



## Detection and Estimation of J-Shaped Risk-Response Relationship

Els J. T. Goetghebeur; Stuart J. Pocock

*Journal of the Royal Statistical Society. Series A (Statistics in Society)*, Vol. 158, No. 1. (1995), pp. 107-121.

Stable URL:

<http://links.jstor.org/sici?sici=0964-1998%281995%29158%3A1%3C107%3ADAE0JR%3E2.0.CO%3B2-3>

*Journal of the Royal Statistical Society. Series A (Statistics in Society)* is currently published by Royal Statistical Society.

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/rss.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

---

The JSTOR Archive is a trusted digital repository providing for long-term preservation and access to leading academic journals and scholarly literature from around the world. The Archive is supported by libraries, scholarly societies, publishers, and foundations. It is an initiative of JSTOR, a not-for-profit organization with a mission to help the scholarly community take advantage of advances in technology. For more information regarding JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

## Detection and Estimation of J-shaped Risk–Response Relationships

By ELS J. T. GOETGHEBEUR† and STUART J. POCOCK

*London School of Hygiene and Tropical Medicine, UK*

[Received January 1993. Revised January 1994]

### SUMMARY

Current statistical approaches for analysing potentially J-shaped relationships between a risk factor and disease outcome can be seriously misleading. For instance, a simple quadratic model is widely used but can substantially exaggerate the statistical evidence for an upturn to the left. Instead, a family of double-quadratic models is proposed in which the relationship between risk factor and disease outcome is represented by two independent quadratic curves (one to the left and one to the right) joined at a low point to be estimated. Asymptotic results are derived for a semiparametric approach that can use standard software to assess the strength of evidence for the existence of a J-shape and estimate the location of the turning point. Alternatively, the minimum  $p$ -value of a sequence of trend tests on subsets of data increasing from the left yields a simple but anticonservative initial screen of the evidence for a linear or quadratic upturn. We indicate how this naïve minimum  $p$ -value can be corrected to a conservative level. For most practical situations, the clear demonstration of a J-shaped relationship needs a much larger amount of data than is generally appreciated. The approaches proposed are illustrated with data on diastolic blood pressure and the risk of coronary death.

**Keywords:** BLOOD PRESSURE AND RISK; CHANGEPOINT; DOUBLE-QUADRATIC MODEL; J-SHAPED RELATIONSHIP; SEQUENTIAL TESTS FOR TREND; SURVIVAL ANALYSIS

### 1. INTRODUCTION

The possibility of J-shaped relationships between risk factors and disease outcomes has become increasingly recognized in medical references. It arises if, for example, a well-established risk factor is suspected of becoming inversely related to risk below a certain point, whereas the more widely accepted positive risk association exists across most of the observed risk factor distribution. Examples include diastolic blood pressure and coronary heart disease (Farnett *et al.*, 1991), alcohol and mortality (Marmot *et al.*, 1981) and cholesterol and mortality (Frank *et al.*, 1992). Birth weight and infant mortality (Wilcox and Russell, 1983) or environmental temperature and daily mortality (Polychronaki *et al.*, 1982) are examples of a mirror image of the J-shape: there is a predominant decreasing risk associated with increasing covariate values and a less pronounced upturn in risk to the right of the risk factor distribution.

Recognition of a J-shape and the subsequent location of a turning point can have major clinical or public health consequences. For instance, is it dangerous to lower blood pressure below a certain level? The statistical methodology presented here has three main objectives:

- (a) to test for a J-shaped relationship;

†Address for correspondence: Biostatistics Department, Limburgs Universitair Centrum, Universitaire Campus, B-3590 Diepenbeek, Belgium.  
E-mail: egoetghe@luc.ac.be

- (b) if the relationship exists, to estimate both the turning point and the shape of the association;
- (c) to display the results in an informative and unbiased manner.

The methods considered can relate either to fully parametric generalized linear models (e.g. a logistic model for a binary response) or to semiparametric models (e.g. the proportional hazards model for time to event). The latter are often more appropriate for medical event data, and we examine them in detail.

Our aim is to provide statistical approaches that can be applied to any data with a possible J-shaped risk. Current approaches, such as fitting a simple quadratic function, are potentially misleading as shown in Section 2. In Section 3, an alternative is proposed based on a family of double-quadratic models. It leads to a direct test for a J-shape and also allows for estimation of the turning point (with its confidence bounds). A simple but anticonservative sequential testing approach is defined in Section 4, while Section 5 describes a more comprehensive analysis for the estimation of the risk function. Section 6 gives an illustrative example on diastolic blood pressure and risk of coronary death and practical consequences are discussed in Section 7.

## 2. CURRENT APPROACHES

An interesting example is from the Swedish primary prevention trial, which studied the relationship between treated diastolic blood pressure and coronary morbidity in 686 treated hypertensive subjects followed for 12 years (Samuelsson *et al.*, 1990). Other relevant examples include D'Agostino *et al.* (1990), Farnett *et al.* (1991) and Frank *et al.* (1992). When investigating a potentially J-shaped relationship, a first descriptive step is to split the  $x$ -axis (blood pressure) into intervals and to plot the observed event rates within each interval, adjusted for other factors (e.g. age) if appropriate. In determining the cut points, data snooping (*post hoc* selection based on the observed data) to enforce an artificial J-shape should be avoided while also ensuring that intervals are not so wide that they obscure a real upturn to the left. Fig. 1(a) shows the unadjusted incidence rates of coronary disease presented in the Swedish trial. Within each blood pressure category, the estimated variance of the log-hazard rate is inversely proportional to the number of events in that category. Drawing confidence intervals to convey the degree of uncertainty in the estimates or indicating the number of events adds valuable information to such plots.

Assuming that there is some suggestion of a J-shape at this stage, we investigate further. In most medical examples, there are many more data to the right of the observed low point (nadir) than to the left; the first two categories in Fig. 1(a) have seven and nine events, which suggests that there is limited statistical power to detect a J-shape. Although data plots of this kind are a useful descriptive tool, they do not provide a firm basis for formal statistical inference. There are also potential problems in the arbitrariness of chosen intervals and plotting positions of extreme groups. Categorizing can waste information, and so to enhance power we use the full risk factor information. A quadratic function, the obvious candidate that allows for an upturn, has become a popular approach for modelling J-shapes. However, it has problems, as follows.

