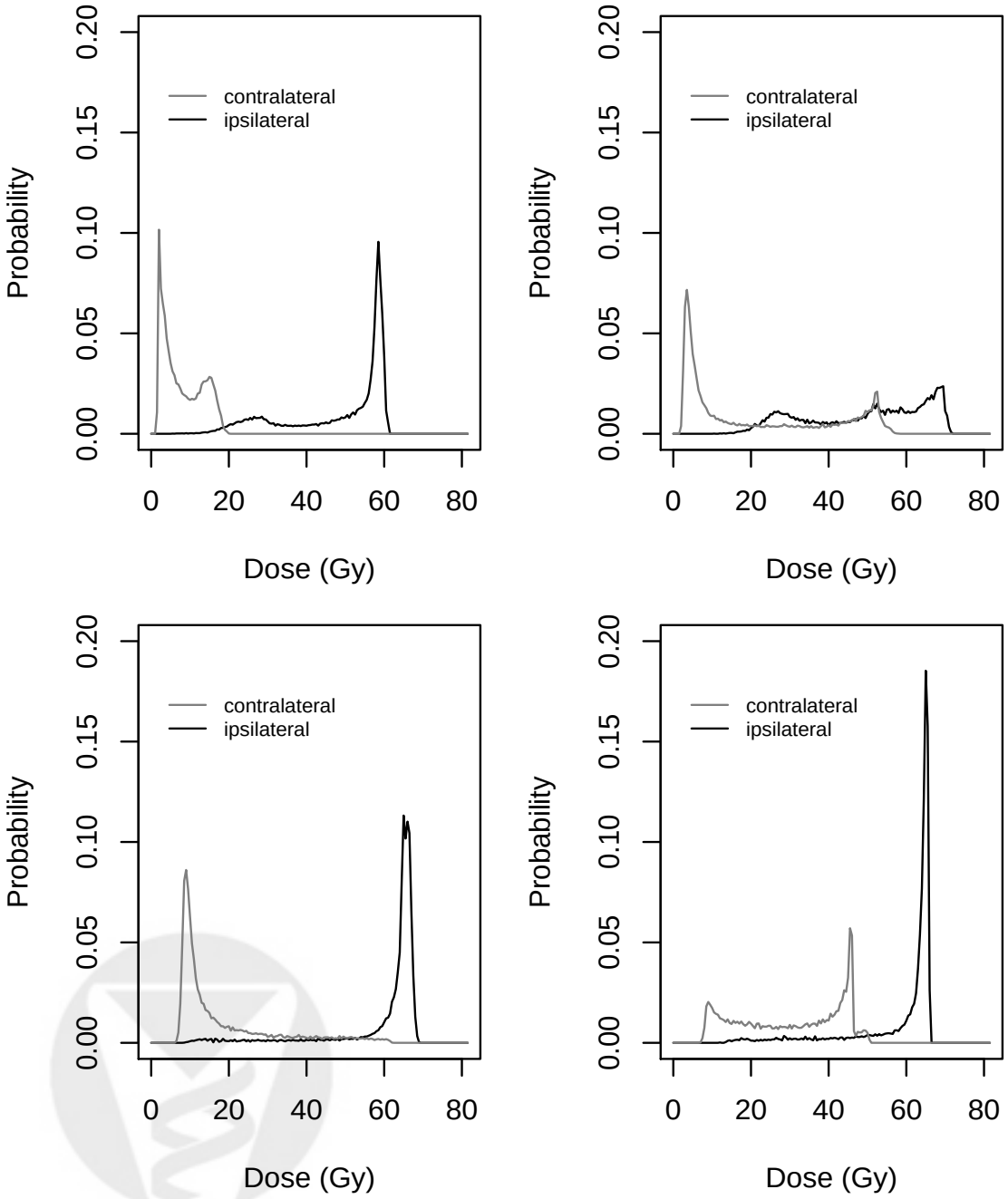

Figure 1: Four examples of the dose volume histogram (DVH). These examples were chosen to show the heterogeneity of the DVHs across patients.

