The International Journal of Biostatistics

Volume 5, Issue 1

2009

Article 26

Using Generalized Additive Models to Detect and Estimate Threshold Associations

Andrea Benedetti, *McGill University*

Michal Abrahamowicz, *McGill University*

Karen Leffondré, *Université de Bordeaux 2*

Mark S. Goldberg, *McGill University*

Robyn Tamblyn, *McGill University*

Recommended Citation:

Benedetti, Andrea; Abrahamowicz, Michal; Leffondré, Karen; Goldberg, Mark S.; and Tamblyn, Robyn (2009) "Using Generalized Additive Models to Detect and Estimate Threshold Associations," *The International Journal of Biostatistics*: Vol. 5: Iss. 1, Article 26.

DOI: 10.2202/1557-4679.1172

Using Generalized Additive Models to Detect and Estimate Threshold Associations

Andrea Benedetti, Michal Abrahamowicz, Karen Leffondré, Mark S. Goldberg,
and Robyn Tamblyn

Abstract

In a variety of research settings, investigators may wish to detect and estimate a threshold in the association between continuous variables. A threshold model implies a non-linear relationship, with the slope changing at an unknown location. Generalized additive models (GAMs) (Hastie and Tibshirani, 1990) estimate the shape of the non-linear relationship directly from the data and, thus, may be useful in this endeavour.

We propose a method based on GAMs to detect and estimate thresholds in the association between a continuous covariate and a continuous dependent variable. Using simulations, we compare it with the maximum likelihood estimation procedure proposed by Hudson (1966).

We search for potential thresholds in a neighbourhood of points whose mean numerical second derivative (a measure of local curvature) of the estimated GAM curve was more than one standard deviation away from 0 across the entire range of the predictor values. A threshold association is declared if an F-test indicates that the threshold model fit significantly better than the linear model.

For each method, type I error for testing the existence of a threshold against the null hypothesis of a linear association was estimated. We also investigated the impact of the position of the true threshold on power, and precision and bias of the estimated threshold.

Finally, we illustrate the methods by considering whether a threshold exists in the association between systolic blood pressure (SBP) and body mass index (BMI) in two data sets.

KEYWORDS: generalized additive models, GAMs, threshold, simulation, linear regression, F-test

Author Notes: This research was supported by grant MA-15669 (Title: Process and Outcome in Primary Care Practice: the Quebec-Ontario Study, P.I.: R. Tamblyn) from the Medical Research Council of Canada and a grant from the Canadian Institutes for Health Research (CIHR, P.I.: M. Abrahamowicz). Dr. Michal Abrahamowicz is a James McGill Professor. Dr. Goldberg gratefully acknowledges support from the Canadian Institutes of Health Research as a CIHR Investigator. Dr. Leffondré was supported by a salary award from the Fond de recherche en santé du Québec

(FRSQ). This research is a part of Andrea Benedetti's Ph.D. thesis in the department of Epidemiology and Biostatistics at McGill University, under the supervision of Michal Abrahamowicz.

1 Introduction

It is sometimes plausible to expect that a continuous risk factor affects the outcome only if its value exceeds a certain, usually unknown, threshold. Alternatively, we may expect that the strength of the association changes at such a threshold. Estimating such thresholds may lead to a better understanding of the biological mechanism underlying the association between a continuous independent variable and a dependent variable, and may be useful from a regulatory perspective in providing a basis for setting acceptable levels of exposure. Indeed, there has been considerable interest in estimating such thresholds in the relationships between air pollution, blood pressure, radiation, carcinogens and occupational exposures and various health outcomes (Hoel and Li, 1998; Hoel and Portier, 1994; Kaufman et al., 1997; Kuchenhoff and Ulm, 1997; Mossman, 1998; Nussbaum, 1998; Schwartz, 1994; Sinclair, 1998; Webster, 1998; Abrahamowicz et al., 2003).

Statistical methods for estimating the point of change in a slope have a long history (Hudson, 1966). In addition to maximum likelihood approaches (Hudson, 1966; Hajat et al., 2001; Kuchenhoff, 1997), a variety of easier-to-implement *ad hoc* methods are used in the medical literature to investigate the existence and location of possible thresholds of clinical interest.

For example, the existence of a threshold in the association between body mass index (BMI) and blood pressure (BP) is the subject of some debate in the medical literature (Kaufman et al., 1997; Bunker et al., 1995; Doll et al., 2002; Stamler et al., 1986). Whereas, generally, BP is expected to increase with increasing BMI, the relationship has been postulated to have a threshold in a lean population at $21\text{kg}/\text{m}^2$ (Kaufman et al., 1997) or at $21.5\text{kg}/\text{m}^2$ (Bunker et al., 1995), such that differences in BMI below these values are not associated with SBP (systolic BP). Different methods were used to investigate this issue. Kaufman (1997) fitted linear regression splines with a single knot, the location of which was chosen by searching iteratively across the range of BMI and choosing the value which minimized the error-sum-of-squares. Bunker (1995) plotted the mean BP of each decile of the BMI distribution against that decile's mean BMI. From the resulting non-parametric curve, a range of possible thresholds was visually identified and threshold models were fit at several points in this neighbourhood (Bunker et al., 1995). The threshold model with the lowest error-sums-of-squares was chosen (Bunker et al., 1995). In contrast, Doll and colleagues (Doll et al., 2002) concluded that a linear model for the BMI-BP association was appropriate after visually examining a loess curve. However, these methods have not been evaluated in terms of the risk of over-fitting bias, type I error for detecting a threshold, and the accuracy and variability of the estimated threshold.

There are two methodologic challenges involved in estimating threshold associations. The first is finding the most likely location of the potential threshold. The second is deciding if there is enough evidence in support of the threshold model over an alternative simpler model that fits a continuous smooth function, usually linear, to describe the association of interest.

In this paper, we propose using generalized additive models (GAMs) (Hastie and Tibshirani, 1990) to identify thresholds and estimate their location. The GAM is a non-parametric regression technique that fits smooth curves to data, without *a priori* assumptions about the functional form of the dose-response relationship. Rather, the shape of the curve is estimated directly from the data. For this reason, GAMs have been used informally to suggest the functional form for the association between two continuous covariates (Hastie and Tibshirani, 1986; Schwartz, 1994; Roland et al., 2000), and might be useful to estimate threshold associations. We investigate several methodological issues and compare the performance of alternative methods for the estimation and detection of thresholds in a simulation study. We also revisit the issue of a threshold in the BMI-BP association, using data from two large epidemiologic studies.

In the next section, we provide an overview of GAMs and briefly describe alternative methods for threshold detection and estimation including maximum likelihood estimation, and a GAM-based approach. Then, section 3 describes the simulations conducted to evaluate these methods. We apply selected methods to re-analyze the BMI-BP association in section 4. Finally, section 5 discusses the limitations of our work, and suggests some directions for future research.

