STATISTICS IN MEDICINE
Statist. Med. 2007; **26**:3735–3752
Published online 30 May 2007 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.2848

Comparing smoothing techniques in Cox models for exposure–response relationships

Usha S. Govindarajulu[1,2,5,*,†], Donna Spiegelman[2], Sally W. Thurston[3], Bhaswati Ganguli[4] and Ellen A. Eisen[5]

[1]*Yale Center for Clinical Investigation, Yale School of Medicine, New Haven, CT 06510, U.S.A.*
[2]*Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, MA 02115, U.S.A.*
[3]*Department of Biostatistics, University of Rochester Medical Center, Rochester, NY 14642, U.S.A.*
[4]*Department of Statistics, University of Calcutta, Calcutta 700 019, India*
[5]*Department of Environmental Health, Harvard School of Public Health, Boston, MA 02115, U.S.A.*

SUMMARY

To allow for non-linear exposure–response relationships, we applied flexible non-parametric smoothing techniques to models of time to lung cancer mortality in two occupational cohorts with skewed exposure distributions. We focused on three different smoothing techniques in Cox models: penalized splines, restricted cubic splines, and fractional polynomials. We compared standard software implementations of these three methods based on their visual representation and criterion for model selection. We propose a measure of the difference between a pair of curves based on the area between them, standardized by the average of the areas under the pair of curves. To capture the variation in the difference over the range of exposure, the area between curves was also calculated at percentiles of exposure and expressed as a percentage of the total difference. The dose–response curves from the three methods were similar in both studies over the denser portion of the exposure range, with the difference between curves up to the 50th percentile less than 1 per cent of the total difference. A comparison of inverse variance weighted areas applied to the data set with a more skewed exposure distribution allowed us to estimate area differences with more precision by reducing the proportion attributed to the upper 1 per cent tail region. Overall, the penalized spline and the restricted cubic spline were closer to each other than either was to the fractional polynomial. Copyright © 2007 John Wiley & Sons, Ltd.

KEY WORDS: penalized spline; restricted cubic spline; fractional polynomial; bootstrapping; environmental epidemiology; dose–response; smoothing

*Correspondence to: Usha S. Govindarajulu, 2, Church St South, Yale Center for Clinical Investigation, Yale School of Medicine, New Haven, CT 06510.
†E-mail: usha.govindarajulu@yale.edu, usha@alum.bu.edu

Contract/grant sponsor: National Cancer Institute; contract/grant number: CA81345

Received 11 January 2006
Accepted 12 January 2007

INTRODUCTION

This paper compares three different methods for fitting non-linear smooth curves in occupational cohort data to model time to disease incidence (or mortality) as a function of exposure. Traditionally, parametric models are fit with either a continuous exposure or categorical exposure using cut-points to define the categories of exposure. The former approach assumes a linear relationship between exposure and the logarithm of relative risk, which may not be plausible. The latter requires the selection of cut-points to define exposure categories, which can affect the shape of the fitted dose–response curve and lead to a loss of power [1–4].

Smoothing is a more flexible approach for exposure–response modeling that avoids major assumptions of a particular dose–response relationship; it essentially estimates the trend of the response to be less variable and smooth [5]. Smoothing methods encompass a broad range of techniques such as kernel smoothing, polynomials and splines. Kernel smoothing uses a set of local weights which are defined by the kernel to produce a smoothed estimate at each target value; this can be mathematically tractable, and in certain applications, computationally intensive [5], but methods to improve efficiency have been developed [6, 7]. Polynomial regression models include terms raised to a power (i.e. $x, x^2, x^3, \ldots$) [5]. Polynomial regression is a global approach that may lack sensitivity to local variation in exposure–response. An alternative is a flexible class of functions called splines, which are used to fit non-linear relationships.

Splines derive their name from physical splines which are thin flexible strips of wood used in nautical engineering. We will discuss two types of the most common splines, regression and smoothing splines. The advantages of using regression and smoothing splines over alternative smoothing procedures such as polynomials, binning and running averages have been discussed by Hastie and Tibshirani [5]. Regression splines allow for a more local fit to the data than polynomial regression because they connect piecewise polynomials at knots. However, they have the disadvantage of being highly sensitive to number of knots and to their location [5]. Like regression splines, smoothing splines also use knots but the knots are located at each unique value of the continuous predictor variable. In addition, smoothing splines incorporate a penalty term for overfitting to which results may be sensitive [8].

Several different smoothing techniques have been applied in environmental and occupational epidemiology. For example, smoothing splines have been used in generalized additive models to quantify the relationship of silica exposure and lung cancer [9] and to model air pollution and mortality [10, 11]. Another common smoothing method, locally weighted regression smoother (LOESS), has also been used to model non-linear exposure–response relationships in generalized additive models relating air pollution and mortality [10, 12]. Penalized splines [13] have recently appeared in several studies of occupational hazards and related health effects [12–18]. Restricted cubic splines (RCS) [19, 20] have also been applied in Cox models in both nutritional [21–23] and cancer epidemiology [18, 24, 25]. Fractional polynomials, a variant on polynomial regression, [26], have been employed in dose–response models of alcohol and mortality [27], trend analysis of human immunodeficiency virus incidence [4], as well as in occupational health studies [28, 29].

In this paper, we compare three methods that have appeared most frequently in the environmental epidemiology literature: penalized splines, restricted cubic splines and fractional polynomials. All three methods are flexible for curve fitting, easily incorporated in Cox models, and available in standard statistical software. Penalized splines have some similarities with both regression splines and smoothing splines. They use many knots like regression splines, where the influence of the knots is downweighted by a penalty term. Penalized splines have much in common with smoothing

DOI: 10.1002/sim

splines but use significantly fewer knots [30]. The restricted cubic spline, also known as a natural cubic spline [5], is a cubic regression spline, whose tails are linearly constrained and contains a pre-specified number of knots located at quantiles of data [19]. Fractional polynomials are neither splines nor piecewise. They are polynomials, with powers estimated in a fixed lower range of fractional powers [26].

Steenland and Deddens [18] recently described both penalized splines and restricted cubic splines in a review of alternative modeling approaches in occupational epidemiology. However, not much is known about how these methods compare to each other or to fractional polynomial when applied to the same data. In this analysis we compare the standard software application of three alternative methods in two epidemiologic data sets and propose a method to measure and compare the difference between smoothed curves.

We first describe the two occupational health studies to which the methods were applied, and then describe the three methods we used for fitting exposure–outcome data within a Cox model. Following this, we propose a method for measuring the difference between curves and fit each model to the data. After presenting results of both applications (in more detail for the first data set than the second), we then draw conclusions regarding this work and future applications.

METHODS

Occupational cohort studies

Two retrospective cohort studies of lung cancer mortality were used to compare the behaviour of the three smoothing techniques. The Silica study, first reported by Checkoway *et al*. [31], was based on 2342 men who worked in the diatomaceous earth industry for at least 12 months, including at least 1 day between 1 January 1942 and 31 December 1987. There were 77 deaths from lung cancer in this cohort during the follow-up period. Cumulative exposure to silica was measured as mg/m^3-years. This study was influential in the recent classification by the International Agency for Research on Cancer (IARC) of occupational exposure to silica as a human carcinogen [32]. A non-linear relationship between cumulative silica exposure and lung cancer mortality has been reported [9, 11].

