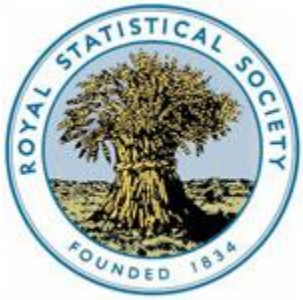

Incorporation of Historical Controls Using Semiparametric Mixed Models
Author(s): Helen Parise, M. P. Wand, David Ruppert and Louise Ryan
Source: *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 50, No. 1 (2001), pp. 31-42
Published by: Wiley for the Royal Statistical Society
Stable URL: http://www.jstor.org/stable/2680839
Accessed: 14-01-2016 01:36 UTC

Appl. Statist. (2001)
50, *Part* 1, *pp.* 31–42

Incorporation of historical controls using semiparametric mixed models

Helen Parise and M. P. Wand,

Harvard School of Public Health, Boston, USA

David Ruppert

Cornell University, Ithaca, USA

and Louise Ryan

Harvard School of Public Health, Boston, and Dana-Farber Cancer Institute, Boston, USA

[Received November 1999. Revised April 2000]

Summary. The analysis of animal carcinogenicity data is complicated by various statistical issues. A topic of recent debate is how to control for the effect of the animals' body weight on the outcome of interest, the onset of tumours. We propose a method which incorporates historical information from the control animals in previously conducted experiments. We allow non-linearity in the effects of body weight by modelling the relationship nonparametrically through a penalized spline. A simple extension of the penalized spline model allows the relationship between weight and onset of the tumour to vary from one experiment to another.

Keywords: Body weight; Carcinogenicity; Penalized spline; Phenolphthalein; Rodent bioassay

1. Introduction

It is quite common for control animals in a rodent carcinogenicity study to weigh substantially more than treated animals throughout the course of an experiment. Several researchers have reported a lower incidence of tumours corresponding to lower body weights (Hart *et al.*, 1995; Haseman *et al.*, 1994; Seilkop, 1995). Thus, dose-related differences in body weights could bias the conclusions drawn from these studies. In fact, a recent editorial in *Science* was highly critical of risk assessment agencies for not controlling the caloric intake of experimental animals (Abelson, 1995). Abelson advocated changing the protocol of rodent experiments to include dietary restrictions such that animals are maintained at some 'ideal' body weight.

Indeed, many studies conducted by the National Toxicology Program (NTP) of the USA have shown apparent protective effects of the chemical being tested on the incidence of certain tumours. These apparent reductions in incidence across dose may be due to differences in body weight (Hart *et al.*, 1995). This is illustrated in Table 1 with data from the NTP study of phenolphthalein — an ingredient in over-the-counter laxatives that has been recently withdrawn by the US Food and Drug Administration (National Toxicology Program, 1995).

Address for correspondence: Helen Parise, Department of Environmental Health, Environmental Science and Engineering Program, Harvard School of Public Health, 655 Huntington Avenue, Boston, MA 02115, USA.
E-mail: hparise@hsph.harvard.edu

 0035–9254/01/50031

Table 1. Summary of mammary tumours and average body weights for the female rats in the phenolphthalein data

	Tumour rates for the following doses:			
	0 ppm	*12000 ppm*	*25000 ppm*	*50000 ppm*
Mammary tumour rate	32/50	25/50	17/50	25/50
Mean body weight†	287	258	254	249
p-value‡	0.043 (N)§	0.021 (N)§§	< 0.001 (N)§§	0.073 (N)§§

†Average body weight at 12 months.
‡N indicates a negative trend.
§For an overall trend based on the logistic regression test.
§§As compared with the control.

Hart and Turturro (1997) suggested mechanisms for the inhibition of spontaneous cancers under dietary restrictions which include decreased body weight leading to decreased cellular proliferation and increased apoptosis in organs which increase and decrease with body size. Although methods of dietary restriction are being researched, this suggestion does not help us to analyse currently available toxicological data. Thus, assessing the risk of tumour from these data may require a statistical adjustment for the confounding effect of weight.

Body weight is difficult to adjust for in a single experiment. Gaylor and Kodell (1999) proposed a method to adjust for body weight in the current experiment which is analogous to the age adjustment of Peto's test (Peto *et al.*, 1980). Animals are stratified into groups based on body weight, dose–response trends are calculated within each group and these test statistics are then pooled to form an overall test. However, it can be unclear how to group the data based on body weight and there simply may not be enough data in each strata to calculate the test statistic efficiently.

The incidence of tumours is generally low in control animals so very few events occur in the range of body weights of the control animals. Even when there is a high background rate of tumours, as in our motivating example, a high correlation between dose and body weight is often observed since the distribution of body weights in the control animals can be quite different from that of the dosed groups. One way to circumvent this problem is to incorporate information from the extensive historical control database maintained by the NTP. In fact, Seilkop (1995) proposed such an approach, although he did not properly account for study-to-study variability.