2 Methods

2.1 Threshold definition

We focus on the relationship between a continuous independent variable (X) and a continuous dependent variable (Y). Two definitions of a threshold have been proposed in the literature, and in this work we investigate both. In the ‘no-effect’ threshold model, the slope before the threshold (t) is constrained to 0, so that the association between X and Y is limited to $X > t$ (Kaufman et al., 1997; Bunker et al., 1995; Daniels et al., 2000; Queiroz-Neto et al., 2001; Ulm, 1991). In the second, broader definition, the threshold is seen as an inflection point, i.e. the slope of the dose-response curve is assumed to change after the threshold, implying a bi-linear model (Chu et al., 1999; Pastor and Guallar, 1998). In the context of multivariable linear regression, these two models can be written, respectively

as:

$$Y = \beta_0 + \sum_{i \in A} \beta_{i_2} [X_i - t_i]_{t_i+} + \sum_{j \in B} \beta_j X_j \quad (1)$$

$$Y = \beta_0 + \sum_{i \in A} \{\beta_{i_1} X_i + \beta_{i_2} [X_i - t_i]_{t_i+}\} + \sum_{j \in B} \beta_j X_j \quad (2)$$

where A is the set of those continuous covariates for which thresholds are assumed; B represents binary variables and those quantitative variables whose effects are modelled as linear; t_i is the threshold in the association between Y and X_i ; and

$$\begin{aligned} [X_i - t_i]_{t_i+} &= X_i - t_i & \text{if } X_i \geq t_i \\ [X_i - t_i]_{t_i+} &= 0 & \text{if } X_i < t_i. \end{aligned}$$

Notice that the ‘no-effect’ threshold model (1) is a special case of model (2), corresponding to the situation where the pre-threshold slope, β_{i_1} , is *a priori* constrained to 0.

2.2 Overview of GAMs

For continuous outcomes, GAMs generalize the conventional multiple linear regression model to avoid *a priori* assumptions about the functional form of the dose-response curve (Hastie and Tibshirani, 1990). Specifically, the effect of a quantitative independent variable, X , on Y is represented by a flexible smooth function $S_k(X)$, the shape of which is estimated directly from the data, without imposing specific parametric assumptions such as e.g. linearity. Smoothing in GAMs can be implemented via several approaches, including smoothing splines and loess (Hastie and Tibshirani, 1990). Accordingly, the GAM extension of the multiple linear regression model for continuous outcomes can be written as:

$$E(Y|X) = \alpha + \sum_{j \in A} S_{j,k}(X_j) + \sum_{i \in B} \beta_i X_i \quad (3)$$

where A represents the set of quantitative variables modelled using non-parametric smoothers ($S_{j,k}(X_j)$), and B represents binary variables and those quantitative variables whose effects are modelled as linear. Here, k represents the user-specified degrees of freedom (df), and this value determines the flexibility of the smooth curve $S_{j,k}$ and, thus, the trade-off between variance and bias of the estimates, or under-fit and over-fit bias (Hastie and Tibshirani, 1990). As in other non-parametric regression models, higher df make it possible to recover more complex relationships but increase the risk of over-fitting bias, especially in small samples, and may result in inflated variance and inefficient tests (Abrahamowicz et al., 1996; Abrahamowicz and Ciampi, 1991).

In GAMs, inference about the nonlinearity of the effect of a particular continuous predictor is accomplished by an analysis of deviance. Specifically, nonlinearity of the effect of a given continuous predictor is tested by comparing the deviance of the k -df smoothed model to the deviance of a 1-df linear model (Hastie and Tibshirani, 1990). The resulting non-parametric likelihood ratio test statistic is approximately distributed as a χ^2 statistic with $k - 1$ dfs (Hastie and Tibshirani, 1990).

2.3 Alternative methods for locating potential thresholds

Investigation of a threshold involves two related steps: (1) locating the potential thresholds and then (2) deciding if indeed a threshold exists. This section presents an overview of methods that may be used to locate the threshold, while section 2.4 proposes some approaches to decide if there is enough evidence to justify a threshold model rather than a simpler alternative form.

2.3.1 Method 1: maximum likelihood estimation (MLE)

The first method investigated is the maximum likelihood estimator of the threshold, proposed by Hudson (1966). In Hudson's (1966) algorithm, we search for a threshold first between, and then at, observed values of the independent variable. At each potential location, t , model (1) or (2) is fitted. The location, t , with the lowest error-sum-of-squares is then chosen as the threshold. The maximum likelihood estimate of the threshold is either at the point of intersection between the pre- and post-threshold lines (in the case that the threshold is located between two data points), or at the data point itself, if the threshold coincides with an observed value of the independent variable.

2.3.2 Method 2: the restricted search

We propose to use the GAM to focus the search for the most likely location for a threshold. We expect that the estimated GAM curve should help identify the neighbourhood within which the threshold is located in the form of increased curvature. This is because around the threshold, the dose-response curve should change, and accordingly, the first derivative of the estimated flexible curve should increase. The increases in the first derivative should be indicated by relatively high 'local' values of the second derivative.

To implement this approach, we first fit a GAM model using the smoothing spline option for the smoother to the data (Hastie and Tibshirani, 1990). A neighbourhood of 'suspicious' points is then determined, consisting of all points at which the numerical local second derivative is more than one standard deviation

(across all points considered) away from 0, indicating relatively high local curvature. Next, only points in this neighbourhood are considered as potential thresholds. At each potential location, t , model (1) or (2) is fitted. The location, t , with the lowest error-sum-of-squares is then chosen as the threshold.

The nonparametric nature of GAM estimates (Hastie and Tibshirani, 1990) makes it impossible to express their derivatives analytically. Thus, the local second derivative is numerically approximated by second differences using the following formula:

$$\frac{((S_{df}(x + \Delta) - S_{df}(x)) - (S_{df}(x) - S_{df}(x - \Delta)))}{\Delta^2}; \quad (4)$$

where $S_{df}(x)$ represents the GAM estimate at x . We refer to Δ as the span, with a larger span resulting in a smoother, less bumpy numerical approximate to the second derivative.

2.4 Criteria for determining if a threshold association exists

Having identified the most likely location for a potential threshold, the data analyst has then to evaluate whether there is enough evidence in favour of the threshold association over a simpler alternative model. In this section, we consider two alternative criteria to compare the best-fitting threshold model with the linear model, which is usually considered a natural alternative.

2.4.1 F-test for threshold

We evaluated the use of an approximate F-test as described by Vieth (1989) to decide between the threshold model and a linear model. Here, the linear model is considered as nested within the threshold model, because the linear model is equivalent to a model with a hypothesized threshold located at the minimum value of the independent variable, and the threshold is considered an additional parameter to be estimated. Because the two models are nested, we can use an F-test for the existence of a threshold. The null hypothesis of this test corresponds to the linear model: $H_0 : Y = \beta_0 + \beta_1 X$, whereas the alternative hypothesis to the threshold model (1) or (2). Thus, for the ‘no-effect’ threshold model (1), the F-statistic is:

$$F = \frac{(SSE_{linear} - SSE_{threshold})/1}{MSE_{threshold}} \sim F_{(1, n-3)}. \quad (5)$$

We propose to accept the H_A , i.e. the existence of a threshold, if the test rejects the H_0 at the $\alpha = 0.05$ level.

The F-statistic for the ‘change-in-slope’ threshold model (2) is similar, except the df in the numerator and denominator are 2 and $n - 4$, respectively, and follows a $F_{(2,n-4)}$.

