Multivariate Dose-Response Meta-Analysis: the dosresmeta R Package

Alessio Crippa
Karolinska Institutet

Nicola Orsini
Karolinska Institutet

Abstract

An increasing number of quantitative reviews of epidemiological data includes a dose-response analysis. Although user-written procedures have been implemented and widely used in common commercial statistical software (Stata, SAS), no package was available for the free software R. Aims of this paper are to describe the main aspects (covariances of correlated outcomes, pooling of study-specific trends, flexible modelling of the exposure, testing hypothesis, statistical heterogeneity, graphical presentation of the pool dose-response trend) of the methodology and to illustrate the novel package **dosresmeta** for performing multivariate dose-response meta-analysis.

Keywords: dose-response, meta-analysis, mixed-effects model, **dosresmeta**, R.

1. Introduction

Epidemiological studies often assess whether the observed relationship between increasing (or decreasing) levels of exposure and the risk of diseases follows a certain dose-response pattern (U-shaped, J-shaped, linear). Quantitative exposures in predicting health events are frequently categorized and modeled with indicator or dummy variables using one exposure level as referent (Turner, Dobson, and Pocock 2010). Using a categorical approach no specific trend is assumed and data as well as results are typically presented in a tabular form. The purpose of a meta-analysis of summarised dose-response data is to describe the overall functional relation and identify exposure intervals associated with higher or lower disease risk.

A method for dose-response meta-analysis was first described in a seminal paper by Greenland and Longnecker (1992) and since then it has been cited about 574 times (citation data available from Web of Science). Methodological articles next investigated how to model non linear dose-response associations (Orsini, Li, Wolk, Khudyakov, and Spiegelman 2012; Bagnardi, Zambon, Quatto, and Corrao 2004; Liu, Cook, Bergström, and Hsieh 2009; Rota, Bellocco, Scotti, Tramacere, Jenab, Corrao, La Vecchia, Boffetta, and Bagnardi 2010), how to deal with covariances of correlated outcomes (Hamling, Lee, Weitkunat, and Ambühl 2008; Berrington and Cox 2003), how to assess publication bias (Shi and Copas 2004), and how to assign a typical dose to an exposure interval (Takahashi and Tango 2010).

The number of published dose-response meta-analyses increased about 20 times over the last decade, from 6 papers in 2002 to 121 papers in 2013 (number of citations of the paper by Greenland and Longnecker (1992) obtained from Web of Science). The increase of publications has been greatly facilitated by the release of user-written procedures for commonly used

Table 1: Summarized dose-response data required for a dose-response meta-analysis.

dose	cases	n	RR	95% CI
x_0	c_0	n_0	1	—
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_n	c_n	n_n	RR_n	$\underline{\mathrm{RR}}_n$, $\overline{\mathrm{RR}}_n$

commercial statistical software; namely the `glst` command (Orsini, Bellocco, and Greenland 2006) developed for Stata and the `metadose` macro (Li and Spiegelman 2010) developed for SAS. No procedure, however, was available in the free software programming language R.

Aims of the current paper are to describe the main aspects (covariances of correlated outcomes, pooling of study-specific trends, flexible modelling of the exposure, testing hypothesis, statistical heterogeneity, predictions, and graphical presentation of the pool dose-response trend) of the methodology and to illustrate the use of the novel R package **dosresmeta** for performing dose-response meta-analysis. Other factors to be considered when conducting quantitative reviews of dose-response data are selection of studies, exposure measurement error, and potential confounding arising in observational studies.

The paper is organized as follows: section 2 introduces the method for trend estimation for single and multiple studies; section 3 describes how to use the **dosresmeta** package; section 4 provides some worked examples; and section 5 contains final comments.

2. Methods

The data required from each study included in a dose-response meta-analysis are displayed in Table 1. Dose values, x, are assigned in place of the n exposure intervals. The assigned dose is typically the median or midpoint value. Depending on the study design, dose-specific odds ratios, rate ratios, or risk ratios (from now on generally referred to as relative risks (RR)) eventually adjusted for potential confounders are reported together with the corresponding 95% confidence intervals $\left(\underline{RR}, \overline{RR}\right)$ using a common reference category x_0. Additionally, information about the number of cases and total number of subjects or person-time within each exposure category is needed.

