

# Multivariate meta-analysis for non-linear and other multi-parameter associations

A. Gasparrini,<sup>a\*†</sup> B. Armstrong<sup>b</sup> and M. G. Kenward<sup>a</sup>

In this paper, we formalize the application of multivariate meta-analysis and meta-regression to synthesize estimates of multi-parameter associations obtained from different studies. This modelling approach extends the standard two-stage analysis used to combine results across different sub-groups or populations. The most straightforward application is for the meta-analysis of non-linear relationships, described for example by regression coefficients of splines or other functions, but the methodology easily generalizes to any setting where complex associations are described by multiple correlated parameters. The modelling framework of multivariate meta-analysis is implemented in the package *mvmeta* within the statistical environment R. As an illustrative example, we propose a two-stage analysis for investigating the non-linear exposure–response relationship between temperature and non-accidental mortality using time-series data from multiple cities. Multivariate meta-analysis represents a useful analytical tool for studying complex associations through a two-stage procedure. Copyright © 2012 John Wiley & Sons, Ltd.

**Keywords:** meta-analysis; multivariate analysis; multivariate meta-analysis; non-linear; splines

## 1. Introduction

Meta-analysis is a standard, well-grounded statistical procedure for combining the evidence from independent studies that address the same research hypothesis [1]. This methodology was developed originally for pooling the results from published observational or experimental studies, for which individual data were not available. Recently, meta-analysis has been described more broadly as a research synthesis method, with the aim of estimating an average association across studies and to explore the degree and sources of heterogeneity [2]. The analytical approach adopted in this context may be described as a two-stage hierarchical procedure: in the first stage, study-specific estimates of the association of interest are derived from individual data, controlling for individual-level covariates; in the second stage, these estimates are combined across studies, optionally exploring the association with study-level predictors. The two-stage approach, a specific form of individual patient data (IPD) meta-analysis, has been shown to be a flexible and computationally efficient method [3] and has been adopted in different contexts: to pool estimates from multiple randomized controlled trials [4]; to combine results from survival models on time-to-event data in multi-centre cohorts [5]; and to synthesize associations from Poisson time-series models in multi-city analyses [6].

The common approach to two-stage meta-analysis consists of summarizing the association in a single parameter estimate from the first stage, optionally controlling for individual-level confounders. This procedure allows standard meta-analytic techniques to be applied. However, complex associations, such as non-linear exposure–responses, are usually described with functions defined by multiple parameters and require more sophisticated meta-analytical approaches capable of handling the multivariate nature of the summary estimates. Multivariate meta-analysis, a method originally developed to pool multiple

<sup>a</sup>Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, U.K.

<sup>b</sup>Department of Social and Environmental Health Research, London School of Hygiene and Tropical Medicine, London, U.K.

\*Correspondence to: Antonio Gasparrini, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, U.K.

†E-mail: antonio.gasparrini@lshtm.ac.uk

correlated outcomes in randomized controlled trials [7–9], provides a platform to extend the standard two-stage meta-analytical approach.

The aim of this article is to formalize the application of multivariate meta-analytic techniques to the synthesis of multi-parameter associations from two-stage hierarchical analyses, describing the statistical framework, methodological issues, limitations and research directions. This contribution originates from a commentary published in this journal [10] to the paper by Jackson and collaborators on multivariate meta-analysis [11]. The article also offers the opportunity to describe the implementation in the package *mvmeta* within the R software [12], designed to perform multivariate meta-analysis and meta-regression in this and other contexts. The document is structured as follows. In Section 2, we introduce an example to illustrate the application of the methodology, consisting of a two-stage meta-analysis of non-linear temperature–mortality associations in 20 US cities. We describe the statistical methodology in the next two sections: in Section 3, we introduce in general terms the first-stage analysis, and we illustrate the modelling framework of multivariate meta-analysis in Section 4, with a specific focus on the setting of multi-parameter associations. We describe the results in Section 5 and emphasize specific methodological issues in Section 6. Finally, we provide a general discussion in Section 7, also reviewing previous research on the topic. The Supplementary Web Appendix contains additional information on the software and the complete R code to replicate the results of the analysis illustrated in Section 5.

## 2. Motivating example

In this section, we describe an example of the application of multivariate meta-analysis for multi-parameter associations. Specifically, we apply the framework to the combination of estimates of the non-linear exposure–response relationship between outdoor temperature and non-accidental mortality, using time-series data from 20 cities in the USA. The example is illustrative, providing a context for the statistical problem, and is not meant to provide substantive evidence on the topic. Subject-specific methodological issues are discussed elsewhere [13]. The illustration is presented in general terms, emphasizing the applicability of the statistical framework beyond the specific research field, first-stage models or multi-parameter functions discussed in this example.

### 2.1. Temperature–health associations

The health effects of exposure to extreme temperatures have been frequently investigated, especially regarding the association with high temperatures and heat waves [14]. Analyses are typically based on a time-series design, comparing averaged daily temperature with aggregated daily outcomes, often mortality or morbidity counts occurring in a specific city. Most studies have reported an increased risk for both cold and hot temperatures and have used a variety of functions to describe such non-linear association in regression models [15]. The association has shown a strong geographical pattern, linked with climatological, socio-economic and demographic factors of each city [16]. This issue has prompted the use of multi-city analyses, where the results from different cities are pooled together, and explanation for differences is sought through reference to study-specific characteristics such as those listed previously. Individual data, corresponding to series of daily observations from different cities, are usually available, and the analytical setting corresponds to the two-stage framework detailed in Section 1. Here the terms *study* and *city* may be used interchangeably to refer to second-stage units.

Different analytical approaches have been proposed for modelling the health effects of temperature in multi-study settings [13]. Traditional methods rely on the simplification of the dependency, assuming a linear association beyond a threshold [17]. An alternative uses non-linear representations such as splines or polynomials but reduces the estimand to a simple comparison between two specific temperatures [18]. Both these approaches are motivated by the need to summarize the study-specific associations in a single parameter for meta-analysis. Such simplifications are, however, based on strong assumptions about the exposure–response shape or, alternatively, only provide a partial picture of possibly complex dependencies. This supports the development of methods allowing the meta-analysis of truly non-linear multi-parameter relationships.

### 2.2. Data

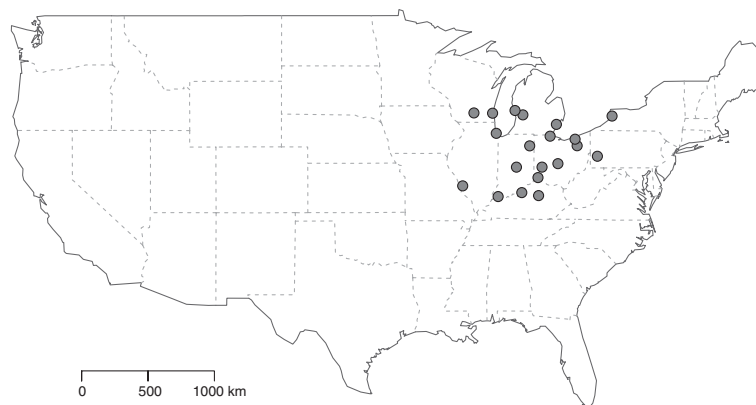
The multi-city time-series data used in this analysis were collected as part of the National Morbidity, Mortality and Air Pollution Study (NMMAPS) (<http://www.ihapss.jhsph.edu>). This publicly available database contains, among other information, daily series of mortality counts and weather and pollution

measurements totalling 5114 observations for the period 1987–2000 in each of 108 cities in the USA. For the purpose of this illustrative example, we restrict the analysis to the 20 cities illustrated in Figure 1, situated in the industrial Midwest region, as defined in the original study.

In addition, the database includes city-level measures of several variables on geographical, climatological, demographic and socio-economic characteristics. A summary of average non-accidental mortality, temperature and other study-specific variables across the 20 cities is reported in Table I.

### 2.3. Modelling strategies

The analysis is performed in two stages, as described in the framework summarized previously. In the first stage, a regression model is applied to individual data for each of the 20 cities included in the analysis in order to derive study-specific estimates of the non-linear exposure–response association, as described in detail in Section 3. The estimated study-specific relationship is entirely defined by the parameters of the function chosen for representing the association, namely, a set of regression coefficients. These coefficients are then used as outcomes for the multivariate meta-analytical model in the second stage, following the methodology described in Section 4. As in conventional univariate meta-analysis, the main aim is to derive a set of regression coefficients defining an average exposure–response association across the studies. In a second step, we aim to test and quantify the amount of heterogeneity and to assess the extent to which this heterogeneity is related to study-specific characteristics, such as those listed in Table I, through multivariate meta-regression.



**Figure 1.** Map of the 20 cities in the USA corresponding to the studies included in the analysis. The cities are located in the industrial Midwest region, as defined in the original National Morbidity, Mortality and Air Pollution Study.

Table I. Mean, range and specific percentiles for study-level variables in 20 US cities, 1987–2000.					
	Mean	Min	25th	75th	Max
Average non-accidental mortality counts	22.3	4.5	9.2	23.3	115.4
Mean temperature (°C)	11.0	8.4	9.9	12.0	14.3
Minimum temperature (°C)	−24.3	−28.6	−25.0	−22.8	−20.8
Maximum temperature (°C)	31.8	29.4	31.0	32.4	34.4
Latitude (degree north)	40.8	38.0	39.6	42.5	43.2
Population size (×100 000)	9.7	1.7	4.1	9.8	53.8
Urbanized population (%)	94.0	73.9	90.7	99.1	100.0
Population with high education (%)	82.8	71.3	81.6	85.7	92.2
Population living in poverty (%)	12.6	8.9	11.2	13.2	24.6
Population unemployed (%)	6.0	3.8	5.0	6.5	11.3

### 3. First-stage model

We suppose that  $N_i$  observations, indexed by  $t = 1, \dots, N_i$ , have been collected in study  $i$  for each of the  $i = 1, \dots, m$  studies. Specifically, the data include observations on a response variable  $Y_{ti}$ , an exposure of interest  $z_{ti}$ , and an optional set of  $J$  confounding variables  $c_{tji}$ , with  $j = 1, \dots, J$ . The aim of the first-stage analysis is to define and estimate the association between  $Y_{ti}$  and  $z_{ti}$ , while controlling for  $c_{tji}$ . This association is specified through a function  $s$ , entirely defined in each study by the  $k$ -length vector of parameters  $\theta_i$ .