2.4.2 Akaike Information Criterion (AIC)

Cakmak et al. (1999) proposed comparing the goodness of fit of the threshold model (1) with a linear model using the AIC (Akaike, 1973). Typically, the AIC is calculated as:

$$2p - 2Ln(L) \quad (6)$$

where L is the likelihood of the estimated model and p is the number of parameters in the model. To calculate the AIC of the threshold models (1) and (2) we let $p = 3$ or $p = 4$, respectively, thus assuming that the threshold location corresponds to an additional parameter to be estimated from the data.

2.5 Discerning among alternative nonlinear functional forms

Finally, there may be situations where a threshold model (1) or (2) fits better than a linear model, but where the true relationship corresponds to a *smooth* nonlinear dose-response curve. Thus we investigated adding an additional criterion, in that we declared a threshold association only if the AIC of the final threshold model (1), or (2), respectively, was lower than that of a 4-df GAM curve. For example, after identifying a statistically significant threshold with the F-test described in section 2.4.1, we further required that the threshold model had lower AIC than a 4-df GAM model.

3 Simulation study

3.1 Data generation

We evaluated the two methods (section 2.3) and two criteria (section 2.4) for identifying a possible threshold in either the ‘no-effect’ model (1) or the ‘change-in-slope’ model (2) in separate simulation studies. The continuous independent variable values $x_i, i = 1, \dots, n$, were generated from a Normal distribution with mean=50 and standard deviation=10. For each x_i , the corresponding value of the dependent variable was generated from a model, $y_i = f(x_i) + \varepsilon_i$, where $\varepsilon_i \sim N[0; \sigma = 12]$ were normally distributed residuals and the pre-specified function, $f(x)$, defined the true $X - Y$ association. Specifically, the following functional forms were considered for $f(x)$ in scenarios where a threshold was assumed not to exist: (i) no association ($f(x) = 0$), (ii) a linear association ($f(x) = 0.75x$), (iii)

a quadratic association ($f(x) = 0.2x + 0.04x^2$), and (iv) an asymmetric J shape ($f(x) = 0.2(x - 47) + 0.04(x - 47)^2$). We also considered two scenarios with an actual threshold (v) a ‘no-effect’ threshold model (1) with $f(x) = 0.75(x - t)_{t+}$, and (vi) a ‘change-in-slope’ threshold model (2) with $f(x) = 0.25x + 1(x - t)_{t+}$. For both threshold models we considered the threshold (t) located at the 10th, 25th, 50th, 75th or 90th percentile of the true distribution of the independent variable.

All simulations were performed using R, version 2.5.0 (R Development Core Team, 2008). Specific parameters used to generate data, and criteria employed to evaluate different aspects of the performance of alternative methods are described in the corresponding subsection, when describing the results of the simulations.

3.2 Simulation results

3.2.1 Parameters for the restricted search

The restricted search depends on three tuning parameters: the df of the GAM curve, the value for the span, Δ , in the calculation of the second derivative and the number of points at which the threshold model is estimated. In fact, we investigated several values for each of these. We saw no systematic advantages in terms of power or bias or variability of the estimated threshold when we increased the df of the GAM curve from 4 to 6 or decreased df to 2. Nor did we see any systematic differences when we searched at every value of X or if we increased the number of points by $n/2$ or n . Furthermore, similar results were obtained with alternative spans corresponding to $\Delta = 0.5$, 1, or 3. (Data not shown.) Thus, in our simulations and empirical analyses we used 4 df, searched for a threshold at every value of X as well as n additional points, and $\Delta = 1$. Both the span, Δ , and the number of points, depend clearly on the units and empirical range of the predictor X , for which the threshold is investigated. Therefore, in each real-life application the data analyst should decide what Δ and n are reasonable; based on our results, selecting Δ of the order of 1-2% of the empirical range of X , and searching at every value of X as well as n additional points should provide a reasonable trade-off between numerical precision and computational efficiency, depending on the size of the dataset.

3.2.2 Type I error rates for different methods of threshold detection

To assess type I error rates, we considered two scenarios, with the data generated from (i) a null, or (ii) a linear association between X and Y , respectively. Accordingly, the proportion of samples in which a threshold association was declared estimated the empirical type I error rate. One thousand samples of size $N=300$ were simulated for each scenario. Each sample was analysed with four different

combinations of (i) a method for threshold location (section 2.3) and (ii) a criterion for ‘testing’ for the existence of the threshold (section 2.4). Furthermore, all data were analyzed twice to estimate, respectively, the ‘no-effect’ threshold model (1) and the ‘change-in-slope’ threshold model (2). The empirical type I error rates and corresponding exact 95% confidence intervals, are presented in Table 1.

Table 1: Type I error (exact 95% confidence interval) by underlying association for each threshold estimation method, based on 1000 simulated datasets with $n=300$

Estimated threshold model ^a	Method	Criteria ^b	True Association	
			No Association	Linear Association
Model (1) (‘No-effect’)	1. MLE	F-test	0.09 (0.07, 0.11)	0.03 (0.02, 0.05)
		AIC	0.24 (0.22, 0.27)	0.10 (0.08, 0.12)
	2. Restricted search	F-test	0.08 (0.06, 0.09)	0.03 (0.02, 0.05)
		AIC	0.20 (0.18, 0.23)	0.10 (0.08, 0.12)
Model (2) (‘Change-in-slope’)	1. MLE	F-test	0.09 (0.07, 0.11)	0.10 (0.08, 0.12)
		AIC	0.27 (0.25, 0.30)	0.28 (0.25, 0.31)
	2. Restricted search	F-test	0.08 (0.06, 0.09)	0.09 (0.07, 0.11)
		AIC	0.24 (0.21, 0.27)	0.25 (0.22, 0.28)

^a Type of threshold model fit.

^b Two criteria were used to determine if a threshold association existed: a significant F-test, or better fit than a linear model as indicated by AIC.

No matter the criteria, type I error was always slightly higher for MLE compared to the restricted search. As expected, using the AIC to decide between a threshold model, either (1) or (2), and the linear model, resulted in inflated type I error for both methods for locating the optimal threshold, especially when there truly was no association. When using the F-test, type I error was also inflated, though to a lesser degree than with the AIC, except there was no inflation when the true model was linear and a threshold model (1) was estimated. When there was no association, fitting both threshold models (1) and (2), and using the F-test, yielded serious, inflated type I error rates. In contrast, when the true association was linear, the rates were much more inflated for the ‘change-in-slope’ model (2).

3.2.3 Empirical power

To estimate the power of detecting a true threshold association we generated data according to the threshold model (1) and (2), as described in section 3.1. When the true association was a ‘change-in-slope’ threshold model (2), we estimated a threshold model (2) only. When the true association was a ‘no-effect’ threshold model (1), we estimated both threshold models. One thousand samples were generated. Each simulated sample was analysed with both the MLE approach and

the restricted GAM-based search, and then the F-test was applied. Because of the serious inflation of type I error (Table 1) the AIC was not employed.

Table 2 presents the empirical power for the F-test combined with each of the two methods for selected locations of the threshold. For both methods, and whether fitting model (1) or (2), power decreased as the location of the true threshold moved away from the median of the independent variable distribution. Power for the GAM-based restricted search was very slightly lower than that obtained using the unrestricted search, likely reflecting the more liberal type I error of the latter method. When the true threshold model was (1) and threshold model (2) was fit, power was lower for both methods than when a model (1) threshold was fit.

