Cubic Spline Regression of J-shaped Dose-Response Curves with Likelihood-based Assignments of Grouped Exposure Levels

Kunihiko Takahashi*1, Hiroyuki Nakao2 and Satoshi Hattori3
1Department of Biostatistics, Nagoya University Graduate School of Medicine, Japan
2Center for Public Health Informatics, National Institute of Public Health, Japan
3Biostatistics Center, Kurume University, Japan

Abstract
In epidemiological studies that measure the risk at different levels of exposure, the data is often only available for the analyses that summarized the response data in grouped exposure intervals. In typical methods, the midpoints are used as the assigned exposure levels for each interval. Results of the analysis with grouped data may be sensitive to the assignment of the exposure levels. In this paper, we propose a procedure for assessing J-shaped associations based on the likelihood-based assignment of values to grouped intervals of exposure, and applying the cubic spline regression models. Numerical illustrations and comparisons based on simulations showed that the proposed procedure can yield better estimates for curves than those obtained using the typical assignment method based on the midpoints of each interval.

Keywords: Dose-response; Spline regression model; Grouped exposure interval; Dose assignment; Likelihood; J-shaped curve

Introduction
In epidemiological studies, it is often necessary to determine the relationship between exposure levels and the risk of disease. However, the data on exposure levels are often available in intervals because they are generally not recorded for each individual subject. For example in studies on the association between alcohol consumption and the risk of disease, researchers often treat the exposure levels as intervals when they interview participants about their consumption levels, but they are unable to obtain the exact values as continuous variables. Also in traditional meta-analysis based on aggregated data, it is not possible to obtain the original data, and the published articles do not include enough data. In such situations, meta-analysis of observational studies often has to rely on the summarized data where the exposure levels are grouped into intervals available from research reports.

Table 1 summarizes data from a study of the association between alcohol consumption and all-cause mortality, which was conducted by Lin et al. [1]. The alcohol intake of current drinkers was classified into five groups: non-drinkers, alcohol intake of 0.1-22.9 g/day, 23.0-45.9 g/day, 46.0-68.9 g/day, and ≥ 69.0 g/day. Table 2 summarizes the characteristics of two studies of coffee consumption and stroke, by Bidel et al. [2] and Grobbee et al. [3]. They were included into a meta-analysis of 11 studies by Larsson and Orsini [4], where the categories used for coffee consumption differed among the studies. Moreover the reference category was assigned to non-drinkers in Grobbee et al. [3], whereas 0-2 cups/day were used in Bidel et al. [2], and hence the meanings of the reported relative risks (RRs) were different and it was inappropriate to combine them directly.

Studies that measure the risk at different levels of exposure are usually analyzed based on a trend estimate by linear (or log-linear) regression analysis. When performing a regression analysis of summarized response data that are grouped into intervals, many researchers use the pre-assigned exposure levels from historical data or the midpoint values of each interval [5]. Results of regression analysis with grouped data may be sensitive to the assignment of the exposure levels. Recently, Takahashi and Tango [6] proposed a method for assigning values by applying the likelihood approach, and they showed the procedure can produce a more accurate linear regression coefficient than the typical procedure which uses the midpoint values.

On the other hand, some studies have reported that the risk of disease has a nonlinear relationship with the exposure level. For example, it is known that the association between alcohol and coronary heart disease [7] or total mortality [8] may be depicted as a J-shaped curve. Some evidence of a J-shaped association has also been reported recently between coffee consumption and the risk of stroke [4]. Curve-fitting methods based on a fractional polynomial model or a spline model has often been applied to these nonlinear dose-response associations for regression [9-11]. Di Catelnuovo et al. [8] applied fractional polynomials to fit the association between alcohol intake and the RR of total mortality in a meta-analysis of 34 prospective studies. On the other hand, for example, Larsson and Orsini [4] performed a dose-response meta-analysis and detected a potentially nonlinear association between coffee consumption and stroke using a cubic spline model, and cubic spline regression models may have many advantages over polynomials [12].