Alternative options are available regarding the choice of the function  $s$ , including different types of splines [19], polynomials or fractional polynomials [20], step functions, segmented regression [21] and other multi-parameter functions. Also, the function  $s$  may describe multi-parameter associations other than non-linearity, although the latter is likely to represent the principal application. Similarly, the estimation of the regression parameters is usually performed through a regression model, such as generalized linear or Cox proportional hazard models, whose specification is dependent on the study design and type of outcome. Whatever the modelling choices for the function  $s$  or type of regression, the purpose is to obtain for each study the  $k$ -length vector of estimated parameters  $\hat{\theta}_i$  and accompanying  $k \times k$  estimated (co)variance matrix  $S_i$ . These are assumed to be derived from an estimator for the unknown parameters  $\theta_i$ , which define the true study-specific associations. The second-stage multivariate meta-analytical framework, described in Section 4, is general and not dependent on choices about the first-stage model and parameterization.

#### 3.1. First-stage model used in our example

In our example, the first-stage analysis is carried out using a generalized linear model with quasi-Poisson family for overdispersed data, following a standard analytical approach for time-series environmental data [15, 22]. The general algebraical definition is given by the following:

$$g(\mu_{ti}) = \alpha + s(z_{ti}; \theta_i) + \sum_{j=1}^J h_{ji}(c_{tji}; \gamma_{ji}), \quad (1)$$

where  $g$  is a log link function of the expectation  $\mu_{ti} \equiv E(Y_{ti})$ , with  $Y_{ti}$  as the series of 5114 non-accidental daily mortality counts. The exposure  $z_{ti}$  corresponds to outdoor mean daily temperature over lag 0–5, computed as the moving average on day  $t$  and on the previous 5 days, in order to account for lagged effects. Here, the exposure–response function  $s$  is chosen as a quadratic B-spline, entirely defined by  $k - 2$  internal knots and 2 boundary knots, where  $k$  corresponds to the dimension of the spline basis and the number of parameters. This is equal to the degrees of freedom ( $df$ ) spent in the first-stage model to estimate the relationship. Number and location of knots are chosen by the selection method detailed in Section 3.3.

The model also includes two potential confounding variables  $c_{tji}$ , included a priori, represented by elapsed time and day of the week, modelled through the functions  $h_{ji}$  and related parameters  $\gamma_{ji}$ . Elapsed time is used to control for seasonal and long-term trends and specified through a natural cubic B-spline with 7  $df$  per year. Day of the week is modelled with six indicator variables through a dummy parameterization.

#### 3.2. Definition of the exposure–response function

In Equation (1), while defining the first-stage regression model, we intentionally omitted the index  $i$  for the function  $s$  to indicate that the same function needs to be applied in all the study-specific models. This requirement assures that the estimated coefficients  $\hat{\theta}_i$  retain the same interpretation in all the  $m$  studies, so that their meta-analysis can provide meaningful and interpretable results. In our example, this condition is met by placing the internal and boundary knots at the same temperatures in all the  $m$  studies. Functions for non-linear associations are usually centred on a specific value, so that the results may be interpreted as the effect of the exposure versus a reference [23]. The choice of the centering value depends on interpretational issues and does not affect the fit of the model in (1).

### 3.3. Control for confounders and model selection

The specification of the study-specific models in (1) can follow two different approaches regarding the control of confounders. In this example, the confounder model, although involving several terms, is conceptually simple, and, consistently with previous time-series environmental studies on the same dataset [22], an identical set of pre-specified predictors is included in all studies. An alternative approach is based on study-specific confounder models. The choice depends, among other issues, on the research setting and the quality of the data. Whatever the strategy adopted to control for confounding in the first stage, the aim is to obtain a valid estimate of the study-specific parameters  $\theta_i$ , and this choice does not affect the development of the second-stage multivariate meta-analysis.

Different selection methods are available for the definition of the function  $s$ . In this illustration, the choice reduces to the identification of the optimal number and location of knots for the quadratic B-spline. In the context of time-series environmental studies, several criteria have been proposed, although consensus as to an optimal approach has not been reached yet. Here we rely on the Q-AIC, a modification of the Akaike information criterion (AIC) for quasi-likelihood models [24]. We first define a list of model candidates, specified by an increasing number of internal knots placed at temperatures corresponding to optimal average percentiles across cities [25, Section 2.4]. Boundary knots are placed at temperatures corresponding to the average minimum and maximum temperature. The selected model has the minimum value of the sum of the Q-AIC in all the 20 studies. As with other issues discussed previously concerning the first-stage model, different selection criteria may be preferred in different settings.

### 3.4. Absolute and relative scales

The meta-analysis of exposure–response relationships requires that the exposure is measured on the same scale in all the studies. Also, in the presence of different exposure ranges, common knots placement may leave some parameters inestimable in some studies. The meta-analysis of their parameters therefore presents additional complications.

The modelling strategy described previously defines the exposure on an *absolute* scale ( $^{\circ}\text{C}$ ). The function  $s$ , common to all studies, is coherently defined with knots placed at the same absolute exposure values, so that the estimated coefficients have the same interpretation across studies. Their synthesis through meta-analysis is therefore meaningful. In our example, selected cities have largely overlapping temperature ranges, as showed in Table I, and the specification of common interior and boundary knots does not raise important issues. However, in analyses including studies with more varied temperature ranges, such as of all the 108 US cities, the definition of common knots would be problematic or impractical.

An alternative is to adopt a *relative* scale, standardizing the study-specific distributions by transforming temperatures to the related study-specific percentiles. This method allows the comparison of studies showing even non-overlapping ranges. The standardization is carried out by selecting the internal knots at the same percentiles in all the studies, which indeed correspond to different absolute temperatures. The interpretation of the estimated coefficients changes, and the comparison must be made on the relative scale of percentiles, as illustrated while commenting on the results in Section 5.4.

## 4. Multivariate meta-analysis

The study-specific estimates obtained from the first-stage model are then combined through multivariate meta-analysis. The theoretical arguments that underpin the definition of this modelling framework closely follow the simple univariate model, recently re-evaluated in detail [26]. A thorough overview on the multivariate extension has been provided in a recent paper [11]. However, in contrast to the original setting of randomized controlled trials, in the current multi-parameter setting it is not necessary that the parameters are individually interpretable, and the association is instead characterized through their joint distribution. This specific feature constitutes the object of the re-assessment we provide in this section, often referring to the example described in Section 2.

### 4.1. The model

The framework we use is nested within that of the multivariate normal linear mixed model and so follows well-developed lines [27]. It will be presented in the specific context of multi-parameter associations.



As anticipated previously, we assume that a first-stage model has been fitted to the data from each of the  $i = 1, \dots, m$  studies, obtaining  $k$ -length vectors of regression coefficients  $\hat{\theta}_i$ , and associated  $k \times k$  estimated (co)variance matrices  $\mathbf{S}_i$ . These regression coefficients are then used as outcomes for the second stage and are termed from here on as outcome parameters in order to distinguish them from the coefficients of the second-stage meta-analytic model.

Following Jackson and colleagues [11], a model for random-effect multivariate meta-analysis can be written as follows:

$$\hat{\theta}_i \sim N_k(\boldsymbol{\theta}, \mathbf{S}_i + \boldsymbol{\Psi}), \quad (2)$$

with  $\mathbf{S}_i + \boldsymbol{\Psi} = \boldsymbol{\Sigma}_i$ . The marginal model in (2) has independent within-study and between-study components. In the within-study component, the estimated  $\hat{\theta}_i$  is assumed to be sampled with error from  $N_k(\boldsymbol{\theta}_i, \mathbf{S}_i)$ , a multivariate normal distribution of dimension  $k$ , where  $\boldsymbol{\theta}_i$  is the vector of true unknown outcome parameters for study  $i$ . In the between-study component,  $\boldsymbol{\theta}_i$  is assumed sampled from  $N_k(\boldsymbol{\theta}, \boldsymbol{\Psi})$ , where  $\boldsymbol{\Psi}$  is the unknown between-study (co)variance matrix. Here  $\boldsymbol{\theta}$  can be interpreted as the population-average outcome parameters, namely the coefficients of the function  $s$  defining the average exposure–response association. Model (2) can be extended to multivariate meta-regression, where the  $k$  outcomes are modelled in terms of study-level variables defining a set of  $p$  meta-predictors  $\mathbf{x}_i = [x_{1i}, x_{2i}, \dots, x_{pi}]^T$  associated with the  $i$ th study, where usually  $x_{1i} = 1$  specifies intercept terms. Algebraically,

$$\hat{\theta}_i \sim N_k(\mathbf{X}_i \boldsymbol{\beta}, \mathbf{S}_i + \boldsymbol{\Psi}). \quad (3)$$

The  $k \times kp$  block-diagonal matrix  $\mathbf{X}_i$ , of rank  $kp$ , is derived by the Kronecker product between an identity matrix  $\mathbf{I}_{(k)}$  of dimension  $k$  and the vector  $\mathbf{x}_i$ , following

$$\mathbf{X}_i = \mathbf{I}_{(k)} \otimes \mathbf{x}_i^T = \begin{Bmatrix} x_{1i} & x_{2i} & \dots & x_{pi} & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & \dots & x_{1i} & x_{2i} & \dots & x_{pi} \end{Bmatrix}. \quad (4)$$

The  $kp$ -dimensional coefficient vector  $\boldsymbol{\beta}$  defines the association of the  $k$  outcomes with  $p$  meta-predictors. They commonly include  $k$  intercepts with a similar interpretation of  $\boldsymbol{\theta}$  in (2) but related to the population average of studies characterized by a zero value of meta-predictors. The other  $k(p-1)$  coefficients in  $\boldsymbol{\beta}$  express how the outcome parameters change in respect to the meta-predictors. The problem can also be re-expressed in the form of a conventional linear mixed model, defining random effects  $\mathbf{u}_i \sim N_k(\mathbf{0}, \boldsymbol{\Psi})$ , which represent study-specific deviations from the average. The model in (3) is then written as follows:

$$\hat{\theta}_i | \mathbf{u}_i \sim N_k(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{u}_i, \mathbf{S}_i). \quad (5)$$

The matrix  $\boldsymbol{\Psi}$  is completely defined by a set of parameters  $\boldsymbol{\xi}$ , dependent on the chosen structure and parameterization. If no a priori structure is assumed,  $k(k+1)/2$  terms are needed. Optionally, under the assumption that each outcome parameter is explained only in terms of a subset of the  $p$  variables, the related columns of  $\mathbf{X}$  and entries of  $\boldsymbol{\beta}$  can be excluded, defining different linear predictors. Fixed-effect meta-analytic models presuppose that no heterogeneity exists in the distribution of the outcome parameters and that the random variability is explained only by sampling error, assuming  $\boldsymbol{\Sigma}_i \equiv \mathbf{S}_i$ . As for the univariate case, estimation procedures treat  $\mathbf{S}$  as known. From here on, the model in (2) for multivariate meta-analysis will be considered a special case of (3)–(5) where  $\mathbf{X} \equiv \mathbf{I}$  and  $\boldsymbol{\beta} \equiv \boldsymbol{\theta}$ . The unknown parameters are therefore  $\boldsymbol{\beta}$  and, for random-effect meta-analytic models,  $\boldsymbol{\xi}$ .