Several methods have been developed for incorporating historical information into tests of trend in the animal carcinogenicity setting. Unfortunately, most focus on the Cochran–Armitage test, which ignores times of death and can be biased in the presence of treatment-related toxicity (Bailer and Portier, 1988). Tarone (1982) was among the first to develop a method to incorporate historical controls into the Cochran–Armitage test by assuming that control probabilities arise from a beta prior. Others have proposed various modifications of Tarone's method. For example, Dempster *et al.* (1983) performed a fully Bayesian analysis within a logistic model. Kikuchi and Tanagawa (1991) used a normal random-effects model with a variance stabilizing transformation. To address the issue of age adjustment, Ibrahim *et al.* (1998) suggested a Bayesian approach within a logistic model that allows for the incorporation of covariates. Also, Ibrahim and Ryan (1996) incorporated historical controls into a survival model.

We propose the use of a logistic mixed model where the logit of the probabilities of the tumour are assumed to come from a normal population. To allow additional flexibility, we propose the use of a generalized additive mixed model to incorporate both a nonparametric

term for body weight and to allow for variability from experiment to experiment. Smoothing is incorporated through the use of penalized splines (Eilers and Marx, 1996). Such smoothers have a simple mixed model representation which allows the entire model to be couched within the generalized mixed model framework. In particular, computation and smoothing parameter selection can be accomplished using existing software for mixed models, specifically the `glmmix` macro in the SAS computing environment.

Recently there has been much research into semiparametric extensions of mixed models. In the Gaussian case, Wang and Taylor (1995) used cubic regression splines to model fixed effects, Anderson and Jones (1995) used smoothing splines to model the random effects and Shi *et al.* (1996) modelled both the random and the fixed effects by using fixed knot *B*-splines. Wang (1998a) applied smoothing splines in the mixed effects analysis-of-variance setting. Also, Wang (1998b) compared smoothing parameter selection techniques for smoothing splines for Gaussian data with correlated errors. Recent work by Lin and Zhang (1999) has focused on the generalized additive mixed model for smoothing splines. Extensions to the generalized case can be thought of as an extension of the generalized additive model developed by Hastie and Tibshirani (1987) to the mixed model framework. Our method has an advantage of conceptual and computational simplicity since it does not rely on Wiener or other sophisticated stochastic processes formulations as do smoothing splines. It is based on finite dimensional random vectors and ordinary mixed models which have simple computational properties and interpretations as described in Section 3. This simplicity also allows for an appealing extension by introducing random variability without fitting a different curve for each experiment.

Data from our phenolphthalein example are described in Section 2. In Section 3 we present the model and in Section 4 we analyse the phenolphthalein data incorporating the historical controls.

2. Description of the data

Recently, the National Cancer Institute of the USA nominated phenolphthalein for testing at the NTP. Although phenolphthalein has been in widespread use as a laxative for nearly a century, there were no animal testing data on the compound. Phenolphthalein had been approved for human use and was considered 'generally recognized as safe' by the Food and Drug Administration in the mid-1970s, on the basis of drug efficacy studies. Although ordinary exposure to this chemical is usually low in humans, laxatives are habit forming and are often abused in dieting attempts, so exposures can sometimes be quite large. In addition to its use as a laxative, phenolphthalein is commonly used as a laboratory reagent and an acid–base indicator, so the potential for exposure in laboratory workers is substantial as well. The NTP report concluded that there was 'clear evidence of carcinogenicity' in all sex–species combinations with the exception of the female rat. Some significant dose-related decreases in the incidence of tumours were also reported for these animals, something sometimes attributed to weight effects. See National Toxicology Program (1995) for more information regarding phenolphthalein.

The phenolphthalein study followed a typical design wherein animals are randomized at about 6 weeks of age to four groups including a control and are followed for approximately 2 years, at which time any surviving animals are sacrificed. There are typically 50 animals in each group. At the time of the death of the animals, a determination of the presence of tumours is made. Body weight measurements are taken throughout the study at 1–4-week intervals.

In the NTP study of phenolphthalein, $B6C3F_1$ mice and F344/N rats were randomized to four dose groups exposed through feed. Table 1 summarizes the results of the study for

mammary tumours. Although many tumour types were examined in this study, we focus on mammary tumours in female rats. These tumours were chosen because there appeared to be a dose-related *decrease*. Dosed animals were significantly lighter than controls in this experiment and Seilkop (1995), among others, found strong correlations between body weight and the incidence of mammary tumours. The NTP report for phenolphthalein suggested that the lower incidence of mammary gland tumours was related to body weight (National Toxicology Program, 1995) but gave no statistical evidence which demonstrated that this was so. In Section 4, we shall concentrate on the comparison of the 25000 parts per million (ppm) dose group with the controls.