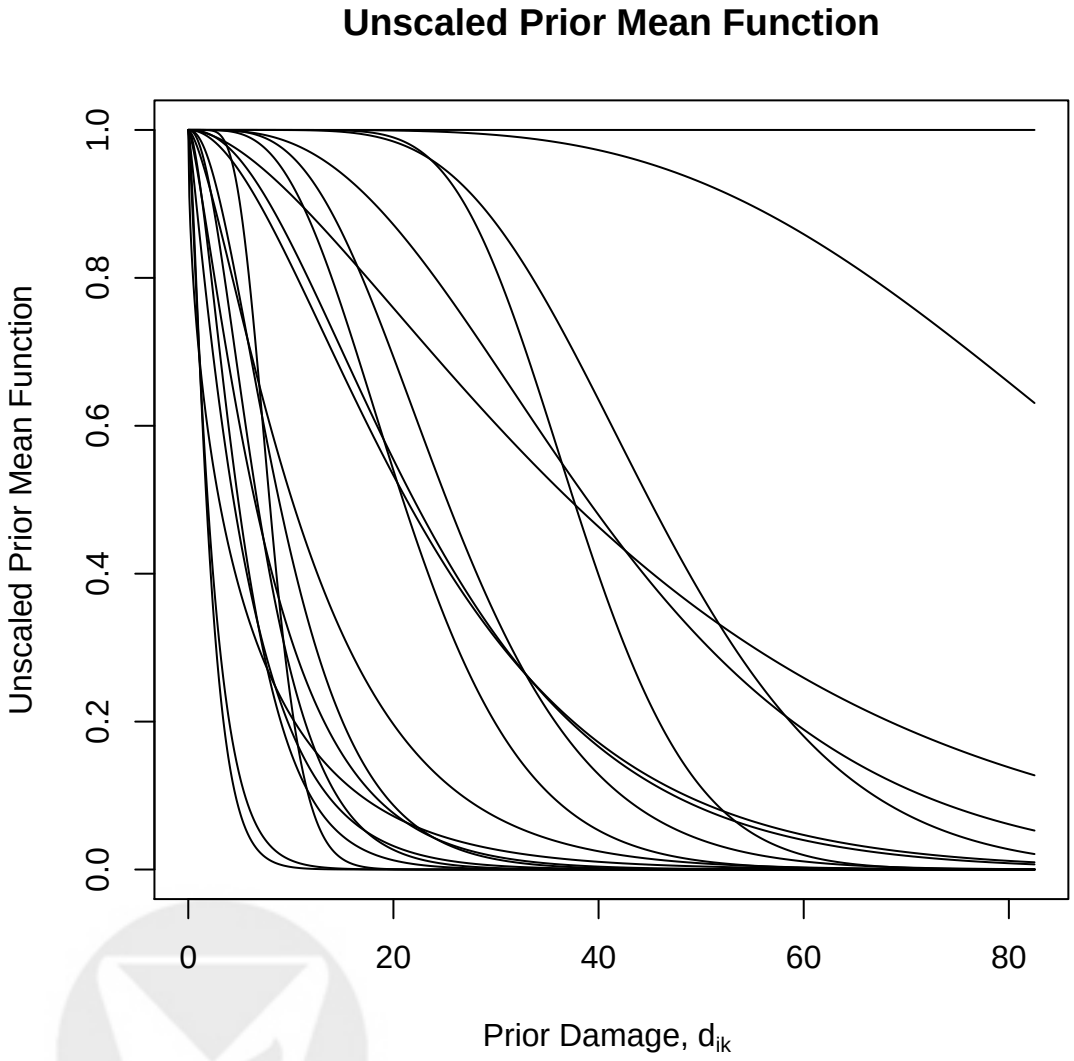

Figure 2: 20 random realizations of the unscaled prior mean function, $\mu_{j\ell}(d_{ik})/\gamma_{ik\ell}$ (equation (2)), based on 20 random draws from the priors for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ (priors (3) and (4)).