When a significant threshold association was found, we compared the fit of the ‘no-effect’ threshold model (1), or the ‘change-in-slope’ threshold model (2) with that of a 4-df GAM curve, using the AIC (Akaike, 1973). No matter the position of the threshold, the threshold model fit better in nearly all simulated samples (Table 2).

Finally, similar patterns of results, though with much lower power, were seen when the magnitude of the post-threshold slope was decreased or when the number of observations in each data set was decreased (data not shown). For example, depending on the location of the true threshold, for the MLE with the F-test power ranged from 19-51% when the sample size decreased to 100. For both methods with the F-test, power was above 60%, for ‘no-effect’ thresholds located at the 10th or 90th percentile when the sample size was increased to 1000 observations. (Data not shown.)

3.2.4 Precision and bias of the estimated thresholds

Tables 3-5 show the mean bias, variability and mean square error (MSE) for the threshold location (t in equation (1) and (2)), estimated with different methods when, respectively, (i) the true association was a ‘no-effect’ threshold model (1) and we estimated the ‘no-effect’ threshold model (1) (Table 3), (ii) the true association was a ‘change-in-slope’ threshold model (2) and we estimated the ‘change-in-slope’ threshold model (2) (Table 4), and (iii) the true association was a ‘no-effect’ threshold model (1) and we estimated the ‘change-in-slope’ threshold model (2) (Table 5). In some applications, the investigator might be sure that a threshold association exists, and will not be interested in testing, but only in estimating the threshold. Therefore, for each of the two estimation methods, we present results corresponding to (a) ‘No selection’, i.e. based on *all* simulated samples and (b) ‘F-test’, i.e. restricted to those samples where the F-test indicated that the threshold model fit significantly better (at $\alpha = 0.05$) than a linear model.

Table 2: Empirical power (95% CI) for each threshold estimation method

Searching for ^b	True Threshold Model ^c	Method	Criterion	Position of the Threshold ^a				
				10	25	50	75	90
Model(1) (‘No-effect’)	Model (1) (‘No-effect’)	1. MLE	F-test	0.47 (0.44, 0.50)	0.78 (0.76, 0.81)	0.90 (0.88, 0.92)	0.80 (0.77, 0.82)	0.47 (0.44, 0.50)
			Better than GAM? ^d	0.99 (0.98, 1.00)	0.99 (0.99, 1.00)	0.99 (0.99, 1.00)	0.99 (0.98, 0.99)	1.00 (0.99, 1.00)
		2. Restricted search	F-test	0.47 (0.44, 0.50)	0.78 (0.75, 0.81)	0.90 (0.88, 0.92)	0.80 (0.77, 0.82)	0.46 (0.43, 0.49)
			Better than GAM? ^d	0.99 (0.99, 1.00)	0.99 (0.99, 1.00)	0.99 (0.99, 1.00)	0.99 (0.98, 0.99)	1.00 (0.99, 1.00)
Model (2) (‘Change-in-slope’)	Model (2) (‘Change-in-slope’)	1. MLE	F-test	0.58 (0.55, 0.61)	0.92 (0.91, 0.94)	0.98 (0.98, 0.99)	0.93 (0.92, 0.95)	0.579 (0.55, 0.58)
			Better than GAM? ^d	1.00 (0.99, 1.00)	1.00 (0.99, 1.00)	1.00 (1.00, 1.00)	1.00 (0.99, 1.00)	1.00 (0.99, 1.00)
		2. Restricted search	F-test	0.57 (0.54, 0.60)	0.92 (0.91, 0.94)	0.98 (0.97, 0.99)	0.93 (0.92, 0.95)	0.57 (0.54, 0.60)
			Better than GAM? ^d	1.00 (0.99, 1.00)	1.00 (0.99, 1.00)	1.00 (1.00, 1.00)	1.00 (0.99, 1.00)	1.00 (0.99, 1.00)
Model (2) (‘Change-in-slope’)	Model (1) (‘No-effect’)	1. MLE	F-test	0.40 (0.37, 0.43)	0.72 (0.69, 0.75)	0.87 (0.85, 0.89)	0.73 (0.70, 0.75)	0.38 (0.37, 0.43)
			Better than GAM? ^d	0.99 (0.98, 1.00)	1.00 (0.99, 1.00)	1.00 (0.99, 1.00)	0.99 (0.99, 1.00)	1.00 (0.99, 1.00)
		2. Restricted search	F-test	0.39 (0.36, 0.42)	0.71 (0.68, 0.74)	0.87 (0.85, 0.891)	0.72 (0.69, 0.75)	0.37 (0.34, 0.40)
			Better than GAM? ^d	0.99 (0.98, 1.00)	1.00 (0.99, 1.00)	1.00 (0.99, 1.00)	0.99 (0.99, 1.00)	1.00 (0.99, 1.00)

^a Percentile of the independent variable at which the true threshold was located.

^b Denotes whether the target was a threshold of type (1) or (2).

^c Denotes whether the true association in the data was threshold model of type (1) or (2).

^d Proportion of samples with a significant threshold that fit better than a 4-df smoothing spline GAM curve according to AIC.

Table 3: Bias, variability (Var) and Mean square error (MSE) of estimated model (1) (‘no-effect’) thresholds for each method by location of true threshold when the true association is the ‘no-effect’ threshold model (1)

Method	Criteria	Method	Position of the threshold ^{a,b}				
			10	25	50	75	90
1. MLE	No selection ^c	Relative Bias (%)	-3.02	-2.02	-1.20	-1.49	-4.26
		Var	34.93	25.75	23.44	59.70	156.76
		MSE	36.20	26.52	23.80	60.42	163.91
1. MLE	F-test ^d	Relative Bias (%)	6.42	1.71	0.51	1.55	2.70
		Var	9.17	8.97	9.65	18.67	30.00
		MSE	14.88	9.51	9.72	19.44	32.88
2. Restricted Search	No selection	Relative Bias (%)	2.60	-0.72	-0.66	-0.94	-4.78
		Var	19.05	16.54	16.45	39.23	104.80
		MSE	19.98	16.64	16.56	39.51	113.81
2. Restricted Search	F-test	Relative Bias (%)	6.32	1.63	0.51	1.38	1.84
		Var	9.33	9.11	10.03	17.04	24.62
		MSE	14.85	9.61	10.10	17.65	25.95

^a Percentile of the independent variable.

^b 1000 samples of size $N=300$ were generated.

^c Results are for all thresholds, whether or not statistically significant.

^d Results are for thresholds statistically significant using the F-test at $\alpha = 0.05$.

When the ‘no-effect’ threshold model (1) was both the true and estimated model, when all thresholds were considered (i.e. no selection criterion), with the exception of when the true threshold was at 50, the estimated threshold was biased toward the median of the independent variable and the amount of bias increased as the true threshold moved away from the median of the independent variable, for the restricted search. Bias was always toward 0 and was more extreme further from the median of the independent variable for the unrestricted MLE method. Variability increased as the threshold increased for both methods (Table 3)

When considering only those samples for which the F-test was statistically significant, bias, and MSE increased as the threshold moved away from the median of the independent variable (Table 3). When all thresholds were considered, the estimated thresholds with the unrestricted MLE search were more biased and variable than with the restricted search. When only statistically significant thresholds were considered the unrestricted MLE search resulted in slightly more biased and variable estimates than the restricted search, especially for thresholds located at the 90th percentile (Table 3).