In this paper, we propose a procedure for assessing nonlinear associations between exposure levels and the risk of disease from a summarized grouped data, which is based on the assignment of levels to grouped exposure intervals by applying the likelihood-based assignment procedure proposed in Takahashi and Tango [6]. In particular, we focus on the restricted cubic spline model that was described in Orsini et al. [13] and Larsson and Orsini [4] for J-shaped dose-response curves. We demonstrate how to estimate a J-shaped curve from the grouped summarized data using only four or five class intervals. Also this procedure can provide the log relative risk on each

*Corresponding author: Department of Biostatistics, Nagoya University Graduate School of Medicine, 65 Tsurumai-cho, Showa-ku, Nagoya, 466-8550, Japan, Tel: +81-52-744-2489; Fax: +81-52-744-2488; E-mail: kunihiko@med.nagoya-u.ac.jp

Received November 18, 2013; Accepted November 28, 2013; Published November 30, 2013

Citation: Takahashi K, Nakao H, Hattori S (2013) Cubic Spline Regression of J-shaped Dose-Response Curves with Likelihood-based Assignments of Grouped Exposure Levels. J Biomet Biostat 4: 181. doi:10.4172/2155-6180.1000181

Copyright: © 2013 Takahashi K, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
Table 1: Relative risks of death from all causes as well as alcohol intake among men in the JACC study in Japan [for details, see Lin et al. [1]].

| Alcohol intake (g/day) | No. of individuals | Person-years | No. of deaths | Adjusted RR (log value) | 95% CI |
|------------------------|--------------------|--------------|---------------|------------------------|-------|
| Nondrinkers            | 7,839              | 75,352       | 1,261         | 1.00 (0.000)           | Reference |
| 0.1-22.9               | 6,444              | 69,708       | 616           | 0.80 (-0.223)          | (0.72, 0.88) |
| 23.0-45.9              | 8,072              | 77,757       | 998           | 0.90 (0.105)           | (0.82, 0.98) |
| 46.0-68.9              | 7,085              | 68,622       | 765           | 0.95 (0.051)           | (0.86, 1.04) |
| ≥ 69.0                | 3,337              | 32,076       | 440           | 1.32 (0.278)           | (1.18, 1.49) |

Table 2: Characteristics of two studies of coffee consumption and stroke that were included in a meta-analysis [4].

| Exposure x | No. of individuals | No. of cases | logRR | SE (logRR) |
|------------|--------------------|--------------|-------|------------|
|            |                    |              |       |            |
|            |                    |              |       |            |
|            |                    |              |       |            |
|            |                    |              |       |            |
|            |                    |              |       |            |

Table 3: Grouped exposure intervals and summarized response data in a cohort study.

| Exposure x | No. of individuals | No. of cases | logRR | SE (logRR) |
|------------|--------------------|--------------|-------|------------|
|            |                    |              |       |            |
|            |                    |              |       |            |
|            |                    |              |       |            |
|            |                    |              |       |            |
|            |                    |              |       |            |

Methods

Likelihood based assignment of levels for grouped exposure intervals

In some cohort studies with N individuals, the data on exposure x of each individual i is summarized in a table of grouped exposure interval \( I_j = (u_j, v_j) \) and the corresponding summarized response data with the frequency \( N_j \), the number of cases \( a_j \), the log value of the relative risk \( y_j \), and its standard error \( s_j \). For each interval \( I_j = (u_j, v_j) \), the lower endpoint \( u_j \) is known but the upper endpoint \( v_j \) is unknown. \( N_j \) is given as the number of individuals, \( a_j \) is sometimes shown as the adjusted values of some covariates, \( s_j \) implies the confidence interval of the risks if the standard error of the logRR is not reported [14]. In case-control studies, the number of controls \( b_j \) is also given and \( y_j \) is the log value of the odds ratio (OR).