#### 4.2. Estimation

Different estimation methods have been proposed for random-effect multivariate meta-analysis: likelihood-based methods [9, 28], estimating equations [29], variants of iterative generalized least squares [8, 30], Bayesian approaches [31] and multivariate extensions of the method of moments [32]. Here we will concentrate on maximum likelihood (ML) and restricted maximum likelihood (REML), following an extensive literature within the framework of linear mixed models [27, 33, 34]. These methods are implemented in the R package *mvmeta* and applied to perform the analysis in Section 5.

The marginal log-likelihood function  $\ell(\boldsymbol{\beta}, \boldsymbol{\xi})$  for model (5) may be written as follows [27]:

$$\ell(\boldsymbol{\beta}, \boldsymbol{\xi}) = -\frac{1}{2}n \log \pi - \frac{1}{2} \sum_{i=1}^m \log |\boldsymbol{\Sigma}_i| - \frac{1}{2} \sum_{i=1}^m \left[ (\hat{\boldsymbol{\theta}}_i - \mathbf{X}_i \boldsymbol{\beta})^\top \boldsymbol{\Sigma}_i^{-1} (\hat{\boldsymbol{\theta}}_i - \mathbf{X}_i \boldsymbol{\beta}) \right], \quad (6)$$

with  $\boldsymbol{\Sigma}_i(\boldsymbol{\xi})$  written here as  $\boldsymbol{\Sigma}_i$  for ease of notation and  $n$  as the total number of observations (usually equal to  $km$  where there are no missing values). Assuming that  $\boldsymbol{\xi}$ , and therefore  $\boldsymbol{\Psi}$  and  $\boldsymbol{\Sigma}$ , are known, the ML estimates for  $\boldsymbol{\beta}$  and its (co)variance matrix  $V(\hat{\boldsymbol{\beta}})$  conditional on  $\boldsymbol{\xi}$  are obtained by maximizing (6). In this case, closed-form equations are given by generalized least squares estimators:

$$\begin{aligned} \hat{\boldsymbol{\beta}}(\boldsymbol{\xi}) &= \left( \sum_{i=1}^m \mathbf{X}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \right)^{-1} \sum_{i=1}^m \mathbf{X}_i^\top \boldsymbol{\Sigma}_i^{-1} \hat{\boldsymbol{\theta}}_i, \\ V(\hat{\boldsymbol{\beta}}) &= \left( \sum_{i=1}^m \mathbf{X}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \right)^{-1}. \end{aligned} \quad (7)$$

When  $\boldsymbol{\Psi}$  is not known, the joint likelihood function in (6) needs to be maximized with respect to both  $\boldsymbol{\beta}$  and  $\boldsymbol{\xi}$ , and iterative methods are required. However, the ML estimator of the (co)variance parameters  $\boldsymbol{\xi}$  is usually biased downward, as it does not account for the loss of degrees of freedom from the estimation of  $\boldsymbol{\beta}$ . An alternative estimator can be obtained from the maximization of the log-likelihood function on the basis of a set of  $n - q$  linearly independent error contrasts, with  $q$  as the number of fixed effects coefficients in  $\boldsymbol{\beta}$ . This restricted log-likelihood (REML) function  $\ell_R(\boldsymbol{\xi})$ , not dependent on  $\boldsymbol{\beta}$ , may be conveniently expressed as follows [27, 33]:

$$\begin{aligned} \ell_R(\boldsymbol{\xi}) &= -\frac{1}{2}(n - q) \log \pi - \frac{1}{2} \sum_{i=1}^m \log |\boldsymbol{\Sigma}_i| - \frac{1}{2} \log \left| \sum_{i=1}^m \mathbf{X}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \right| \\ &\quad - \frac{1}{2} \sum_{i=1}^m \left[ (\hat{\boldsymbol{\theta}}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}})^\top \boldsymbol{\Sigma}_i^{-1} (\hat{\boldsymbol{\theta}}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}) \right], \end{aligned} \quad (8)$$

where  $\hat{\boldsymbol{\beta}}$  is defined in (7) and  $\boldsymbol{\Sigma}_i$  identifying  $\boldsymbol{\Sigma}_i(\boldsymbol{\xi})$  as above.

The ML estimates of  $\boldsymbol{\beta}$  in fixed-effect meta-analysis are simply obtained by (7), given that, as discussed in Section 4.1,  $\boldsymbol{\Sigma}_i$  equals  $\mathbf{S}_i$  and is therefore completely known. The ML and REML estimates in random-effect models can be instead obtained through Newton-type iterative algorithms. For computational purposes, the objective functions in (6) and (8) are both expressed with respect to  $\boldsymbol{\xi}$  only, and maximization of  $\ell(\boldsymbol{\xi})$  and  $\ell_R(\boldsymbol{\xi})$  can be achieved by plugging in at each iteration the conditional estimate of  $\hat{\boldsymbol{\beta}}(\boldsymbol{\xi})$  in (7) using the current estimate of  $\boldsymbol{\xi}$ , until convergence. Additional information on the estimation algorithms used here are provided in the Supplementary Web Appendix.

#### 4.3. Hypothesis testing and model comparison

We can separate inferences about the parameters in models (2) and (3)–(5) into those about the fixed effects  $\boldsymbol{\beta}$ , which will typically be of prime interest, and the between-study (co)variance matrix  $\boldsymbol{\Psi}$ . Inferential procedures, again, follow the theory of linear mixed models [27, Chap. 6].

Regarding fixed effects, under the marginal model and assuming  $\boldsymbol{\xi}$  as known,  $\hat{\boldsymbol{\beta}}$  follows a multivariate normal distribution with mean and (co)variance matrix given in (7). By replacing  $\boldsymbol{\xi}$  by its ML or REML estimator, the corresponding entries of the resulting  $\hat{\boldsymbol{\beta}}$  and  $V(\hat{\boldsymbol{\beta}})$  may be used for obtaining significance tests or confidence intervals for single coefficients. However, in this specific context, the inference is focused on the set of coefficients defining the association of a specific study-level variable with the joint distribution of the outcome parameters. Multivariate extension of Wald or likelihood ratio (LR) tests are easily derived for this purpose. As usual for REML models, general likelihood theory does hold for LR tests only when comparing models with identical fixed effects structures [33]. Suitable adjustments for the underestimation of (co)variance error matrices due to the uncertainty in the estimation of  $\boldsymbol{\Psi}$  are available [35, 36] although not yet been implemented in *mvmeta*.

For random effects, the focus is on comparing models involving different choices about the structure of the between-study (co)variance matrix. In this setting, an interesting hypothesis to test is  $\boldsymbol{\Psi} = \mathbf{0}$ , namely

that no heterogeneity between studies exists beyond that explained by sampling variability. Similarly, a LR test between nested models may be performed, which is appropriate in REML models as well, given the identical fixed effects structures. Note, however, that for alternative hypotheses that constrain (co)variance matrices to be positive-definite (Supplementary Web Appendix), the null value lies on a boundary of the parameter space. Under these conditions, the conventional null asymptotic  $\chi^2_{n-q}$  distribution does not hold, and some adjustment has been proposed [37]. A test for the same null hypothesis and distribution has also been developed as the multivariate extension of the Cochran  $Q$  test for (residual) heterogeneity [7, 29]. The test is based on the statistic:

$$Q = \sum_{i=1}^m \left[ \left( \hat{\theta}_i - \mathbf{X}_i \hat{\beta} \right)^T \mathbf{S}_i^{-1} \left( \hat{\theta}_i - \mathbf{X}_i \hat{\beta} \right) \right], \quad (9)$$

where  $\hat{\beta}$  are estimated by the correspondent fixed-effect model. An extension of this heterogeneity test for a subset of  $\beta$  has also been proposed [29].

In addition, in this meta-analytical setting, the quantification of the heterogeneity among studies, or the residual amount beyond that explained by specific covariates, is also of interest. Indices of heterogeneity analogous to the univariate case may be easily derived, such as the  $H^2$  and  $I^2$  statistics [38]. These measures are interpreted as the relative excess in heterogeneity (above that explained by sampling error) and the proportion of total variation attributable to heterogeneity, respectively. Although recently criticized for being dependent on precision of the estimates from the first-stage model [39], these statistics provide simple summaries on the extent of heterogeneity. These measures may be produced for the joint multivariate distribution of the  $k$  parameters and interpreted as the amount of heterogeneity in the exposure–response relationships. In this case, from (9), we derive the following:

$$\begin{aligned} H^2 &= \max\{1, Q/(n-q)\}, \\ I^2 &= (H^2 - 1)/H^2. \end{aligned} \quad (10)$$

The same statistics may also be defined for single-outcome parameters [40] by simply computing the related contribution to the quantity  $Q$  and  $df = n - q$  while applying (10). However, as previously discussed, the interpretation of these measures for single parameter of a spline function is not very informative.