We now describe the two stage procedure to estimate a pooled exposure-disease curve from such summarized dose-response data.

2.1. First stage: trend estimation for a single study

In the first stage of the analysis the aim is to estimate the dose-response association between the adjusted log relative risks and the levels of a specific exposure in a particular study.

Model definition

The model consists of a log-linear model where the dependent variable is the logarithm of the relative risks (logRR) as function of the dose

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{1}$$

where $\mathbf{y}$ is an $n \times 1$ vector of adjusted log relative risks (not including the reference one) and $\mathbf{X}$ is a $n \times p$ matrix containing the non-referent values of the dose and/or some transforms of it (e.g., splines, polynomials).

$$\mathbf{X} = \begin{bmatrix} g_1(x_{1j}) - g_1(x_{0j}) & \dots & g_p(x_{pj}) - g_p(x_{0j}) \\ \vdots & & \vdots \\ g_1(x_{n_j j}) - g_1(x_{0j}) & \dots & g_p(x_{n_j j}) - g_p(x_{0j}) \end{bmatrix} \tag{2}$$

Of note, the design matrix $\mathbf{X}$ has no intercept because the log relative risk is equal to zero for the reference exposure value x_0. A linear trend ($p = 1$) implies that g_1 is the identity function. A quadratic trend ($p = 2$) implies that g_1 is the identity function and g_2 is the squared function. The β is a $p \times 1$ vector of unknown regression coefficients defining the pooled dose-response association.

Approximating covariances

A particular feature of dose-response data is that the error terms in ϵ are not independent because are constructed using a common reference (unexposed) group. It has been shown that assuming zero covariance or correlation leads to biased estimates of the trend (Greenland and Longnecker 1992; Orsini *et al.* 2012).

The variance-covariance matrix $\mathsf{COV}(\boldsymbol{\epsilon})$ is equal to the following symmetric matrix

$$\mathsf{COV}(\boldsymbol{\epsilon}) = \mathbf{S} = \begin{bmatrix} \sigma_1^2 & & & & \\ \vdots & \ddots & & & \\ \sigma_{i1} & & \sigma_i^2 & & \\ \vdots & & & \ddots & \\ \sigma_{n1} & \dots & \sigma_{ni} & \dots & \sigma_n^2 \end{bmatrix}.$$

where the covariance among (log) relative risks implies the non-diagonal elements of $\mathbf{S}$ are unlikely equal to zero. Briefly, Greenland and Longnecker (1992) approximate the covariances by defining a table of pseudo or effective counts corresponding to the multivariable adjusted log relative risks. A unique solution is guaranteed by keeping the margins of the table of pseudo-counts equal to the margins of the crude or unadjusted data.

More recently Hamling *et al.* (2008) developed an alternative method to approximate the covariances by defining a table of effective counts corresponding to the multivariable adjusted log relative risks as well as their standard errors. A unique solution is guaranteed by keeping the ratio of non-cases to cases and the fraction of unexposed subjects equal to the unadjusted data. An evaluation of those approximations can be found in Orsini *et al.* (2012).

There is no need of such methods, however, when the average covariance is directly published using the floating absolute risk method (Easton, Peto, and Babiker 1991) or the estimated variance/covariance matrix is provided directly by the principal investigator in pooling projects or pooling of standardized analyses.

Estimation

The approximated covariance can be then used to efficiently estimate the vector of regression coefficients $\boldsymbol{\beta}$ of the model in Equation 1 using generalized least squares method. The method involves minimizing $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$. The vector of estimates $\hat{\boldsymbol{\beta}}$ and the estimated covariance matrix $\hat{\mathbf{V}}$ are obtained as:

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}^\top \mathbf{S}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top \mathbf{S}^{-1}\mathbf{y} \\ \hat{\mathbf{V}} &= (\mathbf{X}^\top \mathbf{S}^{-1}\mathbf{X})^{-1}\end{aligned} \tag{3}$$

2.2. Second stage: trend estimation for multiple studies

The aim of the second stage analysis is to combine study-specific estimates using established methods for multivariate meta-analysis (Van Houwelingen, Arends, and Stijnen 2002; White 2009; Jackson, White, and Thompson 2010; Gasparrini, Armstrong, and Kenward 2012; Jackson, White, and Riley 2013).