In typical methods, the midpoints are used as the assigned exposure levels for each interval \( I_j \), \( j=1,2,\ldots,m \) (hereafter likelihood-based assignment). On the other hand, we assign the exposure level based on the summarized data according to the procedure proposed by Takahashi and Tango [6] as follows. We assume that the exposure levels of all individuals in the study are a set of random variables \( x_1, x_2, \ldots, x_N \) and that a power transformation of the exposure, \( x_j^\gamma \), is obtained from a common normal distribution \( N(\mu, \sigma^2) \) with the mean \( \mu \) and variance \( \sigma^2 \), where \( x_j = \log x_j \) for \( \lambda = 0 \). Given this assumption, the frequency \( N_j \) provides a log-likelihood based on a binomial distribution for the distribution of \( x=x_j \). The unknown parameters \( \lambda, \mu, \sigma^2 \) and \( v_j \) can be estimated by maximizing the log-likelihood. Based on the estimated distribution of \( x=x_j \) with the probability density function \( f(x) \) such as \( x^\gamma = N(\mu, \sigma^2) \), the assigned exposure level \( d_j \) for the \( j \)th interval \( I_j = (u_j, v_j) \) is calculated as the mean of its truncated distribution:

\[
d_j = \frac{\int_{u_j}^{v_j} x f(x) dx}{\int_{u_j}^{v_j} f(x) dx} \quad (j=0,1,\ldots,m) \quad \text{(hereafter likelihood-based assignment)}.
\]

Nonlinear association modeling using cubic splines

Splines are smooth functions that can assume virtually any shape, and the most useful type of spline is generally a cubic spline function, which is restricted to be smooth at the junction of each cubic polynomial [12]. In epidemiological studies, a restricted cubic spline model has been often applied to nonlinear dose-response data. As noted by Larsson and Orsini [4], a restricted cubic spline with three knots was recently applied to a potential nonlinear association that was depicted as a J-shaped curve. However, in some studies, such as that in Table 2, the exposure level of the reference group is not assigned on \( x=0 \). Thus, in this paper, we consider a cubic spline model for the log relative risk on the exposure \( x \), \( \logRR(x) \), that satisfies \( \logRR(d_j)=0 \) using assigned exposure levels \( d_j \) for the reference interval \( I_0 = (0, \infty) \) as follows:

\[
\logRR(x) = y = (-\beta_0 d_0 - \beta_1 d_1 + \beta_2 x + \beta_3 x^2) = \beta_0 (x-d_0) + \beta_1 (x-d_1) + \beta_2 x + \beta_3 x^2
\]

where \( x^\gamma = (x-k_1)^{\gamma_1} \gamma (x-k_1)^{\gamma_1} - (1-\gamma)(x-k_1)^{\gamma_1} \), \( d_1 = (d_1-k_1)^{\gamma_1} \gamma (d_1-k_1)^{\gamma_1} - (1-\gamma)(d_1-k_1)^{\gamma_1} \), \( x = (x-k_1)^{\gamma_1} \gamma (x-k_1)^{\gamma_1} - (1-\gamma)(x-k_1)^{\gamma_1} \), \( \gamma = (x_1-k_1)/(x-k_1) \), and \( x = (x-a)^{\gamma_1} \gamma (x-a)^{\gamma_1} - (1-\gamma)(x-a)^{\gamma_1} \).

First, we construct an approximate covariance matrix for the adjusted log relative risks from a fitted model that conforms to the values proposed by Greenland and Longnecker [15], and we construct a variance-covariance matrix \( \Sigma \). In this step, we assume the assigned exposure levels \( d_j \) and standard errors \( s_j \) for each interval \( I_j \) to be fixed. The coefficients \( \hat{b} \) of the restricted cubic spline model (1) are estimated using generalized least-squares regression with \( \Sigma \), i.e.,

\[
\hat{b} = (\hat{\beta}_1, \hat{\beta}_2) (X^\Sigma^{-1} X)^{-1} X^\Sigma^{-1} y
\]

and the estimated variance-covariance matrix of \( \hat{b} \),

\[
V(\hat{b}) = (X^\Sigma^{-1} X)^{-1}
\]

where \( A^T \) and \( A^\dagger \) imply the transpose and inverse matrices of \( A \), and

\[
y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}, \quad X = \begin{bmatrix} d_1 - d_0 & d_1^\gamma - d_0^\gamma \\ d_2 - d_1 & d_2^\gamma - d_1^\gamma \\ \vdots & \vdots \\ d_m - d_{m-1} & d_m^\gamma - d_{m-1}^\gamma \end{bmatrix}
\]