More broadly, non-nested models may be compared using fit statistics, in particular Akaike information criterion,  $AIC = -2 \ell(\hat{\beta}, \hat{\Psi}) + 2q$ , and Bayesian information criterion,  $BIC = -2 \ell(\hat{\beta}, \hat{\Psi}) + q \log(n)$ , where  $\ell(\hat{\beta}, \hat{\Psi})$  is the maximum log-likelihood. These statistics may also be used with REML models, with the additional requirement that fixed effects structure be held constant.

#### 4.4. Prediction

In the context of multi-parameter associations, the general tests and fit criteria described previously, although important, are usually insufficient for interpretation. Coefficients in  $\beta$  refer to single-outcome parameters that are rarely interpretable on their own, and the tests only offer a statistical belief on *whether* the multivariate distribution of outcome parameters depends on study-level covariates. However, these procedures fail to inform on *how* the latter modifies the former.

In the current setting, prediction represents an important tool to extend the inference from multivariate meta-regression models, offering a method to link specific values of study-level variables with outcome parameters expectations. Given a set of meta-predictor values  $\mathbf{x}_0$ , the model-predicted mean  $\hat{\theta}_0$  and (co)variance matrix  $V(\hat{\theta}_0)$  are obtained by the following:

$$\begin{aligned} \hat{\theta}_0 &= \mathbf{X}_0 \hat{\beta}, \\ V(\hat{\theta}_0) &= \mathbf{X}_0 V(\hat{\beta}) \mathbf{X}_0^T, \end{aligned} \quad (11)$$

with  $\mathbf{X}_0$  computed from  $\mathbf{x}_0$  following (4). The equations in (11) may be used to predict specific parameters  $\hat{\theta}_0$ , which are used to define the average exposure–response curve predicted for the set of meta-predictor values  $\mathbf{x}_0$ . The same equations may be used to predict the association in a new study characterized by a specific set of study-level variables by simply increasing the uncertainty in the estimates by adding  $\hat{\Psi}$  to  $V(\hat{\theta}_0)$  in (11) [41].



In addition, the assumptions outlined in Section 4.1 regarding the random-effect multivariate distribution may be exploited to extend the inference regarding study-specific outcome parameters  $\theta_i$  estimated in the first-stage model, computing the (asymptotic) best linear unbiased prediction (BLUP) [27, Section 7.4]. The predicted  $\hat{\theta}_{b(i)}$  and associated asymptotic (co)variance matrix  $V(\hat{\theta}_{b(i)})$  are as follows:

$$\begin{aligned}\hat{\theta}_{b(i)} &= \mathbf{X}_i \hat{\beta} + \hat{\Psi} \hat{\Sigma}_i^{-1} (\hat{\theta}_i - \mathbf{X}_i \hat{\beta}), \\ V(\hat{\theta}_{b(i)}) &= \mathbf{X}_i V(\hat{\beta}) \mathbf{X}_i^T + \hat{\Psi} - \hat{\Psi} \hat{\Sigma}_i^{-1} \hat{\Psi},\end{aligned}\quad (12)$$

for  $\hat{\Sigma}_i = \mathbf{S}_i + \hat{\Psi}$ . The BLUP equations in (12) represent the sum of two components: the predicted averaged outcome parameters in (11) and study-specific deviations, predicted as random effects  $u_i$  in (5). Exposure–response associations predicted with BLUP in (12) represent a trade-off between city-specific and average estimates, with weights inversely proportional to the two components  $\Psi$  and  $\mathbf{S}_i$  of the total variability  $\Sigma_i$ , respectively. It is noteworthy, in this multivariate setting, that the BLUP estimates of missing parameters from the first stage exploit the information about the other study-specific parameters obtained through the between-study (co)variance matrix  $\hat{\Psi}$  and may be therefore different from predicted values from (11).

## 5. Results

The first-stage generalized linear model in (1) is fitted to the data from each city, considered as independent studies. The selection procedure indicates a preference for a quadratic B-spline with 6 *df*, with 4 interior knots at  $-4.7^\circ\text{C}$ ,  $6.0^\circ\text{C}$ ,  $16.7^\circ\text{C}$  and  $24.8^\circ\text{C}$ , corresponding to the 5th, 35th, 65th and 95th average distribution percentiles across studies. The boundary knots are set to the average minimum and maximum, ranging from  $-18.3^\circ\text{C}$  to  $29.2^\circ\text{C}$ . All the spline basis variables are centred at  $20^\circ\text{C}$ . The vectors  $\hat{\theta}_i$  and (co)variance matrices  $\mathbf{S}_i$  for the six parameters are then pooled across studies in multivariate meta-analysis and meta-regression. All the models are fitted through ML. Given the illustrative purpose of this example, in the description of the results we favour methodological matters over subject-specific details. All the results described in this section may be reproduced using the R code available in the Supplementary Web Appendix.<sup>‡</sup>

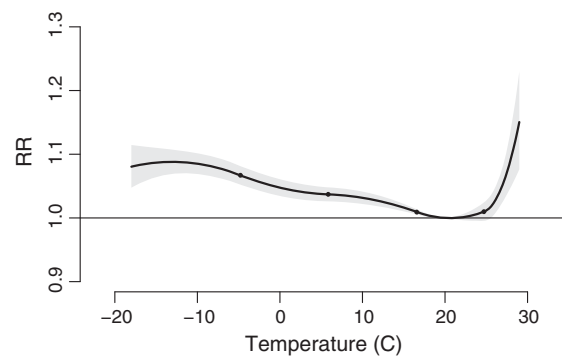
### 5.1. Pooled exposure–response relationship

The main result of multivariate meta-analysis, applying the model in (2), consists of an estimate  $\hat{\theta}$  of the outcome parameters  $\theta$ , representing the population-average coefficients of the quadratic B-spline. These are used to compute the average exposure–response relationship depicted in Figure 2, together with confidence intervals obtained through  $V(\hat{\theta})$ , over the average exposure range. As expected, the plot indicates an increase in risk for non-accidental mortality for both high and low temperatures, with a point of minimum mortality at about  $20.6^\circ\text{C}$ . The risk increases approximately linearly in the cold tail but shows a steep non-linear increase at high temperatures. The model estimates a relative risk (RR) of 1.085 (95%CI: 1.069–1.102) and 1.150 (95%CI: 1.076–1.230) for temperatures of  $-10^\circ\text{C}$  and  $29^\circ\text{C}$  versus  $20^\circ\text{C}$ , respectively. The multivariate Cochran  $Q$  test and  $I^2$  obtained from (9)–(10) are reported in Table II. These statistics reveal a significant heterogeneity across studies, where 70.2% of the variability in the multivariate outcome  $\hat{\theta}_i$ , and therefore in the derived exposure–response association, is attributed to true between-study differences.

### 5.2. Best linear unbiased prediction

As discussed in Section 4.4, the assumptions about the between-city variability, namely the distribution of the random effects in (5), can be used to extend the inference regarding study-specific estimates. Variability around the average is due to both heterogeneity between cities and uncertainty in the first-stage model. Applying (12), we derived the BLUP estimates  $\hat{\theta}_{b(i)}$  of the unknown city-specific parameters  $\theta_i$  of the quadratic spline, representing a shrunk version of the estimated  $\hat{\theta}_i$  towards the population average  $\hat{\theta}$ . Figure 3 illustrates the associated exposure–response curves: the left panel shows the original

<sup>‡</sup>Supporting information may be found in the online version of this article.

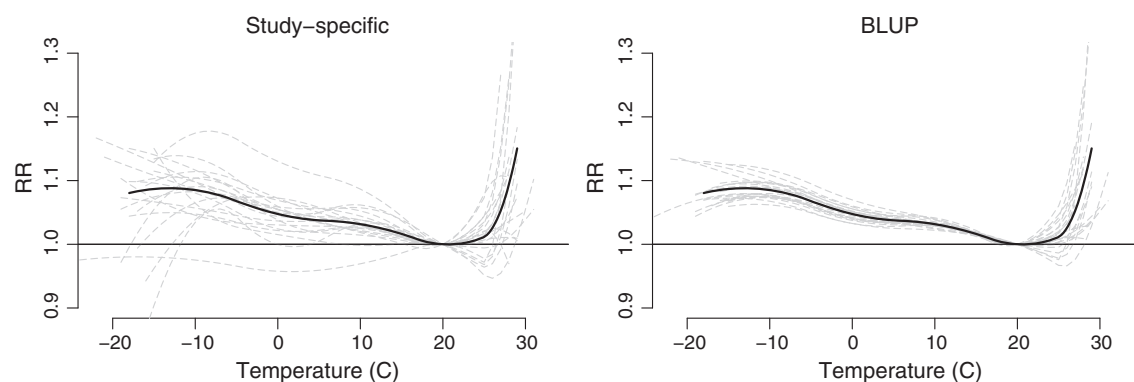


**Figure 2.** Pooled exposure–response relationship in relative risk between temperature and non-accidental mortality in 20 US cities, 1987–2000. The grey area represents 95% confidence intervals. Dots indicate knots location. Reference at 20°C.

**Table II.** Cochran  $Q$  test,  $I^2$  and information criteria for second-stage multivariate models based on an absolute or relative scale.