Let  $x_i$  denote the risk factor value of interest, i.e. diastolic blood pressure, and let the hazard rate at time  $t$  for the  $i$ th individual be

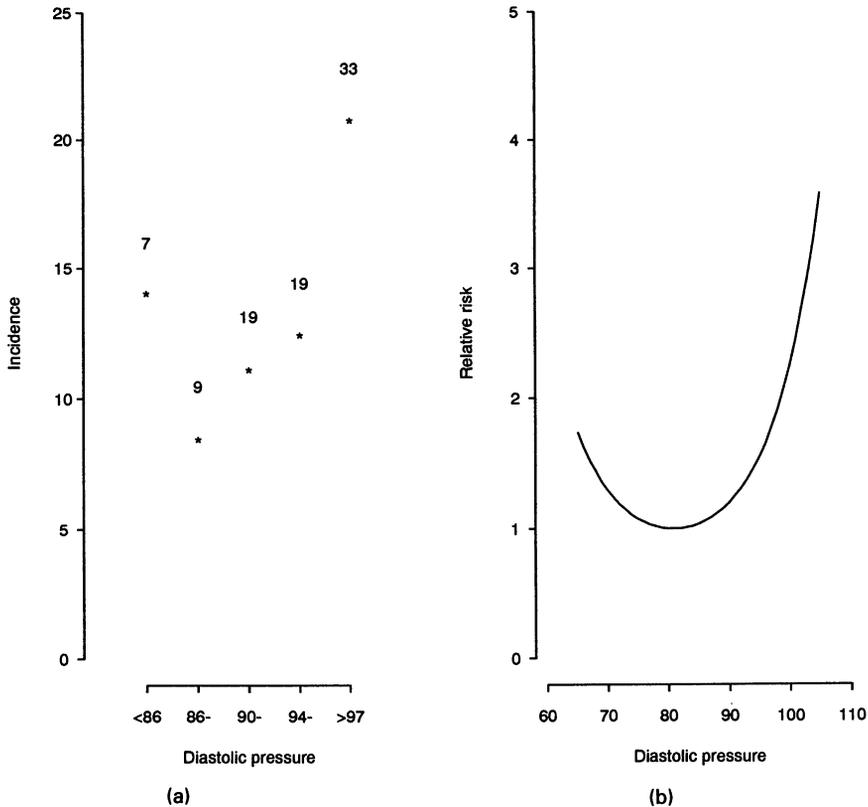


Fig. 1. Treated diastolic blood pressure and evidence of coronary heart disease in 12 years follow-up in the primary prevention trial (87 cases in 686 subjects followed): (a) observed percentage 12-year incidence of coronary heart disease against mean in-study diastolic blood pressure; (b) fitted relative hazards from the simple quadratic model against mean in-study diastolic blood pressure

$$\lambda_i(t) = \lambda_0(t) \exp(\beta_1 x_i + \beta_2 x_i^2 + \beta^{*T} \mathbf{Z}_i^*)$$

where  $\mathbf{Z}_i^*$  is the vector of all other covariates. Hence, the log-relative-hazard is a quadratic function of  $x_i$ . This model was fitted by Samuelsson *et al.* (1990) for the Swedish data, with  $x_i$  representing diastolic blood pressure during treatment and  $\mathbf{Z}_i^*$  comprising base-line blood pressure, serum cholesterol, smoking habits and age. The estimated function is shown in Fig. 1(b). However, the test of  $H_0: \beta_2 = 0$  versus  $H_1: \beta_2 > 0$  as evidence for the existence of an upturn is potentially misleading. In the Swedish trial this test has a  $p$ -value of 0.002, correctly indicating that inclusion of the quadratic term in the model provides a significantly better fit to the data than the linear term alone. However, the nadir of the fitted curve is  $-\hat{\beta}_1/2\hat{\beta}_2 = 81$  mmHg, to the extreme left of the observed blood pressure data, so that the estimated quadratic may predominantly reflect non-linearity in the positive association to the right rather than specifically an upturn to the left. As also shown by Roy and Potthoff (1958), Fieller's theorem would allow for calculation of an approximate confidence interval for the nadir by using the estimated covariance

matrix of  $\hat{\beta}_1$  and  $\hat{\beta}_2$ . In this example, it is likely that the confidence interval goes well below the observed blood pressure range, as only two events had been observed with a diastolic blood pressure below 81 mmHg.

Generally, the  $p$ -value for  $\beta_2$  in a simple quadratic regression model is not a reliable measure of evidence of an upturn *to the left*. Also, the location of the estimated nadir and the length of the confidence interval around it depend on the precision of the two estimated shape parameters and can easily be driven by data to the right of the estimated nadir. To avoid confounding between evidence for left and right upturn we model both sides explicitly by independent parameters, as considered next.

### 3. MODEL WITH NON-FIXED CHANGEPOINT

For simplicity, we focus on a single risk factor but adjustment for other risk factors can be made in the usual way. Consider a proportional hazards model from a 'double-quadratic family', i.e. two (different) quadratic arms meeting at a change-point  $\eta$  as follows:

$$\begin{aligned} \lambda_i(t) &= \lambda_0(t) \times \begin{cases} \exp\{\gamma_1(x_i - \eta)^2\} & \text{if } x_i \leq \eta, \\ \exp\{\gamma_2(x_i - \eta)^2\} & \text{if } x_i > \eta \end{cases} \\ &= \lambda_0(t) \exp\{\gamma_1(x_i - \eta)^2 I(x_i \leq \eta) + \gamma_2(x_i - \eta)^2 I(x_i > \eta)\}, \end{aligned} \quad (1)$$

where  $I$  is the indicator function. So the log-relative-hazard is a function of three unknowns: two quadratic coefficients  $\gamma_1$  and  $\gamma_2$ , and the turning point (or nadir)  $\eta$ . Let the true underlying parameter values be  $(\gamma_1^0, \gamma_2^0, \eta^0)$ . Note that for any fixed cut-off point  $\eta$  the magnitude of the two quadratic coefficients  $\gamma_1$  and  $\gamma_2$  can be different, which is reasonable, given that

- (a) we are not even sure that an upturn in risk to the left exists and
- (b) it is desirable to have separate estimates of the strength of evidence for left and right upturns.

An equivalent model could be used with binary event data and logistic models or Poisson counts and log-linear models: indeed any generalized linear model.

The simple null hypothesis  $H_0: \gamma_1^0 = 0$  versus  $H_1: \gamma_1^0 > 0$ , with  $\gamma_2$  left undefined, tests for an upturn to the left. In the blood pressure example, as long as  $\gamma_1^0$  is not positive, no harm is done by lowering blood pressure below  $\eta^0$ . If  $\gamma_1^0 > 0$ , the parameter  $\eta$  directly represents the nadir that we are interested in. The option to model a piecewise quadratic function rather than a piecewise linear function is motivated by the smoother change in log-relative-hazard rate at  $\eta$  which is biologically more plausible, whereas both cover the same null model to the left of the nadir. Other options are considered in the discussion.