To broaden the scope of the comparisons, we also applied the three methods to a second data set. The Uranium study, initially reported by Samet *et al*. [33], included 2153 underground uranium miners exposed to radon (Rn progeny) in New Mexico, who worked anytime from 31 December 1976 until 31 December 1985. There were 66 deaths due to lung cancer in this cohort during the study period of 1976–1985. Exposure to radon was measured in Working Level Months (WLM) units, which is used to express a worker's exposure to radon daughters. Results are presented in less detail for this second application.

For both cohorts, Cox models included the same covariates as had been used in the previously published studies. Samet *et al*. [33] adjusted for age, ethnicity (white *versus* Hispanic), calendar year at risk and smoking status in his exposure–response models for lung cancer and cumulative radon exposure. Checkoway *et al*. [31] adjusted for age, ethnicity (white *versus* Hispanic), calendar year and duration of follow-up. Duration of follow-up was included to address bias due to healthy worker survivor effect [34]. Smoking status was only available for 50 per cent of the cohort in the Silica study, and was reported not to be acting as a confounder in the original analysis. The Silica cohort was dynamic, with subjects entering the study over a range of calendar years, whereas

DOI: 10.1002/sim

the Uranium cohort was closed, with all subjects entering the study at a fixed point in time. All subjects at risk to have become a case at the age of the case's lung cancer death were included in the risk sets for both studies.

Methods for fitting splines

We modeled non-linear exposure–response relationships using restricted cubic splines, penalized splines and fractional polynomials. We used the Cox proportional hazards regression model [13], where the model for the lung cancer mortality rate for the ith subject at time t is

$$\lambda(t|Z_i, X_i) = \lambda_0(t)\exp(\beta_z' Z_i(t) + s(X_i(t))) \tag{1}$$

where β_z' is a vector of coefficients for the covariates, Z_i at time t, and s is a smooth function of the cumulative exposure X_i at time t defined by the particular smoothing method. Each smoothing method is described below in more detail.

1. Restricted cubic spline (RCS)

The RCS is a cubic regression spline constrained to have continuous first and second derivatives at the knots [5] for visual smoothness [19]. There is a restriction in RCS that for any value of the independent variable, x, below the first knot or above the last knot, the spline function, $s(X_i(t))$, is linear [19]. The linearity in the tails allows for an even more parsimonious model.

To model the dose–response relationship using a RCS transformation, we first select H values, say $(\kappa_1 < \kappa_2 < \cdots < \kappa_H)$, within the observed range of the exposure. In the standard software implementation, these values, or knots, correspond to a pre-specified number of evenly spaced quantiles of the exposure distribution. We then assume the model in equation (1), where

$$s(X_i(t)) = \beta_0 X_i(t) + \sum_{h=1}^{H-2} \beta_h \cdot X_{ih}(t) \tag{2}$$

and the X_{ih}'s are non-linear functions where the value of $X_{ih}(t)$ is determined by the position of the knots as

$$\begin{aligned} &\text{if } X_i(t) < \kappa_h \text{ or if } X_i(t) > \kappa_h \text{ then } X_{ih}(t) = 0 \\ &\text{if } \kappa_h < X_i(t) < \kappa_{H-1} \text{ then } X_{ih}(t) = (X_i(t) - \kappa_h)^3 \\ &\text{if } \kappa_{H-1} < X_i(t) < \kappa_H \text{ then } X_{ih}(t) \text{ is as is in equation (4)} \end{aligned} \tag{3}$$

and the third possible value of X_{ih} in equation (3) is

$$\begin{aligned} X_{ih}(t) = {} & (X_i(t) - \kappa_h)_+^3 - (X_i(t) - \kappa_{H-1})_+^3 (\kappa_H - \kappa_h)/(\kappa_H - \kappa_{H-1}) \\ & + (X_i(t) - \kappa_H)_+^3 (\kappa_{H-1} - \kappa_h)/(\kappa_H - \kappa_{H-1}) \end{aligned} \tag{4}$$

where $h = 1, \ldots, H-2$ and u_+ is the 'plus' function defined as $u_+ = u$ if $u > 0$ and $u_+ = 0$ if $u \leqslant 0$.

The $H-2$ spline variables created by the RCS function are included in the Cox proportional hazard regression model and standard modeling techniques can then be applied. We implemented RCS within a SAS macro written by one of us (D. Spiegelman and colleagues)

DOI: 10.1002/sim

(`http://www.hsph.harvard.edu/faculty/spiegelman/lgtphcurv8.html`). In the default implementation of this procedure, the macro begins with 25 knots located at the evenly spaced quantiles of the exposure distribution, i.e. at the 3.5th percentile, the 12.1th percentile, etc. A stepwise selection procedure then selects those spline variables that satisfy the default entry and exit significance levels of 0.05. The final model includes the exposure variable, other covariates, and whatever spline variables were retained using the stepwise selection procedure. By starting with a large number of knots, any important non-linearity in relative risk over the range of the exposure distribution should be detected. The macro allows for the significance level of the exit and entry criteria to be changed by the user, and for the default value of $H-2=25$ to change.

2. Penalized spline (P-spline)

Penalized splines were fit using the standard software implementation in Splus [35]. For a data set which contains a sudden feature such as a changepoint, the choice of a particular knot location may either enhance or mask the feature [36]. While too many knots can lead to overfitting, too few can lead to underfitting. Penalized splines offer an approach to selecting knots that is relatively robust to the location and number by choosing a relatively large number of knots and modelling the smooth function, s, as defined in equation (1) as

$$s(X_i(t)) = \beta_0 X_i(t) + \sum_{h=1}^{H-2} \beta_h \cdot X_{ih}(t) \tag{5}$$

The $X_{ih}(t)$, where $h=1,\ldots,H-2$, are non-linear basis functions corresponding to a large number of knots, H. They may be the piecewise cubic polynomials used in the preceding section or other types of piecewise polynomials such as B-splines which we shall discuss below. In contrast to RCS, where the knot coefficients, $\beta_h, h=1,\ldots,H-2$, are estimated by maximizing the partial likelihood function, in penalized splines, the knot coefficients are estimated by adding a penalty, λ, on the second derivative of s where the objective function is:

$$l_{\mathrm{p}} - \lambda \int_0^{\infty} \{s''(x(t))\}^2 \, \mathrm{d}x(t) \tag{6}$$

and l_{p} denotes the log partial likelihood.

We shall describe a variant of the above procedure called P-splines due to Eilers and Marx [13], which is implemented by the `pspline` function in Splus [35]. To model the dose–response relationship using P-splines, we first need to select a large number, say n of equi-spaced knots. We then compute the cubic B-spline basis functions corresponding to these knots and represent s as

$$s(X_i(t)) = \sum_{h=1}^{n} \beta_h \cdot B_h(t) \tag{7}$$

where the $B_h(t)$'s are the B-spline basis functions. These basis functions are typically piecewise cubic polynomials and are preferred to truncated polynomial bases because they lead to more stable numerical fits. The variant used by Eilers and Marx [13] then computes the coefficients, $\beta_h, h=1,\ldots,n$ by approximating the integral of the squared second derivative

$$\int_0^{\infty} \{s''(x(t))\}^2 \, \mathrm{d}x(t) = \int_{X_{\min}}^{X_{\max}} \left\{ \sum_{h=1}^{n} \beta_h \cdot B_h''(x) \right\}^2 \mathrm{d}x \tag{8}$$

DOI: 10.1002/sim

where $X_{\min}$ and $X_{\max}$ denote the minimum and maximum exposures for the observed data by the simpler sum of squared second differences of adjacent B-spline coefficients,

$$\sum_{h=3}^{n} (\Delta^2 \beta_h)^2 \tag{9}$$

where $\Delta\beta_h = \beta_h - \beta_{h-1}$ and $\Delta^2\beta_h = (\beta_h - \beta_{h-1}) - (\beta_{h-1} - \beta_{h-2})$. The number and location of knots have little influence on the shape of the penalized spline curve, as long as the knots are adequately spaced and the number of knots is sufficiently large [30]. Alternative definitions for the effective number of degrees of freedom (df) of a model and the standard errors of the estimated coefficients have been worked out by Verweij and van Houwelingen [37] and Gray [38]. We use the former set of definitions as these are the Splus defaults.

