Epidemiologic Methods

Volume 1, *Issue* 1 2012 *Article* 2

Effects of Classical Exposure Measurement Error on the Shape of Exposure-Disease Associations

Ruth H. Keogh, *MRC Biostatistics Unit, Cambridge, UK*
Alexander D. Strawbridge, *MRC Biostatistics Unit, Cambridge, UK*
Ian R. White, *MRC Biostatistics Unit, Cambridge, UK*

Recommended Citation:

Keogh, Ruth H.; Strawbridge, Alexander D.; and White, Ian R. (2012) "Effects of Classical Exposure Measurement Error on the Shape of Exposure-Disease Associations," *Epidemiologic Methods*: Vol. 1: Iss. 1, Article 2.
DOI: 10.1515/2161-962X.1007

Effects of Classical Exposure Measurement Error on the Shape of Exposure-Disease Associations

Ruth H. Keogh, Alexander D. Strawbridge, and Ian R. White

Abstract

In epidemiology many exposures of interest are measured with error. Random, or 'classical', error in exposure measurements attenuates linear exposure-disease associations. However, its precise effects on different nonlinear associations are not well known. We use simulation studies to assess how classical measurement error affects observed association shapes and power to detect nonlinearity. We focus on a proportional hazards model for the exposure-disease association and consider six true association shapes of relevance in epidemiology: linear, threshold, U-shaped, J-shaped, increasing quadratic, asymptotic. The association shapes are modeled using three popular methods: grouped exposure analyses, fractional polynomials, P-splines. Under each true association shape and each method we illustrate the effects of classical exposure measurement error, considering varying degrees of random error. We also assess what we refer to as MacMahon's method for correcting for classical exposure measurement error under grouped exposure analyses, which uses replicate measurements to estimate usual exposure within observed exposure groups. The validity of this method for nonlinear associations has not previously been investigated. Under nonlinear exposure-disease associations, classical measurement error results in increasingly linear shapes and not always an attenuated association at a given exposure level. Fractional polynomials and P-splines give similar results and offer advantages over grouped exposure analyses by providing realistic models. P-splines offer greatest power to detect nonlinearity, however random exposure measurement error results in a potential considerable loss of power to detect nonlinearity under all methods. MacMahon's method performs well for quadratic associations, but does not in general recover nonlinear shapes.

KEYWORDS: classical measurement error, dose-response, grouped exposure analysis, fractional polynomials, P-splines

Author Notes: This work was supported by the Medical Research Council [Unit Programme number U105260558]. Ruth Keogh is now at the Department of Medical Statistics, London School of Hygiene and Tropical Medicine (ruth.keogh@lshtm.ac.uk). The authors are grateful to Professor Simon Thompson for his comments on an earlier draft of this paper.

1 Introduction

Many epidemiological exposures are subject to measurement error. For example, biological measurements, e.g. blood pressure (MacMahon et al., 1990, Clarke et al., 1999), are subject to within-subject fluctuations (Toniolo et al., 1997). Hence, when the exposure of interest is 'usual' level, a measurement made on a single occasion is subject to random variability. Some biological measurements are also subject to error due to assay variability. Measurement error is a particular problem in nutritional epidemiology, because measurements of long term dietary intake are subject to error due to variability over time and limitations of measurement instruments (Willett, 1998). We focus on classical measurement error, which may result from random within-person variability or random error in the measurement process. Under classical measurement error the observed exposure, W say, can be written as $W = X + U$ where X denotes the true underlying exposure and U is a random error term, that is U is independent of X, of any characteristics of the individual to whom the measurement pertains, and of any outcome such as disease status whose association with the exposure is of interest.

It is well known that use of exposures subject to classical measurement error gives attenuated exposure-outcome associations when the association is linear (Carroll et al., 2006, Rosner et al., 1989, Gardner and Heady, 1973, Rosner et al., 1992, Hughes, 1993, Spiegelman et al., 1997). In linear, logistic and proportional hazards regressions it is respectively the estimated slope, log odds ratio, and log hazard ratio (HR) which is attenuated. We focus on disease outcomes in proportional hazards regression. When the shape of the exposure-disease association is nonlinear, random error in exposure measurements can mask the features of the association so that the observed shape is not the true shape (Carroll et al., 2006). The association shape can have important implications for public health policy: for example, it may be important to know whether an increase in exposure at a higher level of exposure results in a greater change in risk than the same change at a lower level, or whether there exists a threshold of exposure below which there is no change in risk.

The aim of this paper is to clearly outline for the first time the precise effects of classical measurement error in nonlinear models and to quantify what is lost in terms of power to detect nonlinearity. Simulation studies are used to illustrate the effects of random error in exposure measurements on six exposure-disease association shapes representing some of the shapes of interest in epidemiology. Shapes considered are linear, threshold, U-shaped, J-shaped, increasing quadratic and asymptotic. Examples of nonlinear shapes from the literature include an observed threshold association between systolic blood pressure and risk of both cardiovascular and all-cause death in the Framingham Heart Study (Port et al., 2000). There is evidence of a U- or J-shaped association between alcohol intake and total

mortality, with both heavy- and non-drinkers having higher mortality than moderate drinkers (Poikolainen, 1995, Shaper et al., 1988, Bagnardi et al., 2004). U- and J-shaped associations have also been observed between diastolic blood pressure and risk of cardiac events (D'Agostino et al., 1991, Farnett et al., 1991). An asymptotic association has been found between plasma 25-hydroxyvitamin D and premenopausal breast cancer risk, showing a flattening of the risk for higher concentrations (Abbas et al., 2009) and in a study of the relationship between ethanol intake and risk of cancer of the aero-digestive tract a flattening of the association was observed for high ethanol intake (Polesel et al., 2005).

We investigate the effects of different degrees of random error and how its effects are manifested under three methods for modelling exposure in proportional hazards models: a grouped exposure analysis, fractional polynomials (FP) and P-splines. We consider a simple method of correcting the exposure-disease association in a grouped exposure analysis using repeated measurements, referred to here as MacMahon's method (MacMahon et al., 1990), which has been used especially in large pooling projects (Prospective Studies Collaboration, 2002, Allen et al., 2009, Liu et al., 2009, Asia-Pacific Cohort Studies Collaboration, 2004). The validity of MacMahon's method for nonlinear association shapes has not been previously assessed. However, our aim is not to assess correction methods in general.

In the following sections we describe the statistical methods, simulation study and results, and we conclude with a discussion.

2 Methods for Investigating the Shape of Exposure-Disease Associations

Let X denote a continuous exposure of interest, t denote the time to disease diagnosis (e.g. age), and $h(t|X)$ denote the hazard function at t for an individual with exposure X. Under the proportional hazards assumption, when the exposure-disease association is linear the log hazard is

$$\log\{h(t|X)\} = \log\{h_0(t)\} + \beta X \tag{1}$$

where $h_0(t)$ is the baseline hazard at t and β is the log HR, representing the linear change in $\log\{h(t|X)\}$ when X increases by 1 unit. Model (1) is typically extended to include adjustment for a vector of confounders which are assumed to be measured without error. To simplify the notation, we do not include confounders in the models here.