When model (2) was both the true and the estimated model, absolute bias, variance and mean squared error of the estimated thresholds increased as the threshold moved away from the median of the data. Estimated thresholds were biased toward the median of the data. Thresholds estimated via the restricted search were

Table 4: Bias, variability (Var) and Mean square error (MSE) of estimated model (2) ('change-in-slope') thresholds for each method by location of true threshold when the true association is the 'change-in-slope' threshold model (2)

Method	Criteria	Method	Position of the threshold ^{a,b}				
			10	25	50	75	90
1. MLE	No selection ^c	Relative Bias (%)	9.24	0.72	0.12	-0.10	-5.83
		Var	139.06	37.63	20.10	32.08	139.48
		MSE	150.89	37.73	20.11	32.08	152.91
1. MLE	F-test ^d	Relative Bias (%)	3.98	0.21	-0.04	-0.02	-2.33
		Var	68.92	29.38	18.91	26.99	81.32
		MSE	71.11	29.39	18.91	26.99	83.47
2. Restricted Search	No selection	Relative Bias (%)	9.64	1.10	0.23	-0.19	-5.91
		Var	110.72	34.49	19.00	28.29	114.62
		MSE	123.57	34.71	19.01	28.30	128.38
2. Restricted Search	F-test	Relative Bias (%)	4.70	0.40	0.07	-0.09	-2.52
		Var	62.15	27.81	17.79	24.84	67.86
		MSE	65.20	27.84	17.79	24.84	70.36

^a Percentile of the independent variable.

^b 1000 samples of size N=300 were generated.

^c Results are for all thresholds, whether or not statistically significant.

^d Results are for thresholds statistically significant using the F-test at $\alpha = 0.05$.

comparably biased but were much less variable than MLE-estimated thresholds (Table 4)

When the true model was (1), and we estimated model (2), the resulting thresholds were more biased and variable than when estimating model (1) thresholds (Table 5 vs. Table 3).

3.2.5 Discerning among smooth nonlinear functional forms

When the true association corresponded to a smooth quadratic function (scenario (iii) in section 3.1), a 'no-effect' threshold model (1) compared to a linear model was incorrectly declared in around 67% of samples when the MLE approach was used with the F-test, and 60% of samples when the restricted search was used with the F-test (Table 4). Among those samples in which a threshold was statistically significant, the AIC of the threshold model indicated a better fit than a smooth GAM curve in 51% or 55% of samples when the unrestricted MLE or restricted GAM-based method was used, respectively. Results were worse when the 'change-in-slope' threshold model (2) was fit: type I error was 96% for both methods.

When the true association was J-shaped (scenario (iv) in section 3.1), a threshold model (1) or (2) was declared in around 98% of samples when either the MLE

Table 5: Bias, variability (Var) and Mean square error (MSE) of estimated model (2) ('change-in-slope') thresholds for each method by location of true threshold when the true association is the 'no-effect' threshold model (1)

Method	Criteria	Method	Position of the threshold ^{a,b}				
			10	25	50	75	90
1. MLE	No selection ^c	Relative Bias (%)	16.74	3.41	-0.18	-2.02	-9.91
		Var	207.55	91.52	55.38	91.89	225.61
		MSE	246.31	93.70	55.39	93.21	264.36
1. MLE	F-test ^d	Relative Bias (%)	8.96	1.92	-0.26	-0.86	-4.01
		Var	136.64	64.72	42.09	65.93	128.27
		MSE	147.74	65.41	42.11	66.17	134.62
2. Restricted Search	No selection	Relative Bias (%)	15.93	3.62	-0.14	-2.88	-9.77
		Var	164.80	76.37	49.07	83.24	170.80
		MSE	199.91	78.82	49.07	85.92	208.46
2. Restricted Search	F-test	Relative Bias (%)	8.45	1.86	-0.22	-1.01	-3.59
		Var	113.23	53.83	41.84	57.54	96.95
		MSE	123.11	54.47	41.86	57.88	102.03

^a Percentile of the independent variable.

^b 1000 samples of size N=300 were generated.

^c Results are for all thresholds, whether or not statistically significant.

^d Results are for thresholds statistically significant using the F-test at $\alpha = 0.05$.

or restricted search was used with the F-test (Table 4). Among those samples in which a threshold was statistically significant, the AIC of the threshold model indicated a better fit than a GAM model in almost all samples.

4 Real-life illustration: searching for a threshold in the effect of BMI on blood pressure

To illustrate these methods, we investigated the relationship between body mass index (BMI) and blood pressure. Previous work has focused on the existence of a 'no-effect' threshold model (1) in this association (Kaufman et al., 1997; Bunker et al., 1995; Doll et al., 2002; Stamler et al., 1986), and so we restrict our focus to model (1). We re-analysed data from two large cohort studies of cardiovascular disease (CVD) mortality and morbidity: the Framingham Heart Study (n=5209) and the Lipid Research Clinics Prevalence and Follow-Up Study (LRC) (n=2512). The Framingham Heart Study assessed risk factors for CVD morbidity and death by following a group of participants who had not yet developed overt symptoms of CVD at baseline (Dawber et al., 1951). The LRC study followed 2512 men, randomly selected from 81,926 initially contacted subjects, for CVD mortality (Central Patient Registry and Coordinating Center for the Lipid Research Clinics,

Table 6: Empirical type I error (95% CI) by threshold estimation method, when the true association was nonlinear, but not threshold

Estimated threshold model ^a	Method	Criteria	True Association	
			Asymmetric J	Quadratic
Model (1) (‘No-effect’)	1. MLE	F-test	0.98 (0.97, 0.99)	0.67 (0.64, 0.70)
		Better than GAM? ^b	0.88 (0.86, 0.90)	0.51 (0.48, 0.55)
	2. Restricted search	F-test	0.98 (0.97, 0.99)	0.60 (0.57, 0.63)
		Better than GAM?	0.87 (0.85, 0.89)	0.55 (0.51, 0.59)
Model (2) (‘Change-in-slope’)	1. MLE	F-test	0.98 (0.97, 0.99)	0.96 (0.95, 0.97)
		Better than GAM?	1.00 (0.99, 1.00)	0.99 (0.99, 1.00)
	2. Restricted search	F-test	0.98 (0.97, 0.99)	0.96 (0.95, 0.97)
		Better than GAM?	1.00 (0.99, 1.00)	0.99 (0.99, 1.00)

^a Type of threshold model fit.

^b Proportion of samples with a significant threshold that fit better than a 4-df smoothing spline GAM curve according to AIC..