Model definition

Let us indicate each study included in the meta-analysis with the index $j = 1, \ldots, m$. The first stage provides p-length vector of parameters $\hat{\boldsymbol{\beta}}_j$ and the accompanying $p \times p$ estimated covariance matrix $\hat{\mathbf{V}}_j$. The coefficients $\hat{\boldsymbol{\beta}}_j$ obtained in the first stage analysis are now used as outcome in a multivariate meta-analysis

$$\hat{\boldsymbol{\beta}}_j \sim N_p(\boldsymbol{\beta}, \hat{\mathbf{V}}_j + \boldsymbol{\psi}) \tag{4}$$

where $\hat{\mathbf{V}}_j + \boldsymbol{\psi} = \boldsymbol{\Sigma}_j$. The marginal model defined in Equation 4 has independent within-study and between-study components. In the within-study component, the estimated $\hat{\boldsymbol{\beta}}_j$ is assumed to be sampled with error from $N_p(\boldsymbol{\beta}_j, \hat{\mathbf{V}}_j)$, a multivariate normal distribution of dimension p, where $\boldsymbol{\beta}_j$ is the vector of true unknown outcome parameters for study j. In the between-study component, $\boldsymbol{\beta}_j$ is assumed sampled from $N_p(\boldsymbol{\beta}, \boldsymbol{\psi})$, where $\boldsymbol{\psi}$ is the unknown between-study covariance matrix. Here $\boldsymbol{\beta}$ can be interpreted as the population-average outcome parameters, namely the coefficients defining the pooled dose-response trend. The **dosresmeta** package refers to the **mvmeta** package for the second-stage analysis (Gasparrini *et al.* 2012). Different methods are available to estimate the parameters: fixed effects ($\boldsymbol{\psi}$ sets to $\mathbf{0}$), maximum likelihood, restricted maximum likelihood, and methods of moments.

Hypothesis testing and heterogeneity

There are two questions of interest in a dose-response meta-analysis: is there an association between the quantitative exposure and the relative risk of disease? Is there significant statistical heterogeneity across the study-specific trends?

The overall exposure-disease curve is specified by the vector of coefficients $\boldsymbol{\beta}$. Thus the first question can be addressed by testing $H_0 : \boldsymbol{\beta} = \mathbf{0}$. If we reject H_0 we may be interested in evaluating any possible departure from a log-linear model. This can be done by testing $H_0^* : \boldsymbol{\beta}^* = \mathbf{0}$, where $\boldsymbol{\beta}^*$ refers to the vector of coefficients defining non-linearity (e.g. quadratic term, spline transformations). Wald-type confidence intervals and tests of hypothesis for $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^*$ can be based on $\hat{\boldsymbol{\beta}}$ and its covariance matrix (Harville 1977).

Variation of dose-response trends across studies, instead, is captured by the variance components included in $\boldsymbol{\psi}$. The hypothesis of no statistical heterogeneity, $H_0 : \boldsymbol{\psi} = \mathbf{0}$, can be tested by adopting the multivariate extension of the Cochran Q-test (Berkey, Anderson, and Hoaglin 1996). The test is defined as

$$Q = \sum_{j=1}^{m} \left[\left(\hat{\boldsymbol{\beta}}_j - \hat{\boldsymbol{\beta}} \right)^\top \hat{\mathbf{V}}_j^{-1} \left(\hat{\boldsymbol{\beta}}_j - \hat{\boldsymbol{\beta}} \right) \right] \tag{5}$$

where $\hat{\boldsymbol{\beta}}$ is the vector of regression coefficients estimated assuming $\boldsymbol{\psi} = \mathbf{0}$. Under the null hypothesis of no statistical heterogeneity among studies the Q statistic follow a χ^2_{m-p} distribution. It has been shown, however, that the test is likely not to detect heterogeneity in meta-analyses of small number of studies, and conversely can lead to significant results even for negligible discrepancies in case of many studies (Higgins and Thompson 2002).