We caution against naïve use of partial likelihood results to make inference about  $\gamma_1$  and  $\eta$ . The log-hazard rate is a linear expression in  $\gamma_1$  and  $\gamma_2$ , but  $\eta$  enters the likelihood in a non-linear (non-smooth) way. Only in the theoretical situation with  $\eta^0$  known, and by defining the covariates

$$\begin{aligned} z_{i1} &= (x_i - \eta^0)^2 I(x_i \leq \eta^0), \\ z_{i2} &= (x_i - \eta^0)^2 I(x_i > \eta^0) \end{aligned} \quad (2)$$

for each individual, is the problem reduced to a conventional proportional hazards model. Testing and estimation of  $\gamma_1$  is then straightforward with standard software. In practice  $\eta^0$  is unknown. Alternative analysis strategies are explored in the next two sections.

#### 4. SEQUENTIAL QUADRATIC TESTS

Consider repeated significance tests for a downward trend in risk on increasingly large subsets of the data, where each test is confined to subjects whose risk factor  $x_i$  is below a certain cut-off point. The subset of data grows as the cut-off point is moved to the right and the smallest  $p$ -value of the sequence is examined. This has some analogy with sequential analyses in clinical trials (Geller and Pocock, 1987; Tsiatis, 1982), but moves from left to right on the risk factor axis, rather than on the time axis. It also relates to work by Davies (1977, 1987) as discussed in Section 7.1.

Let the parameter  $\eta$  move from the lowest meaningful  $x$ , which we call  $\eta_l$ , to a value  $\eta_u$ , beyond the point where its true value  $\eta^0$  might reasonably be sought. For each  $\eta$ , restrict the data set to the observations  $i$  for which  $x_i \leq \eta$ . Then fit the model  $\lambda_i(t) = \lambda_0(t) \exp\{\gamma_1(x_i - \eta)\}^2$  to the reduced data set and record the naïve standardized normal deviate for  $\hat{\gamma}_1$ , say  $Z_\eta$ .

Let  $Z_{\max} = \max\{Z_\eta, \eta_l < \eta < \eta_u\}$ . Under model (1),  $Z_{\eta^0} \leq Z_{\max}$ , where  $Z_{\eta^0}$  results from a fit using the known changepoint as explained at equations (2). Hence, compared with the correctly informed analysis the evidence for an upturn is overestimated if the naïve  $N(0, 1)$   $p$ -value associated with  $Z_{\max}$  is used. With such an anticonservative test, lack of statistical significance can be trusted as insufficient evidence to confirm a J-shape. However, a small naïve  $p$ -value for  $Z_{\max}$  should first be interpreted as a marker for more detailed analysis, rather than confirmation of a J-shape. Also, insufficient evidence in favour of a J-shape does not constitute evidence against it unless the study was sufficiently large (see Section 7). In practice,  $Z_\eta$  may only be computed for a limited number of equally spaced values (e.g. 2 mmHg intervals of diastolic blood pressure). Using martingale arguments we can obtain consistent estimators of the covariance between  $Z_\eta$ s for different values of  $\eta$  and derive an adjusted conservative  $p$ -value for the group sequential test. A description of this method is available on request. Unfortunately, the work involved is substantial and compared with the method of Section 5 it gives lower power and no consequent estimators. Hence we consider this naïve sequential testing as a simple preliminary technique to see whether further analysis is warranted.

#### 5. FULL ANALYSIS

The changepoint problem in model (1) has been extensively studied in normal error models. It exemplifies a segmented polynomial model within the broader framework of multiphase regression models (Seber and Wild, 1990). When the errors are independently normal with constant variance, proofs of asymptotic normality of the least squares estimates have been developed with difficulty. Gallant (1971) was the first to prove asymptotic normality of the maximum likelihood estimates under fairly tolerant conditions: continuous *first-order* derivatives of the

likelihood (which includes the double-quadratic model) and some additional restrictions on the covariate distribution. He also studied the power of the likelihood ratio test of location in this setting (Gallant, 1975).

As far as we know, none of the equivalent properties has been proved for generalized linear models or proportional hazards models. Nevertheless they have been used in practice (see for instance Rigby and Stasinopoulos (1991) and Stasinopoulos and Rigby (1992a)) and GLIM macros to fit such models are available (Stasinopoulos and Rigby, 1992b). A proof of asymptotic normality of the partial likelihood estimates for model (1) under the condition that no subject's risk factor level is too close to the true cut-off point is given in Appendix A. Under such assumptions, we now derive an asymptotic confidence interval for  $\eta$  and an asymptotic  $p$ -value for  $\gamma_1$  by using standard software.

First, the maximum partial likelihood estimates are obtained from the profile likelihood as follows. A range of possible values  $\eta$  for the nadir is scanned. (For instance with diastolic blood pressure we might usefully start with values 5 mmHg apart.) For each chosen  $\eta$ , we proceed as if it were the true known nadir and fit the full model (1) as in Section 3 to obtain the maximum log-likelihood value at that point,  $l(\eta) = l(\hat{\gamma}_1(\eta), \hat{\gamma}_2(\eta), \eta)$ , along with the estimates  $\hat{\gamma}_1(\eta)$  and  $\hat{\gamma}_2(\eta)$ . By a search across the values of  $\eta$ , we obtain the maximum of this profile log-partial-likelihood at  $\eta = \hat{\eta}$  with corresponding estimates  $\hat{\gamma}_1(\hat{\eta})$  and  $\hat{\gamma}_2(\hat{\eta})$ . We denote this global maximizing value by  $(\hat{\gamma}_1, \hat{\gamma}_2, \hat{\eta})$ .

An asymptotically correct  $p$ -value for  $\gamma_1$  and confidence interval for  $\eta$  can be obtained from likelihood ratio statistics as follows. Construct the profile log-partial-likelihood for model (1) subject to  $\gamma_1 \equiv 0$ . Let the maximum under this reduced model be reached at  $\tilde{\gamma}_2, \tilde{\eta}$  and denoted by  $l(0, \tilde{\gamma}_2, \tilde{\eta})$ . Then, under  $H_0: \gamma_1 = 0$ ,  $2l(\hat{\gamma}_1, \hat{\gamma}_2, \hat{\eta}) - 2l(0, \tilde{\gamma}_2, \tilde{\eta})$  is asymptotically distributed as  $\chi_1^2$ .