1974). More detailed descriptions of these studies and their data collection methods have been reported elsewhere (Dawber et al., 1951; Central Patient Registry and Coordinating Center for the Lipid Research Clinics, 1974; Heiss et al., 1980; The Lipid Research Clinics Program Epidemiology Committee, 1979). For the purposes of this example, baseline values of systolic blood pressure (SBP) and BMI were analysed. We adjusted the relationship between BMI and SBP for current cigarette smoking status, age and use of blood pressure medication. The latter was available only for the LRC men. For the Framingham study, separate analyses were conducted for men and women. Only subjects for whom complete information was available were included in the analyses. There were 5209 subjects, of whom 4761 (91%) had complete information from the Framingham study, including 2132 women and 2629 men, and 2512 men from the LRC. The ‘no-effect’ threshold model (1) (considered in previous studies of the BMI-SBP association (Kaufman et al., 1997; Bunker et al., 1995; Doll et al., 2002; Stamler et al., 1986)), was fitted, as well as a linear model and a smooth GAM curve. SBP was used as the dependent variable and age and BMI were examined in separate analyses for possible thresholds using the two methods (section 2), each with the F-test.

In contrast to the populations studied by Bunker et al. (1995) and Kaufman et al. (1997), the Framingham and LRC populations were older, less lean and had higher blood pressure (Table 7). The LRC men were relatively similar to the Framingham men in terms of age and BMI, though they had slightly lower SBP. Also, lower interquartile ranges for BMI and SBP suggest a narrower range of values for the LRC men than the Framingham men (Table 7).

Table 7: Baseline characteristics of LRC and Framingham populations in comparison to those used by Kaufman and Bunker

Characteristic	Subpopulation	Framingham	LRC	Bunker ^a	Kaufman ^b
N	Men	2629	2512	5062	500
	Women	2132	0 6173	299	
Age (IQR ^c)	Men	44.1 (14.0)	46.6 (17.0)	43.3	42.3
(Years)	Women	44.1 (14.0)	–	41.4	38.6
BMI ^d (IQR)	Men	25.6 (5.3)	26.3 (4.2)	22.1	22.0
(kg/m^2)	Women	25.6 (5.2)	–	23.8	24.3
Current smoking	Men	41.1%	25.7%	–	–
(%)	Women	78.1%	–	–	–
DBP ^e (IQR)	Men	84.2 (16.0)	–	74.7	77.1
(mm Hg)	Women	86.3 (15.0)	–	73.9	72.1
SBP ^f (IQR)	Men	135.1 (27.0)	125.8 (21.0)	122.2	120.3
(mm Hg)	Women	135.6 (21.5)	–	120.5	112.5

^a (Bunker et al., 1995)^b (Kaufman et al., 1997)^c Numbers in brackets denote the interquartile range (IQR).^d Body Mass Index.^e Diastolic blood pressure.^f Systolic blood pressure.

The upper part of Table 8 presents the results for the association between BMI and SBP. The first section of the table shows that the standard linear model revealed a slight association between BMI and SBP in all three data sets (Framingham men and women, and LRC men) even if the slope was much lower for the LRC men. In the Framingham men, using the restricted search (section 3.2.2) we estimated a ‘no-effect’ model (1) threshold (at $21.8kg/m^2$), which significantly improved fit over the linear model. This threshold was located at the 21st percentile of the empirical distribution of BMI. However, comparing the threshold model with a smooth 4-df GAM curve indicated that the association may be non-linear, but also not threshold. In contrast, for the Framingham women and for the LRC men the ‘optimal’ ‘no-effect’ thresholds were estimated at 20.6 and 19.2, respectively, were definitely non-significant, and were located at the 7th and 1st percentile of the empirical distribution of BMI indicating that the association exists across almost the entire range of BMI values. Overall, it appears that no threshold exists for the Framingham women or the LRC population or that there are not enough subjects with low BMI to estimate it accurately. Using Method 1 (MLE) yielded almost identical results in each group (data not shown).

The lower half of Table 8 presents the results for the association between age and SBP. The GAM-based restricted search estimates of the ‘no-effect’ threshold, all statistically significant, were at 37 years for Framingham women, and at 36 and 45 years for Framingham and LRC men respectively. These correspond to

Table 8: Investigating the associations between BMI vs. SBP and age vs. SBP, separately in men and women

Method	Measure	Framingham Women ^{a, b}	Framingham Men ^{a,b}	LRC Men ^{a, b}
BMI				
Linear Model	Slope ^f	1.62	1.46	1.06
	95% CI	(1.40, 1.83)	(1.29, 1.63)	(0.84, 1.18)
	Threshold (p val ^d)	20.61 (0.19)	21.83 (<0.01)	19.20 (0.45)
	β ₂	1.68	1.62	1.02
	95% CI for β ₂	(1.45, 1.90)	(1.43, 1.80)	(0.85, 1.19)
Better than GAM? ^e		No		
AGE				
Linear Model	Slope	0.51	1.12	0.53
	95% CI	(0.42, 0.59)	(1.03, 1.22)	(0.48, 0.59)
	Threshold (p val)	37.00 (0.03)	36.00 (<0.01)	45.00 (<0.01)
	β ₂	0.59	1.25	0.83
	95% CI for β ₂	(0.49, 0.69)	(1.14, 1.35)	(0.76, 0.91)
Better than GAM?		Yes	Yes	Yes

^a When independent variable was BMI, these values are adjusted for age and smoking. When the independent variable was age, these values are adjusted for BMI and smoking.

^b When independent variable was BMI, these values are adjusted for age, smoking and BP medication. When the independent variable was age, these values are adjusted for BMI, smoking and BP medication.

^c A 'no-effect' threshold model (1) was fit using the GAM-based restricted search.

^d P-value refers to the p-value obtained using the F-test.

^e In the case of a significant threshold models ($p < 0.05$ for the F-test), we compared its AIC to that of a smooth 4-df GAM curve.

^f Estimated adjusted change in SBP associated with 1 unit increase in BMI.

the 25th, 20th and 50th percentile of the distribution of age. When compared to a smooth GAM curve, the threshold model fit better for all three groups, providing further support for the conjecture that the age-SBP association may start only in middle age. Similar results were found using the MLE approach (data not shown).

Figure 1 shows the estimated 4-df GAM curves and scatter plot for the associations between BMI and SBP (left-hand side), and age and SBP (right-hand side) for the three groups. Similar to other epidemiological studies that typically focus on a putative threshold in the effect of a single independent variable (Kaufman et al., 1997; Bunker et al., 1995), we estimated the smoothed effects of one variable at a time, with other variables modeled parametrically as linear or binary effects (e.g. $SBP = s(\text{Age}) + \text{BMI} + \text{other variables described above}$). The location of the estimated 'no-effect' threshold is indicated by the vertical line, suggesting that BMI or age, respectively, have no impact on SBP below the estimated threshold.

The strongest evidence in favour of a threshold is for the association between age and SBP. In all groups the threshold was statistically significant, and the GAM curve did not fit better than the threshold model. In sensitivity analyses, when the effects of both BMI and age were simultaneously modeled with smooth GAM curves, for all three subgroups the estimated curves were quite similar to those shown in the respective panels of Figure 1 (data not shown). This helped estimating the potential risk that the estimated curves for BMI-SBP relationships are not affected by residual confounding due to a failure to account for the non-linear effect of age on SBP, and vice versa (Benedetti and Abrahamowicz, 2004).

5 Discussion

In this simulation study, we investigated the statistical properties of two methods to estimate the threshold location and two criteria to assess if a threshold model fits better than a linear model. Both the ‘no-effect’ threshold model (1) and, a more general, ‘change-in-slope’ model (2) were considered. The primary objective was to find an approach with acceptably low type I error and reasonable power that would yield relatively unbiased and stable estimates of the threshold location.