To investigate whether $\log\{h(t|X)\}$ is nonlinear in X, quadratic or higher order exposure effects can be added to (1). However, this is restrictive since the

exposure-disease association is assumed to take a particular form. Below we outline three more flexible methods for investigating nonlinearity in the exposure-disease association.

2.1 A grouped exposure analysis

To avoid making assumptions about the association shape, the exposure is commonly divided into groups, and the log HR is estimated within each group compared with a reference group. We call this the grouped exposure analysis. Suppose that the continuous exposure X is grouped into K categories. Let $X^{(k)}$ take value 1 if the exposure is in the kth category and value 0 otherwise $(k = 1, \ldots, K)$. Under the grouped exposure analysis the log hazard at time t is modelled as

$$\log\left\{h(t|X^{(2)},\ldots,X^{(K)})\right\} = \log\{h_0(t)\} + \beta_2 X^{(2)} + \beta_3 X^{(3)} + \cdots + \beta_K X^{(K)} \quad (2)$$

where β_k is the log HR for an individual in the kth category of exposure $(k = 2, \ldots, K)$ relative to an individual in the first category. For some exposures fixed cutpoints are typically used, for example body mass index which is often classified according to < 18.5 (underweight), $18.5- < 25$ (healthy weight), $25- < 30$ (overweight), ≥ 30 (obese). For exposures for which there are no commonly used pre-defined cutpoints or for which the distribution may differ considerably in different populations, quantile cutpoints are often used, with the number of categories depending on sample size and exposure distribution. Inferences about the shape of the exposure-disease association are often made based on a plot of the log HR for each exposure category against the mean exposure within that category.

A grouped exposure analysis is attractive because it is simple to implement and does not make assumptions about the exposure-disease association shape. However, this method does not provide a realistic model for disease risk. The observed association can strongly depend on the number of exposure groups and positioning of cut-points. The method will perform badly if most subjects are exposed within a narrow range and exposure effects are limited to extreme ends of the scale (Greenland, 1995a,b).

The null hypothesis that the exposure-disease association is linear can be tested by comparing the grouped exposure model (2) with a model which assumes a linear association. The latter model is $\log\{h(t|X_G)\} = \log\{h_0(t)\} + \beta X_G$, where for each individual in exposure category k, X_G takes the mean exposure in that group. Under the null hypothesis, twice the difference between the the log likelihoods under the two models has a χ^2 distribution with $K-2$ degrees of freedom.

MacMahon et al. (1990) proposed a method to allow us to observe graphically the true HR estimates within exposure categories when the exposure is subject

to classical measurement error. They noted that when using an exposure measured with random error in a grouped exposure analysis, the lowest/highest category will include disproportionately many individuals whose single exposure measurement happened to be lower/higher than their 'usual' exposure, resulting in an observed diluted association when the association is linear. In this correction method the log HR within each original exposure category is plotted against an unbiased estimate of the mean 'usual' exposure in that category. Such an estimate can be obtained using one or more repeated exposure measurements available for at least a subset of the population: the means of the replicate measurements within categories defined by the original exposure measurements provide unbiased estimates of mean usual exposure within categories.

2.2 Fractional polynomials (FP)

It is often desirable to observe the shape of the exposure-disease association continuously across the range of exposure. FPs can give more plausible shapes with the use of fewer regression parameters than standard polynomial analyses, which make *a priori* assumptions about the shape (Royston and Altman, 1994, Royston et al., 1999). A FP analysis uses maximum likelihood to choose the 'best' set of power transformations of X, from a limited set, for inclusion in the exposure-disease model. Using a FP of degree 2 we select the best powers p_1 and p_2 in the model

$$\log\{h(t|X)\} = \begin{cases} \log\{h_0(t)\} + \xi_1 X^{p_1} + \xi_2 X^{p_2} & \text{if } p_2 \neq p_1 \\ \log\{h_0(t)\} + \xi_1 X^{p_1} + \xi_2 X^{p_1}\log(X) & \text{if } p_2 = p_1 \end{cases} \tag{3}$$

from the set of potential powers $P = \{-2,-1,-0.5,0,0.5,1,2,3\}$, where a power of 0 denotes the log transformation ($\log(X)$). This set is usually sufficiently rich and FPs of degree higher than 2 are rarely required (Royston et al., 1999).

For a FP of degree 2 we can perform an approximate test of the null hypothesis that the exposure-disease association is linear by comparing $-2\{l(1,1) - l(2,(\tilde{p}_1,\tilde{p}_2))\}$ with the χ^2 distribution with 3 degrees of freedom (Royston and Altman, 1994, Ambler and Royston, 2001), where $l(1,1)$ denotes the log likelihood from the linear model and $l(2,(\tilde{p}_1,\tilde{p}_2))$ denotes the log likelihood from the best fitting FP of degree 2 with powers $(\tilde{p}_1,\tilde{p}_2)$. The test is approximate because estimation of the powers from the set P does not consume 2 degrees of freedom, hence the true degrees of freedom for the test is less than 3 (Royston and Altman, 1994, Ambler and Royston, 2001).

2.3 P-splines

Splines are continuous smooth curves which are constructed by piecing together a series of polynomials within exposure intervals. They provide a flexible approach to exploring the exposure-disease association because they fit models within exposure intervals divided by 'knots', rather than assuming a parametric model across the entire range of exposure. Here we use P-splines (Eilers and Marx, 1996, Govindarajulu et al., 2007) and model the log hazard as

$$\log\{h(t|X)\} = \log\{h_0(t)\} + \sum_{j=1}^{n} \beta_j S_j(t) \tag{4}$$

where n is the number of knots, which are evenly spaced, $S_j(t)$ are B-spline basis functions (De Boor, 2001), and β_j are parameters to be estimated. The method applies a penalty to the log partial likelihood $l(X;\beta)$ arising from the proportional hazards regression model in (4) to reduce the influence of the number and position of the knots. The penalized log likelihood is

$$l_p(X;\beta,\lambda) = l(X;\beta) - \lambda \sum_{j=1}^{n} \left(\Delta^2 \beta_j\right)^2 \tag{5}$$

where Δ is the difference function such that $\Delta^2\beta_j = (\beta_j - \beta_{j-1}) - (\beta_{j-1} - \beta_{j-2})$. We use a common approach to analysis using P-splines, which is to specify the degrees of freedom, leading directly to the value of λ (Malloy et al., 2009, Hurvich et al., 1998, Ruppert, 2002).