The standard implementation of the `pspline` function within a Cox model in Splus uses df = 4 as the criterion for selecting the optimal amount smoothing. In addition, we have considered an alternative model selection criterion based on minimizing Akaike's information criteria (AIC) [36, 39]. The AIC criterion, as implemented in the `pspline` function in Splus begins with a default of 15 spline terms in the B-spline basis expansion [40]. AIC then selects the optimal smoothing parameter that is used in the penalized partial likelihood fit, which is equivalent to selecting the optimal df [36]. The chosen knots are then evenly spaced across the range of X, i.e the exposure variable.

3. Fractional polynomial (*Fracpoly*)

Like P-splines and RCS, fractional polynomials may be used with any generalized linear model or Cox model for survival data [41]. Although a global approach, an advantage of the fractional polynomial model is that it has a simpler form than the other two options. By incorporating a wider range of functional forms than that permitted by the standard polynomial family, a greater range of possible dose–response relationships can potentially be accommodated. A fractional polynomial of degree m is defined as follows [26]:

$$s_m(X_i(t); \beta, \mathbf{p}) = \sum_{j=0}^{m} \beta_j V_j(X_i(t)) \tag{10}$$

where m is generally taken to be either 1 or 2, $\mathbf{p} = \{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$ is a set of powers with $p_1 < \cdots < p_m$, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)$. An $m = 1$ model would use a single value from $\mathbf{p}$, call it p_j so in equation (10), $V_j(X_i(t))$ would be a single term, $X_i(t)^{p_j}$, except when $p_j = 0$, then it would be $\ln(X_i(t))$ [26, 41].

For an $m = 2$ model, two values are selected from $\mathbf{p}$, p_j and p_k. For example, if $p_j = -2$ then p_k can take on any of the values, $\{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$, to create all possible pairs with p_j or if $p_j = 1$ then p_k can take any of the values, $\{1, 2, 3\}$ to create the pairs, $\{(1, 1), (1, 2)$ or $(1, 3)\}$ with p_j. In equation (10), $V_j(X_i(t))$ would now contain two terms. If $p_j \neq p_k$, the two terms in $V_j(X_i(t))$ are $X_i(t)^{p_j}$ and $X_i(t)^{p_k}$. If, $p_j = p_k \neq 0$, then the two terms in $V_j(X_i(t))$ are $X_i(t)^{p_j}$ and $X_i(t)^{p_k} \ln(X_i(t))$. Finally, if $p_j = p_k = 0$, then the two terms in $V_j(X_i(t))$ are then $\ln(X_i(t))$ and $\ln(X_i(t))^2$ [26, 41].

There are 44 possible combinations of $m = 1$ and 2 models from which the STATA function [42], `fracpoly`, can select in the default invocation, where the highest degree fractional polynomial considered is $m = 2$. For a given $m = 1$ or 2 model, the best model is chosen to be the one with

DOI: 10.1002/sim

the lowest deviance. Since we are fitting Cox models we cannot simply define the deviance as $-2 *$ (log-likelihood) [43]. Therefore, in this context, each Cox model is first ranked by the value of its partial log-likelihood, which we will denote as $D(m, \mathbf{p})$. The value of $D(m, \mathbf{p})$ associated with a straight-line model (i.e. $m = 1$ and $p = 1$) is $D(1, 1)$. Therefore, the gain, G, for a given model is obtained as $G = D(1, 1) - D(m, \mathbf{p})$, where a larger gain denotes a better model fit.

In the standard implementation in STATA [42], the best $m = 1$ model and best $m = 2$ model are chosen (each with largest G) and compared. The final model is the one with the largest gain. In our analyses, we transformed the exposure variable to a fractional polynomial, $s_m(X_i(t); \beta, \mathbf{p})$, which is then used as a predictor within the Cox model, adjusting for the other covariates. Continuous covariates are mean-centered while dichotomous covariates are not.

Calculating area between smoothed curves

The area under a smoothed curve was obtained by calculating the areas in 500 successive non-overlapping rectangles of the same width, and then summing each of these areas to obtain the total area. As depicted in the hypothetical figure (Figure 1), each rectangle was constructed *beneath* the curve. Our calculation of the area *between* two curves fit by different smoothing methods extends this approach, by subtracting the area of one of the two overlapping rectangles of the same width from the other in each successive interval along the x-axis. The area differences (shown as hatched areas in Figure 1) were then summed to obtain a final area difference.

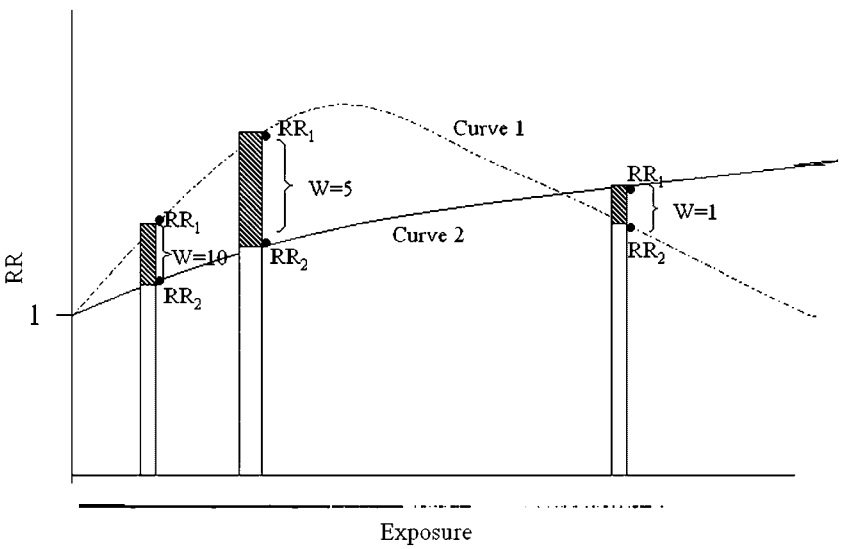

Figure 1. Hypothetical rectangles used to calculate area between curves with weights. Calculation of area between two curves based on summing the differences between 500 pairs of overlapping rectangles. The exposure rug is presented below the x-axis. Three rectangle pairs are presented; one in the highest density region of exposure, one in the mid range and one in the sparse tail region. For each pair of rectangles, RR_1 and RR_2, were estimated from curves 1 and 2, respectively, at the right endpoint of the rectangle. Each difference, i.e. hatched rectangle, was weighted by w (equation (11)). Hypothetical values for w, 10, 5 and 1, are presented to demonstrate that the rectangles in the denser exposure region had higher weights than rectangles in the tail.