The main goal of the NTP studies is to assess the effect of dose on the occurrence of tumours while controlling for confounders such as weight and time of death. However, the rarity of the tumour events makes this difficult. A possible remedy is to incorporate the large amount of historical data on the same measurements for control animals that the NTP has collected over time. The historical data that are used in this paper consist of 1042 observations from 21 feed experiments. This data structure is described by Fig. 1. Fig. 1(a) corresponds to the current experiment, which consists of several dose groups and measurements on weight and time. We shall call this *experiment 0*. Fig. 1(b) corresponds to the historical controls data.

Consider for the moment only the current experiment. The primary statistical method used by the NTP to detect dose-related trends is the logistic regression test (Dinse and Lagakos, 1983). The probability of tumour in this case is modelled as a linear function of time (and possibly the square of time if it improves the model fit) and dose as follows:

$$\text{logit}\{\Pr(y_j = 1)\} = \beta_0 + \beta_2 t_j + \beta_3 t_j^2 + \gamma d_j, \tag{1}$$

where $y_j = 1$ if animal j develops a tumour at some time during the experiment (and $y_j = 0$ otherwise), t_j is the time of animal j's death and d_j is the dose level to which animal j was randomly assigned. The test of H_0: $\gamma = 0$ then corresponds to the desired trend test for carcinogenicity.

A term for body weight could easily be added to model (1); however, there are several reasons why this would not be appropriate. Typically there are very few tumours of any particular type in a study. Thus, tumours may not be seen at the full range of body weights. Also, since the controls are often heavier than the experimental groups there is considerable confounding of body weight and dose level. We believe that incorporating the historical information will alleviate both of these pitfalls. Seilkop (1995) proposed a method to incorporate the historical information on body weight into the current experiment by using the historical database to establish the relationship between the probability of tumour and body weight, adjusting for age at death by a logistic regression. He then treated the parameters from this model as fixed and calculated expected probabilities for the treatment groups in the current experiment given the body weights and times of death of those animals

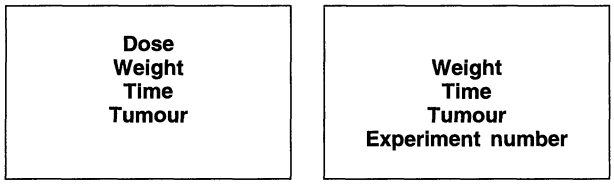

Fig. 1. Structure of the data: (a) current experiment (experiment 0); (b) historical controls data (dose = 0)

in the current experiment. Like Seilkop (1995), we restrict attention to the body weight measurements taken at 12 months post randomization. This time point is used since correlations between body weight and the occurrence of tumours have been determined for this time point (Haseman *et al.*, 1997; Seilkop, 1995) and it is unlikely that the animal will have developed a tumour by this stage in the experiment; thus the presence of a tumour is not inherent in the body weight measurement at 12 months. Any animals who die before 12 months, however, are excluded from the analysis.

Fig. 2 illustrates the distributions of the body weights at 12 months on study for the phenolphthalein experiment and the historical controls. It is evident that the animals in the control group in the phenolphthalein experiment are heavier than the exposed groups, since the distribution is shifted substantially to the left in the exposed animals. However, the distribution for the historical group roughly covers all groups for the phenolphthalein data. Having the additional information from the historical controls should alleviate the confounding issue.

Fig. 3 shows unadjusted nonparametric estimates of the probability of four tumour types based on a smooth of the body weights at 12 months in a logistic regression for the NTP set of data on controls from the feed experiments. These plots suggest that non-linear relationships exist and that semiparametric models for the incorporation of body weight data would be beneficial.

3. Model

We propose the use of a simple linear penalized regression spline to model the non-linear effects of body weight in this problem. We first describe the use of penalized splines for ordinary linear regression.

3.1. *Ordinary nonparametric regression*

Consider data (x_i, y_i), $i = 1, \ldots, n$, representing covariates and outcomes for n subjects. The ordinary nonparametric regression model can be written

$$y_i = f(x_i) + \epsilon_i, \tag{2}$$

where the ϵ_i are independent $N(0, \sigma^2)$ variables and $f(x)$ is an arbitrary smooth function. The spline approach involves defining $\kappa_1, \ldots, \kappa_K$ to be a set of distinct numbers within the range of the x_i and work with the set of *truncated functions* $(x - \kappa_k)_+$, where x_+ denotes the minimum of 0 and x. Their addition to the model allows the non-linear structure to be estimated. The spline model for f is

$$f(x_i) = \beta_0 + \beta_1 x_i + \sum_{k=1}^{K} b_k (x_i - \kappa_k)_+, \tag{3}$$

representing piecewise linear functions with knots at the κ_k. A large number of knots is included by default to capture fine structure in general scatterplots. A reasonable allocation rule is one knot for every four or five observations, up to a maximum of 40 knots. If this model is fitted using ordinary least squares then it overfits the data, rather than smoothing it. A remedy is to treat the $(x_i - \kappa_k)_+$ as *random effects*, resulting in the *mixed model*

$$f(x_i) = \beta_0 + \beta_1 x_i + \sum_{k=1}^{K} b_k (x_i - \kappa_k)_+, \qquad b_k \text{ independent and identically distributed } N(0, \sigma_b^2).$$