	Cochran $Q$ test			$I^2$	Information criteria		LR test			Wald test		
	$Q$	$df$	$p$	%	AIC	BIC	Stat	$df$	$p$	Stat	$df$	$p$
Absolute scale												
Intercept only	382.8	114	0.000	70.2	−322.8	−247.5						
Latitude	186.0	108	0.000	41.9	−326.5	−234.5	15.7	6	0.015	30.9	6	0.000
Urban population	342.8	108	0.000	68.5	−317.9	−226.0	7.1	6	0.309	8.7	6	0.191
High education	376.4	108	0.000	71.3	−316.1	−224.1	5.3	6	0.506	8.2	6	0.227
Living in poverty	365.8	108	0.000	70.5	−314.9	−222.9	4.1	6	0.661	5.4	6	0.493
Relative scale												
Intercept only	304.5	114	0.000	62.6	−344.2	−268.9						
Latitude	258.7	108	0.000	58.3	−338.4	−246.4	6.3	6	0.393	7.7	6	0.265

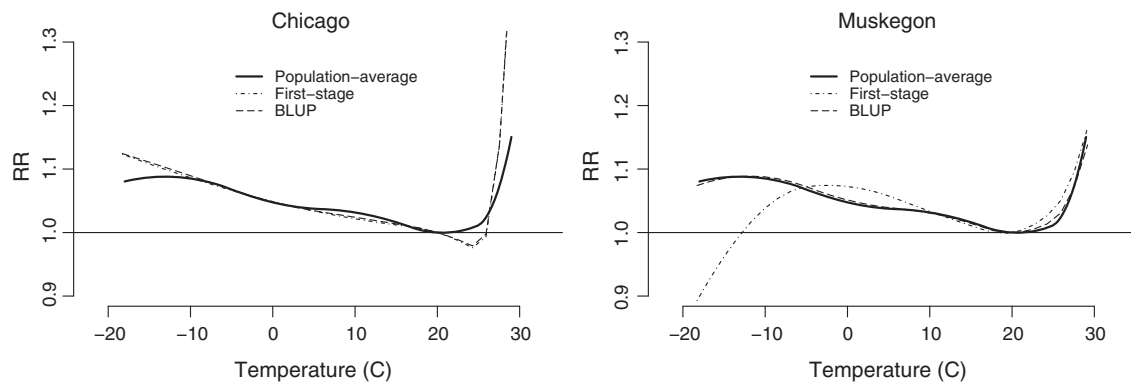
Likelihood ratio (LR) and Wald tests, also included, refer to the association with meta-predictors in meta-regression models.



**Figure 3.** First-stage (left) and best linear unbiased predicted (right) estimates of the exposure–response relationships in relative risk between temperature and non-accidental mortality in 20 US cities, 1987–2000. The continuous bold black line represents the population-average curve, whereas the long-dashed grey lines are the study-specific estimates. Reference at 20°C.

estimates from the first-stage models in the 20 studies, whereas the related BLUP relationships are displayed in the right panel. The shrinkage effect is stronger in the cold extreme, possibly because of the higher uncertainty in the within-study estimates.

Given the large difference in population size shown in Table I and in the related within-study precision, the shrinkage effect is expected to vary considerably across studies. Figure 4 shows the first-stage and BLUP temperature–mortality relationships in two studies, together with the population average as



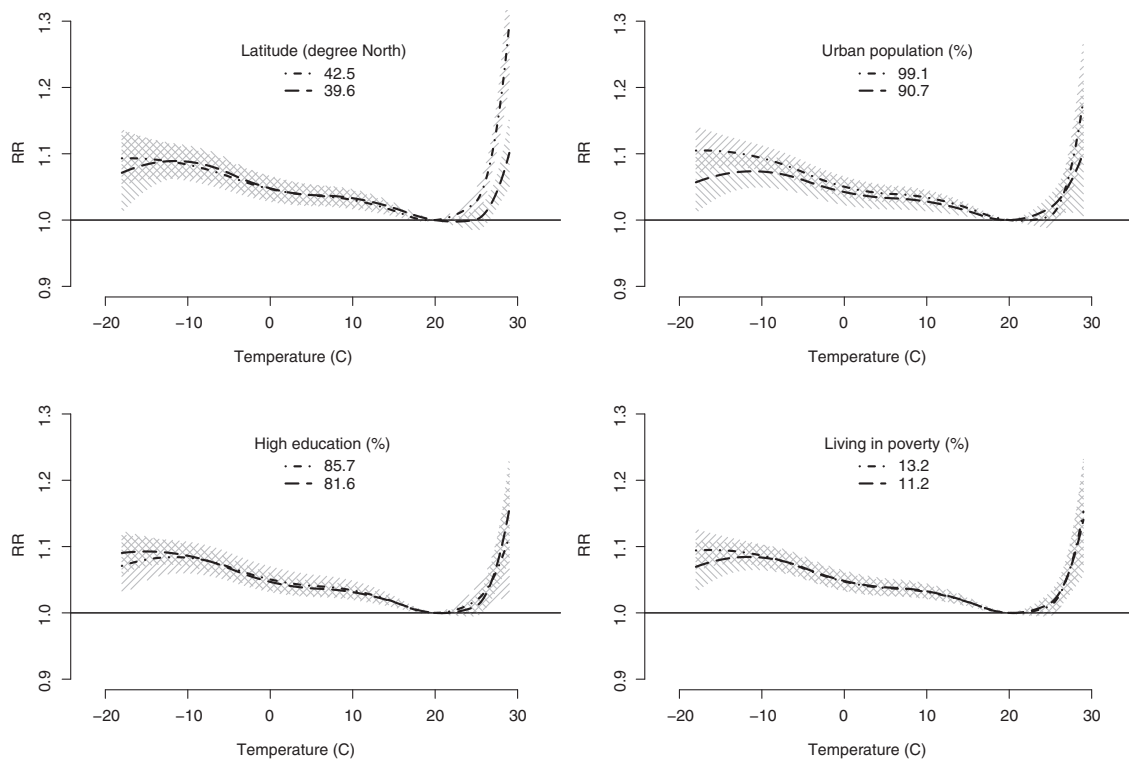
**Figure 4.** Population-average (continuous bold line), first-stage (dash-dot line) and best linear unbiased predicted (dash line) exposure-response relationships in relative risk between temperature and non-accidental mortality in two US cities, 1987–2000. The figure illustrates the largest (Chicago, left) and smallest (Muskegon, right) studies included in the analysis. Reference at 20°C.

depicted in Figure 2. As expected, the BLUP estimate is closer to the original curve in Chicago, the largest city characterized by a high number of daily deaths (above 5 million inhabitants, 115 deaths on average per day). By contrast, the study in the smallest city of Muskegon (170 000 inhabitants, five deaths per day) produces very imprecise estimates (confidence intervals not shown), in particular for cold, which are heavily shrunk toward the population average.

### 5.3. Multivariate meta-regression

Part of the large heterogeneity in temperature–mortality curves between studies, displayed in Figure 3, may be explained by different patterns of study-level factors that modify the association. Multivariate meta-regression models in (3)–(5), including meta-predictors, are applied to extend the analysis and partly characterize such differences. From a fitted model, it is possible to compute the outcome parameters  $\hat{\theta}_0$  predicted for a set of meta-predictor values  $\mathbf{x}_0$  using (11). As discussed in Section 4.4, these parameters are interpreted as the average coefficients of the function  $s$  for the subset of studies from the hypothetical population that are characterized by  $\mathbf{x} = \mathbf{x}_0$ . A test on the significance of the multivariate association between the outcome parameters and each study-level variable is carried out by applying a Wald test on the related subset of coefficients  $\beta$ . Alternatively, an LR test or information criteria such as AIC and BIC can be used to compare the models with and without specific meta-predictors.

As an illustration, univariable multivariate meta-regression models, each including a single meta-predictor, are specified for four study-level variables, namely latitude, percentage of population living in urban settings, percentage of population with high education (high school or higher) and percentage of population living in poverty. The exposure-response associations are predicted for the values of the approximate 25th and 75th percentiles of the study-level variables (reported in Table I) and shown in Figure 5. The top-left panel illustrates the two predicted curves for latitude. The effect of hot temperature, on average, is higher in studies for northern cities, with a steeper raise in risk: the RR at 29°C versus 20°C increases from 1.101 (95%CI: 1.049–1.156) for the 25th percentile of latitude to 1.297 (95%CI: 1.205–1.396) for the 75th percentile, although there is also a suggestion of a lower point of minimum mortality when increasing latitude. No substantial effect modification is suggested for cold temperatures. The importance of the overall association with the study-level variable can be assessed by the comparison of the AIC and BIC criteria, reported in Table II, with the corresponding model with no meta-predictor in Figure 2. The lower AIC, from –322.8 to –326.5, suggests a better fit of the model with the predictor, whereas the BIC, highly penalized by the number of observations in the model, shows a preference for the more parsimonious model. The Wald and LR tests, also reported in Table II, are both specified with 6 *df*, corresponding to the six additional coefficients used to model the linear effects of latitude on each outcome parameters. The tests are both significant at the standard 5% level, although the *p*-values differ somewhat. Latitude seems to explain a substantial part of heterogeneity between studies, with an overall  $I^2$  of 41.9% compared with the 70.2% of the model with no predictor. However, the test for the residual amount of heterogeneity is still significant.



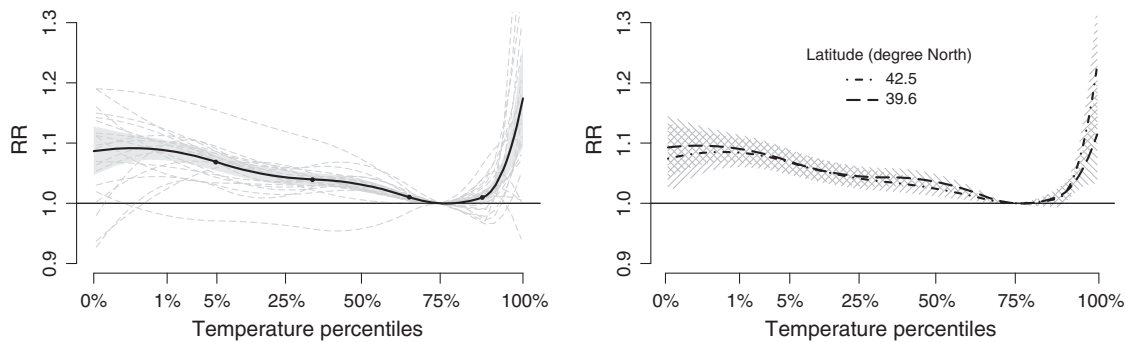
**Figure 5.** Predicted exposure–response relationships in relative risk between temperature and non-accidental mortality in 20 US cities, 1987–2000. The figure illustrates the predicted curves from meta-regression for the 25th (dash line) and 75th (dash–dot line) percentiles of latitude (top left), percentage of population living in urban settings (top right), percentage of population with high education (bottom left) and percentage of population living in poverty (bottom right). The shaded areas represent 95% confidence intervals. Reference at 20°.

The other three panels in Figure 5 show the association with the other study-level variables. The percentage of population living in urban settings seems to increase the effect of both cold and hot temperatures, although fit statistics and tests in Table II suggest that this effect modification is not significant. Also, the decrease in heterogeneity is negligible. There is no evidence that percentage of population with higher education or living in poverty modify the temperature–mortality association, as shown by the two identical curves predicted for their 25th and 75th percentiles in the bottom panels of Figure 5 and related tests and fit statistics in Table II.