For a P-spline with d degrees of freedom we can perform an approximate test of the null hypothesis that the exposure-disease association is linear by comparing $-2\{l(1,1) - l_p(X;\hat{\beta},\hat{\lambda})\}$ with the χ^2 distribution with $d-1$ degrees of freedom.

3 Simulation Study

Let X_i denote true exposure for individual i, which is unobserved, and W_{ij} their observed exposure on measurement occasion j. Under the classical measurement error model we have

$$W_{ij} = X_i + U_{ij} \tag{6}$$

where errors U_{ij} have mean 0 and are independent of X_i and each other. For a cohort of 15,000 individuals, values X_i were randomly sampled from a normal distribution with mean 10 and variance $\sigma_X^2 = 1$. We first focus on observed exposure measurements made on one occasion, W_{i1}. These were obtained using (6), with the errors

U_{i1} being generated from a normal distribution with mean 0 and variance σ_U^2. We consider $\sigma_U^2 = 0, 0.25, 0.5, 1$, where $\sigma_U^2 = 0$ represents no measurement error. Under a linear exposure-disease association the ratio of the log HR estimated using W_1 to that estimated using X is called the regression dilution ratio (RDR) and is approximated by $\text{var}(X)/\text{var}(W_1) = \sigma_X^2/(\sigma_X^2 + \sigma_U^2)$ (Rosner et al., 1992, Spiegelman et al., 1997). Error variances $\sigma_U^2 = 0, 0.25, 0.5, 1$ correspond to RDRs 1, 4/5, 2/3, 1/2.

Table 1 shows the form of the log hazard for six exposure-disease association shapes. We used $\beta_1 = 0.3$ in the linear model and $\beta_2, \beta_3, \beta_4, \beta_5, \beta_6$ were chosen so that the degree of nonlinearity in the association is the same across all nonlinear shapes. Degree of nonlinearity was defined as the squared difference, averaged over the distribution of X, between the true log hazard and that found by fitting the linear model (1). We chose $\beta_6 = 10$ and calculated values $\beta_2 = 0.28, \beta_3 = \beta_4 = \beta_5 = 0.06$. The aim is to compare different methods and the effects of different RDRs for the same true shape, not to compare the impact of measurement error *between* true shapes, hence our results are likely to be similar for other choices of $\beta_1, \ldots, \beta_6$.

Table 1: Simulation study: Models for shapes of association between true exposure X and the log hazard under a proportional hazards model for disease risk. ($I_{X>10}$ is an indicator which takes value 1 if $X > 10$ and value 0 otherwise.)

Shape	Form of $\log\{h(t\|X)\}$
Linear	$\log\{h_0(t)\} + \beta_1(X-10)$
Threshold	$\log\{h_0(t)\} + \beta_2 I_{X>10}(X-10)$
J-shaped	$\log\{h_0(t)\} + \beta_4(X-9)^2$
U-shaped	$\log\{h_0(t)\} + \beta_3(X-10)^2$
Increasing quadratic	$\log\{h_0(t)\} + \beta_5(X-7)^2$
Asymptotic	$\log\{h_0(t)\} + \beta_6/(4-X)$,

Survival times t were simulated under each model in Table 1 using a constant baseline hazard, chosen so that approximately 10% of individuals die during 10 years of follow-up. For each RDR and each shape, we generated 1000 simulated data sets containing $[X_i, W_{i1}, t_i]$, $i = 1, \ldots, 15,000$.

In each simulated data set, grouped exposure, FP and P-spline analyses were used to fit proportional hazards models, first using X_i and then using W_{i1}. For grouped exposure analyses individuals were grouped by quintiles. In FP analyses we selected the best FP of degree 2 in each simulated data set. For P-spline analyses the exposure was divided into 10 equally spaced intervals and the model was fitted with 4 degrees of freedom, the defaults in the `pspline` function of the survival package in R.

Power to detect nonlinearity under each nonlinear association shape was estimated as the proportion of simulations in which the null hypothesis of a linear exposure-disease association was rejected at the 5% level. For comparison, we also estimated the probability of detecting a nonlinear association when the true association is linear, that is the type I error rate, using a χ^2 test with 3 degrees of freedom. The sample size of 15,000 was chosen to give approximately 80% power to reject the null hypothesis of linearity at the 5% level using a FP analysis when there is no measurement error and the true exposure-disease association is quadratic. We also estimated power to detect nonlinearity for smaller sample sizes of 10,000, 5000 and 1000 to illustrate the effects of exposure measurement error under a situation in which the power would be low using the true exposure. The other results remain with a sample size of 15,000 as they relate to bias and are unaffected by sample size.

We investigated the performance of MacMahon's method by extending the simulation study to include a second exposure measurement W_{i2} for all individuals. An unbiased estimate of mean usual exposure within quintile k of the categorized exposure W_{i1} was calculated as the mean of W_{i2} within that quintile.

4 Results

For each true exposure-disease association shape and each analysis method we graphed the association found using true exposure (RDR=1) and using observed exposure with RDRs 4/5, 2/3, 1/2. Figure 1 shows the results from grouped exposure analyses. Log HR estimates for each quintile relative to quintile 3 were averaged across 1000 simulations and plotted against the mean exposure within each quintile. With the exception of the threshold association, the results from FP and P-spline analyses were very similar so we show only the FP results here for the other shapes (Figure 2). Results from P-spline analyses under a threshold association are shown in Figure 3. Different FP models were selected in different simulations; in each simulation we found the fitted value of the log HR at each value of the observed exposure relative to the mean of the exposure, and averaged these across 1000 simulations to give the average association shape. Parameter estimates from the P-spline analyses were averaged across the 1000 simulations and the P-spline resulting from the mean estimates was found. The results from using MacMahon's correction method are shown in Figure 4: we obtained the average log HR estimate within each quintile and plotted it against the mean usual exposure within each quintile. All figures also show the true association shape.

Table 2 shows estimated powers to reject the null hypothesis of a linear association under the three analysis methods, six shapes, and different degrees of

measurement error for sample sizes of 15,000, 10,000, 5000 and 1000. Powers for MacMahon's method are not given in Table 2 because they are the same as for the grouped exposure method.

4.1 Performance of the methods when there is no exposure measurement error

Except for a linear association, HR estimates from the grouped exposure analyses do not lie exactly on the true association curve even when there is no exposure measurement error (Figure 1). This is particularly marked for the U-shaped association, and is explained in Appendix A. The reason for this is that the grouped exposure method does not provide a correct model for the exposure-disease associations, which is rarely mentioned in results from such analyses. The observed association shape under the grouped exposure method lies closer to the true shape on average when the shape within an exposure group is more linear and the number of groups is increased.