The ordinary least squares fit corresponds to $\sigma_b = \infty$, where the b_k are unrestricted. Taking

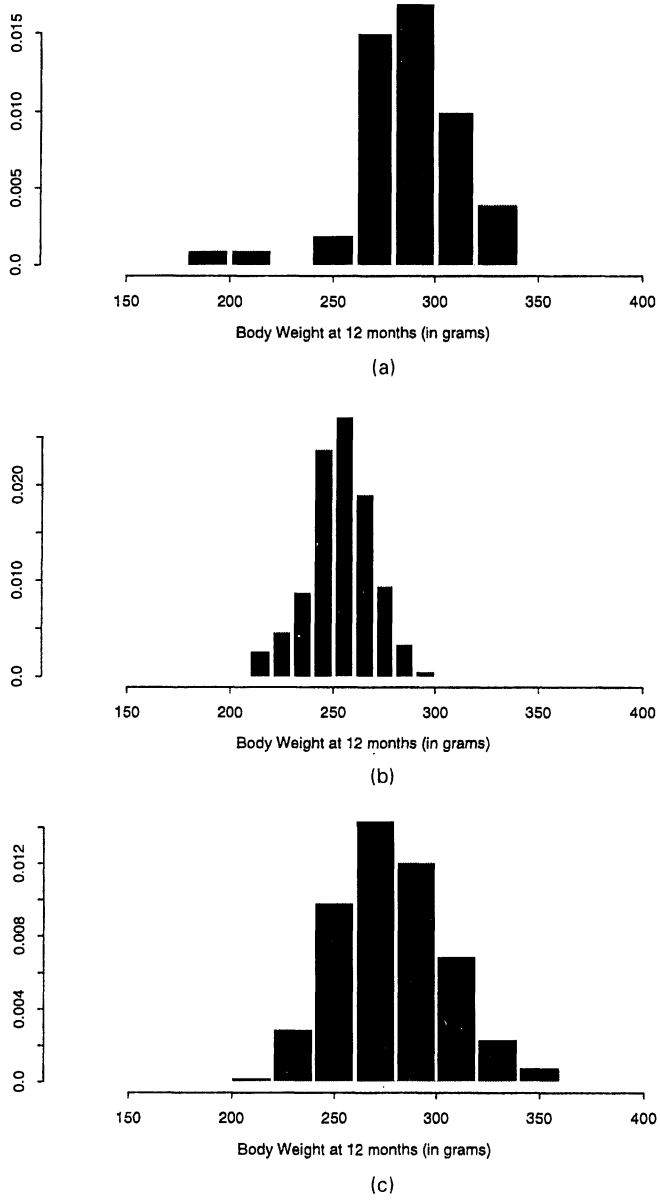

Fig. 2. Distribution of body weights at 12 months for female rats in (a) the control group (dose = 0; experiment 0; mean, 287.3 (25.58); minimum, 194.4; maximum, 330.0) and (b) the exposed animals (dose > 0; experiment 0; mean, 253.5 (15.58); minimum, 211.0; maximum, 295.6), and (c) the historical data feed experiments (mean, 278.1 (27.00); minimum, 195.9; maximum, 383.8)

σ_b to be finite leads to smaller estimates of the b_k and the effect of the $(x_i - \kappa_k)_+$ being diminished. Truncated lines are used here for their simplicity. Equivalent results can be achieved by working with alternative spline bases such as B-splines (e.g. Eilers and Marx (1996)). Higher degree splines may also be used. However, the motivating application did not benefit from this extension.