The extension to multivariable multivariate meta-regression is straightforward: tests and statistics are defined in exactly the same way, and predicted effects as those showed in Figure 5, controlled for the effect of other study-level variables, may be computed in a similar way. The interpretation, as with standard multivariable models, refers to the effect of a meta-predictor for constant values of the other ones. Standard model selection procedures for regression models may be applied to choose the set of meta-predictor to be included. An example is included in the code provided in the Supplementary Web Appendix.

## 5.4. Adopting a relative scale

As shown in Table I and in Figure 3, the exposure ranges are slightly different in each study. In this specific example, range variation was too small to cause problems, but it is likely to do so when exposure ranges differ more substantially. An alternative is to adopt a relative scale based on percentiles, as described in Section 3.4. As an illustration, we repeat the analysis placing the knots at the 5th, 35th, 65th and 95th percentiles of study-specific temperature distribution and centering on the 75th percentile. The predicted exposure–response relationship is displayed in Figure 6. To aid interpretation in this standardized scale, the x-axis is scaled so that percentiles match those of the average temperature distribution across all the cities.



**Figure 6.** Pooled and predicted exposure–response relationships in relative risk between relative temperature (percentiles) and non-accidental mortality in 20 US cities, 1987–2000. The  $x$ -axis is scaled so that percentiles match those of the average temperature distribution of all the cities. The figure illustrates the population-average curve from meta-analysis and study-specific estimates (left panel, dots indicate knots location) and the predicted curves from meta-regression for the 25th (dash line) and 75th (dash-dot line) percentiles of latitude (right panel). Reference at the 75th percentile of temperature.

Not surprisingly, given the similar temperature ranges, the shape of the pooled association shown in the left panel of Figure 6 is very close to that predicted in absolute scale in Figure 2. However, the results from the two approaches are not easily comparable because of the different interpretation of the results. The exposure–response relationship predicted in an absolute scale informs about the average RR between specific temperatures, for example 30°C versus 20°C. By contrast, the curve predicted in relative scale measures the RR of relative temperatures, for example 95th versus 75th percentiles, based on study-specific temperature distributions. The right panel of Figure 6 illustrates the results from the meta-regression model including latitude. Although the risk of hot temperatures is still higher in northern cities, the difference is smaller if compared with the related plot predicted in absolute scale in Figure 5, top-left panel, and no longer significant, as shown in Table II. Also, the point of minimum mortality is similar in cities at different latitudes, approximately at the 76th percentile. This may be explained by a partial adaptation of populations living in different cities to their own climate.

### 5.5. Additional analysis

The selection procedure regarding the function  $s$  described in Section 3.3 indicates a preference for a spline with 4 knots. As a sensitivity analysis, the results in Section 5.1 are compared with those obtained using a spline with 3 knots, placed at  $-1.8^{\circ}\text{C}$ ,  $11.3^{\circ}\text{C}$  and  $23.3^{\circ}\text{C}$ , corresponding to the 10th, 50th and 90th average distribution percentiles across studies. The exposure–response curve is very similar to that showed in Figure 2 (result not shown), suggesting that the choice of number and position of knots is not critical in this case.

All the models presented previously have been estimated through ML. The choice of this estimator allows the comparison of models with different fixed effects structures through LR test and information criteria, as described in Sections 4.2–4.3. The estimation can be potentially improved through the use of REML, which accounts for some additional uncertainty in the estimation of the between-study (co)variance matrix  $\Psi$  not accounted by ML estimates. However, even with this relatively limited number of studies, standard errors of coefficients for REML models only increase in the range of 1–2%, and the two predicted exposure–response curves and confidence intervals are almost identical (Supplementary Web Appendix).

As previously pointed out, estimation difficulties may arise in this complex modelling framework, particularly regarding between-study correlations close to the boundary of their parameter space of  $-1$  or  $1$  [11]. The procedure implemented in *mvmeta*, based on Cholesky decomposition to assure positive-definiteness, seems to perform well under several modelling choices using these data, and estimation problems have not occurred.

In meta-regression models presented in Section 5.3, 120 correlated outcomes from 20 studies are used to estimate 12 coefficients and 21 (co)variance parameters. These sophisticated models require several assumptions that are commonly hard to verify, such as the multivariate normality of the random effects  $u_i$  in (5). The results are therefore expected to be potentially sensitive to model mis-specifications.



In particular, we have noticed some discrepancies between the Wald and LR tests in meta-regression models in Figure 5, probably because of different assumptions of the two tests. In a preliminary assessment, the quadratic approximation to the likelihood surface of the Wald test seems to perform poorly in badly conditioned models, for example meta-regression in the presence of outliers. Fitted models may be checked with appropriate regression diagnostics, such as residuals and influence measures already available for univariate meta-analysis [42], which still need to be developed in the multivariate meta-analytical techniques.

## 6. Further considerations

As mentioned earlier, the methodology of multivariate meta-analysis has been largely developed in the context of randomized controlled trials to pool estimates of multiple outcomes. Moreover, as shown in Section 4, the statistical framework may be placed within that of the general linear mixed model, although with particular characteristics. Nevertheless, specific issues arise when this methodology is applied to multi-parameter associations in two-stage analyses. Here we provide some comments on these aspects, highlighting advantages and limitations and directions for future research.

*Advantages of multi-parameter synthesis.* As anticipated in Section 1, the application of multivariate meta-analysis extends the standard two-stage design, where the data on the associations of interest are usually summarized in the estimate of a single parameter. For complex associations, the traditional approach may be too limited to characterize the phenomenon under study. Referring to the examples illustrated in Section 5, a standard analysis can be based on the pooling of single estimated effects at specific percentiles for temperature [18], or on simplification of the relationship through linear-threshold parameterization [17]. The estimate of the truly non-linear exposure–response relationship is more informative and flexible and may be preferred for the investigation of complex associations. This approach may be more broadly described in the context of multi-parameter evidence synthesis [43, 44].

*Dealing with complexity.* The two-stage approach discussed previously provides tools for the analysis of complex associations. In the first step, the estimate is controlled for potential confounders while reducing the relationship to a limited number of parameters of a chosen function, which are subsequently used as outcomes for the second-stage meta-analytic model. The amount of complexity retained in the first stage represents a trade-off between the synthesis of simple approaches and details offered by sophisticated models. Ideally, this balance should be fine-tuned only to the purpose of the analysis. However, in practice, concrete problems such as mathematical and statistical properties of the function or the number of parameters need to be taken into account, as discussed in the following paragraphs. Moreover, this complex framework involves important modelling assumptions and is particularly sensitive to model mis-specification, as pointed out in Section 5.5. Further research is needed to develop this approach as a standard analytical tool.

*Dimensionality and sample size.* Using this type of approach, the dimension  $k$  of the parameter vector that can be accommodated in multivariate meta-analysis is of course limited, as  $k$   $p$  fixed effects coefficients and  $k(k + 1)/2$  (co)variance parameters need to be estimated. Guidance on the sample size (i.e. number of studies) for multivariate meta-analytical models is not yet available, and additional research is required to better characterize this complex modelling framework [11]. A possible solution to reduce the number of parameters is to structure the between-study (co)variance matrix  $\Psi$ , for example imposing compound-symmetry, diagonal or autoregressive forms. Robust estimation to account for wrong correlation structures has been proposed for meta-regression of correlated outcomes [45], although further research is needed for this approach.

*The two-stage design.* In the setting of randomized controlled trials, the two-stage approach is often compared with the so-called one-stage IPD analysis, usually performed through a single multilevel model. Although the latter has been advocated as more efficient and less prone to bias if compared with the meta-analysis of published studies [46, 47], the two-stage alternative has been proved as competitive when applied through IPD analysis [48, 49]. In addition, a single multilevel development is not always feasible or advisable, especially in the presence of many individual-level covariates, which would require the definition of intricate study-specific dependencies. In our example,  $7 \times 14 = 98$  parameters are used in the first stage to model within-study seasonal and long-term trends. In the two-stage framework, parameters not related to the association of interest are treated as nuisance terms in the first-stage model, offering computational efficiency and flexibility regarding model specification and assumptions.

*Analysis of published studies.* Although the modelling framework proposed here is focused on two-stage analysis of complete study-specific datasets, most of the original development of meta-analysis

is based on the combination of estimates from published results. This also applies to the multivariate extension, as described in Section 7. The meta-analysis of published studies poses additional problems. First, the outcome parameters defining the association in each study may not be comparable, for instance if estimated from different functions. Referring to the application in which non-linear dependencies are modelled, exposure categories may be defined with different cut-offs or spline functions with different knots. Solutions have been previously proposed to retrieve estimates of comparable outcome parameters from available study-specific information, as discussed in Section 7. Another issue is that correlation between estimated outcome parameters are rarely reported. Methods to deal with missing correlations have been developed [50,51], although this issue needs to be explored further, especially for multi-parameter associations.

*Definition of the function  $s$ .* As discussed previously, the choice of the function is primarily dictated by the purpose of the analysis and the amount of complexity needed to represent the association. Also, the exact specification of the function may be the result of a data-led selection procedure. A main limitation of the proposed approach is the requirement to specify the same function in all the studies, a constraint that could have a strong influence on study-specific estimates. In particular, the selected function might not fit the data well in some studies, while producing the best fit overall. As showed in Section 5.5, the sensitivity of the results to modelling choices and selection should always be inspected.

*Exposure ranges and absolute or relative scale.* In our illustrative example, the 20 studies showed similar exposure ranges, and the meta-analysis was first performed adopting an absolute scale. However, as already mentioned in Section 3.4, this approach is not feasible to pool studies characterized by more varied exposures ranges. In such an analysis, the definition of a common function  $s$  is not straightforward. In the specific example illustrated here, the selection of common knots may result in some parameters of the function being inestimable. The interpretation of the other city-specific outcome parameters would change, and the meta-analysis, although feasible in the presence of missing outcomes, may produce results of uncertain meaning. The analysis using a relative scale represents an interesting and feasible alternative, although it involves a different interpretation of the results, as described in Section 5.4. In temperature–mortality associations, the relative scale accounts for adaptation of populations to their own climate, but this approach may not provide meaningful results when applied for investigating other associations. This issue represents an important limitation of the proposed methodology and needs to be addressed in future research.