The grouped exposure analysis does not lend itself to identifying the sharp turning point in the threshold association, and for a J-shaped association the upward curve for low exposures is easily missed. FPs and P-splines give association curves very close to the true shapes, except for the threshold association (Figures 2,3). FPs perform particularly badly under a threshold association, instead showing a J-shape. Because FPs are used to fit a smooth shape across the range of exposure it is not unexpected that this method cannot provide a fit to the turning point. P-splines perform considerably better (Figure 3).

P-splines offer the highest power to detect departures from linearity. FPs offer higher power compared with a grouped exposure analysis, except for the threshold association where both methods suffer from low power, and except when the sample size is small so as to offer little power to detect nonlinearity under any method (sample size 1000) (Table 2). Decreasing the sample size by a third from 15,000 to 10,000 resulted in a proportional reduction in power of between 26% (asymptotic) and 35% (increasing quadratic) using the grouped exposure analysis, between 29% (U-shaped) and 37% (asymptotic) using FPs, but only of between 19% (U-shaped) and 29% (asymptotic) using P-splines. Under the FP analyses, type I error rates for true linear associations are lower than 0.05 because estimation of power parameters in a FP does not consume two degrees of freedom (Royston and Altman, 1994, Ambler and Royston, 2001).

Table 2: Estimated power to detect nonlinearity under each nonlinear exposure-outcome association shape and estimated type I error in a test of the null hypotheses of linearity when the true association is linear, using each analysis method.

			Shape of exposure-outcome association					
Method	Sample Size	RDR	Linear	Threshold	U-shaped	J-shaped	Increasing quadratic	Asymptotic
Grouped exposure	15000	1	0.04	0.63	0.47	0.45	0.52	0.23
		4/5	0.05	0.40	0.32	0.31	0.34	0.16
		2/3	0.04	0.26	0.23	0.24	0.26	0.12
		1/2	0.04	0.14	0.14	0.15	0.14	0.10
	10000	1	0.05	0.46	0.32	0.30	0.34	0.17
		4/5	0.05	0.26	0.21	0.21	0.23	0.13
		2/3	0.05	0.18	0.16	0.16	0.15	0.10
		1/2	0.05	0.11	0.11	0.10	0.12	0.08
	5000	1	0.04	0.20	0.14	0.14	0.16	0.09
		4/5	0.04	0.12	0.09	0.10	0.10	0.08
		2/3	0.04	0.09	0.08	0.09	0.08	0.06
		1/2	0.04	0.07	0.07	0.07	0.06	0.06
	1000	1	0.06	0.08	0.07	0.07	0.07	0.07
		4/5	0.06	0.07	0.06	0.07	0.06	0.06
		2/3	0.06	0.06	0.05	0.06	0.07	0.06
		1/2	0.06	0.06	0.05	0.05	0.06	0.06
Fractional polynomial	15000	1	0.01	0.64	0.76	0.77	0.76	0.41
		4/5	0.01	0.42	0.54	0.52	0.51	0.24
		2/3	0.01	0.26	0.34	0.34	0.29	0.15
		1/2	0.01	0.10	0.14	0.16	0.12	0.10
	10000	1	0.01	0.41	0.54	0.53	0.52	0.26
		4/5	0.01	0.25	0.33	0.33	0.30	0.15
		2/3	0.01	0.17	0.24	0.21	0.20	0.10
		1/2	0.02	0.09	0.11	0.12	0.11	0.06
	5000	1	0.01	0.18	0.22	0.22	0.21	0.09
		4/5	0.02	0.11	0.14	0.13	0.14	0.06
		2/3	0.02	0.09	0.09	0.09	0.10	0.05
		1/2	0.02	0.05	0.06	0.05	0.05	0.02
	1000	1	0.02	0.03	0.04	0.03	0.03	0.02
		4/5	0.01	0.02	0.03	0.02	0.02	0.03
		2/3	0.01	0.02	0.03	0.03	0.02	0.02
		1/2	0.02	0.02	0.02	0.02	0.02	0.02
P-spline	15000	1	0.08	0.84	0.86	0.87	0.87	0.62
		4/5	0.08	0.64	0.69	0.68	0.68	0.43
		2/3	0.08	0.47	0.53	0.53	0.48	0.33
		1/2	0.08	0.27	0.30	0.33	0.26	0.22
	10000	1	0.07	0.65	0.70	0.69	0.68	0.44
		4/5	0.08	0.44	0.50	0.48	0.48	0.31
		2/3	0.07	0.33	0.38	0.38	0.36	0.22
		1/2	0.07	0.21	0.24	0.25	0.23	0.15
	5000	1	0.07	0.38	0.40	0.38	0.38	0.23
		4/5	0.08	0.24	0.27	0.28	0.27	0.17
		2/3	0.08	0.20	0.21	0.21	0.21	0.14
		1/2	0.08	0.14	0.15	0.15	0.15	0.11
	1000	1	0.10	0.14	0.15	0.15	0.14	0.12
		4/5	0.09	0.11	0.12	0.12	0.10	0.10
		2/3	0.09	0.11	0.12	0.11	0.09	0.11
		1/2	0.09	0.10	0.09	0.10	0.08	0.10