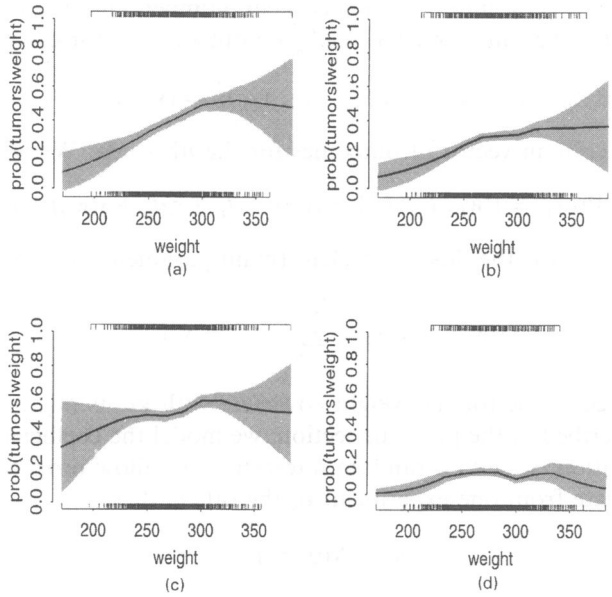

Fig. 3. Estimated probabilities of (a) mammary tumours, (b) leukaemia, (c) pituitary tumours and (d) thyroid tumours as a function of weight for a set of NTP historical controls: the shaded regions represent plus and minus twice the estimated (pointwise) standard errors; the bars at the top and bottom of the plots are the actual data points, where a 1 indicates the event (a tumour or leukaemia discovered at the time of death) and 0 indicates otherwise

We can rewrite model (2) as a simple normal mixed model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \epsilon, \tag{4}$$

where $\mathbf{y}$ is the vector of outcomes, $\mathbf{X}$ corresponds to the fixed effect covariate matrix and $\mathbf{Z}$ corresponds to the random-effect covariate matrix containing the truncated polynomials. For given σ_b and σ, the solution is well known as the best linear unbiased predictor estimators:

$$\hat{\mathbf{f}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\mathbf{b}}, \tag{5}$$

where

$$\hat{\boldsymbol{\beta}} = \{\mathbf{X}^{\mathrm{T}}(\sigma^2\mathbf{I} + \sigma_b^2\mathbf{Z}\mathbf{Z}^{\mathrm{T}})^{-1}\mathbf{X}\}^{-1}\mathbf{X}^{\mathrm{T}}(\sigma^2\mathbf{I} + \sigma_b^2\mathbf{Z}\mathbf{Z}^{\mathrm{T}})^{-1}\mathbf{y},$$
$$\hat{\mathbf{b}} = (\sigma^2\mathbf{I} + \sigma_b^2\mathbf{Z}\mathbf{Z}^{\mathrm{T}})^{-1}\mathbf{Z}^{\mathrm{T}}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

(Robinson, 1991).

As shown by Brumback *et al.* (1999), if we let $\mathbf{D}$ be the $(K+2) \times (K+2)$ matrix

$$\mathbf{D} = \begin{pmatrix} \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times K} \\ \mathbf{0}_{K\times 2} & \mathbf{I}_{K\times K} \end{pmatrix},$$

$\mathbf{C} = (\mathbf{X}, \mathbf{Z})$ and $\lambda = \sigma^2/\sigma_b^2$, then $\hat{\mathbf{f}}$ can be rewritten as $\mathbf{C}(\mathbf{C}^{\mathrm{T}}\mathbf{C} + \lambda\mathbf{D})^{-1}\mathbf{C}^{\mathrm{T}}\mathbf{y}$ which is equivalent to the penalized spline smoother of Eilers and Marx (1996).

3.2. Carcinogenicity data

Let us now consider the full data which include both the current experiment and the historical

controls. Letting $i = 0, \ldots, I$ index the experiment number and $j = 1, \ldots, n_i$ index the animals in experiment i, the full $N \times 1$ ($N = \Sigma_{i=0}^{I} n_i$) outcome vector can be written as

$$\mathbf{y} = (y_{01}, \ldots, y_{0n_1}, \ldots, y_{I1}, \ldots, y_{In_I})^{\mathrm{T}} = (\mathbf{y}_0^{\mathrm{T}}, \ldots, \mathbf{y}_I^{\mathrm{T}})^{\mathrm{T}},$$

where $\mathbf{y}_i$ is the $n_i \times 1$ column vector of outcomes for the ith study. We wish to model

$$\mathrm{logit}\{\Pr(y_{ij} = 1 | w_{ij}, t_{ij}, d_{ij})\} = \alpha_0 + \alpha_1 t_{ij} + \gamma d_{ij} + s(w_{ij}) + u_i,$$

where α_0 is an intercept, γ is the dose coefficient (main parameter of interest) and

$$s(w_{ij}) = \beta_1 w_{ij} + \sum_{k=1}^{K} b_k (w_{ij} - \kappa_k)_+$$

represents the penalized spline for the weight covariate with knots $(\kappa_1, \ldots, \kappa_K)$ indexed by $k = 1, \ldots, K$. As described in the previous section, we model the coefficients of the knots as $b_k \sim N(0, \sigma_b^2)$. In addition, we add a random intercept u_i to allow properly for background shifts in the tumour rate from one experiment to the other where

$$u_i \sim N(0, \sigma_u^2).$$

Let $\mathbf{t}$, $\mathbf{d}$ and $\mathbf{w}$ be the vectors of observed death times, dose and body weights respectively (e.g. $\mathbf{t} = (t_{01}, \ldots, t_{0n_0}, \ldots, t_{I1}, \ldots, t_{In_I})^{\mathrm{T}}$), and let $\mathbf{1}$ be the $N \times 1$ column vector of 1s. Then, let