DOI: 10.1002/sim

In addition, a weighted area difference was calculated to account for the varying precision in the estimates of the relative risk across the range of exposure. The inverse variance at the right endpoint of the height of each rectangle was used in the weightings. Each weight was computed as

$$w_s = \frac{\dfrac{1}{\mathrm{Var}(\hat{\mathrm{RR}}_{1s} - \hat{\mathrm{RR}}_{2s})}}{\sum_{s=1}^{S} \dfrac{1}{\mathrm{Var}(\hat{\mathrm{RR}}_{1s} - \hat{\mathrm{RR}}_{2s})}} \tag{11}$$

where $\mathrm{Var}(\hat{\mathrm{RR}}_{1s} - \hat{\mathrm{RR}}_{2s}) = \mathrm{Var}(\hat{\mathrm{RR}}_{1s}) + \mathrm{Var}(\hat{\mathrm{RR}}_{2s}) - 2\,\mathrm{Cov}(\hat{\mathrm{RR}}_{1s}, \hat{\mathrm{RR}}_{2s})$, where s indexes rectangles ($s = 1$ to 501), and $\hat{\mathrm{RR}}_{1s}$ and $\hat{\mathrm{RR}}_{2s}$ are the estimated relative risk values of the curve at the right-hand side of the sth rectangle on curve 1 and curve 2, respectively. In order to obtain the $\mathrm{Var}^{-1}(\hat{\mathrm{RR}}_{1s} - \hat{\mathrm{RR}}_{2s})$ term for each w_s, we used bootstrapping to generate a number of independent bootstrap samples each of the same size. Efron and Tibshirani [44] note that the number of bootstrap samples needed to obtain a reasonable estimate for the standard error or variance will usually be between 25 and 200 samples. We created bootstrap samples by sampling with replacement from the original data set 50 times. For each sample, we recalculated risksets and then fit each of the three Cox models (with P-spline-AIC, Fracpoly and RCS) separately. For each sample and each model, we calculated the predicted relative risks at evenly spaced x values, and took the difference between the pairs of relative risks. Finally, from the sample of 50 bootstrapped differences at each x value, we calculated the empirical variance of each difference, using the `var` function in Splus [45]. By using the bootstrap for this estimation, we avoided the need to derive the asymptotic covariances of the relative risk differences. We then used these empirical variance estimates in the weighting.

For unweighted and weighted calculations between each pair of curves, we estimated the area differences, $\hat{D}$, as

$$\hat{D} = \sum_{s=1}^{S} w_s D_s \Big/ \sum_{s=1}^{S} w_s \tag{12}$$

where S is the total number of rectangles, D_s is the area difference between curves in the sth rectangle, and for the unweighted method, $w_s = 1$ for all s and for the weighted method, the w_s's are given in equation (11). With the w_s's from equation (11) and when multiplied by S, equation (12) gives the weighted total area difference. These equations were used to compare area differences between the methods up to particular quantiles of exposure, to assess how much of any difference observed was concentrated in particular regions of the exposure distribution. The calculation of the weighted area difference is illustrated in Figure 1.

RESULTS

Descriptive characteristics of the occupational cohort studies

Descriptions of the two data sets are presented in Table I. They had approximately the same sample size, although the Silica study included a longer follow-up period and more lung cancer deaths. In both studies, most of the exposures were low although some were very high, as is typical of

DOI: 10.1002/sim

Table I. Descriptive characteristics of occupational cohort studies used to illustrate smoothing techniques applied to dose–response models [31].

Characteristics	Silica	Uranium
Outcome	Lung cancer mortality	Lung cancer mortality
Cumulative exposure	Silica	Radon
Number of subjects	2342	2153
Number of cases	77	66
Number of person-years	66 060	57 685
Follow-up period	1942–1994	1958–1985
Length of follow-up (years)		
Median (IQR)	29.9 (22.3)	26.0 (11.0)
Age at start of follow-up (years)		
Median (IQR)	27 (10.7)	27 (11.0)
Calendar year at start of follow-up		
Median (IQR)	1954 (18)	1960 (0)
Smokers (per cent current)	37 per cent*	75 per cent
Race		
Caucasian	77 per cent	53 per cent
Hispanic	23 per cent	40 per cent
Native American	—	7 per cent
Cumulative exposure:		
Mean (units)	2.16 (mg/m^3-years)	78.72 (WLM)
Standard deviation	3.51	111.22
Coefficient of variation	1.62	1.41
Skewness	5.62	2.27
Percentiles:		
25 per cent 50 per cent 75 per cent	0.03 1.06 2.48	7.30 26.80 104.30
90 per cent 95 per cent 99 per cent	5.14 7.75 16.80	230.50 321.60 492.70
min–max	0–62.56	0–1080.80

*Reported for only half of cohort with known smoking status.

environmental studies. The mean and median exposures were 2.16 and 1.06 mg/m^3-years for silica and 78.72 and 26.8 WLM for radon, respectively. Both exposure distributions were skewed, and silica more so than radon, with skewness parameters of 5.62 and 2.27, respectively.

Evaluating potential confounding by smoking in Silica cohort

To examine evidence for confounding by smoking, exposure–response models were fit among the subset of the Silica cohort with complete smoking information. The relative risk for smoking (ever *versus* never smoking) was 3.3 in a log-linear model adjusting for other covariates. However, the coefficient for exposure did not change with the removal of the smoking term, indicating the absence of confounding (for both analyses, RR = 1.2, 95 per cent CI:(1.1,1.3)). This suggests that if we could adjust for smoking in the entire Silica cohort, the estimated silica dose–response curve would be similar to the unadjusted estimate. Since smoking information was available for only half the Silica cohort and did not act as a confounder, we did not include it in the models, consistent with the previous analyses of this cohort study [31]. Smoking was, however, included in the original analysis of the Uranium study as well as this reanalysis.

DOI: 10.1002/sim

Table II. Features of selected smoothed functions of exposure in Cox models for lung cancer in two occupational cohorts.

Type of smooth	Model selection criterion*	Knot location	AIC¶		Degrees of freedom (df)	
			Silica	Uranium	Silica	Uranium
P-spline	AIC	Evenly spaced across range	115.8	47.0	2.6*	1.9*
	df = 4		115.0	44.4	4	4
RCS	LRT†	Selected from knots located at quantiles	118.2	47.6	2‡	1‡
Fractional polynomial	LRT†	NA	116.6	48.1	2§	2§

*Selected by Akaike's information criteria (AIC).
†LRT: likelihood ratio test.
‡Selected *via* stepwise selection (sle = sls = 0.05) where sle is the significance level for entry of a variable into the model and sls is the significance level for a variable to stay in the model.
§See Methods section.
¶AIC = $-2 \times$ (log-partial likelihood − df).

Features of fitted smoothed curves

For each of the alternative smoothing techniques, the specific criterion for model fit in the standard software is presented in Table II, along with knot location and df. For P-splines, two model selection criteria are used namely, AIC (P-spline-AIC) and default 4 df (P-spline-df = 4). AIC was used as a goodness-of-fit measure for comparing models fit with different smoothing techniques to each other. AIC has been re-defined for penalized splines [5, 36] with df substituted for number of parameters. Based on this and refined for comparing Cox models, we defined AIC = $-2 \times$ (log partial likelihood − df). This is different from the AIC method of selecting knots for the P-spline described in the Methods section. The AIC scores for both the Silica and Uranium studies show that the P-spline-default models have the best fit. However, AIC scores are not dramatically different and it is not apparent that any one method is superior to the others.