The  $100(1 - \alpha)\%$  confidence interval for  $\eta$  contains all  $\eta$ -values for which the hypothesis  $H_0: \eta^0 = \eta$  is not rejected by the likelihood ratio test and is obtained as follows. Plot the profile likelihood for  $\eta$  under the full unrestricted model (1). Next, subtract  $\chi_\alpha^2$ , the  $\chi^2$ -deviate associated with probability  $\alpha$ , from  $2l(\hat{\gamma}_1, \hat{\gamma}_2, \hat{\eta})$ . The confidence interval consists of all  $\eta$ -values for which  $2l(\hat{\gamma}_1(\eta), \hat{\gamma}_2(\eta), \eta)$  lies between the maximum,  $2l(\hat{\gamma}_1, \hat{\gamma}_2, \hat{\eta})$ , and  $2l(\hat{\gamma}_1, \hat{\gamma}_2, \hat{\eta}) - \chi_\alpha^2$ . In the examples that we have analysed this procedure yielded a simple connected interval.

From the assumptions explained in Appendix A, there is concern that the asymptotic results might not apply very quickly in finite sample data sets or with general covariate distributions. Because of the extensive computing time required we have not yet explored this issue with a large scale simulation study but have done some checks of plausibility using simulated data in S-PLUS as follows.

We generated 1000 blood pressure data points from an  $N(83, 12^2)$  distribution. This approximates to the blood pressure distribution in a typical British middle-aged male population. We supposed that the true probability of an event followed a logistic model with intercept 0 and double-quadratic arms meeting at changepoint  $\eta^0 = 75$ . For efficiency of computing, the event rate at  $\eta = \eta^0$  was set rather high, i.e. at probability  $\frac{1}{2}$ . That is, for comparable power in a realistic blood pressure setting we would need more subjects, but with a lower event probability. Next, we generated 2000 data sets each of size 1000 subjects from the null model:  $\gamma_1^0 = 0$  and  $\gamma_2^0 = 0.003$ .

The analysis of each such simulated data set scanned discrete changepoints  $\eta$

equal to 65, 70, 75, 80 and 85 and derived likelihood ratio statistics on  $\gamma_1$  as described above. Owing to our computational restrictions another level of approximation to the method was introduced through the scanning of relatively few potential changepoints. On our simulated data sets, the analyses appeared to be slightly anticonservative: the  $\chi_1^2$ -statistic for an upturn exceeded the 5% point 3.84 in 6.26% of simulations (95% confidence interval 5.2–7.3%).

Simulations under the alternative situation  $\gamma_1 = \gamma_2 = 0.003$  found a significant quadratic effect to the left in 63.3% of the 400 simulations. This compared well with a power of 63.8% (255/400) of the informed analysis (i.e. knowing  $\eta^0 = 75$ ), but the scanning of just five potential changepoints would enhance comparability here.

The corresponding estimator of  $\eta$  was centred around 75 with a much smaller variance than for the null situation, as expected. For this truncated estimation procedure the likelihood ratio confidence interval for  $\eta$  had 75 as lower or upper limit in 6.5% (26/400) of the cases (close to the desired 5%). No higher lower limit or lower upper limit was obtained.

## 6. EXAMPLE: FRAMINGHAM DATA

In this section we apply the proposed methods to data from the Framingham study where the issue of a possible J-shaped relationship between diastolic blood pressure and risk of coronary death was of interest.

On the data tape available to us 5209 people were examined at 2-year intervals over 30 years of follow-up. D'Agostino *et al.* (1991) provide a detailed study of the same question on 34-year follow-up data. Their interesting findings stimulated us to apply our methods to the Framingham study data. One method of analysis is to relate each measured diastolic blood pressure to the occurrence of coronary death within the next 2 years for the 15 episodes recorded. D'Agostino *et al.* (1990) have argued why this 'person-years' approach is close to the time-dependent covariate Cox regression analysis. We examine the relationship between the risk of dying from coronary heart disease within 2 years and the diastolic blood pressure at the start of those 2 years for patients aged between 45 and 84 years who had a previous myocardial infarction and no previous congestive heart failure. The models used here adjust for age only since the covariates cholesterol, smoking status and diabetic status showed no significant association with risk in this subgroup. There are 1604 person examinations and 104 coronary deaths for this analysis.

Fig. 2 shows the observed proportion of events in blood pressure categories of 10 mmHg, with the number of events indicated. We need the confidence intervals to realize that the evidence for a J-shape is limited: the point estimates have a great level of uncertainty.

Using logistic regression, we first examine the simple quadratic fit (Section 2), which yields the results

$$\begin{aligned}\hat{\beta}_1 &= -0.13, & SE &= 0.057, & p &= 0.023, \\ \hat{\beta}_2 &= 0.00081, & SE &= 0.00032, & p &= 0.011,\end{aligned}$$

with nadir  $-\hat{\beta}_1/2\hat{\beta}_2 = 80.1$  mmHg. On the basis of Fieller's theorem the calculated confidence interval for the nadir is very wide: 46.6–89.0 mmHg. However, note our previous caution against the interpretation of the significance of the quadratic

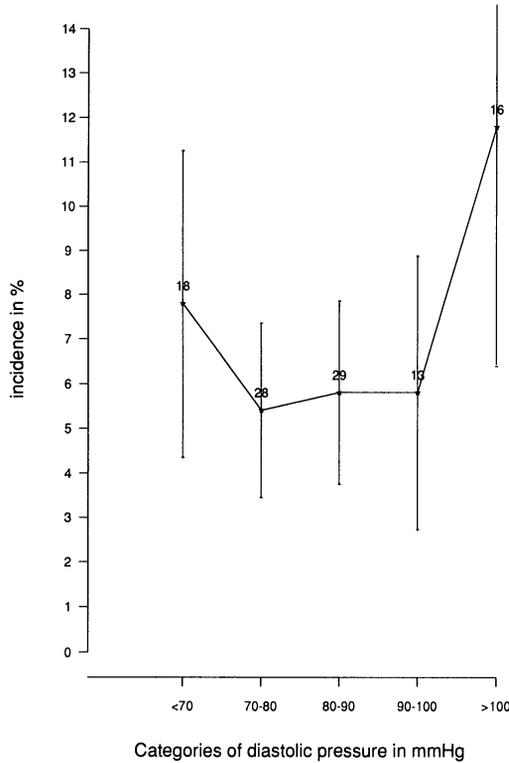


Fig. 2. Diastolic blood pressure in categories at the start of a 2-year period and proportion of cardiovascular deaths in the next 2 years in a subgroup of Framingham study subjects with a previous myocardial infarction: \*, observed incidence rates within blood pressure categories and confidence intervals (the numbers of coronary deaths are given alongside)

coefficient  $\hat{\beta}_2$  as evidence of an upturn to the left. Instead, let us explore the proposed double-quadratic model (Sections 3-5) for these same data.