respectively. Note that, if we need to estimate a spline model for the log relative risk on \( x \) relative to \( x=0 \), \( \logRR(x) \), when the reported \( y \) implies that relative to the reference \( x=d_0 \), we can determine the spline model as

\[
\logRR(x) = \hat{\beta}_1 x + \hat{\beta}_2 x^2
\]
using the same \( \hat{b} \) estimated from (1), because \( y = \log(p_i / p_0) = \log(p_i / p_{d_j}) = -\beta d - \beta d^2 \) and \( \log(p_i / p_{d_j}) = -\hat{\beta} d - \hat{\beta} d^2 \) where \( p_0 \) is the probability of being a case with an exposure of \( x = \alpha \).

A crucial problem in spline regression is knot placement [10]. One simple approach is to have the observations \( x \) determine the positions of the knots [16]. Some studies such as those by Larson and Orsini [4], Harrell et al. [12] and Orsini et al. [13], placed knots at fixed percentiles in the data. Therefore, we examine a procedure here for selecting the positions of three knots among the assigned exposure levels \( d_j \) (\( j=1,2,...,m \)) with the exception of \( d_1 \) of the reference group \( I \). Thus, we assign the knots \( k_j = d_j, k_j = d_j, k_j = d_j \) when \( m=3 \). If \( m>3 \), we require a procedure that is based on likelihood. Under the normal assumption of generalized linear regression, \( y \sim N_\Sigma(\beta Xb, \Sigma) \), we can derive the log-likelihood function of \( y \) and evaluate the likelihood of each resulting model for several candidate knot positions. Thus, the putatively better knots \( k_1, k_2, \) and \( k_3 \) are placed in a position that maximizes

\[
I(\hat{b}|k_1, k_2, k_3) = \frac{m}{2} \log(2\pi) - \frac{1}{2} \log|\Sigma| - \frac{1}{2} (y - \hat{X}\hat{b})^T \Sigma^{-1} (y - \hat{X}\hat{b})
\]  

(3)

**Application**

First for the data in Table 1 with person-years for \( N \), the likelihood-based assignment procedure assigned the exposure levels as \( d=0 \), 14.26, 34.20, 56.09, and 86.41 g/day with the estimated parameters \( \hat{\lambda} = 0.5 \) and \( \hat{\nu} = 139.90 \). When using the number of individuals for \( N \), they are assigned as \( d=0 \), 14.27, 34.21, 56.09, and 86.41 g/day with the estimated parameters \( \hat{\lambda} = 0.5 \) and \( \hat{\nu} = 140.00 \). Thus, they differed from the midpoints assignment of \( d=0 \), 11.5, 34.45, 57.45, and 82.8 (as 1.2 times the lower boundary for the open-ended upper category) or 80.45 (as assuming the open-ended upper category has the same amplitude as the adjacent category) g/day for each \( I \). By using the exposure levels of \( d=0 \), 14.26, 34.20, 56.09, and 86.41 g/day, likelihood procedure (3) selected \( k_1 = 14.26, k_2 = 34.20, \) and \( k_3 = 86.41 \) as the knots, and the coefficients were estimated as \( \hat{\beta}_1 = 4.88 \times 10^{-4} \) and \( \hat{\beta}_2 = 5.46 \times 10^{-6} \), respectively, while the midpoints assignment with \( d=80.45 \) estimated the coefficients as \( \hat{\beta}_1 = -2.51 \times 10^{-3} \) and \( \hat{\beta}_2 = -1.36 \times 10^{-5} \) with \( k_1 = 11.5, k_2 = 57.45 \), and \( k_3 = 80.45 \) (Figure 1).

Next for the data in Table 2, the exposure levels were assigned by the likelihood-based assignment as \( d=1.65, 3.56, 5.44, \) and 7.94 cups/day with \( \hat{\lambda} = 2/3 \) and \( \hat{\nu} = 14.4 \) in Bidel et al. [2] where the intervals were assumed as \( 0 \leq x < 2, 2 \leq x < 4, 4 \leq x < 6, \) and \( 6 \leq x < 10, \) and \( 6.5 \leq x, \) respectively.