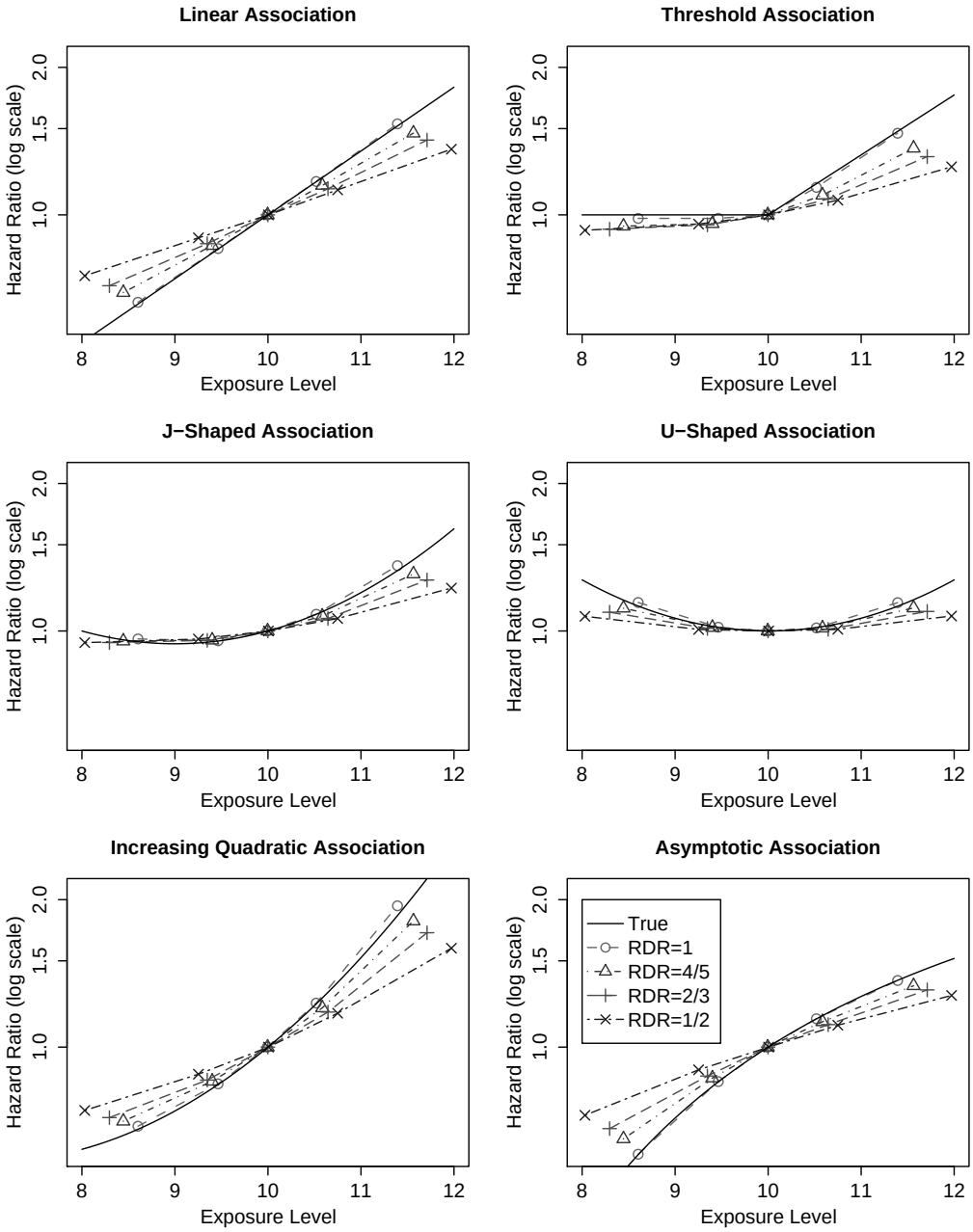

Figure 1: Grouped exposure analysis using quintiles: The effect of random error in exposure measurements on the observed exposure-disease association.

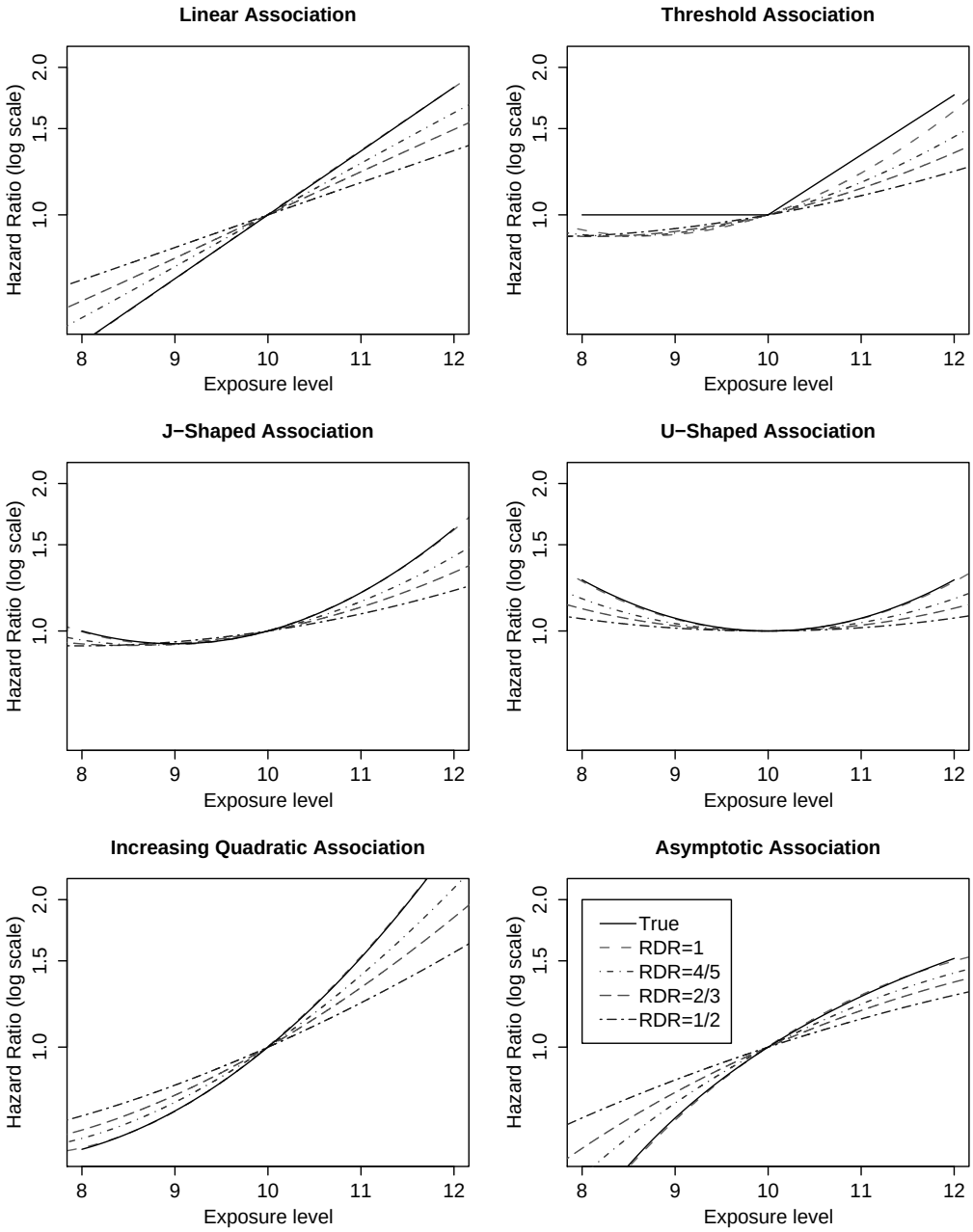

Figure 2: Fractional polynomial analysis: The effect of random error in exposure measurements on the observed exposure-disease association.