The result of performing sequential quadratic tests to the left (Section 4) are displayed in Table 1. The minimum  $p$ -value obtained equals 0.064. Even this anti-conservative test shows only marginally significant evidence of an upturn to the left.

However, for completeness of presentation here, the statistical evidence for a J-shape is assessed by fitting the full model. The profile partial likelihood for model (1) is shown in Fig. 3, from which we obtain the estimate of the cut-off point  $\hat{\eta} = 79$  with maximized 2 times the log-likelihood equal to  $-761.43$ . The maximum likelihood analysis (Section 3) yields estimates

TABLE 1  
Sequential quadratic tests

Cut-off point	70	75	80	85	86	87	88	89	90	95	100
$p$ -value for $\gamma_1$	0.96	0.81	0.31	0.11	0.083	0.064	0.21	0.28	0.26	0.26	0.54
$10000\hat{\gamma}_1$	1.7	3.8	9.2	9.4	9.5	9.4	6.0	4.9	4.7	3.5	1.5

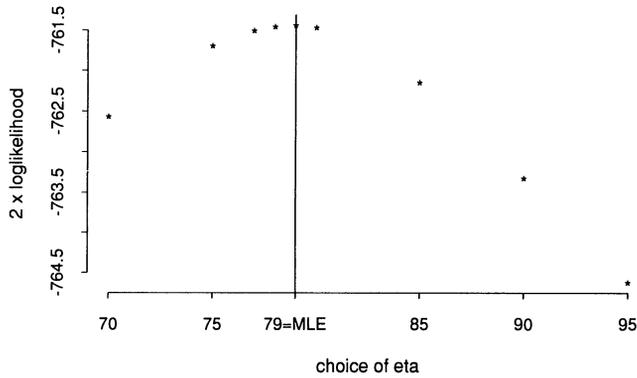


Fig. 3. Twice the profile partial likelihood for the nadir in diastolic blood pressure ('eta') for coronary heart disease risk in a subgroup of Framingham study subjects with a previous myocardial infarction

$$\hat{\gamma}_1 = 0.00103, \quad SE = 0.00091, \quad p = 0.26,$$

$$\hat{\gamma}_2 = 0.00077, \quad SE = 0.00025, \quad p = 0.002.$$

The adjusted  $p$ -value for  $\gamma_1$  from the likelihood ratio test described in Section 5 gives  $\chi_1^2 = 1.13$  with  $p = 0.29$ , i.e. insufficient evidence of an upturn to the left even though the point estimate represents an observed substantial upturn in that range. The significance of the quadratic term  $\beta_2$  appears to contradict this, but perhaps it is too dominated by the rise at the right-hand side. This is also reflected in the small changes in the profile likelihood when moving  $\eta$  below  $\hat{\eta}$ . No useful confidence interval for  $\eta$  can be obtained, essentially because there is insufficient evidence to support the J-shaped relationship. It would be interesting to explore whether the application of our methods to the more-extensive 34-year follow-up data in D'Agostino *et al.* (1991) would confirm their conclusions.

## 7. DISCUSSION

We have demonstrated that the statistical problem of detecting and locating a J-shape usually requires more specific techniques than descriptive plots or fitting a simple quadratic association. In particular, the simple quadratic fit is unable to disentangle evidence of an upturn to the left from evidence of a curvilinear (quadratic) upturn to the right. We have provided a double-quadratic model which enables us to address directly the real question: is there evidence of an upturn to the left? Although small sample properties may warrant further investigation, the model helps to clarify what evidence of a J-shape really exists in any given data set. Before summarizing the practical conclusions, some methodological issues are discussed next.

### 7.1. Methodological Issues

The methods proposed in this paper are (semi)parametric and built on quadratic assumptions. In many applications this will be appropriate and especially so when

a preliminary simple quadratic fit turned out to be significant. In other circumstances we may be looking for a linear or perhaps nonparametric upturn to the left. Both ideas can be explored.

A *linear trend from the left* can be investigated through either a sequential linear trend test or a linear-quadratic model. The former is a linear version of the (group) sequential test approach in Section 4 and gains some appeal because we then have a true submodel for each  $\eta < \eta^0$ , i.e. in each step with  $\eta$  below  $\eta^0$  we have an unbiased estimate of the fixed but unknown linear slope. However, for  $\eta^0$  unknown it remains unclear how this approach can derive an optimal overall estimate of any linear downward trend from the left.

A related problem of testing for a change in linear trend at some unspecified point was studied by Davies (1977, 1987). His approach is similar to our sequential testing but starts from a set of finite sample Gaussian test statistics. The supremum is taken over a continuous set and an upper bound on the  $p$ -value derived. Furthermore, in the setting considered by Davies, the changepoint disappears under the null hypothesis of interest and it is not estimated as it is considered a nuisance parameter.

The respective abilities of linear and quadratic tests to detect a significant trend will depend on the true underlying relationship and the distribution of the risk factor in the left-hand tail. In 400 simulations based on the true quadratic model as described under the alternative hypothesis in Section 5, the tests gave similar results with the linear test rejecting the null hypothesis in 70.5% of the simulations and the quadratic test rejecting in 66.3%.

The *linear-quadratic model* fits a global linear function with a quadratic changepoint to the log-hazard rates as follows:

$$\lambda_i(t) = \lambda_0(t) \exp\{\zeta_1 x_i + \zeta_2 (x_i - \eta)^2 I(x_i > \eta)\}. \quad (3)$$

Similar conditions as for the quadratic-quadratic fit would need to be fulfilled to justify the asymptotic properties of the corresponding partial likelihood approach. In model (3), evidence for a J-shape requires  $\zeta_1 < 0$ . The nadir is no longer directly parameterized by  $\eta$ , but reached at  $x = \eta - \zeta_1/2\zeta_2$ .

Other parametric models can describe risks that level off or go to an asymptote, e.g. inverse polynomials (McCullagh and Nelder (1989), p. 291). Unfortunately, their parameters are globally defined as for the simple quadratic test, and the specific test that we require is not readily derived. When the emphasis shifts towards model fitting, however, they deserve to be explored and techniques as in Royston and Thompson (1993) could be used to compare non-nested models.