Each method used approximately 2 df to smooth RR as a function of exposure, except for the P-spline-default criterion which is defined as df = 4. As mentioned before, using the AIC to select optimal df did not change the number of terms or knots with which AIC began, but rather affected the constraint on the slopes of the terms, which is λ in equation (6). Even with roughly the same df, the P-spline will have more knots than the RCS. P-splines typically have up to 40 knots [30], and their influence is downweighted by the penalty function. The RCS was linear on the log scale for the Uranium cohort.

Cox model results

The three different smoothing techniques were fit within Cox models for each study. For the Cox models in both studies, time to event was defined by age and censoring was defined by case status. Regression parameter estimates and p-values are presented only for the Silica study (Table III).

DOI: 10.1002/sim

Table III. Results for Cox models fit with each smoothing method in a study of lung cancer and silica exposure in diatomaceous earth workers.

Variable	Mortality rate ratio	Coefficient	Standard error	p-value
P-spline-AIC				
P-spline(cumexp) linear*,†	1.07	0.056	0.021	0.0080
P-spline(cumexp) non*,†,‡				0.0048
Hispanic	0.27	−1.308	0.518	0.0120
Calendar year	1.04	0.041	0.011	0.0021
Duration of follow-up	0.88	−0.125	0.013	0.0000
P-spline-df = 4				
P-spline(cumexp) linear*,†	1.05	0.047	0.024	0.0490
P-spline(cumexp) non*,†,‡				0.0016
Hispanic	0.27	−1.321	0.518	0.0110
Calendar year	1.04	0.043	0.011	0.0001
Duration of follow-up	0.88	−0.127	0.013	0.0000
RCS				
Cumulative exposure‡	1.24	0.218	0.050	<0.0001
P-spline(cumexp)*,‡	0.79	−0.239	0.067	0.0004
Hispanic	0.27	−1.318	0.518	0.0109
Calendar year	1.04	0.043	0.011	0.0001
Duration of follow-up	0.88	−0.128	0.013	<0.0001
Fractional polynomial				
cumexp1‡,§	7.43	2.006	0.512	<0.0001
cumexp2‡,§	0.80	−0.217	0.112	0.0532
Hispanic	0.26	−1.343	0.518	0.0094
Calendar year§	1.04	0.043	0.011	0.0001
Duration of follow-up§	0.88	−0.128	0.013	<0.0001

*P-spline(cumexp) refers to the particular spline as a function of the cumulative exposure variable.

†For the P-spline, the first term tests if the linear spline function is significant and the second term tests whether the non-linear component of the spline function is significant.

‡These estimates are uninterpretable.

§These variables have been transformed as follows: $x = (\text{cumulative exposure} + 2.38 \times 10^{-7})/10$, $\text{cumexp1} = x^{0.5} - 0.5227$, $\text{cumexp2} = x^2 - 0.0746$, calendar year—1978, duration of follow-up—38.92.

The regression coefficients and standard errors were approximately the same for the three potential confounders (Hispanic ethnicity, calendar year and duration of follow-up) in all three models.

Splus provides separate results for linear and non-linear terms for the P-spline function in the Cox model. Due to the symmetry of the P-spline basis functions, the chi-square test for linearity is a test for zero slope of the regression of the spline coefficients on the centeres of the basis functions [36]. The slope was significantly different from zero for both P-splines, though slightly more so for P-spline-AIC. For the non-linear term, an approximate p-value is presented from the chi-square test for non-linearity, which tests the significance of the deviation from linearity. There was a significant non-linear effect for silica exposure on the RR for lung cancer in both P-spline models.

DOI: 10.1002/sim

The standard software for RCS also includes a test for both non-linear and the linear relationships between exposure and outcome. As with the P-spline, the test for non-linearity was significant (p-value<0.001). The final RCS model for the Silica study included the exposure variable plus one spline variable created by the 25 knot RCS transformation and selected to remain in the model through stepwise selection. While the coefficients of the exposure and spline variables are not directly interpretable, we can obtain the predicted hazard rate ratio at particular values of exposure.

Finally, the Fracpoly model selected the powers of exposure from the best $m=2$ model with powers of $(0.5, 2)$. These transformations were used in the Cox model, adjusting for the other covariates, which are mean-centered. Since exposure has been transformed, the coefficients are, as with RCS, not easily interpretable.

Dose–response plots

Figures 2 and 3 show the predicted RR plotted against the exposure for each smoothing method, in the Silica and Uranium studies, respectively. In addition, we present a step function based on a model with a categorized exposure variable to illustrate the standard epidemiologic approach. For the Silica study, we used the same exposure cut-points to define the categories as Checkoway *et al*. [31] did for their Poisson model. For the Uranium study, the categories were defined by quintiles of the exposure of the cases. A rugplot, or histogram, is displayed along the x-axis to illustrate the exposure distribution.

In each figure, the predicted RRs are fairly similar across models over most of the exposure range, but it is difficult to discern just how discrepant the curves are without a quantitative summary of the difference, accounting for precision of estimated RR's. In the Silica study, the RR is higher at lower exposure values for the Fracpoly than for the other two smoothing methods, while in the Uranium study, the three smoothing methods are fairly similar until the 99th percentile of exposure, when the Fracpoly and P-spline ($\mathrm{df}=4$) both drop while the others continue to increase across range of exposure. In both studies, the estimated RR is lower for the categorical approach than for the smoothed functions, in most of the exposure intervals. Moreover, the traditional categorical approach obscures both the peak in RR and the subsequent decline in the highest regions of exposure.

Bootstrapping results

To calculate weighted area differences, we needed to bootstrap to obtain the inverse variance weights for the calculations. We demonstrate these calculations for the Silica study only, since it had a more skewed exposure distribution. Overall, the estimated curves for each smoothing method in the bootstrap samples were similar to each other, with some degree of variation due to resampling from the original data. The bootstrap resamples had many tied exposure values because there were many unexposed subjects in the original study. However, this does not have much impact on the fit or knot selection for any of the smoothing techniques we considered because of the way in which knots are defined for each method. The P-spline uses a default number of knots which are evenly spaced across the range of exposure. The RCS uses knots located at quantiles of exposure, and the fractional polynomial does not use knots at all.

For all samples, we estimated the RCS and the Fracpoly using the default software settings (Table II). For the P-spline, however, AIC in the standard implementation led to a very wiggly curve in 21 of the 50 bootstrap samples, consistent with Therneau and Grambsch's [36] observation that the AIC under-penalizes, i.e. chooses fits that are less smooth (more wiggly). We did not bootstrap