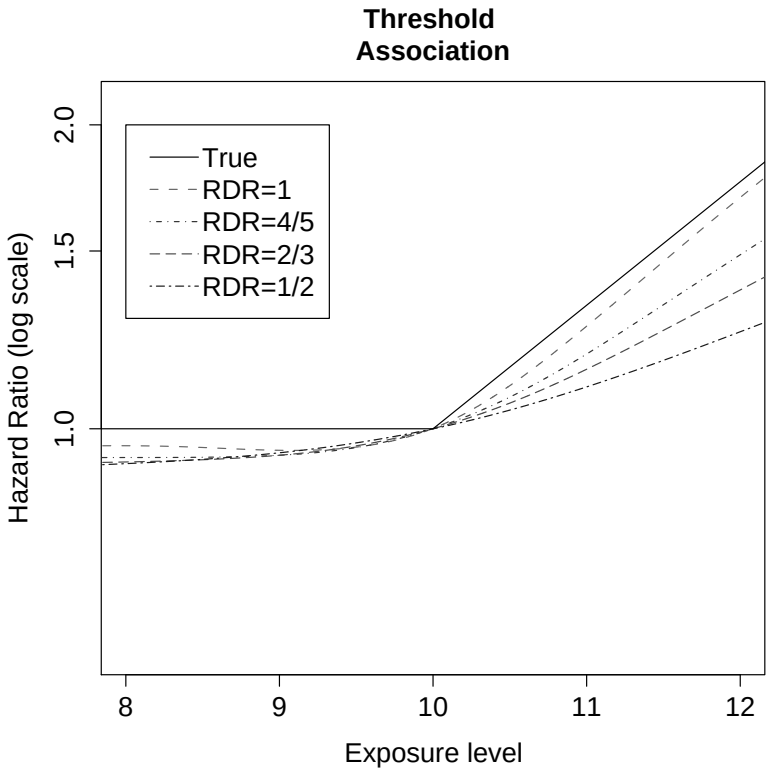

Figure 3: P-spline analysis: The effect of random error in exposure measurements on the observed exposure-disease association under a true threshold association.

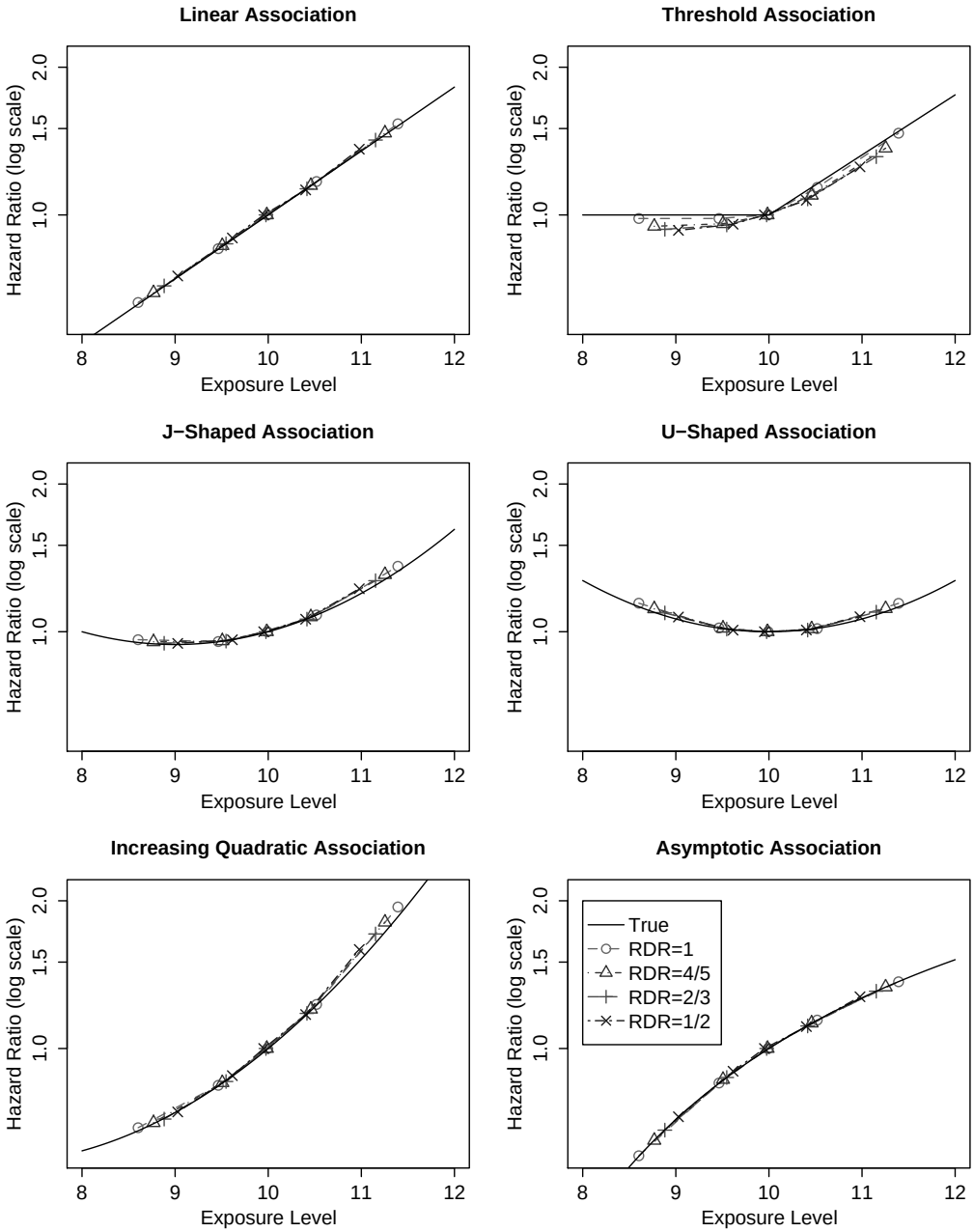

Figure 4: MacMahon's method for a grouped exposure analysis using quintiles: The results from correcting for random error in exposure measurements in the grouped exposure method using repeat exposure measurements.

4.2 Effects of exposure measurement error on exposure-disease association shapes

Random measurement error does not affect the linearity of the exposure-disease association when it is genuinely linear, though this is an approximation under a proportional hazards model (Carroll et al., 2006). As is well known, when the exposure-disease association is linear the effect of random exposure measurement error is to attenuate the degree of association, with the attenuation becoming more severe as the measurement error becomes more severe (Figures 1, 2).

When the true association is nonlinear, the effect of classical measurement error is to make it appear more linear. Nonlinearity becomes less obvious as the severity of measurement error increases, i.e. as the RDR decreases (Figures 1-3). This comes with a potential considerable loss of power to detect departures from linearity (Table 2). For smaller sample sizes, where the power to detect nonlinearity would be low using the true exposure, the proportional effect of measurement error on power is somewhat less in general. P-splines continue to offer the highest power to detect nonlinearity across a range of sample sizes when the exposure is subject to measurement error. Power to detect nonlinearity using FPs is more severely diminished by measurement error, and this method performs little better or sometimes worse than the grouped exposure analysis when the degree of error becomes severe.

The effect of classical measurement error on nonlinear associations is not always an attenuation of the log HR estimates at a given exposure level; the bias can go in *either* direction depending on the true shape. For example, using a grouped exposure analysis the effect of classical measurement error on a true threshold association is to attenuate the log HR estimates in the linear part of the association beyond the threshold, while estimates below the threshold move away from the null. There is a similar effect for the J-shaped association. When a continuous exposure is subject to classical measurement error, categorization can result in differential misclassification, that is misclassification depending on the outcome (Flegal et al., 1991), which can result in bias in either direction (Dosemeci et al., 1990).