We investigated the potential advantage of using generalized additive models in this context. Our expectation was that the GAM-based restricted search would reduce the risk of the estimated threshold falling very far from the ‘true’ threshold. In the unrestricted search, across the entire empirical range of the predictor values (X), the global minimum MSE may correspond to such an ‘outlying’ value, simply due to sampling error, especially in small to moderate samples. In contrast, the restricted search limits the potential threshold locations to those sub-intervals where the smooth GAM estimate of the dose-response function bends considerably, as indicated by the elevated second derivative. Because the GAM curve is modeled through smoothing splines, which explicitly penalize for increased (squared) curvature (Wahba, 1978), such bending will not occur unless there is a systematic change of the slope over a reasonably wide sub-interval. (This should considerably reduce the risk of unstable estimates, due to sampling error producing an ‘optimal’ MSE at a specific value of X , outside the interval where the ‘true’ curve actually does bend.) The *a posteriori* results of, at least some, our simulations corroborated, in the general ‘qualitative’ sense, the above *a priori* expectations, but quantitative gains were smaller than we expected. It is difficult to assess to what extent larger gains may be obtained with different sample sizes and/or different assumptions regarding the true location of the threshold and the extent to which the slope changes at that location.

Figure 1: Estimated 4-df GAM curves for the associations between BMI and SBP (left hand side), and AGE and SBP (right hand side) for Framingham women and men and LRC men. The vertical line indicates the location of the estimated 'no-effect' threshold (model (1)). We estimated the smoothed effects of one variable at a time, with other variables modeled parametrically as linear or binary effects (e.g. $SBP = s(Age) + BMI + \text{other variables}$ described in Section 4.)

Results were generally similar for the ML-based estimation and the GAM-based method. Still, the restricted GAM-based search helped reduce the risk of type I error, especially when AIC was used as the criterion to decide if the data support the threshold model (Table 1). Furthermore, the restricted search resulted in more accurate estimation of the threshold location, as reflected by a

lower MSE, especially when combined with ‘no selection’ strategy (Tables 3, 4 and 5). Further research, including a larger variation of the simulated scenarios, sample sizes, and underlying assumptions, will be necessary to establish to what extent, and in which situations, the restricted GAM-based search may be expected to improve the estimation of, and/or the inference about, thresholds. Nevertheless, the approach based on a restricted search of ‘suspicious’ points identified from the GAM fit resulted in estimated thresholds that were less biased, on average, and more stable.

In this paper, we mainly focused on the relatively simple scenario of deciding between a linear or null model and a threshold model. In reality, one may also have to consider the possibility of a nonlinear non-threshold association. Both approaches were likely to misclassify a smooth curve as a ‘no-effect’ threshold model (1), or a ‘change-in-slope’ threshold model (2) – though the restricted search performed somewhat better in this regard. These results still suggest that, except in very large data sets, it will likely remain difficult to distinguish between a threshold association and a J-shaped quadratic or other non-linear monotone association with gradually increasing second derivative. Results indicated that using AIC to compare the fit of the threshold model to a smooth GAM curve after identification of a statistically significant threshold could help a little here. However, *a priori* identification of the functional form of the association of interest is probably the best approach. In practice, the difference between the two types of associations may not be that important because similar conclusions will be reached based on e.g. a smooth J-shaped function and a no-effect threshold model (1). In both cases, for subjects with low exposure there is weak or no association between X and Y, while the association becomes evident for subjects with higher exposure values. On the other hand, it may be easier to interpret the results from a threshold model.

Identifying a threshold located at the extremes of the data remains a challenge. Power decreased considerably as the true threshold moved away from the mean. However, identification of thresholds at the extremes became much easier in large data sets (e.g. for sample size $N=1000$). Moreover, estimated thresholds were biased toward the mean of the data, and the amount of bias increased as the true threshold moved away from the mean, when all identified thresholds were considered (i.e. not just those that were statistically significant) with the restricted search. Among statistically significant thresholds, bias was more pronounced for thresholds near the lower end of the range of the independent variable. If the threshold was near the mean of the data, both methods performed well, estimated the threshold with a negligible amount of bias, and offered high power. The implications for design of studies aiming at estimation or testing of a threshold expected to be far from the mean of the data are crucial. In order to estimate thresholds at

the extremes, it is important to have many subjects with independent variable values both below and above the expected threshold location.

We investigated estimating thresholds from both the ‘no-effect’ model (1) and the ‘change-in-slope’ model (2). Model (2) is more general and includes model (1) as a special case. However, if the true association corresponds actually to the simpler model (1), the price to pay for using model (2) is lower power to detect the threshold, and more biased and variable estimated thresholds (Table 2-5).

We have investigated the existence of a threshold in the BMI-SBP association. In a previous study, by searching iteratively over a neighbourhood of suspicious points identified by plotting BP vs. sextiles of BMI, a threshold was estimated at $21.5\text{kg}/\text{m}^2$ (Bunker et al., 1995). The authors remarked that the threshold was more apparent in men (among whom about 50% had BMI less than $21.5\text{kg}/\text{m}^2$) than among women, (among whom around 30% had BMI less than $21.5\text{kg}/\text{m}^2$). This agrees with the results of our simulations, which showed the difficulty in identifying thresholds closer to the extremes of the data. However, our investigation did not provide strong support for a threshold in the BMI-SBP association.

Using the second derivative of the GAM estimate offered an interesting tool to screen for potential threshold locations, and formalised the visual method proposed by Cakmak et al. (1999), as well as the method used by Bunker et al. (1995). Inspection of plots of the GAM curves, and the corresponding second derivatives, suggests that points more than one standard deviation away from 0 seemed to include those that our eyes recognised as ‘suspicious’ (data not shown). Still, future research may consider alternative measures of local curvature.

Cakmak et al. (1999) also evaluated the ability of the criterion based on comparing the AIC (Akaike, 1973) between a bi-linear threshold model and a linear model, to detect threshold associations. Estimated power was lowest when the true threshold was at the 10th percentile and highest when it was at the 75th percentile. Cakmak et al. (1999) also concluded, on the basis of six loess models fit to threshold data (varying the position of the threshold), that the use of non-parametric smoothers to identify threshold associations is feasible and acceptably accurate. Our simulations provide quantifiable, systematic evidence for this conclusion. On the other hand, using a comparison of AIC between the threshold model and a linear model, as Cakmak et al. (1999) did, resulted in type I error that is definitely too high.

The consequences of incorrectly modelling threshold associations as linear include an overestimation of the predictor effect for the group of subjects with an independent variable value below the threshold, and an underestimation in the group of subjects with values above the threshold. Moreover, it can result in a crucial misrepresentation of the nature of the association which could have broad implications for policy makers and/or for etiologic research. Yet, declaring a spu-

rious threshold may also lead to incorrect conclusions and ill-advised public health policies. This calls for an accurate, efficient and reliable method to both locate the potential threshold and assess to what extent the data support the existence of a threshold.