DOI: 10.1002/sim

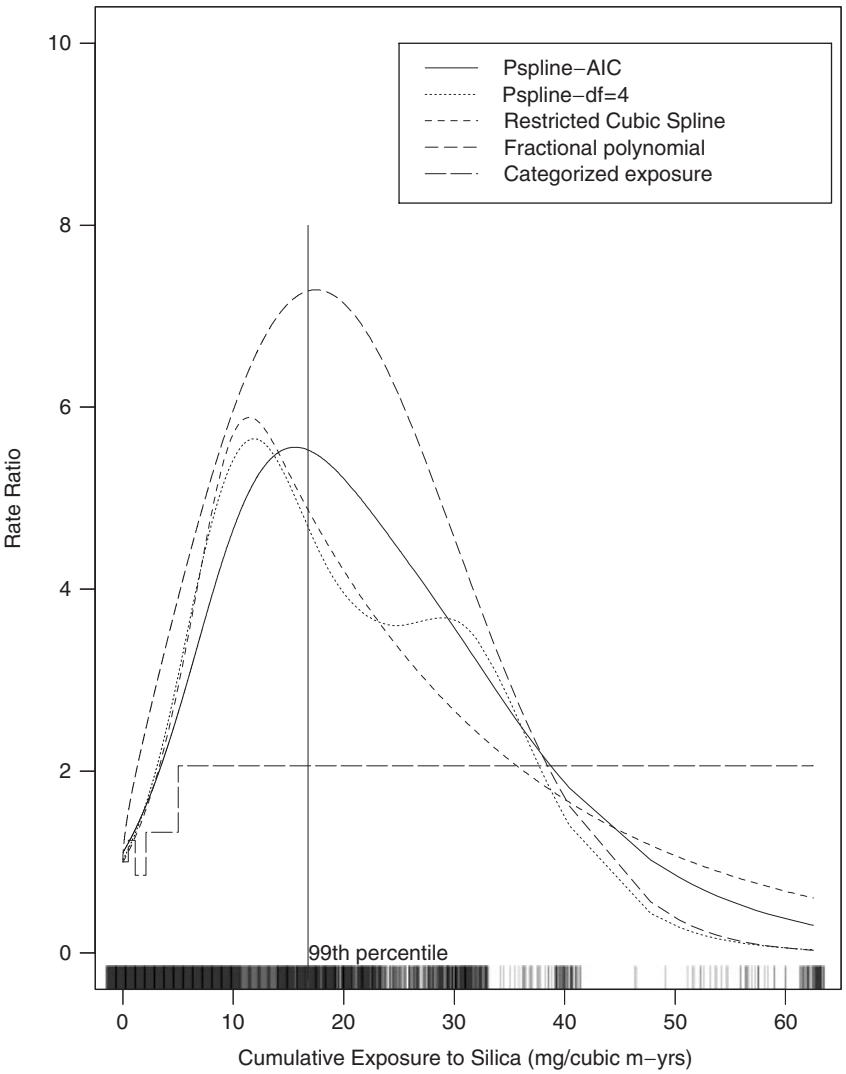

Figure 2. Different smoothing methods fit to lung cancer and silica exposure in cohort of diatomaceous earth workers.

the P-spline-df = 4 since they would have been at least as wiggly as the P-spline-AIC selected curves. These results suggest that neither AIC nor df = 4 is a reliable criterion for selecting the df.

This raises an issue for Splus software design which currently implements 4 df as the default for model selection in the `pspline` function. There is an AIC option as well as a corrected AIC (cAIC) option available for the `pspline` function in Splus. cAIC uses $n(\mathrm{df}+1)/(n-(\mathrm{df}+2))$ in place of df in the AIC equation [36]. When we used cAIC for samples with the most wiggly curves, the P-splines were no smoother and had only a minor decrease in df. Thus, we hand picked df based on a compromise between AIC and *a priori* ideas of biologic plausibility, which would support a more monotonic curve. Previously, Eisen *et al*. [15] determined that P-splines with df = 3

DOI: 10.1002/sim

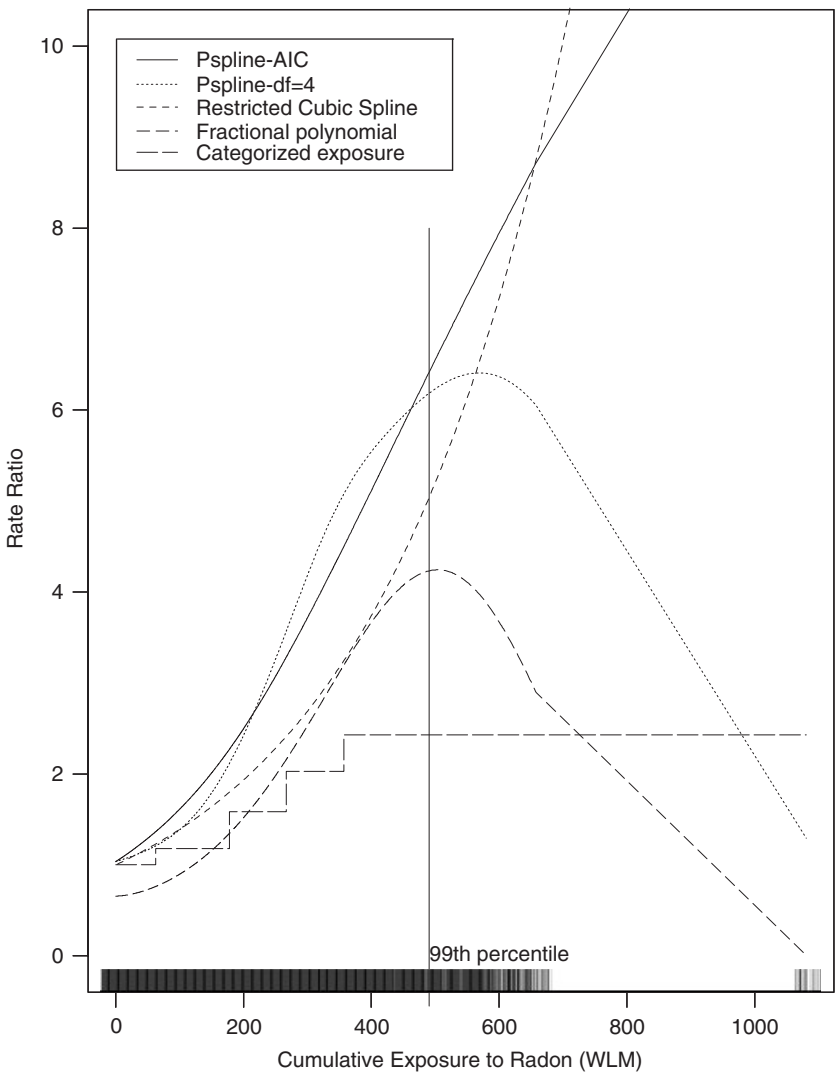

Figure 3. Different smoothing methods fit to lung cancer and radon exposure in Uranium cohort.

had the best balance between goodness of fit and biologic plausibility for this data set. For most bootstrap samples, we chose curves with df equal to 2 or 3, using the highest df value which was no longer wiggly. The average df for the 'by hand' criteria was 2.8, which is very close to the df of 2.6 estimated for the penalized spline on the data. Using all bootstrap samples for each pair of splines, we then computed the weights for the weighted area calculations described below.

Area differences between pairs of smoothed curves

Calculations for unweighted differences between curves are presented in Table IV(a) (Silica and Uranium studies) and weighted area differences in Table IV(b) (Silica study). The pair of curves

DOI: 10.1002/sim

Table IV(a). Unweighted area under and between pairs of smoothed curves for exposure in proportional hazards models for lung cancer in two occupational cohorts.