**Simulation**

In this section, we discuss the simulation studies conducted to assess our proposed procedure, wherein we used cubic spline regression for the nonlinear association with the likelihood-based assignments in grouped exposure intervals. Two cubic spline curves

\[
y = \log RR_0(x) = \log \frac{p_i}{p_0} = -0.2x - 0.05x^2
\]

(4)

with three knots \( (k_1, k_2, k_3) = (1) \) (3, 5, 7) and (II) (2.3, 3.2, 10.5) were assumed to be the true model for the association between exposure \( x \) and the log relative risk \( y \) in a cohort study, respectively, where \( p_0 \) and \( p_i \) are the probabilities of being a case with the exposure \( x \) and \( x=0 \), respectively (Figure 3). We considered interval for grouping, \( 0 \leq x < 2 \), \( 2 \leq x < 4 \), \( 4 \leq x < 6 \), and \( x \geq 6 \), such as that the number of knots in each interval \( I \) was at most one for curve I, whereas the interval \( 2 \leq x < 4 \) had two knots, i.e., \( k_1 = 2.3 \) and \( k_2 = 3.2 \), for curve II.
We set the population size \( N=2000 \) and the probability \( p=0.05 \) for the reference \( x=0 \). We generated a set of exposures \( x=[x_1, x_2, \ldots, x_{2000}] \) from a truncated normal distribution of \( N(3.5, 8.0) \) with the interval \( 0 \leq x \leq 12 \). By calculating \( p_i \), we generated a set of 1 or 0 Bernoulli random numbers \( w=[w_1, w_2, \ldots, w_{2000}] \) using \( \Pr\{W_i=1\}=p_i \) for each \( x_i \), where the sample was counted as a case when \( w=1 \), and we summarize the generated data \( (x, w) \) in Table 4. Note that each relative risk was calculated relative to the reference group \( 0 \leq x \leq 2 \).

We compared the following four procedures:

(i) The likelihood-based assignment and the three knots were set at them:
\[
(d_0, d_1, d_2, d_3)=(1.22, 3.0, 4.91, 7.96)
\]

(ii) The midpoints assignment with \( d_0=1.0 \) and the three knots were set at them:
\[
(d_0, d_1, d_2, d_3)=(1.0, 3.0, 5.0, 7.0)
\]

(iii) The midpoints assignment with \( d_0=0.0 \) and the three knots were set at them:
\[
(d_0, d_1, d_2, d_3)=(0.0, 3.0, 5.0, 7.0)
\]

(iv) The likelihood-based assignment \( (d_0, d_1, d_2, d_3)=(1.22, 3.0, 4.91, 7.96) \), but the three knots were fixed on the “true knots” \( (k_0, k_1, k_2)=(3.0, 5.0, 7.0) \) for true curve I and \( (k_0, k_1, k_2)=(2.3, 3.2, 10.5) \) for true curve II.

Procedure (i) is our proposed method, and procedures (ii) and (iii) are typical methods that use midpoints, where the highest interval was assigned by assuming that the boundary had the same amplitude as the adjacent category [4]. In addition, we compared the proposed method with procedure (iv), which has true knots. Note that for true curve I, the positions of the three knots in curves according to procedures (ii) and (iii), \( (k_0, k_1, k_2)=(3.0, 5.0, 7.0) \), were in the same positions as the true knots.

We generated \( B=1000 \) sets of \( w_1, w_2, \ldots, w_B \) for the fixed \( x \). The curves were estimated using procedures (i-iv) for each set, and we estimated \( \log RR(x) \) as \( \hat{y}_b(x) = y_b(x) \) for each point of \( x=1.2, \ldots, 12 \). Comparing with the values derived from the true model (4), Tables 5 and 6 show the bias \( \text{Bias}(\hat{y}(x)) = \frac{1}{B} \sum_{b=1}^{B} \hat{y}_b(x) - y(x) \).

and the mean squared error \( \text{MSE}(\hat{y}(x)) = \frac{1}{B} \sum_{b=1}^{B} (\hat{y}_b(x) - y(x))^2 \).