We compared several methods to estimate and detect thresholds in small and moderate samples. By using either the maximum likelihood estimation, or the restricted GAM-based search, in combination with the F-test we obtained similar power. However, the restricted iterative search with the F-test yielded more accurate threshold estimates. The performance of these methods when the dependent variable is dichotomous, or represents a time-to-event, rather than continuous, should be explored in the future. We believe that the present analyses will enable researchers to search for, and estimate, thresholds more effectively, and will give them better insight into the precision and bias of those estimates, as well as into the related methodological challenges.

References

- Abrahamowicz, M., Ciampi, A., 1991. Information theoretic criteria in non-parametric density estimation: bias and variance in the infinite dimensional case. *Computational Statistics and Data Analysis* 12, 239–247.
- Abrahamowicz, M., MacKenzie, T., Esdaile, J., 1996. Time-dependent hazard ratio: modeling and hypothesis testing with application in lupus nephritis. *Journal of the American Statistical Association* 91, 1432–1439.
- Abrahamowicz, M., Schopflocher, T., Leffondre, K., du Berger, R., Krewski, D., 2003. Flexible modeling of exposure-response relationship between long-term average levels of particulate air pollution and mortality in the american cancer society studyn. *Journal of Toxicology and Environmental Health A* 66 (16-19), 1625–1654.
- Akaike, H., 1973. in *Second International Symposium on Information Theory*. Akademia Kiado, Budapest.
- Benedetti, A., Abrahamowicz, M., 2004. Using generalized additive models to reduce residual confounding. *Statistics in Medicine* 23 (24), 3781–3801.
- Bunker, C., Ukoli, F., Matthews, K., Kriska, A., Huston, S., Kuller, L., 1995. Weight threshold and blood pressure in a lean black population. *Hypertension* 26 (4), 616–623.

- Cakmak, S., Burnett, R. T., Krewski, D., 1999. Methods for detecting and estimating population threshold concentrations for air pollution-related mortality with exposure measurement error. *Risk Analysis* 19 (3), 487–496.
- Central Patient Registry and Coordinating Center for the Lipid Research Clinics, 1974. Reference Manual for the Lipid Research Clinics Prevalence Study. Vol. 1 and 2. Department of Biostatistics, University of North Carolina, Chapel Hill, NC.
- Chu, K. C., Baker, S. G., Tarone, R. E., 1999. A method for identifying abrupt changes in u.s. cancer mortality trends. *Cancer* 86 (1), 157–169.
- Daniels, M. J., Dominici, F., Samet, J. M., Zeger, S. L., 2000. Estimating particulate matter-mortality dose-response curves and threshold levels: an analysis of daily time-series for the 20 largest us cities [see comments]. *American Journal of Epidemiology* 152 (5), 397–406.
- Dawber, T. R., Meadors, G. F., Moore, F. E. J., 1951. Epidemiological approaches to heart disease: the Framingham Study. *American Journal of Public Health* 41, 279–286.
- Doll, S., Paccaud, F., Bovet, P., Burnier, M., Wietlisbach, V., 2002. Body mass index, abdominal adiposity and blood pressure: consistency of their association across developing countries and developed countries. *International Journal of Obesity* 26, 48–57.
- Hajat, S., Haines, A., Atkinson, R. W., Bremner, S. A., Anderson, H. R., Emberlin, J., 2001. Association between air pollution and daily consultations with general practitioners for allergic rhinitis in london, united kingdom. *American Journal of Epidemiology* 153 (7), 704–714.
- Hastie, T., Tibshirani, R., 1986. Generalized additive models. *Statistical Science* 1 (3), 297–310.
- Hastie, T., Tibshirani, R., 1990. Generalized Additive Models. Monographs on Statistics and Applied Probability. Chapman and Hall, London.
- Heiss, G., Tamir, I., Davis, C., Tyroler, H., Rifkind, B., Schonfeld, G., et al., 1980. Lipoprotein-cholesterol distributions in selected north american populations: the lipid research clinics program prevalence study. *Circulation* 61, 302–315.
- Hoel, D. G., Li, P., 1998. Threshold models in radiation carcinogenesis [see comments]. *Health Physics* 75 (3), 241–250.

- Hoel, D. G., Portier, C. J., 1994. Nonlinearity of dose-response functions for carcinogenicity. *Environmental Health Perspectives Supplement* 102 Supplement, 109–113.
- Hudson, D. J., 1966. Fitting segmented curves whose join points have to be estimated. *Journal of the American Statistical Association* 61 (316), 1097–1129.
- Kaufman, J. S., Asuzu, M. C., Mufunda, J., Forrester, T., Wilks, R., Luke, A., Long, A. E., Cooper, R. S., 1997. Relationship between blood pressure and body mass index in lean populations. *Hypertension* 30 (6), 1511–1516.
- Kuchenhoff, H., 1997. An exact algorithm for estimating breakpoints in segmented generalized linear models. *Computational Statistics* 12, 235–247.
- Kuchenhoff, H., Ulm, K., 1997. Comparison of statistical methods for assessing threshold limiting values in occupational epidemiology. *Computational Statistics* 12, 249–264.
- Mossman, K. L., 1998. The linear no-threshold debate: where do we go from here? *Medical Physics* 25 (3), 279–284.
- Nussbaum, R. H., 1998. The linear no-threshold dose-effect relation: is it relevant to radiation protection regulation? *Medical Physics* 25 (3), 291–299.
- Pastor, R., Guallar, E., 1998. Use of two-segmented logistic regression to estimate change-points in epidemiologic studies [see comments]. *American Journal of Epidemiology* 148 (7), 631–642.
- Queiroz-Neto, A., Zamur, G., Carregaro, A. B., Mataqueiro, M. I., Salvadori, M. C., Azevedo, C. P., Harkins, J. D., Tobin, T., 2001. Effects of caffeine on locomotor activity of horses: determination of the no-effect threshold. *Journal of Applied Toxicology* 21 (3), 229–234.
- R Development Core Team, 2008. R: A language and environment for statistical computing. Version 2.5.0.
- Roland, J., Keyghobadi, N., Fownes, S., 2000. Alpine parnassius butterfly dispersal: effects of landscape and population size. *Ecology* 81 (6), 1642–1653.
- Schwartz, J., 1994. Nonparametric smoothing in the analysis of air pollution and respiratory disease. *Canadian journal of statistics* 22, 471–487.

- Sinclair, W. K., 1998. The linear no-threshold response: why not linearity? *Medical Physics* 25 (3), 285–290.
- Stamler, J., Wentworth, D., Neaton, J., 1986. Is relationship between serum cholesterol and risk of premature death from coronary heart disease continuous and graded? *Journal of the American Medical Association* 256, 2823–2828.
- The Lipid Research Clinics Program Epidemiology Committee, 1979. Plasma lipid distributions in selected north american populations: the lipid research clinics prevalence study. *Circulation* 60, 427–439.
- Ulm, K., 1991. A statistical method for assessing a threshold in epidemiological studies. *Statistics in Medicine* 10, 341–349.
- Vieth, E., 1989. Fitting piecewise linear regression functions to biological responses. *Journal of Applied Physiology* 67 (1), 390–396.
- Wahba, G., 1978. Improper priors, spline smoothing and the problem of guarding against model errors in regression. *Journal of the Royal Statistical Society, Series B, Methodological* 40, 364–372.
- Webster, E. W., 1998. The linear no-threshold debate: a summary. *Medical Physics* 25 (3), 300–300.