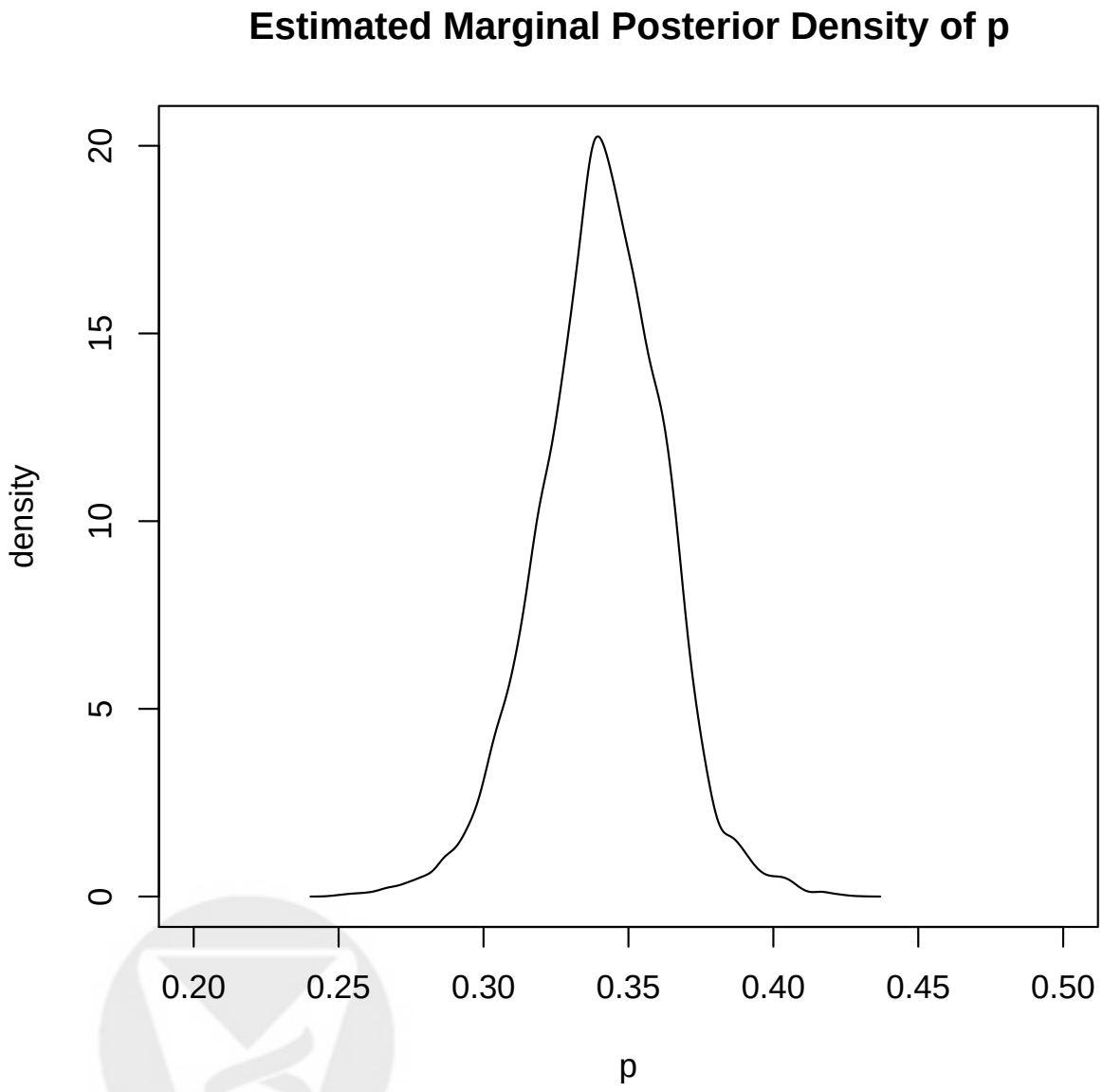

Figure 3: The marginal, estimated posterior density of p. From this figure we see that the marginal posterior distribution is fairly symmetric and not too variable.