and the coverage probability \( \text{CP}(x) \) for a 95% confidence interval of \( B=1000 \) sets for each \( x=1.2, \ldots, 12 \) based on the fitted curves produced by each procedure using the summarized data for true curves I and II, respectively. The means and standard deviations of estimated coefficients \( \hat{b} = (\hat{\beta}_1, \hat{\beta}_2) \) are also shown as \( \text{E}(\hat{b}) \) and \( \text{SD}(\hat{b}) \) in the tables. Note that the accuracy of the fitted curves cannot be measured based only on \( \hat{b} \), because \( \hat{b} \) and the accuracy must be affected directly by the positions of the knots. Figure 4 shows the curves joining the mean values
\[
\frac{1}{B} \sum_{b=1}^{B} \hat{y}_b(x).
\]

First, procedure (iv), which used the likelihood-based assignment and the true knots, had a small bias and small MSE values in each scenario. It also gave closer coverage probabilities to 0.95, and \( E(b) \) was close to true \( b \). Overall the curve produced using this procedure had a good fit to the true curve.

Next, the proposed procedure (i), which used the likelihood-based assignment and the selected knots, also had a small bias, small MSE values for each \( x \). In particular, when \( x \geq 9 \) for true curve I, it was shown that the curve produced using this procedure had a better fit to the true curve than that produced using procedure (iv). When \( x < 5 \) for true curve II, this procedure had a slightly higher bias than procedure (iv). However, the coverage probabilities remained higher than 90%.

Procedure (ii), which used the midpoints assignment, was very similar to procedure (i) when \( x < 5 \) for curves I and II. When \( x > 5 \), the bias and MSE increased gradually, so this procedure could not deliver stable estimates with a large \( x \).

Procedure (iii), which used the midpoints assignment and \( d_0=0 \), estimated the value slightly higher with a low \( x \) compared with the other procedures. With a high \( x \), it showed similar behavior to procedure (ii).

**Discussion**

In this paper, we proposed a procedure for assessing the nonlinear association between exposure levels and the risk of disease by assigning exposure levels to grouped exposure intervals. In particular, we focused on a restricted cubic spline model for J-shaped dose-response curves. The procedure can be applied to the log relative risks when they are given relative to the reference point \( x=0 \) and also to the interval \( x \in I_b \).
Our simulation results showed that the estimated curve was sensitive to the assignment, and the likelihood-based assignment could estimate the nonlinear association accurately. It showed that the proposed procedure using the likelihood-based assignment had a lower bias when used for estimation compared with other procedures that used the midpoints assignment.

One of the applications to estimate a dose-response association from summarized data is that it should be applicable to a meta-analysis. In general, the exposure categories were different in the
studies so they should not be combined. Furthermore, in some of the meta-analysis studies described in Table 2, the reference category was also different, so it was inappropriate to combine them directly. Some methods have been discussed for meta-analysis to obtain the pooled estimate without estimating the association within individual studies. For example, Greenland and Longnecker [15] described the pool-first method for meta-analysis of trend involving data pooling before trend analysis. On the other hand, Rota et al. [11] proposed a random-effects meta-regression model for nonlinear dose-response relationship fitting second-order fractional polynomial models, where the two-step procedures requires initially fitting second-order fractional polynomial models within each study, and then pooling the study-specific two trend components. They tried to fit a pool-first method on the data of a small number of studies, and they obtained the identical results achieved by using their random-effects approach. It may also be possible to estimate pooled curves using our proposed procedure by a multivariate random effect meta-analysis [17], as discussed in Larsson and Orsini [4]. However, it has to be noted that both pool-first method and two-step procedure use pre-assigned exposure levels for grouped exposure intervals, and they assigned values by the midpoints assignment. Thus, the likelihood-based assignment could give different results in the pooled estimates. Moreover, in situations such as those for assessment of publication or other availability bias by the funnel plot, it is important to accurately estimate individually.