A *nonparametric trend test* for an upturn to the left has the obvious appeal of avoiding any parametric assumptions. An adaptation of the approach discussed by Stone (1988) and Takacs (1962) works as follows. Concentrate on the data set within the risk factor range of interest, i.e. from  $\eta_l$  to  $\eta_u$  as described in Section 4. Rank the observed risk factor values from small (1) to large ( $N$ ) and condition on the total number  $d$  of events accrued over the set. If the risk factor has no effect on the event rate, the number of events accumulated over all  $x$ -values with rank below  $r$ , say  $d_r$ , is expected to be  $(r/N)d$ . Hence, the accumulated number of events plotted against the ranks tends to follow a straight line through the origin with constant slope  $d/N$ . If smaller  $x$ -values have a higher event rate, then more events will be accrued among the low ranks and less among the higher ranks. The initial slope of the cumulative number of events  $d_r/r$  is the rate at which events appear

and will tend to be higher for smaller  $r$  than the expected  $d/N$ . An exact null distribution of the maximum slope over the ranks,  $\max_r \{d_r/r\}$ , can be calculated. This approach is worth further exploration, though it may lack power and does not readily extend to censored survival data and to adjustment for other risk factors.

The general problem of low power suffered by all methods suggests a meta-analysis in this field. Specific methodology to combine evidence on observed risk factor associations from several studies has not been widely considered (one example can be found in MacMahon *et al.* (1990)). Only when the full data sets of all studies are made available can a stratified proportional hazards analysis be performed which allows for different base-line hazards but uses the same parameters for the parametric part of the model in all studies. We should not naïvely combine quadratic upturn *estimates* based on different changepoint estimates. However, it is conceivable to combine related independent *test* results in an overall measure with more power (Greenland (1987), p. 18).

### 7.2. Practical Conclusions

One practical motivation for our research has been the need to control the plethora of 'false positive' claims (type I errors) for J-shape relationships which have arisen through inappropriate statistical analysis and an overenthusiasm to find interesting and positive relationships. This has been particularly apparent in studies of treated diastolic blood pressure and coronary heart disease as discussed by Fletcher and Bulpitt (1992). However, it is equally important to recognize that type II errors (i.e. failure to detect a true J-shape) are a serious possibility given the small size of many data sets. For instance, in the Framingham data that we examined there was insufficient evidence of an upturn ( $p=0.29$ ) though the estimated double-quadratic curve (Fig. 2) is still compatible with a sizable true upturn in risk for diastolic blood pressure below 79 mmHg. The problem is that in too many epidemiological applications the nadir is liable to be well into the tail of the risk factor distribution. Thus, it requires very large studies indeed to have sufficient subjects (and events) to the left of the nadir so that realistic J-shape relationships can be detected and described.

Only if we both use appropriate statistical methods and have data sets with adequate statistical power can we reliably determine the statistical validity of any particular J-shape hypothesis. Until then, arguments about the causal interpretation and practical meaning will either be of only academic interest or may lead to misguided decisions in clinical and public health policy.

### ACKNOWLEDGEMENTS

This paper uses data supplied by the National Heart, Lung, and Blood Institute, National Institutes of Health, DHHS, from the Framingham Heart Study. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the National Heart, Lung, and Blood Institute or of the Framingham study. We would also like to thank Simon Thompson and Peter Sasieni for helpful comments on a draft of this paper. This research was supported by a grant from the Wellcome Trust.

## APPENDIX A: ASYMPTOTIC THEORY FOR THE FULL MODEL

We need to investigate the properties of the (partial) maximum likelihood statistics derived from model (1) because  $\eta$  enters the expression in a non-smooth way, i.e. for a given sample of size  $n$  with covariates  $\{x_i, i = 1, \dots, n\}$  the second derivative of the log-partial-likelihood with respect to  $\eta$  will not exist at each point  $\eta = x_i$ . At those points the left-hand second derivative will generally differ from the right-hand second derivative unless  $\gamma_1 = \gamma_2$ , and likewise for higher order derivatives. We assume that  $\eta$  lies strictly within the range of the observed  $x$ -values.

We first prove asymptotic normality for the score statistic in the next section. Next, our proof of the existence of a sequence of consistent solutions to the score equations assumes that there is a neighbourhood of the true  $\eta^0$  in which no  $x_i$ s lie. Following this, there is a neighbourhood of  $\theta^0 = (\gamma_1^0, \gamma_2^0, \eta^0)$  in which the theorem of Taylor can be used to yield asymptotic normality of the solution to the score equations and the asymptotic  $\chi^2$ -distribution under the true model of twice the partial likelihood ratio statistic.

The condition that the  $x$ s cannot come arbitrarily close to  $\theta^0$  is unfortunate, even though it is naturally met in certain situations where it is implied by the study design, e.g. a discrete  $x$ -distribution as in trials with few well-chosen target blood pressures (Dahlof *et al.*, 1991; HOT Study Group, 1992). However, Gallant's (1973) proof of asymptotic results of the quadratic changepoint model with normal errors does not need such heavy restriction on the  $x$ -distribution and this gives rise to the conjecture that here also the condition can be relaxed.

## A.1. Asymptotic Normality of Score Vector

For the double-quadratic model we can construct a partial likelihood in the usual way. At each event time the contribution consists of the hazard from the failed individual at that time, divided by the sum of the hazards of the individuals still at risk at that time. Thus, for any fixed value of the three parameters  $\theta = (\gamma_1, \gamma_2, \eta)$  we obtain the log-partial-likelihood:

$$\sum_{i \in \mathcal{D}_l} \gamma_1 (x_i - \eta)^2 + \sum_{i \in \mathcal{D}_r} \gamma_2 (x_i - \eta)^2 - \sum_{i \in \mathcal{D}} \log \left[ \sum_{j \in \mathcal{R}_{il}} \exp\{\gamma_1 (x_j - \eta)^2\} + \sum_{j \in \mathcal{R}_{ir}} \exp\{\gamma_2 (x_j - \eta)^2\} \right]. \quad (4)$$

$\mathcal{D}$ ,  $\mathcal{D}_l$  and  $\mathcal{D}_r$  indicate respectively the sets of individuals who died and those who died with  $x$ -value to the left and to the right of the point  $\eta$ . In a similar way  $\mathcal{R}_i$  is the risk set at the event time of the  $i$ th subject and  $\mathcal{R}_{il}$  and  $\mathcal{R}_{ir}$  are those subsets with  $x$ -values to the left and to the right of  $\eta$  respectively.