Under a true J-shape the effect of classical measurement error on the results from FP and P-spline analyses is to shift the turning point so that the lowest risk appears to occur at a lower exposure as the error variance increases, which has been shown previously in quadratic regression models (Kuha and Temple, 2003). The results obtained under true threshold, J-shaped and increasing quadratic association shapes look very similar when the exposure measurement error is severe (Figure 2).

MacMahon's method deattenuates the slope under a linear association and also appears to perform well for quadratic associations, where the exposure-disease association within groups is reasonably linear (Figure 4). However, MacMahon's

result does not extend in general to nonlinear exposure-disease associations and we provide an explanation for this in Appendix B. We also note that under this method there is a smaller difference between the mean exposures within exposure groups compared with the situation where there is no exposure measurement error, with the result being that the exposure-outcome association is viewed across a narrower exposure range, exacerbating the difficulty of seeing the association shape in the extremes of exposure.

5 Discussion

We have assessed the effect of classical exposure measurement error on the shape of exposure-disease associations using grouped exposure, FP and P-spline analyses.

Random error in exposure measurements results in a change in the shape of the exposure-disease association when it is nonlinear, with observed associations appearing increasingly linear as measurement error increases. The effect is not necessarily an attenuation of the association. The change in shape due to classical measurement error is seen under all three methods of analysis, with only the U-shape being clearly visually detectable under severe measurement error.

Threshold associations could be difficult to detect even when there is no exposure measurement error because of limitations of the analysis methods. Under a grouped exposure analysis there is greater ability to identify the turning point when the number of exposure categories is increased, though this is restricted by sample size. Similarly for a J-shaped association the upward curve for low exposures is easily missed using the grouped exposure analysis and this problem may be lessened by using more exposure categories, but sufficient numbers of individuals with very low exposure would be required. P-splines provided the closest approximation to a threshold shape.

Even in the absence of measurement error, detecting nonlinearity of associations typically requires large sample sizes, as is shown by Table 2. Random measurement error further decreases power to detect nonlinearity. We found that P-splines offered greater power to detect nonlinear associations compared with the grouped exposure and FP methods when the exposure is measured without error, and that for power is severely diminished by random exposure measurement error under all three methods. The relative effect of measurement error on power to detect nonlinearity is somewhat decreased in situations where power would be low using the true exposure.

We showed that MacMahon's method can perform well for quadratic associations, but does not recover true nonlinear exposure-disease associations in general, including threshold associations. To our knowledge these findings are new. In the

original application of MacMahon's method there was concluded to be no evidence of a threshold association between DBP and stroke or coronary heart disease mortality. However, we have shown that the method may not be able to reveal such an association if the error variation is large.

Methods for correcting for exposure measurement error include regression calibration, in which the relevant function of X_i, $g(X_i)$ say, is replaced with $E(g(X_i)|W_{i1})$ in the exposure-disease model (Rosner et al., 1989, 1992, Carroll et al., 2006). The expectation can be found using replicate measurements. Regression calibration has been particularly widely used for linear associations but also extends to nonlinear associations (Carroll et al., 2006, Cheng and Schneeweiss, 1998). Other correction methods for nonlinear models include simulation extrapolation (SIMEX) (Cook and Stefanski, 1994, Carroll et al., 1999, Staudenmayer and Ruppert, 2004), Bayesian methods for P-splines (Berry et al., 2002, Cheng and Crainiceanu, 2009), and approaches using local polynomial estimators (Fan and Truong, 1993, Delaigle et al., 2009).

We have focused on classical measurement error. Different types of error, such as error depending on true exposure, multiplicative error, differential error, that is error in the exposure measurement which depends on the outcome, or misclassification in a categorical exposure, could have different effects. The presence of additional covariates measured with error in the exposure-disease model could also result in different effects on the observed exposure-disease association. We can partly test whether error follows the classical model using replicate measurements (Carroll et al., 2006).

To conclude, investigators should be aware that random error in exposure measurements can mask nonlinear exposure-disease associations and severely decrease power to detect nonlinearity. In light of our results, the use of more complex methods for correcting for the effects of classical measurement error in nonlinear exposure-disease associations should be encouraged.

References

Abbas, S., Chang-Claude, J., and Linseisen, J. (2009). Plasma 25-hydroxyvitamin D and premenopausal breast cancer risk in a German case-control study. *International Journal of Cancer*, 124: 250–255.

Allen, N. E., Beral, V., Casabonne, D., Kan, S. W., Reeves, G. K., Brown, A., and Green, J. (2009). Moderate alcohol intake and cancer incidence in women. *Journal of the National Cancer Institute*, 101(5): 296–305.

Ambler, G. and Royston, P. (2001). Fractional polynomial model selection procedures: Investigation of type I error rate. *Journal of Statistical Computation and Simulation*, 69: 89–108.

Asia-Pacific Cohort Studies Collaboration (2004). Blood glucose and risk of cardiovascular disease in the Asia Pacific region. *Diabetes Care*, 27: 2836–2842.

Bagnardi, V., Zambon, A., Quatto, P., and Carrao, G. (2004). Flexible meta-regression functions for modelling aggregate dose-response data, with an application to alcohol and mortality. *American Journal of Epidemiology*, 159: 1077–1086.

Berry, S. M., Carroll, R. J., and Ruppert, D. (2002). Bayesian smoothing and regression splines for measurement error problems. *Journal of the American Statistical Association*, 97: 160–169.

Carroll, R. J., Maca, J. D., and Ruppert, D. (1999). Nonparametric regression in the presence of measurement error. *Biometrika*, 86: 541–554.

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*, London: Chapman & Hall/CRC, second edition.

Cheng, C. L. and Schneeweiss, H. (1998). Polynomial regression with errors in the variables. *Journal of the Royal Statistical Society (Series B)*, 60: 189–199.

Cheng, Y.-J. and Crainiceanu, C. H. (2009). Cox models with smooth functional effect of covariates measured with error. *Journal of the American Statistical Association*, 104: 1144–1154.

Clarke, R., Shipley, M., Lewington, S., Youngman, L., Collins, R., Marmot, M., and Peto, R. (1999). Underestimation of risk associations due to regression dilution in long-term follow-up of prospective studies. *American Journal of Epidemiology*, 150: 341–353.

Cook, J. R. and Stefanski, L. A. (1994). Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association*, 89: 1314–1328.

D'Agostino, R. B., Belanger, A. J., Kannel, W. B., and Cruikshank, J. M. (1991). Relation of low diastolic blood pressure to coronary heart disease death in presence of myocardial infarction: The Framingham study. *British Medical Journal*, 303: 385–389.