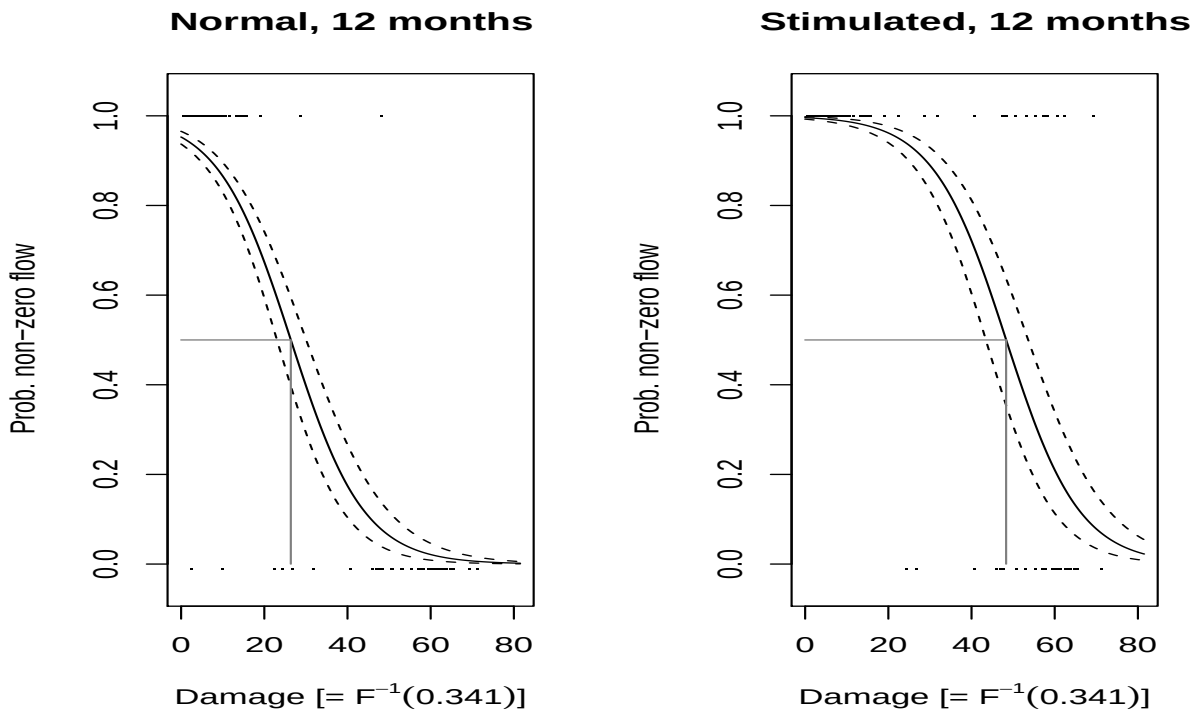

Figure 4: The probability of non-zero saliva flow rate as a function of damage at 1 year post-radiation therapy. The solid black line is the median probability. The dashed lines are the 95% Bayesian point-wise credible intervals. Observations are shown as dots at 0 on the vertical axis (those with zero saliva flow and at 1 (those with positive saliva flow).