$$\mathbf{X} = (\mathbf{1}, \mathbf{t}, \mathbf{d}, \mathbf{w})$$

be the $N \times 5$ matrix corresponding to the fixed effects and

$$\mathbf{Z} = (\mathbf{1}_0, \ldots, \mathbf{1}_I, (\mathbf{w} - \kappa_1)_+, \ldots, (\mathbf{w} - \kappa_k)_+, \ldots, (\mathbf{w} - \kappa_K)_+)$$

be the $N \times (K + I)$ matrix corresponding to the random effects where $\mathbf{1}_i$ indicates a vector of 1s and 0s with 1s only in the rows corresponding to experiment i. Then,

$$\mathrm{logit}\{\Pr(\mathbf{y} = 1 | \mathbf{w}, \mathbf{t}, \mathbf{d})\} = \mathbf{X} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \gamma \\ \beta_1 \end{pmatrix} + \mathbf{Z} \begin{pmatrix} u_0 \\ \vdots \\ u_I \\ b_1 \\ \vdots \\ b_K \end{pmatrix}. \tag{6}$$

Thus, as in the ordinary regression, estimates of the unknown parameters, including the smoothing parameter, can be obtained by solving the generalized mixed model equation (6). The solution can be obtained via the `glmmix` macro available for the SAS software system.

Having only a random intercept may be too restrictive. Although it is possible to estimate a unique spline for each of the experiments, this may lead to considerable overfitting. Since the tumour rates are not necessarily very high, we would like to draw more information for the relationship between weight and the incidence of tumours from consistencies across experiments. We propose that most of the variability from experiment to experiment can be accommodated by the first two terms of the polynomial, i.e. by the fixed effects — the intercept and slope. We had already considered the intercept to be random, so this extension requires that we additionally allow the slope to vary by experiment, i.e. we let

$$s(w_{ij}) = \beta_1 w_{ij} + \tilde{b}_{1i} w_{ij} + \sum_{k=1}^{K} b_k (w_{ij} - \kappa_k)_+,$$

where

$$\tilde{\mathbf{b}}_{1i} \sim N(0, \sigma^2_{\tilde{b}1})$$

and $\sigma^2_{\tilde{b}1}$ is the variance associated with the added random effect. We write this extended model as

$$\text{logit}\{\Pr(\mathbf{y} = 1|\mathbf{w}, \mathbf{t}, \mathbf{d})\} = \mathbf{X} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \gamma \\ \beta_1 \end{pmatrix} + \tilde{\mathbf{Z}} \begin{pmatrix} u_0 \\ \vdots \\ u_I \\ \tilde{b}_{10} \\ \vdots \\ \tilde{b}_{1I} \\ b_1 \\ \vdots \\ b_K \end{pmatrix},$$

where $\tilde{\mathbf{Z}}$ is now given by

$$\tilde{\mathbf{Z}} = (\mathbf{1}_0, \ldots, \mathbf{1}_I, \mathbf{1}_0 \odot \mathbf{w}, \ldots, \mathbf{1}_I \odot \mathbf{w}, (\mathbf{w} - \kappa_1)_+, \ldots, (\mathbf{w} - \kappa_K)_+),$$

and $\odot$ indicates elementwise multiplication, i.e. $\mathbf{1}_i \odot \mathbf{w}$ will be a vector containing the body weight measurements in the rows corresponding to the ith experiment and 0s elsewhere.

4. Example: phenolphthalein

Table 2 contains the results of fitting the logistic regression tumour model under four alternatives for comparing the 25000 ppm dose group with the control animals. The first is the original analysis in which no historical data are incorporated. The second includes a weight term in the logistic regression model for the current experiment only. The third incorporates the historical data but does not adjust for weight and the last is the full model

Table 2. Summary of the dose effect for various models of the phenolphthalein data and various smoothing parameters

	σ_b	$\hat{\gamma}$	*Standard error*	*p-value*
Current experiment only				
No weight term (original)		−0.0763	0.0210	<0.001
Including weight		−0.0424	0.0298	0.1556
Including historical data				
No weight (random intercept)		−0.0462	0.0156	0.003
Model proposed	0	−0.0240	0.0152	0.115
	10	−0.0235	0.0153	0.124
	100	−0.0226	0.0154	0.144
	1000	−0.0223	0.0153	0.145
	10000	−0.0226	0.0151	0.136
	100000	−0.0232	0.0152	0.128

incorporating the historical data as outlined in this paper and adjusting for weight via the spline for various choices of the smoothing parameter σ_b. Knots were placed at the deciles of the weight data.