De Boor, C. (2001). *A practical guide to splines*, New York: Springer, revised edition.

Delaigle, A., Fan, J., and Carroll, R. J. (2009). A design-adaptive local polynomial estimator for the errors-in-variables problem. *Journal of the American Statistical Association*, 104: 348–359.

Dosemeci, M., Wacholder, S., and Lubin, J. (1990). Does non-differential misclassification of exposure always bias a true effect toward the null value? *American Journal of Epidemiology*, 132: 746–748.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11: 89–121.

Fan, J. and Truong, Y. K. (1993). Nonparametric regression with errors in variables. *Annals of Statistics*, 21: 1900–1925.

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 beyond which pressure-reduction is dangerous? *Journal of the American Medical Assoiation*, 265: 489–495.

Flegal, K., Keyl, P., and Neito, F. (1991). Differential misclassification arising from nondifferential errors in exposure measurement. *American Journal of Epidemiology*, 134: 1233–1244.

Gardner, M. J. and Heady, J. A. (1973). Some effects of within person variability in epidemiological studies. *Journal of Chronic Diseases*, 26: 781–795.

Govindarajulu, U. S., Spiegelman, D., Thurston, S. W., Ganguli, B., and Eisen, E. A. (2007). Comparing smoothing techniques in Cox models for exposure-response relationships. *Statistics in Medicine*, 26: 2735–2752.

Greenland, S. (1995a). Avoiding power loss associated with categorization and ordinal scores in dose-response and trend analysis. *Epidemiology*, 6: 450–454.

Greenland, S. (1995b). Dose-response and trend analysis in epidemiology: alternatives to categorical analysis. *Epidemiology*, 6: 356–365.

Hughes, M. D. (1993). Regression dilution in the proportional hazards model. *Biometrics*, 49: 1056–1066.

Hurvich, C. M., Simonoff, J. S., and Tsai, C. L. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society (Series B)*, 60: 271–293.

Kuha, J. and Temple, J. (2003). Covariate measurement error in quadratic regression. *International Statistical Review*, 71: 131–150.

Liu, B., Balkwill, A., Roddam, A., Brown, A., and Beral, V. (2009). Separate and joint effects of alcohol and smoking on the risks of cirrhosis and gallbladder disease in middle-aged women. *American Journal of Epidemiology*, 169: 153–160.

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. Part 1, Prolonged differences in blood pressure: Prospective observational studies corrected for the regression dilution bias. *Lancet*, 335: 765–774.

Malloy, E. J., Spiegelman, D., and Eisen, E. A. (2009). Comparing measures of model selection for penalized splines in Cox models. *Computational Statistics and Data Analysis*, 53: 2605–2616.

Poikolainen, K. (1995). Alcohol and mortality: A review. *Journal of Clinical Epidemiology*, 48: 455–465.

Polesel, J., Dal Maso, L., Bagnardi, V., Zucchetto, A., Zambon, A., Levi, F., La Vecchia, C., and Franceschi, S. (2005). Estimating dose-response relationship between ethanol and risk of cancer using regression spline models. *International Journal of Cancer*, 114: 836–841.

Port, S., Demer, L., Jennrish, R., Walter, D., and Garfinkel, A. (2000). Systolic blood pressure and mortality. *Lancet*, 355: 175–180.

Prospective Studies Collaboration (2002). Age-specific relevance of usual blood pressure to vascular mortality: A meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet*, 360: 1903–1913.

Rosner, B., Spiegelman, D., and Willett, W. C. (1992). Correction of logistic regression relative risk estimates and confidence intervals for random within-person measurement error. *American Journal of Epidemiology*, 136: 1400–1413.

Rosner, B., Willett, W. C., and Spiegelman, D. (1989). Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Statistics in Medicine*, 8: 1051–1069.

Royston, P. and Altman, D. G. (1994). Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Applied Statistics*, 43: 429–467.

Royston, P., Ambler, G., and Sauerbrei, W. (1999). The use of fractional polynomials to model continuous risk variables in epidemiology. *International Journal of Epidemiology*, 28: 964–974.

Ruppert, D. (2002). Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics*, 11: 735–757.

Shaper, A. G., Wannamethee, G., and Walker, M. (1988). Alcohol and mortality in British men: explaining the U-shaped curve. *Lancet*, 332: 1267–1273.

Spiegelman, D., McDermott, A., and Rosner, B. (1997). Regression calibration method for correcting measurement-error bias in nutritional epidemiology. *American Journal of Clinical Nutrition*, 65: 1179S–1186S.

Staudenmayer, J. and Ruppert, D. (2004). Local polynomial regression and simulation-extrapolation. *Journal of the Royal Statistical Society (Series B)*, 66: 17–30.

Toniolo, P., Bofetta, P., Shuker, D. E. G., Rothman, N., Hulka, B., and Pearce, N. (1997). *Application of biomarkers in cancer epidemiology*, Lyon: International Agency for Research on Cancer.

Willett, W. (1998). *Nutritional Epidemiology*, Oxford: Oxford University Press, second edition.

Appendices

A Grouped Exposure Analysis for Nonlinear Associations

Consider the model

$$\log\{h(t|X)\} = \log\{h_0(t)\} + \beta g(X), \tag{7}$$

where $g(X)$ is a function of X. Under this model the true log HR at the average exposure level in the kth exposure category is $\beta g\left\{E(X|X^{(k)}=1)\right\}$ (relative to exposure 0 ($g(0)=0$) for simplicity), where $E(X|X^{(k)}=1)$ is the average exposure in category k. However, under the grouped exposure analysis we observe $\beta E\left\{g(X)|X^{(k)}=1\right\}$, which equals $\beta g\left\{E(X|X^{(k)}=1)\right\}$ only when $g(X)$ is a linear function of X.

B MacMahon's Method for Nonlinear Associations

Let $W_1^{(k)}$ equal 1 if W_1 is in the kth exposure category and 0 otherwise. An unbiased estimate of mean exposure for individuals with $W_1^{(k)}=1$ is $E(W_2|W_1^{(k)}=1)$, where W_2 is a repeated measurement. Under model (7), the true log HR for an individual with exposure $E(W_2|W_1^{(k)}=1)$ is $\beta g\{E(W_2|W_1^{(k)}=1)\}$, while the log HR observed using MacMahon's plot is $\beta E\{g(W_2)|W_1^{(k)}=1\}$. The terms $g\{E(W_2|W_1^{(k)}=1)\}$ and $E\{g(W_2)|W_1^{(k)}=1\}$ are equal only when $g(X)$ is a linear function or the replicate W_2 is constant within each exposure category defined by $W_1^{(k)}$, i.e. $\text{var}(W_2|W_1^{(k)}=1)=0$. Errors in 'corrected' log HR estimates plotted under MacMahon's method therefore depend on $\text{var}(W_2|W_1^{(k)}=1)$.