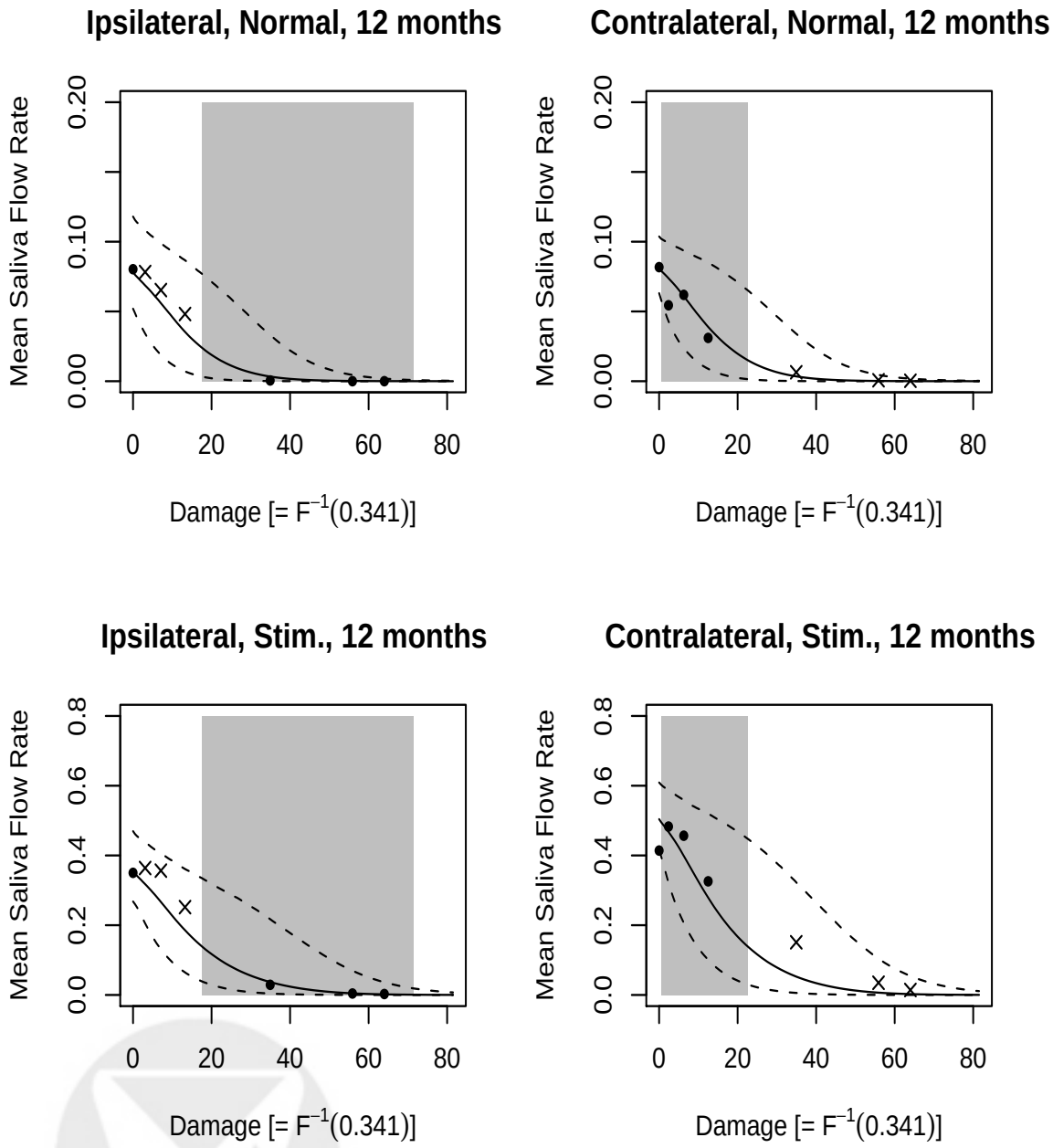

Figure 5: The marginal population mean saliva flow rate at 1 year post-radiation therapy. The solid black line is the median of the estimated marginal population mean saliva flow rate. The dashed lines are the 95% Bayesian confidence bands. The solid dot at zero represents the mean of the observed baseline data. The gray shaded region represents the range of estimated damage, evaluated at $p = 0.341$, from observed data. The solid dots within the gray regions are grouped observed mean saliva flow rates evaluated at the mean level of damage within each group. Groups were formed by taking the lower, middle and upper 1/3 quantiles. The Xs represent the mean values of grouped mean latent saliva flow rates.