Including a term for body weight using only the current experiment does change the significance of the relationship (i.e. the null hypothesis would no longer be rejected). However, the weight term is not significant, and it is highly likely that collinearity may be present. Although it is true that the current experiment has a control rate of mammary tumours that is higher than normal (relative to the historical controls), including the historical data without adjusting for body weight does little to change the p-value. Also note that the magnitude of the estimated coefficient for dose under the model proposed, $\hat{\gamma}$, is approximately a third of the estimate in the original analysis, and approximately a half of the estimate in either of the other two cases. The estimated dose effect remains virtually unchanged by the magnitude of the smoothing parameter, however. In addition, the incorporation of the historical control data reduces the standard error of the estimated coefficient for dose. Thus, the incorporation of the historical data appears to improve the precision of our estimate.

Fig. 4 contains a plot of the p-value for rejecting the null hypothesis of no dose effect *versus* a wide range of smoothing parameters. Bowman and Azzalini (1997), page 89, referred to such a plot as a *significance trace* and advocated its use on the grounds that it allows us to assess evidence over a wide range of smooth function fits and avoids the need to settle on a particular amount of smoothing. In this instance, we see that there is no strong evidence of a dose effect for all smooth functions of weight that can be generated from the penalized spline model.

After controlling for body weight, allowing for this possibility of non-linearity in its relationship with the incidence of tumours, we would no longer reject the null hypothesis, which had originally suggested a dose-related decrease in the incidence.

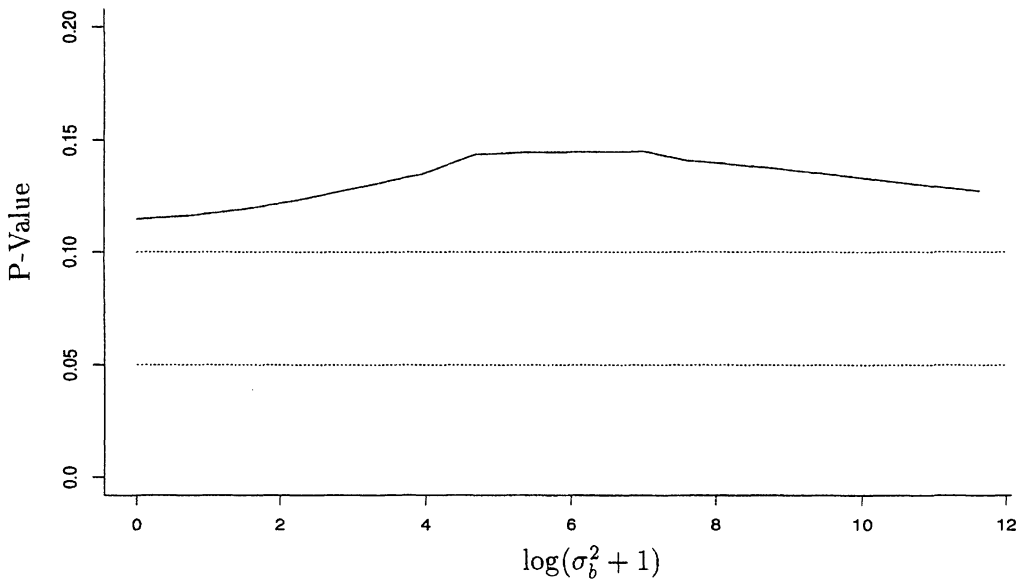

Fig. 4. p-value for rejecting the null hypothesis of no dose effect (H_0: $\gamma = 0$) as a function of the smoothing parameter

5. Discussion

We have presented a method for incorporating historical information into animal carcinogenicity studies. Existing software can be used to implement the procedure.

We focused on experiments where the chemical being tested is thought to be carcinogenic, although the method may be particularly useful for tests of anticarcinogens, since in this case a false decreasing trend is much more serious. Many cancer fighting drugs are initially tested in rodent experiments, so it is important that the issue of body weight is addressed there.

One must always use caution when using historical data. The current experiment is, of course, the most relevant and care should be taken to include relevant historical data (Haseman, 1994). We chose only to use those historical experiments where animals were exposed through feed since the phenolphthalein study was also a 'feed' study.

Additive models are continuing to gain widespread use in applied statistics. We believe that the method proposed here has the advantage of conceptual simplicity and computational savings over some of the smoothing spline methods that are currently available. It also allows for simple extensions such as that presented here allowing just the intercept and slope of the polynomial basis to vary by experiment.

Acknowledgement

This work was supported in part by grant 5 T32 ES07142 from the National Institutes of Health, USA.

References

Abelson, P. (1995) Flaws in risk assessment. *Science*, **270**, 215.

Anderson, S. J. and Jones, R. H. (1995) Smoothing splines for longitudinal data. *Statist. Med.*, **14**, 1235–1248.

Bailer, A. J. and Portier, C. J. (1988) Effects of treatment-induced mortality and tumor-induced mortality on tests for carcinogenicity in small samples. *Biometrics*, **44**, 417–431.