In the procedure reported herein, we fixed three knots in the J-shaped curve. The choice of the location of the knots is a crucial problem and the estimate of the curve was sensitive to their positions. Although we showed simulation results only for a situation of m=3 here, we also examined a procedure where the choice was based on likelihood for m=3, and the results of simulation studies could show that it produced well-fitted curves. In a similar manner to the procedure proposed herein, we can apply restricted cubic spline regression using other numbers of fixed knots to produce a more flexible curve shape. However, the choice of the number of knots is generally a crucial problem in spline regression. In this situation, the model with the best fit can be selected using a similar procedure by evaluating Akaike’s Information Criterion (AIC), which is a penalized likelihood that takes into account the number of parameters estimated in the model based on likelihood (3). Also non-cubic spline models have been discussed in epidemiological studies. Further discussions of such models, including evaluations of different methods, are required in the future.

Our work, moreover, could be located in the errors-in-variable field, which aims to correct for bias that arises if measurement error in x is ignored. The statistical approaches developed in those fields might be applied in the situation discussed in this paper. We would also like to leave such a study including comparisons with the proposed procedure here, in our future work.

References
1. Lin Y, Kikuchi S, Tamakoshi A, Wakai K, Kawamura T, et al. (2005) Alcohol consumption and mortality among middle-aged and elderly Japanese men and women. Ann Epidemiol 15: 590-597.
2. Bidel S, Hu G, Qiao Q, Jousilahti P, Antikainen R, et al. (2006) Coffee consumption and risk of total and cardiovascular mortality among patients with type 2 diabetes. Diabetologia 49: 2618-2625.
3. Grobbee DE, Rimm EB, Giovannucci E, Colditz G, Stampfer M, et al. (1990) Coffee, caffeine, and cardiovascular disease in men. N Engl J Med 323: 1026-1032.
4. Larsson SC, Orsini N (2011) Coffee consumption and risk of stroke: a dose-response meta-analysis of prospective studies. Am J Epidemiol 174: 993-1001.
5. Berlin JA, Longnecker MP, Greenland S (1993) Meta-analysis of epidemiologic dose-response data. Epidemiology 4: 218-228.
6. Takahashi K, Tango T (2010) Assignment of grouped exposure levels for trend estimation in a regression analysis of summarized data. Stat Med 29: 2605-2616.
7. Corrao G, Rubbiati L, Bagnardi V, Zambon A, Poikolainen K (2000) Alcohol and coronary heart disease: a meta-analysis. Addiction 95: 1505-1523.
8. Di Castelnuovo A, Costanzo S, Bagnardi V, Donati MB, Iacoviello L, et al. (2006) Alcohol dosing and total mortality in men and women: an updated meta-analysis of 34 prospective studies. Arch Intern Med 169: 2437-2445.
9. Greenland S (1995) Dose-response and trend analysis in epidemiology: alternatives to categorical analysis. Epidemiology 6: 356-365.
10. Bagnardi V, Zambon A, Quatto P, Corrao G (2004) Flexible meta-regression functions for modeling aggregate dose-response data, with an application to alcohol and mortality. Am J Epidemiol 159: 1077-1086.
11. Rota M, Bellocco R, Scotti L, Tramacere I, Jenab M, et al. (2010) Random-effects meta-regression models for studying nonlinear dose-response relationship, with an application to alcohol and esophageal squamous cell carcinoma. Stat Med 29: 2679-2687.
12. Harrell FE Jr, Lee KL, Pollock BG (1988) Regression models in clinical studies: determining relationships between predictors and response. J Natl Cancer Inst 80: 1198-1202.
13. Orsini N, Li R, Wool A, Khudyakov P, Spiegelman D (2012) Meta-analysis for linear and nonlinear dose-response relations: examples, an evaluation of approximations, and software. Am J Epidemiol 175: 66-73.
14. Greenland S (1987) Quantitative methods in the review of epidemiologic literature. Epidemiol Rev 9: 1-30.
15. Greenland S, Longnecker MP (1992) Methods for trend estimation from summarized dose-response data, with applications to meta-analysis. Am J Epidemiol 135: 1301-1309.
16. Hastie T, Tibshirani R, Friedman J (2009) The Elements of Statistical Learning. (2nd edn), Springer, New York, USA.
17. Jackson D, White IR, Thompson SG (2010) Extending DerSimonian and Laird’s methodology to perform multivariate random effects meta-analyses. Stat Med 29: 1282-1297.