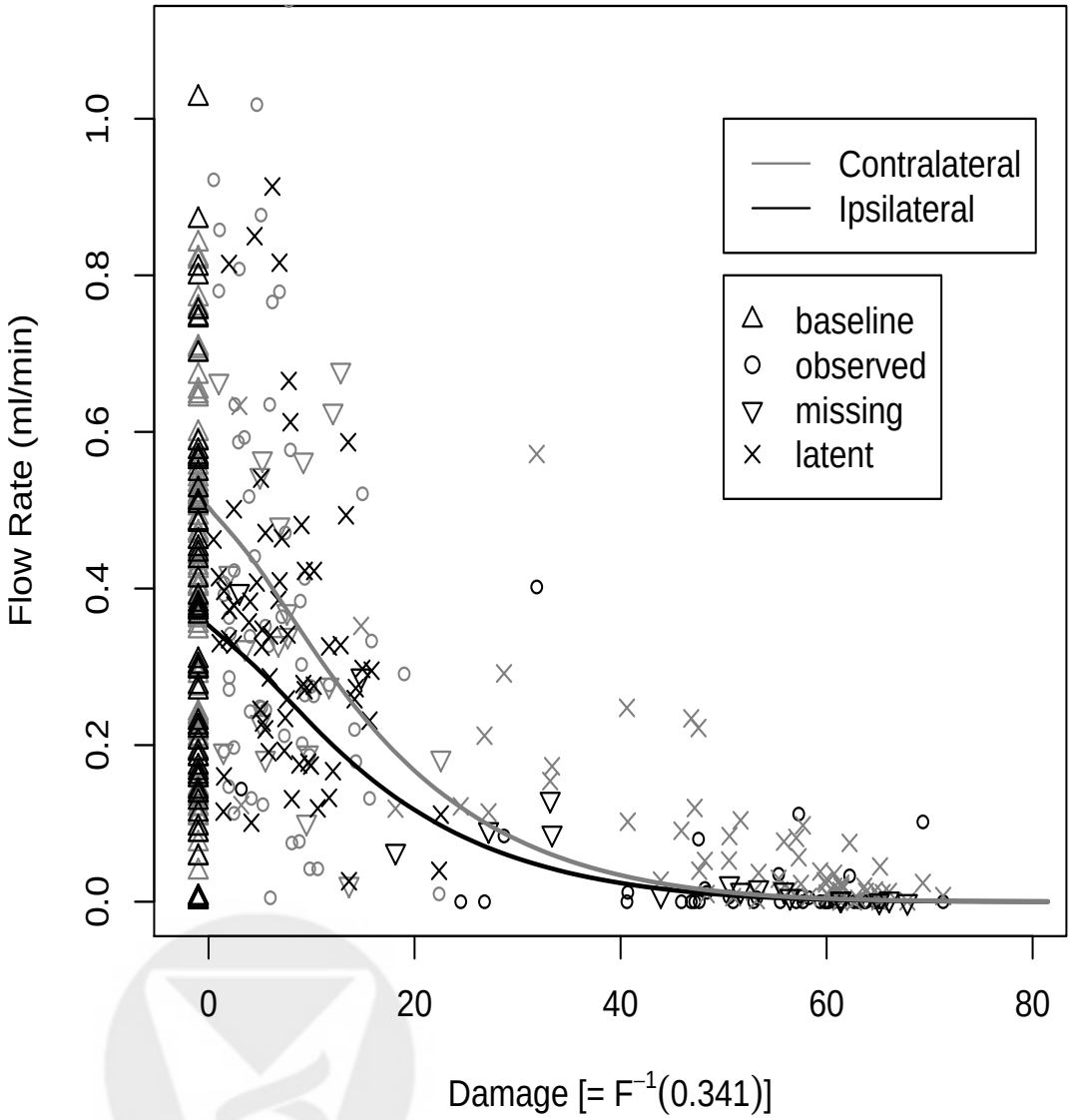

Figure 6: The median of the estimated marginal population mean stimulated saliva flow rate posterior distribution for both contraleral and ipsilateral parotid glands (gray and black lines respectively) at 12 months. Baseline data are indicated by △, observed data by ∘, missing data (means) by ▽ and latent variables (means) by ×. All baseline data are observed and plotted at -1 to assist in visual clarity.