Bowman, A. and Azzalini, A. (1997) *Applied Smoothing Techniques for Data Analysis: the Kernel Approach with S-Plus Illustrations*. Oxford: Oxford University Press.

Brumback, B., Ruppert, D. and Wand, M. (1999) Comment on paper by Shively, Kohn and Wood. *J. Am. Statist. Ass.*, **94**, 794–797.

Dempster, A. P., Selwyn, M. R. and Weeks, B. J. (1983) Combining historical and randomized controls for assessing trends in proportions. *J. Am. Statist. Ass.*, **78**, 221–227.

Dinse, G. E. and Lagakos, S. W. (1983) Regression analysis of tumour prevalence data. *Appl. Statist.*, **32**, 236–248; correction, **33** (1984), 79–80.

Eilers, P. H. C. and Marx, B. D. (1996) Reply to comments on 'Flexible smoothing with B-splines and penalties'. *Statist. Sci.*, **11**, 115–121.

Gaylor, D. W. and Kodell, R. L. (1999) Dose-response trend tests for tumorigenesis, adjusted for body weight. *Toxicol. Sci.*, **49**, 318–323.

Hart, R. W., Keenan, K., Turturro, A., Abdo, K. M., Leakey, J. and Lyn-Cook, B. (1995) Symposium overview: caloric restriction and toxicity. *Fund. Appl. Toxicol.*, **25**, 184–195.

Hart, R. W. and Turturro, A. (1997) Dietary restrictions and cancer. *Environ. Hlth Perspect.*, **105**, suppl. 5, 989–992.

Haseman, J. K. (1994) Data analysis: statistical analysis and use of historical control data. *Regul. Toxicol. Pharmacol.*, **21**, 52–59.

Haseman, J. K., Bourbina, J. and Eustis, S. L. (1994) The effect of individual housing and other experimental design factors on tumor incidence in B6C3F1 mice. *Fund. Appl. Toxicol.*, **23**, 44–52.

Haseman, J. K., Young, E., Eustis, S. and Hailey, J. (1997) Body weight-tumor incidence correlations in long-term rodent carcinogenicity studies. *Toxicol. Pathol.*, **25**, 256–263.

Hastie, T. and Tibshirani, R. (1987) Non-parametric logistic and proportional odds regression. *Appl. Statist.*, **36**, 260–276.

Ibrahim, J. G. and Ryan, L. M. (1996) Use of historical controls in time-adjusted trend tests for carcinogenicity. *Biometrics*, **52**, 1478–1485.

Ibrahim, J. G., Ryan, L. M. and Chen, M.-H. (1998) Use of historical controls to adjust for covariates in trend tests for binary data. *J. Am. Statist. Ass.*, **93**, 1282–1293.

Kikuchi, Y. and Tanagawa, T. (1991) Incorporating historical controls using a random effects model with a normal prior. *Communs Statist. Theory Meth.*, **20**, 1273–1291.

Lin, X. and Zhang, D. (1999) Inference in generalized additive mixed models by using smoothing splines. *J. R. Statist. Soc.* B, **61**, 381–400.

National Toxicology Program (1995) Toxicology and carcinogenesis studies of phenolphthalein in F344/N rats and B6C3F_1 mice. *Technical Report NTP TR 465*. National Toxicology Program, Research Triangle Park.

Peto, R., Pike, M. C., Day, N. E., Gray, R. G., Lee, P. N., Parish, S., Peto, J., Richards, S. and Wahrendorf, J. (1980) Guidelines for simple, sensitive significance tests for carcinogenic effects in long-term animal experiments. In *IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans*, suppl. 2, pp. 311–346. Lyon: International Agency for Research on Cancer.

Robinson, G. K. (1991) That Blup is a good thing: the estimation of random effects. *Statist. Sci.*, **6**, 15–32.

Seilkop, S. K. (1995) The effect of body weight on tumor incidence and carcinogenicity testing in B6C3F1 mice and F344 rats. *Fund. Appl. Toxicol.*, **24**, 247–259.

Shi, M., Weiss, R. E. and Taylor, J. M. G. (1996) An analysis of paediatric CD4 counts for acquired immune deficiency syndrome using flexible random curves. *Appl. Statist.*, **45**, 151–163.

Tarone, R. E. (1982) The use of historical control information in testing for a trend in proportions. *Biometrics*, **38**, 215–220.

Wang, Y. (1998a) Mixed effects smoothing spline analysis of variance. *J. R. Statist. Soc.* B, **60**, 159–174.

———(1998b) Smoothing spline models with correlated random errors. *J. Am. Statist. Ass.*, **93**, 341–348.

Wang, Y. and Taylor, J. M. G. (1995) Inference for smooth curves in longitudinal data with application to an Aids clinical trial. *Statist. Med.*, **14**, 1205–1218.