Smooth1 Smooth2	Areas under and between pairs of smoothed curves	Silica	Uranium
P-spline-AIC P-spline-df = 4	Average area under curves	241.16	10 786.78
	Total area difference	29.01 (12.0 per cent)*	3492.00 (32.4 per cent)
	Up to 50th percentile	0.06 (0.2 per cent)†	0.74 (0.02 per cent)
	Up to 99th percentile	7.09 (24.4 per cent)†	122.81 (3.5 per cent)
P-spline-AIC RCS	Average area under curves	225.79	15 858.20
	Total area difference	30.91 (13.7 per cent)	3868.61 (24.4 per cent)
	Up to 50th percentile	0.22 (0.7 per cent)	3.23 (0.1 per cent)
	Up to 99th percentile	7.91 (25.6 per cent)	370.62 (9.6 per cent)
P-spline-AIC Fracpoly	Average area under curves	294.00	7933.35
	Total area difference	55.18 (18.8 per cent)	5086.66 (64.1 per cent)
	Up to 50th percentile	0.49 (0.9 per cent)	7.09 (0.1 per cent)
	Up to 99th percentile	21.25 (38.5 per cent)	500.37 (9.8 per cent)
RCS Fracpoly	Average area under curves	277.11	10 786.78
	Total area difference	70.73 (25.5 per cent)	7948.88 (73.7 per cent)
	Up to 50th percentile	0.46 (0.6 per cent)	7.48 (0.1 per cent)
	Up to 99th percentile	15.72 (22.2 per cent)	138.54 (1.7 per cent)
P-spline-df = 4 RCS	Average area under curves	224.27	12 541.26
	Total area difference	24.74 (11.0 per cent)	7107.34 (56.7 per cent)
	Up to 50th percentile	0.15 (0.61 per cent)	2.49 (0.03 per cent)
	Up to 99th percentile	2.34 (9.5 per cent)	426.20 (6.0 per cent)
P-spline-df = 4 Fracpoly	Average area under curves	292.48	4616.41
	Total area difference	54.82 (18.7 per cent)	1818.89 (39.4 per cent)
	Up to 50th percentile	0.48 (0.87 per cent)	7.83 (0.4 per cent)
	Up to 99th percentile	16.68 (30.4 per cent)	560.64 (30.8 per cent)

Table IV(b). Weighted area under and between pairs of smoothed curves for exposure in proportional hazards models for lung cancer in two occupational cohorts.

Smooth1 Smooth2	Areas under and between pairs of smoothed curves	Silica
P-spline-AIC RCS	Average area under curves	110.97
	Total area difference between curves	16.41* (14.8 per cent)
	Up to 50th percentile	3.00† (18.3 per cent)
	Up to 99th percentile	5.56† (33.9 per cent)
P-spline-AIC Fracpoly	Average area under curves	128.13
	Total area difference between curves	43.73 (34.1 per cent)
	Up to 50th percentile	7.96 (18.2 per cent)
	Up to 99th percentile	37.59 (86.0 per cent)
RCS Fracpoly	Average area under curves	138.80
	Total area difference between curves	50.23 (36.2 per cent)
	Up to 50th percentile	8.45 (16.8 per cent)
	Up to 99th percentile	25.84 (51.4 per cent)

*Total area difference as a per cent of average area under smoothed curves.
†Per cent of total area difference.

DOI: 10.1002/sim

from the P-spline's and RCS were the closest in the Silica study, with area between them of 11.0 per cent (P-spline-df = 4 and RCS) and 13.7 per cent (P-spline-AIC and RCS) of the area under the curves. The fractional polynomial was farthest from the others. For the Uranium study, P-spline-AIC and RCS were the closest pair (24.4 per cent) and both curves were farthest from the pair, Fracpoly and P-spline-df = 4, which were also close to each other. Looking at differences up to percentiles of exposure, all pairs of curves were close over the denser portion of the exposure range in both studies. Less than 1 per cent of the unweighted difference was found up to the median, whereas beyond the median, the area differences were smallest for P-spline-df = 4 and RCS in the Silica study and RCS and Fracpoly in the Uranium study. It was interesting to note that P-spline-df = 4 and P-spline-AIC were not the closest pair of curves in either study.

The weighted area differences for the Silica study are presented in Table IV(b). The P-spline-AIC and RCS were again the closest pair and the distance of each from Fracpoly was more pronounced than for unweighted differences. The weighted per cent area differences at the median were much larger than unweighted differences, with 10–20 per cent of the total difference occurring in the lower (denser) half of the exposure range. This is because the curves in the upper tail of exposure, estimated with less precision, were downweighted. Beyond the median, the distance of both P-spline-AIC and RCS from Fracpoly was more pronounced (34.1 per cent and 36.2 per cent, respectively). The P-spline-AIC was farthest from the Fracpoly with 86 per cent of the total difference occurring by the 99th percentile. Thus, less of the differences appear to be concentrated in the tail when using weighted areas.

DISCUSSION

We modeled time to lung cancer mortality in two occupational cohort studies of lung cancer mortality and compared penalized splines, restricted cubic splines and fractional polynomials, each fit within Cox models and implemented *via* standard software. In the Silica study, fractional polynomials estimated a higher relative risk than the other methods over the entire range of exposure. Overall in both studies, the curves predicted by the P-spline (selected by df or AIC) and RCS curves were closer, as measured by the area difference between them as a per cent of the average area under the two curves, than either was to the curve predicted by fractional polynomials. This may be due to the global nature of the fractional polynomials in contrast to the more local fit of the splines.

In all cases, most of the unweighted difference between each pair of curves was concentrated in the top 1 per cent of the exposure distribution. Differences in the distribution of exposure may impact the disparity between pairs of curves in the top tail. Weighting by the inverse variance gave more weight to rectangles in the denser portions of the exposure range where predicted rate ratios have narrower confidence bands. The weighted area calculations then made it possible to take into account regions of the curves with higher variances, and substantially reduced the influence of the upper tail of the exposure distribution, where the curves for all three methods were estimated with the least precision. The more uniform the exposure distribution, the less important it would be to use the weighted area correction in comparisons of this kind since the weighting would also be more uniform and area calculations would be similar to unweighted. In situations where the exposure distribution is highly skewed, the correction will be most useful.

Bootstrapping was used to estimate the variance between splines needed for the weighted calculations because an analytic derivation of the asymptotic covariance was not feasible. In bootstrapping, we observed a familiar problem with the AIC criterion for the P-spline in Splus,

DOI: 10.1002/sim

in which it selected a wiggly curve with more df than may be necessary, almost half the time. Bootstrapping became a useful tool for estimating the weights as well as for observing the nature of the results of the AIC criterion for the P-spline. Since the default of df = 4 also resulted in less smoothed curves than biologically plausible, we recommend df = 2 or 3 as a default in analysis of biological endpoints.

In order to generalize these results, future work is needed to compare smoothing methods by simulations, where the fit of each smoothing method could be compared to the known exposure–response curve. By generating data from different dose–response curves, we could examine how well each of the smoothing methods captures the truth under varying scenarios and thus provide better information for judging the relative merits of each.

ACKNOWLEDGEMENTS

We are grateful for the generosity of both Harvey Checkoway and Jonathan Samet for providing their data for analysis for this paper.