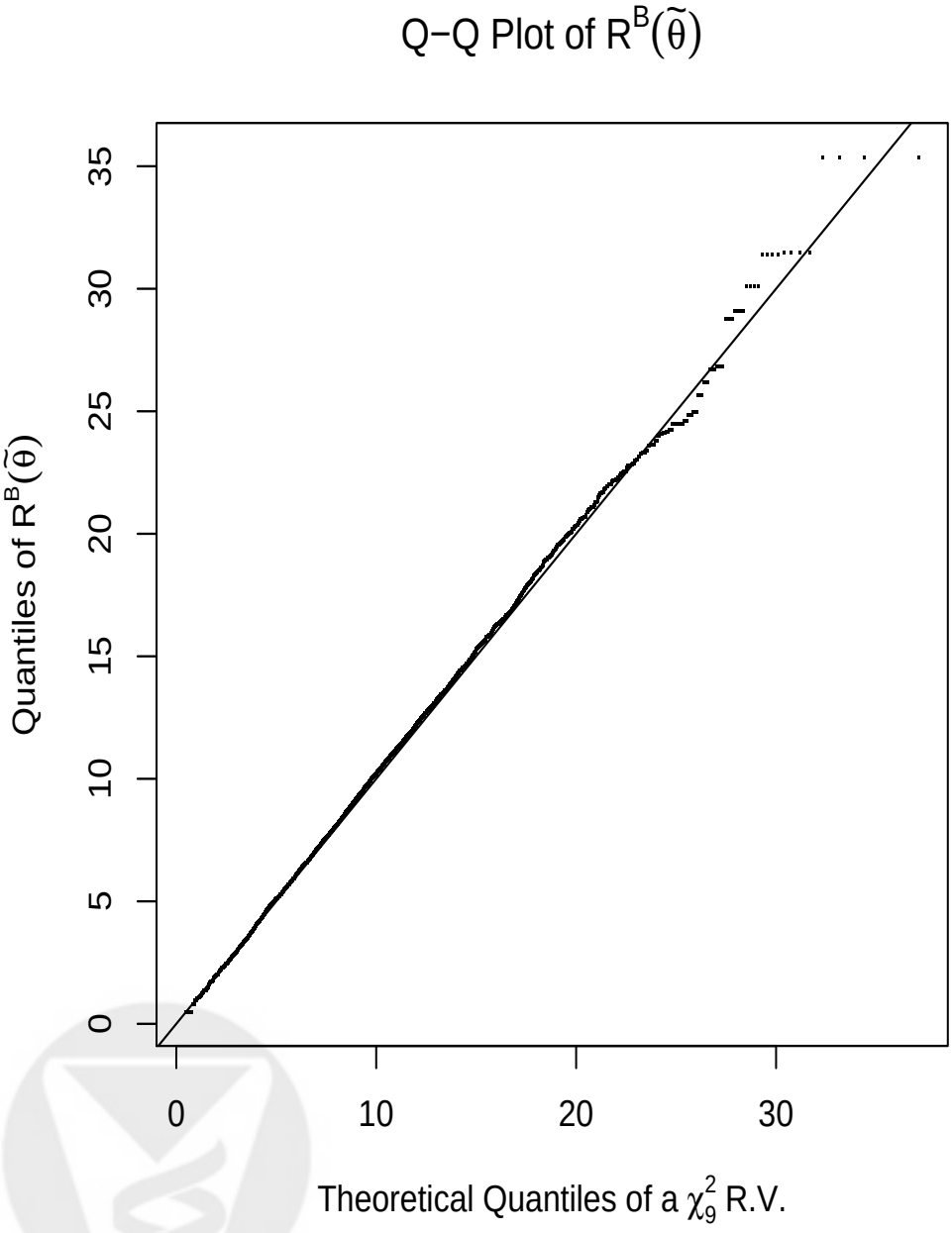

Figure 7: Quantile-quantile plot of R^B values for 20,000 draws from the posterior distribution of our model. Observed values were sorted into 10 equally spaced bins. The expected order statistics are from a χ^2 distribution with 9 degrees of freedom.