The first-order derivatives with respect to all the parameters exist and are continuous functions of those parameters. Introducing some extra notation, we can simplify the form of the score function. Let  $N_i(t) = 1$  when the  $i$ th subject's event happened before  $t$  and let  $N_i(t) = 0$  otherwise. For  $i = 1, \dots, n$ :  $I_{il}(\eta) = 1$  if  $x_i \leq \eta$  and  $I_{il}(\eta) = 0$  otherwise;  $I_{ir}(\eta) = 1 - I_{il}$ ;  $Y_i(t) = 1$  if the individual is at risk at time  $t$ , and  $Y_i(t) = 0$  otherwise. Finally,

$$Z_{i1}(\theta) = (x_i - \eta)^2 I_{il}(\eta),$$

$$Z_{i2}(\theta) = (x_i - \eta)^2 I_{ir}(\eta),$$

$$Z_{i3}(\theta) = -2\gamma_1 (x_i - \eta) I_{il}(\theta) - 2\gamma_2 (x_i - \eta) I_{ir}(\theta),$$

$$w_i(\theta, t) = Y_i(t) \exp\{\gamma_1 (x_i - \eta)^2 I_{il}(\eta) + \gamma_2 (x_i - \eta)^2 I_{ir}(\eta)\}.$$

With the vector  $\mathbf{Z}_i(\theta) = (Z_{i1}(\theta), Z_{i2}(\theta), Z_{i3}(\theta))$  the score functions  $d\ell(\theta, 1)/d\theta$  now are (assuming that we rescaled time on the interval  $[0, 1]$ )

$$\sum_i^n \int_0^{t=1} \left\{ Z_i(\theta) - \frac{\sum_j Z_j(\theta) w_j(\theta, u)}{\sum_j w_j(\theta, u)} \right\} dN_i(u). \quad (5)$$

Let the intensity process of  $N_i(t)$  be  $\lambda_i(\theta^0, t) Y_i(t)$  with  $\theta^0$  equal to the true parameter  $(\gamma_1^0, \gamma_2^0, \eta^0)$  and  $\lambda_i(\theta^0, t)$  as in Section 4. We assume that the 'normal' boundedness conditions hold, i.e.

$$\int_0^1 \lambda_0(t) dt < \infty$$

and the  $x$ s are bounded with probability 1. It is now easy to calculate that the compensator of the three components of  $dl(\theta^0, t)/d\theta$  (with  $t \in [0, 1]$ ) is 0, so they are in fact (square integrable) martingales.

The  $(k, m)$ th element of the covariation process of

$$n^{-1/2} \frac{dl}{d\theta}(\theta^0, t)$$

takes the form

$$n^{-1} \int_0^t \sum_i \left( Z_{ik} - \frac{\sum_j Z_{jk} w_j}{\sum_j w_j} \right) \left( Z_{im} - \frac{\sum_j Z_{jm} w_j}{\sum_j w_j} \right) w_i \lambda_0 du, \quad (6)$$

where  $k$  and  $m$  take values 1, 2 and 3, and all parameter values equal the true value:  $\theta^0$ .

Under limiting assumptions similar to those made in the classical proportional hazards model (see Andersen and Gill (1982)), we can use the inequality of Lengart to prove that expression (6) converges in probability to the matrix

$$\mathcal{I}(\theta^0, t) = \int_0^t v_{z,z}(\theta^0, u) s^0(\theta^0, u) \lambda_0(u) du$$

where  $v_{z,z}(\theta^0, u)$  is the limit in probability for  $n \rightarrow \infty$  of the urn model (Miller (1981), p. 85) variance  $V_{z,z}(\theta^0, u)$  of  $Z$  with weights  $w_i(\theta^0, u)$  and  $s^0(\theta^0, u)$  is the probability limit of  $(1/n) \sum w_j(\theta^0, u)$ .

The conditions are then fulfilled to apply the martingale central limit theorem yielding asymptotic normality  $N\{\theta, \mathcal{I}(\theta^0, 1)\}$  of the score function at the true parameters  $(\gamma_1^0, \gamma_2^0, \eta^0)$ . A consistent estimator of  $\mathcal{I}(\theta^0, 1)$  is obtained in the usual way by replacing  $w_i \lambda_0(u) du$  by  $dN_i(u)$  in expression (6). No special restrictions on the  $x$ -distribution are needed at this stage.

### A.2. Consistency of Maximum Likelihood Estimate

We can prove the existence of a consistent sequence of solutions to the score equations from concavity of the log-partial-likelihood function  $l$  in a neighbourhood of the true parameters with arbitrarily large probability. We assume that  $\mathcal{I}(\theta^0, 1)$  is invertible.

Focus on a finite sample size  $n$  and therefore a finite number of  $x$ -values. This function  $l$  is continuous and has continuous first-order derivatives at each point  $(\gamma_1, \gamma_2, \eta)$ . It also has continuous second-order derivatives with respect to the  $\gamma$ s for each fixed  $\eta$ . Furthermore, the second-order derivative with respect to  $\eta$  exists and is continuous at each  $\eta$  different from an observed  $x_i$ .

Fix any point  $(\gamma_1, \gamma_2, \eta)$  with  $\eta \notin \{x_i, i = 1, \dots, n\}$ . The scaled second-order derivative with respect to the three parameters at that point is minus the following:

$$\Gamma_n(\theta, t=1) = \frac{1}{n} \sum_i \int_0^{t=1} \left\{ \mathbf{V}_{z,z}(\theta, u) + \begin{pmatrix} 0 & 0 & b_{i1}(\theta, u) \\ 0 & 0 & b_{i2}(\theta, u) \\ b_{i1}(\theta, u) & b_{i2}(\theta, u) & b_{i3}(\theta, u) \end{pmatrix} \right\} dN_i(u).$$

The matrix  $\mathbf{V}_{z,z}(\theta, u)$  indicates the urn model covariance of the  $\mathbf{Z}_i(\theta)$  at time  $u$  with weights  $w_i(\theta, u)$ . Hence,  $\mathbf{V}_{z,z}(\theta, u)$  is automatically positive semidefinite. The  $b$ s in the second matrix take the form of observed minus expected values under that same urn model:

$$b_{ik}(\theta, u) = X_{ik}(\theta) - \frac{\sum_j X_{jk}(\theta) w_j(\theta, u)}{\sum_j w_j(\theta, u)}$$

where  $X_{i1} = 2(x_i - \eta) I_{i1}(\eta)$ ,  $X_{i2} = 2(x_i - \eta) I_{i\bar{r}}(\eta)$  and  $X_{i3} = 2\gamma_1 I_{i1} + 2\gamma_2 I_{i\bar{r}}$ . Again, under assumptions similar to those in Andersen and Gill (1982), it follows that the integrals over the  $b_j(\theta)$ s in  $\Gamma_n(\theta, 1)$  are continuous functions of  $\theta$  at  $\theta^0$ , even though  $X_{j3}(\theta)$  itself is not. The proof uses the inequality of Lengart, boundedness of the  $x$ s and convergence in probability of

$$\frac{1}{n} \sum_i \int_0^1 b_{ij}(\theta, u) w_i(\theta^0, u) du$$

to 0 as  $\theta$  tends to  $\theta^0$ .