REFERENCES

1. Greenland S. Modeling and variable selection in epidemiologic analysis. *American Journal of Public Health* 1989; **79**(3):340–349.
2. Greenland S. Dose–response and trend analysis in epidemiology: alternatives to categorical analysis. *Epidemiology* 1995; **6**(4):345–347.
3. Greenland S. Avoiding power loss associated with categorization and ordinal scores in dose–response and trend analysis. *Epidemiology* 1995; **6**(4):450–454.
4. Wartenberg D, Northridge M. Defining exposure in case–control studies: a new approach. *American Journal of Epidemiology* 1991; **133**(10):1058–1071.
5. Hastie TJ, Tibshirani RJ. *Generalized Additive Models*. Chapman & Hall: New York, 1990.
6. Fan J, Marron JS. Fast implementations of nonparametric curve estimators. *Journal of Computational and Graphical Statistics* 1994; **3**:35–56.
7. Wand MP, Jones MC. *Kernel Smoothing*. Chapman & Hall: London, 1994.
8. De Boor C. *A Practical Guide to Splines* (Rev. edn). Springer: New York, 2001.
9. Rice FL, Park R, Stayner L, Smith R, Gilbert S, Checkoway H. Crystalline silica exposure and lung cancer mortality in diatomaceous earth industry workers: a quantitative risk assessment. *Occupational and Environmental Medicine* 2001; **58**:38–45.
10. Dominici F, McDermott A, Zeger SL, Samet JM. On the use of generalized additive models in time-series studies of air pollution and health. *American Journal of Epidemiology* 2002; **156**(3):193–203.
11. Schwartz J, Laden F, Zanobetti A. The concentration–response relation between $PM_{2.5}$ and daily deaths. *Environmental Health Perspectives* 2002; **110**(10):1025–1029.
12. Braga ALF, Zanobetti A, Schwartz J. The lag structure between particulate air pollution and respiratory and cardiovascular deaths in 10 US cities. *Journal of Occupational and Environmental Medicine* 2001; **43**(11):927–933.
13. Eilers PHC, Marx BD. Flexible smoothing with B-splines and penalties. *Statistical Science* 1996; **11**:89–121.
14. Peng RD, Dominici F, Louis TA. Model choice in time series studies of air pollution and mortality. *Working Papers*, Department of Biostatistics, Johns Hopkins University, 2005, paper 55.
15. Eisen EA, Agalliu I, Coull BA, Thurston SW, Checkoway H. Smoothing in occupational cohort studies: an illustration based on penalized splines. *Occupational and Environmental Medicine* 2004; **61**:854–860.
16. André C, de Barros AP, Pereira LAA, Saldiva PHN. The distributed lag between air pollution and intrauterine mortality. *Epidemiology* 2004; **15**(4):S51.
17. Thurston SW, Eisen EA, Schwartz J. Smoothing in survival models: an application to workers exposed to metalworking fluids. *Epidemiology* 2002; **13**:685–692.
18. Steenland K, Deddens JA. A practical guide to dose–response analyses and risk assessment in occupational epidemiology. *Epidemiology* 2004; **15**(1):63–70.
19. Durrleman S, Simon R. Flexible regression models with cubic splines. *Statistics in Medicine* 1989; **8**:551–561.

DOI: 10.1002/sim

20. Herndon 2nd JE, Harrell Jr FE. The restricted cubic spline as baseline hazard in the proportional hazards model with step function time-dependent covariables. *Statistics in Medicine* 1995; **14**(19):2119–2129.
21. Jiang R, Hu FB, Giovannucci EL, Rimm EB, Stampfer MJ, Spiegelman D, Rosner BA, Willett WC. Joint association of alcohol and folate index with risk of major chronic disease in women. *American Journal of Epidemiology* 2003; **158**(8):760–771.
22. Woods MN, Spiegelman D, Knox TA, Forrester JE, Connors JL, Skinner SC, Silva M, Kim JH, Gorbach SL. Nutrient intake and body weight in a large HIV cohort that includes women and minorities. *American Dietetic Association* 2003; **102**(2):203–211.
23. Smith-Warner SA, Spiegelman D, Yaun S, van den Brandt PA, Folsom AR, Goldbohm A, Graham S, Holmberg L, Howe GR, Marshall JR, Miller AB, Potter JD, Speizer FE, Willett WC, Wolk A, Hunter DJ. Alcohol and breast cancer in woman. *The Journal of the American Medical Association* 1998; **279**(7):535–540.
24. Wise LA, Palmer JR, Harlow BL, Spiegelman D, Stewart EA, Adams-Campbell LL, Rosenberg L. Reproductive factors, hormonal contraception, and risk of uterine leiomyomata in African-American women: a prospective study. *American Journal of Epidemiology* 2003; **159**(2):113–123.
25. Bolard P, Quantin C, Abrahamowicz M, Esteve J, Giorgi R, Chadha-Boreham H, Binquet C, Faivre J. Assessing time-by-covariate interactions in relative survival models using restrictive cubic spline functions. *Journal of Cancer Epidemiology and Prevention* 2002; **7**(3):113–122.
26. Royston P, Altman DG. Using fractional polynomials to model curved regression relationship. *STATA Technical Bulletin No. 21*, *sg26*, STATA Corporation, College Station, TX, 1994.
27. Bagnardi V, Zambon A, Quatto P, Giovanni C. Flexible meta-regression functions for modeling aggregate dose–response data, with an application to alcohol and mortality. *American Journal of Epidemiology* 2004; **159**:1077–1086.
28. Stone RA, Youk AO, Marsh GM, Buchanich JM, McHenry MB, Smith TJ. Historical cohort study of US man-made vitreous fiber production workers: IV. Quantitative exposure-response analysis of the nested case–control study of respiratory system cancer. *Journal of Occupational and Environmental Medicine* 2001; **43**(9):779–792.
29. van Wijngaarden E. A graphical method to evaluate exposure-response relationships in epidemiologic studies using standardized mortality or morbidity studies. *Dose Response* 2005; **3**:465–473.
30. Ruppert D. Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics* 2002; **11**:735–737.
31. Checkoway H, Heyer NJ, Seixas NS, Welp EAE, Demers PA, Hughes JM *et al*. Dose–response associations of silica and nonmalignant respiratory disease and lung cancer mortality in diatomaceous earth industry. *American Journal of Epidemiology* 1997; **145**:680–688.
32. IARC. *Silica*, *Some Silicates*, *Coal Dust*, and *Para-Aramid Fibrils*, vol. 68. IARC Publications: Lyon, France, 1997.
33. Samet J, Pathak DR, Morgan MV, Key CR, Vadiva AA, Lubin JJ. Lung cancer mortality and exposure to radon progeny in a cohort of New Mexico underground uranium miners. *Health Physics* 1991; **61**:745–752.
34. Monson RR. Observations on the healthy worker effect. *Journal of Occupational Medicine* 1986; **28**:425.
35. Insightful Corp. *Insightful Corp*. *Copyright* (*c*) *1988*, *2003*. Insightful Corp.: Seattle, Washington.
36. Therneau TM, Grambsch PM. Penalized Cox models and frailty. *Technical Report*, Division of Biostatistics, Mayo Clinic, Rochester, MN, 1998.
37. Verweij P, van Houwelingen H. Penalized likelihood in Cox regression. *Statistics in Medicine* 1994; **13**:2427–2436.
38. Gray RJ. Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. *Journal of the American Statistical Association* 1992; **87**:942–951.
39. Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 1974; **19**(6):716–723.
40. Therneau TM, Grambsch PM. *Modeling Survival Data*: *Extending the Cox Model*. Springer: New York, 2002.
41. Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Applied Statistics* 1994; **43**(3):429–467.
42. StataCorp. *STATA 9.0 Copyright (c)* (*1984–2005*). StataCorp.: College Station, TX.
43. Collett D. *Modelling Survival Data in Medical Research* (2nd edn). CRC Press: Boca Raton, FL, 2003.
44. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. Chapman & Hall/CRC: New York, 1993.
45. Chan TF, Golub GH, LeVeue RJ. Algorithms for computing the sample variance: analysis and recommendations. *The American Statistician* 1983; **37**(3):242–247.

DOI: 10.1002/sim