*Interpretational issues.* The point discussed previously is closely linked to the more general problem of interpretation of estimates of complex associations. Although the results illustrated in Section 5 are described on the original scale of the first-stage model, estimation is carried out in the multivariate dimension of the spline parameters. In practice, we read the association in the usual exposure–response frame, but we model it through coefficients of a function. We presuppose that these parameters, in this multi-study assessment, still preserve their interpretation and that the way we model the relationship between study-level variables and the multivariate distribution of the outcome parameters reflects the association of interest. If, for example, different combinations of parameters define exactly the same association, this link vanishes, and interpretation of the results would be less straightforward. This issue requires further consideration.

*Model selection and control for confounding.* The modelling framework presented in this contribution requires a complex two-stage model selection. In the first stage, the selection of a suitable model involves the choices for the function  $s$  to define the association and for the control for potential confounders. As previously discussed, this procedure is usually specific to the data, study design and regression method applied in the first stage. Also, this selection procedure should always be accompanied by a sensitivity analysis. The selection in the second stage is limited to the choice of the meta-predictors to be included in the meta-analytical model and may follow traditional methods used in regression analysis, exploiting the multivariate extension of tests and statistics defined in Section 4.3. However, further research is required to provide a general methodological approach for this two-stage selection procedure.

## 7. Discussion

In this contribution, we provide a methodological overview of the application of multivariate meta-analysis and meta-regression for the investigation of complex associations that are described by multiple parameters. The modelling framework we propose is general and potentially applicable to different

research fields, such as multi-site clinical trials, multi-centre cohorts, studies on multiple cancer registries or multi-city time-series studies, each involving different first-stage regression models. The methodology requires the specification of a common function to model the association, which nevertheless may be selected among several potential alternatives, like splines, step functions, fractional polynomials or threshold functions, among others. The statistical framework is nested within that of the general linear mixed model, which offers the basis for the specification of tests and statistics for inferential procedures, model comparison and model checking.

Previous work explored methods to pool non-linear dose–response relationships within IPD two-stage analyses. A recent paper has proposed a method to obtain pooled estimates through a series of univariate meta-analyses of estimated effects on a grid of exposure values in order to re-construct the pooled non-linear relationships [52]. This flexible approach was previously developed and termed *meta-smoothing* in the context of multi-city time-series studies [53] and potentially allows the use of different study-specific functions to model the association. However, the meta-analytic estimates of the effects in the grid are treated as independent, although obtained from the same first-stage model, an approach that could affect the validity of the estimates for standard errors and confidence intervals. Although the framework we propose is limited by the definition of the same function in all the studies, it accommodates in an appropriate way the correlation between estimates from the first stage, establishing valid inferential procedures about the pooled association. Also, it formalizes the meta-analytic procedure within a proper statistical model, providing statistics and significance tests for model fit and comparison, heterogeneity and prediction. Other investigators have previously applied multivariate meta-analysis for IPD analysis of non-linear exposure–response relationships, although not focusing on methodological aspects. A recent paper has discussed the statistical methods for two-stage analysis of multi-site cohorts, also illustrating the use of multivariate meta-analysis for pooling dose–response associations that have been estimated using multiple categories [5]. Other examples include applications in multi-city time-series studies to assess potential non-linear effects of air pollution [17, 54, 55], using approaches similar to the example in Section 5.

Methods for obtaining pooled dose–response dependencies from published epidemiological studies, rather than IPD analyses, have been also investigated in previous research. Pioneering work [56, 57] described an analysis based on log-RR estimates for different exposure categories compared with a common reference, in which the whole within-study (co)variance matrix is reconstructed using ad hoc approximations. The estimates of linear and (optionally) quadratic terms were then combined using fixed meta-analytic methods and then the random counterpart based on method of moments. This approach has also been applied with splines or fractional polynomials to model non-linearity [58–60]. More recently, a general methodological treatment of the meta-analysis of published estimates for non-linear associations has been provided [61, 62], also on the basis of multivariate meta-analysis [63, 64].

The methodology illustrated here is not limited to model multi-parameterized non-linear exposure–response dependencies: investigators have also applied multivariate meta-analysis to synthesize survival curves [30, 65], longitudinal profiles [66], ROC curves [67] and heat wave effects [68]. Other studies have adopted multivariate meta-analysis to explore the distributed lag effects of air pollution [69–71] and temperature [72]. Finally, the same methods have also been applied to pool main and interaction terms across studies [57, 73].

In its traditional setting for pooling multiple health end points in randomized controlled trials, multivariate meta-analysis offers parameter estimates with better statistical properties, in particular a potentially increased precision from accommodating the estimated between-study covariance structure [11]. Nonetheless, the analysis could be carried out with multiple univariate meta-analysis, although often less efficiently. In the application we have described, instead, estimates of complex associations, such as those illustrated in Figures 2–6, cannot be provided by simple univariate models without important limitations or additional assumptions. In this context, multivariate meta-analysis offers clear advantages. As discussed in Section 4, this modelling framework can be seen as an example of a multivariate linear mixed model. The extensive body of research defining this statistical framework may therefore be exploited in this context, for example for the definition of tests discussed in Section 4.3. There are, of course, specific issues that deserve further research, for example statistics for heterogeneity, methods for handling missing correlations or a critical comparison of estimation methods. Other important issues specific to multivariate meta-analysis have been illustrated and discussed by Jackson and colleagues [11] and need to be addressed in future research in order to develop this methodology as a standard analytical tool.

## Software

All the analyses presented in this contribution are performed in R (version 2.15.0). The package *mvmeta* (version 0.2.4) is used to run multivariate meta-analysis and meta-regression. The package *dlnm* (version 1.6.0) [74] is used to specify the basis matrices for the quadratic spline for temperature and to predict and plot the results of a fitted model. The data are accessed through functions in the package *NMMAPlite*. The R code to replicate all the results of Section 5 is available in the Supplementary Web Appendix, together with additional information on the package *mvmeta*. An alternative software implementation of multivariate meta-analysis is available in *Stata* [28, 40], whereas a *SAS* code has been previously presented [75].

## Acknowledgements

Our research has been motivated and encouraged by the event ‘Multivariate meta-analysis’, organized by Ian White, Dan Jackson and Richard Riley at the Royal Statistical Society, London, on 26 January 2010. The development of the R package *mvmeta* has been substantially expedited by the comparison with a previous implementation in *Stata* provided by Ian White, cited previously. The analysis presented in this paper are completely reproducible given the public availability of the data, thanks to the NMMAPlite researchers. The final version of this paper has been substantially improved by the comments received from two anonymous reviewers. This study was sponsored by the Medical Research Council (UK) under grants G1002296 and G0701030.

## References

1. Normand SLT. Meta-analysis: formulating, evaluating, combining, and reporting. *Statistics in Medicine* 1999; **18**(3):321–359.
2. Sutton AJ, Higgins JP. Recent developments in meta-analysis. *Statistics in Medicine* 2008; **27**(5):625–650.
3. Stukel TA, Demidenko E, Dykes J, Karagas MR. Two-stage methods for the analysis of pooled data. *Statistics in Medicine* 2001; **20**(14):2115–2130.
4. Simmonds MC, Higgins J, Stewart LA, Tierney JF, Clarke MJ, Thompson SG. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. *Clinical Trials* 2005; **2**(3):209–217.
5. Thompson S, Kaptoge S, White I, Wood A, Perry P, Danesh J, Emerging Risk Factors Collaboration. Statistical methods for the time-to-event analysis of individual participant data from multiple epidemiological studies. *International Journal of Epidemiology* 2010; **39**:1345–1359.
6. Dominici F, Samet JM, Zeger SL. Combining evidence on air pollution and daily mortality from the 20 largest US cities: a hierarchical modelling strategy. *Journal of the Royal Statistical Society: Series A* 2000; **163**(3):263–302.
7. Berkey CS, Anderson JJ, Hoaglin DC. Multiple-outcome meta-analysis of clinical trials. *Statistics in Medicine* 1996; **15**(5):537–557.
8. Berkey CS, Hoaglin DC, Antczak-Bouckoms A, Mosteller F, Colditz GA. Meta-analysis of multiple outcomes by regression with random effects. *Statistics in Medicine* 1998; **17**(22):2537–2550.
9. Arends LR, Vokó Z, Stijnen T. Combining multiple outcome measures in a meta-analysis: an application. *Statistics in Medicine* 2003; **22**(8):1335–1353.
10. Gasparrini A, Armstrong B, Kenward MG. Multivariate meta-analysis: a method to summarize non-linear associations. *Statistics in Medicine* 2011; **30**(20):2504–2506.
11. Jackson D, Riley R, White IR. Multivariate meta-analysis: potential and promise. *Statistics in Medicine* 2011; **30**(20):2481–2498.
12. R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria, 2012. URL <http://www.R-project.org/>, ISBN 3-900051-07-0.
13. Gasparrini A, Armstrong B. Time series analysis on the health effects of temperature: advancements and limitations. *Environmental Research* 2010; **110**(6):633–638.
14. Basu R, Samet JM. Relation between elevated ambient temperature and mortality: a review of the epidemiologic evidence. *Epidemiologic Reviews* 2002; **24**(2):190–202.
15. Armstrong B. Models for the relationship between ambient temperature and daily mortality. *Epidemiology* 2006; **17**(6):624–631.
16. McMichael AJ, Wilkinson P, Kovats RS, Pattenden S, Hajat S, Armstrong B, Vajanapoom N, Niciu EM, Mahomed H, Kingeow C, Kosnik M, O’Neill MS, Romieu I, Ramirez-Aguilar M, Barreto ML, Gouveia N, Nikiforov B. International study of temperature, heat and urban mortality: the ‘ISOTHURM’ project. *International Journal of Epidemiology* 2008; **37**(5):1121.
17. Baccini M, Biggeri A, Accetta G, Kosatsky T, Katsouyanni K, Analitis A, Anderson HR, Bisanti L, D’Ippoliti D, Danova J, Forsberg B, Medina S, Paldy A, Rabczenko D, Schindler C, Michelozzi P. Heat effects on mortality in 15 European cities. *Epidemiology* 2008; **19**(5):711–719.
18. Anderson BG, Bell ML. Weather-related mortality: how heat, cold, and heat waves affect mortality in the United States. *Epidemiology* 2009; **20**(2):205–213.
19. Ruppert D, Wand MP, Carroll RJ. *Semiparametric Regression*. Cambridge University Press: New York, 2003.



20. Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Journal of the Royal Statistical Society: Series C* 1994; **43**(3):429–467.
21. Muggeo VM. Estimating regression models with unknown break-points. *Statistics in Medicine* 2003; **22**(19):3055–3071.
22. Dominici F, Sheppard L, Clyde M. Health effects of air pollution: a statistical review. *International Statistical Review* 2003; **71**(2):243–276.
23. Cao J, Valois MF, Goldberg MS. An S-Plus function to calculate relative risks and adjusted means for regression models using natural splines. *Computer Methods and Programs in Biomedicine* 2006; **84**(1):58–62.
24. Peng RD, Dominici F, Louis TA. Model choice in time series studies of air pollution and mortality. *Journal of the Royal Statistical Society: Series A* 2006; **169**(2):179–203.
25. Harrell FE. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer Verlag: New York, 2001.
26. Higgins J, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society: Series A* 2009; **172**(1):137–159.
27. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data*. Springer Verlag: New York, 2000.
28. White IR. Multivariate random-effects meta-analysis. *Stata Journal* 2009; **9**(1):40–56.
29. Ritz J, Demidenko E, Spiegelman D. Multivariate meta-analysis for data consortia, individual patient meta-analysis, and pooling projects. *Journal of Statistical Planning and Inference* 2008; **139**(7):1919–1933.
30. Dear KBG. Iterative generalized least squares for meta-analysis of survival data at multiple times. *Biometrics* 1994; **50**:989–1002.
31. Nam IS, Mengersen K, Garthwaite P. Multivariate meta-analysis. *Statistics in Medicine* 2003; **22**(14):2309–2333.
32. Jackson D, White IR, Thompson SG. Extending DerSimonian and Laird's methodology to perform multivariate random effects meta-analyses. *Statistics in Medicine* 2010; **29**(12):1282–1297.
33. Harville DA. Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* 1977; **72**(358):320–338.
34. Pinheiro JC, Bates DM. *Mixed-Effects Models in S and S-PLUS*. Springer Verlag: New York, 2000.
35. Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* 1997; **53**(3):983–997.
36. Kenward MG, Roger JH. An improved approximation to the precision of fixed effects from restricted maximum likelihood. *Computational Statistics and Data Analysis* 2009; **53**(7):2583–2595.
37. Stram DO, Lee JW. Variance components testing in the longitudinal mixed effects model. *Biometrics* 1994; **50**(4):1171–1177.
38. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* 2002; **21**(11):1539–1558.
39. Rucker G, Schwarzer G, Carpenter JR, Schumacher M. Undue reliance on  $I^2$  in assessing heterogeneity may mislead. *BMC Medical Research Methodology* 2008; **8**(1):79.
40. White IR. Multivariate random-effects meta-regression: updates to mvmeta. *Stata Journal* 2011; **11**(2):255–270.
41. Riley R, Higgins J, Deeks JJ. The interpretation of random-effects meta-analysis. *British Medical Journal* 2011; **342**:d549.
42. Viechtbauer W, Cheung MWL. Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods* 2010; **1**(2):112–125.
43. Ades AE. A chain of evidence with mixed comparisons: models for multi-parameter synthesis and consistency of evidence. *Statistics in Medicine* 2003; **22**(19):2995–3016.
44. Ades AE, Sutton AJ. Multiparameter evidence synthesis in epidemiology and medical decision-making: current approaches. *Journal of the Royal Statistical Society: Series A* 2006; **169**(1):5–35.
45. Hedges LV, Tipton E, Johnson MC. Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods* 2010; **1**(1):39–65.
46. Berlin JA, Santanna J, Schmid CH, Szczech LA, Feldman HI. Individual patient- versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. *Statistics in Medicine* 2002; **21**(3):371–387.
47. Lambert PC, Sutton AJ, Abrams KR, Jones DR. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. *Journal of Clinical Epidemiology* 2002; **55**(1):86–94.
48. Olkin I, Sampson A. Comparison of meta-analysis versus analysis of variance of individual patient data. *Biometrics* 1998; **54**(1):317–322.
49. Mathew T, Nordstrom K. Comparison of one-step and two-step meta-analysis models using individual patient data. *Biometrical Journal* 2010; **52**(2):271–287.
50. Riley RD, Thompson JR, Abrams KR. An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown. *Biostatistics* 2008; **9**(1):172.
51. Riley RD. Multivariate meta-analysis: the effect of ignoring within-study correlation. *Journal of the Royal Statistical Society: Series A* 2009; **172**(4):789–811.
52. Sauerbrei W, Royston P. A new strategy for meta-analysis of continuous covariates in observational studies. *Statistics in Medicine* 2011; **30**(28):3341–3360.
53. Schwartz J, Zanobetti A. Using meta-smoothing to estimate dose–response trends across multiple studies, with application to air pollution and daily death. *Epidemiology* 2000; **11**(6):666–672.
54. Dominici F, Daniels MJ, Zeger SL, Samet JM. Air pollution and mortality: estimating regional and national dose–response relationships. *Journal of the American Statistical Association* 2002; **97**:100–111.
55. Samoli E, Analitis A, Touloumi G, Schwartz J, Anderson HR, Sunyer J, Bisanti L, Zmirou D, Vonk JM, Pekkanen J, Goodman P, Paldy A, Schindler C, Katsouyanni K. Estimating the exposure–response relationships between particulate matter and mortality within the APHEA multicity project. *Environmental Health Perspectives* 2005; **113**(1):88–97.



56. Greenland S, Longnecker MP. Methods for trend estimation from summarized dose-response data, with applications to meta-analysis. *American Journal of Epidemiology* 1992; **135**(11):1301–1309.
57. Berlin JA, Longnecker MP, Greenland S. Meta-analysis of epidemiologic dose-response data. *Epidemiology* 1993; **4**(3):218–228.
58. Bagnardi V, Zamboni A, Quatto P, Corrao G. Flexible meta-regression functions for modeling aggregate dose-response data, with an application to alcohol and mortality. *American Journal of Epidemiology* 2004; **159**(11):1077–1086.
59. Hartemink N, Boshuizen HC, Nagelkerke NJ, Jacobs MA, van Houwelingen HC. Combining risk estimates from observational studies with different exposure cutpoints: a meta-analysis on body mass index and diabetes type 2. *American Journal of Epidemiology* 2006; **163**(11):1042–1052.
60. Vlaanderen J, Portengen L, Rothman N, Lan Q, Kromhout H, Vermeulen R. Flexible meta-regression to assess the shape of the benzene-leukemia exposure-response curve. *Environmental Health Perspectives* 2010; **118**(4):526–532.
61. Shi JQ, Copas JB. Meta-analysis for trend estimation. *Statistics in Medicine* 2004; **23**(1):3–19.
62. Orsini N. Meta-analysis for linear and nonlinear dose-response relations: examples, an evaluation of approximations, and software. *American Journal of Epidemiology* 2012; **175**(1):66–73.
63. Liu Q, Cook NR, Bergstrom A, Hsieh CC. A two-stage hierarchical regression model for meta-analysis of epidemiologic nonlinear dose-response data. *Computational Statistics and Data Analysis* 2009; **53**(12):4157–4167.
64. Rota M, Bellocco R, Scotti L, Tramacere I, Jenab M, Corrao G, La Vecchia C, Boffetta P, Bagnardi V. Random-effects meta-regression models for studying nonlinear dose-response relationship, with an application to alcohol and esophageal squamous cell carcinoma. *Statistics in Medicine* 2010; **29**:2679–2687.
65. Arends LR, Hunink MG, Stijnen T. Meta-analysis of summary survival curve data. *Statistics in Medicine* 2008; **27**(22):4381–4396.
66. Ishak KJ, Platt RW, Joseph L, Hanley JA, Caro JJ. Meta-analysis of longitudinal studies. *Clinical Trials* 2007; **4**(5):525–539.
67. Arends LR, Hamza TH, Van Houwelingen JC, Heijnenbroek-Kal MH, Hunink MGM, Stijnen T. Bivariate random effects meta-analysis of ROC curves. *Medical Decision Making* 2008; **28**(5):621–638.
68. Gasparrini A, Armstrong B. The impact of heat waves on mortality. *Epidemiology* 2011; **22**(1):68–73.
69. Zanobetti A, Schwartz J, Samoli E, Gryparis A, Touloumi G, Atkinson R, Le Tertre A, Bobros J, Celko M, Goren A, Forsberg B, Michelozzi P, Rabczenko D, Aranguez Ruiz E, Katsouyanni K. The temporal pattern of mortality responses to air pollution: a multicity assessment of mortality displacement. *Epidemiology* 2002; **13**(1):87–93.
70. Zanobetti A, Schwartz J. Mortality displacement in the association of ozone with mortality: an analysis of 48 cities in the United States. *American Journal of Respiratory and Critical Care Medicine* 2008; **177**(2):184–189.
71. Samoli E, Zanobetti A, Schwartz J, Atkinson R, Le Tertre A, Schindler C, Perez L, Cadum E, Pekkanen J, Paldy A, Touloumi G, Katsouyanni K. The temporal pattern of mortality responses to ambient ozone in the APHEA project. *Journal of Epidemiology and Community Health* 2009; **63**:960–966.
72. Analitis A, Katsouyanni K, Biggeri A, Baccini M, Forsberg B, Bisanti L, Kirchmayer U, Ballester F, Cadum E, Goodman PG, Hojs A, Sunyer J, Tiittanen P, Michelozzi P. Effects of cold weather on mortality: results from 15 European cities within the PHEWE Project. *American Journal of Epidemiology* 2008; **168**(12):1397.
73. Zanobetti A, Schwartz J. Temperature and mortality in nine US cities. *Epidemiology* 2008; **19**(4):563–570.
74. Gasparrini A. Distributed lag linear and non-linear models in R: the package dlnm. *Journal of Statistical Software* 2011; **43**(8):1–20.
75. van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in Medicine* 2002; **21**(4):589–624.