Combined with convergence in probability of the estimator corresponding to expression (6) to  $\mathcal{J}(\theta^0, 1)$ , this means that  $\Gamma_n(\theta, t=1)$  becomes positive definite in a neighbourhood of  $\theta^0$  with arbitrarily large probability as  $n$  becomes big. A combination of the inverse function theorem as in Foutz (1977) and the inequality of Lengart then yields the desired result (Goetghebeur (1990), pages 64–65).

REFERENCES

Andersen, P. K. and Gill, R. D. (1982) Cox's regression model for counting processes: a large sample study (with commentary). *Ann. Statist.*, **10**, 1100–1124.  
 D'Agostino, R., Belanger, A. J., Kannel, W. B. and Cruickshank, J. M. (1991) Relation of low diastolic blood pressure to coronary heart disease death in presence of myocardial infarction: the Framingham study: *Br. Med. J.*, **303**, 385–389.  
 D'Agostino, R., Lee, M.-L., Belanger, A., Cupples, L., Anderson, K. and Kannel, W. (1990) Relation of pooled logistic regression to time dependent Cox regression analysis: the Framingham Heart Study. *Statist. Med.*, **9**, 1501–1515.  
 Dahlof, B., Lindholm, L., Hansson, L., Schersten, B., Ekbom, T. and Wester, P.-O. (1991) Morbidity and mortality in the Swedish Trial in Old Patients with Hypertension (STOP-hypertension). *Lancet*, **ii**, 1281–1285.  
 Davies, R. B. (1977) Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, **64**, 247–254.  
 ——— (1987) Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, **74**, 33–43.  
 Farnett, L., Mulrow, C. D., Linn, W. D., Lucey, C. R. and Tuley, M. R. (1991) The J-curve phenomenon and the treatment of hypertension: is there a point below which pressure reduction is dangerous? *J. Am. Med. Ass.*, **265**, 489–495.  
 Fletcher, A. E. and Bulpitt, C. J. (1992) How far should blood pressure be lowered? *New Engl. J. Med.*, **326**, 251–254.

- Foutz, R. V. (1977) On the unique consistent solution to the likelihood equations. *J. Am. Statist. Ass.*, **72**, 147-148.
- Frank, J. W., Dwayne, M. R., Grove, J. S. and Benfante, R. (1992) Will lowering population levels of serum cholesterol affect total mortality? *J. Clin. Epidem.*, **45**, 333-346.
- Gallant, A. R. (1971) Statistical inference from non-linear regression models. *PhD Thesis*. Iowa State University, Iowa.
- (1975) The power of the likelihood ratio test of location in nonlinear regression models. *J. Am. Statist. Ass.*, **70**, 198-203.
- Gallant, A. R. and Fuller, W. A. (1973) Fitting segmented polynomial regression models whose joint points have to be estimated. *J. Am. Statist. Ass.*, **68**, 144-147.
- Geller, N. L. and Pocock, S. J. (1987) Interim analyses in randomized clinical trials: ramifications and guidelines for practitioners. *Biometrics*, **43**, 213-224.
- Goetghebeur, E. (1990) Analyzing competing risks survival data when some failure types are missing. *PhD Thesis*. University of Limburg, Diepenbeek.
- Greenland, S. (1987) Quantitative methods in the review of epidemiologic literature. *Epidem. Rev.*, **9**, 1-30.
- HOT Study Group (1992) *The Hypertension Optimal Treatment Study*. Göteborg: Astra.
- MacMahon, S., Peto, R., Cutler, J., Collins, R., Sorlie, P., Neaton, J., Abbott, R., Godwin, J., Dyer, A. and Stamler, J. (1990) Blood pressure, stroke and coronary heart disease, parts I and II. *Lancet*, **335**, 765-774, 827-838.
- Marmot, M. G., Rose, G., Shipley, M. J. and Thomas, B. J. (1981) Alcohol and mortality: a U-shaped curve. *Lancet*, **i**, 580-583.
- McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models*, 2nd edn. London: Chapman and Hall.
- Miller, Jr, R. G. (1981) *Survival Analysis*. New York: Wiley.
- Polychronaki, H., Hatzakis, A., Vatopoulos, A., Katsouyanni, K., Tzonou, A. and Trichopoulos, D. (1982) Association of coronary mortality with temperature and air pollution in Athens. *Haemostasis*, **12**, 133.
- Rigby, R. A. and Stasinopoulos, D. M. (1991) Detecting break points in the hazard functions in survival analysis. In *Statistical Modelling*. Utrecht: Elsevier.
- Roy, S. N. and Pothoff, R. F. (1958) Confidence bounds on vector analogues of the "ratio of means" and the "ratio of variances" for two correlated normal variates and some associated tests. *Ann. Math. Statist.*, **28**, 829-841.
- Royston, P. and Thompson, S. G. (1993) Comparing non-nested regression models. *Biometrics*, to be published.
- Samuelsson, O. G., Wilhelmsen, L. W., Pennert, K. M., Wedel, H. and Berglund, G. L. (1990) The J-shaped relationship between coronary heart disease and achieved blood pressure level in treated hypertension: further analysis of 12 years of follow-up of treated hypertensives in the Primary Prevention Trial in Gothenburg, Sweden. *J. Hypertensn*, **8**, 547-555.
- Seber, G. A. F. and Wild, C. J. (1990) *Nonlinear Regression*, ch. 9. New York: Wiley.
- Stasinopoulos, D. M. and Rigby, R. A. (1992a) Detecting break points in generalized linear models. *Comput. Statist. Data Anal.*, **13**, 461-471.
- (1992b) GLIM macro's to fit piece-wise polynomials with one break point. *GLIM Newslett.*, **20**, 18-24.
- Stone, R. A. (1988) Investigations of excess environmental risks around putative sources: statistical problems and a proposed test. *Statist. Med.*, **7**, 649-660.
- Takacs, L. A. (1962) A generalization of the ballot problem and its application in the theory of queues. *J. Am. Statist. Ass.*, **57**, 327-337.
- Tsiatis, A. A. (1982) Repeated significance testing for a general class of statistics in censored survival analysis. *J. Am. Statist. Ass.*, **77**, 855-861.
- Wilcox, A. J. and Russell, I. T. (1983) Birthweight and perinatal mortality: II, On weight-specific mortality. *Int. J. Epidem.*, **12**, 319-325.