A measure of heterogeneity $I^2 = max\left\{0, (Q - df)/Q\right\}$ can be derived from the Q statistic and its degrees of freedom ($df = m - p$) (Higgins and Thompson 2002). Its use, along with the Q, has been recommended because it describes the impact rather than the extent of heterogeneity.

Prediction

Obtaining predictions is an important step to present the results of a dose-response meta-analysis in either a tabular or graphical form. The prediction of interest in a dose-response analysis is the relative risk for the disease comparing two exposure values. Given a range of exposure x and a chosen reference value x_{ref}, the predicted pooled dose-response association can be obtained as follows

$$\widehat{\text{RR}}_{\text{ref}} = \exp\left\{ (\mathbf{X} - \mathbf{X}_{\text{ref}}) \hat{\boldsymbol{\beta}} \right\} \tag{6}$$

where $\mathbf{X}$ and $\mathbf{X}_{\text{ref}}$ are the design matrices evaluated, respectively, in x and x_{ref}.

A $(1 - \alpha/2)\,\%$ confidence interval for the predict pooled dose-response curve is given by

$$\exp\left\{ \log(\widehat{\text{RR}}_{\text{ref}}) \mp z_{\alpha/2} \text{diag} \left((\mathbf{X} - \mathbf{X}_{\text{ref}}) \hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}) (\mathbf{X} - \mathbf{X}_{\text{ref}})^\top) \right)^{1/2} \right\} \tag{7}$$

where $\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}})$ is the estimated covariance matrix of $\hat{\boldsymbol{\beta}}$. Of note, by construction the confidence intervals limits for the pooled relative risks are equal to 1 for the reference exposure x_{ref}.

3. The dosresmeta package

The **dosresmeta** package performs multivariate dose-response meta-analysis. The package is available via at `http://CRAN.R-project.org/package=dosresmeta` and can be installed directly within R by typing `install.packages("dosresmeta")`.

The function `dosresmeta` estimates a dose-response model for either a single or multiple studies. We now describe the different arguments of the function.

```
dosresmeta(formula, id, type, v, cases, n, data, intercept = F, center = T,
  se, lb, ub, covariance = "gl", method = "reml", fcov, ucov,
  alpha = 0.05, ...)
```

The argument `formula` defines the relation between the outcome and the dose.

The argument `id` requires an identification variable for the study while the argument `type` specifies the study-specific design. The codes for the study design are `"cc"` (case-control data), `"ir"` (incidence-rate data), and `"ci"` (cumulative incidence data).

The argument `v` requires the variances of the log relative risks. Alternatively one can provide the corresponding standard errors in the `se` argument, or specify the confidence interval for the relative risks in the `lb` and `ub` arguments.

The arguments `cases` and `n` requires the variables needed to approximate the covariance matrix of the log relative risks: number of cases and total number of subject for each exposure level. For incidence-rate data `n` requires the amount of Person-Time for each exposure level.

The argument `data` specifies the name of the data set containing the variables in the previous arguments.

The logical argument `intercept`, `FALSE` by default, indicates if an intercept term needs to be included in the model. As mentioned earlier the model in Equation 1 typically does not contain the constant term. The logical argument `center`, `TRUE` by default, specifies if the design matrix of the model should be constructed as defined in Equation 2.

The `method` is a string that specifies the estimation method: `"fixed"` for fixed-effects models, `"ml"` or `"reml"` for random-effects models fitted through maximum likelihood or restricted maximum likelihood (default), and `"mm"` for random-effects models fitted through method of moments.

The argument `covariance`, instead, is a string specifying how to approximate the covariance matrix for the log relative risks: `"gl"` for Greenland and Longnecker method (default), `"h"` for Hamling method, `"fl"` for floating absolute risks, `"independent"` for assuming independence, and `"user"` if the covariance matrices are provided by the user.

The output of the `dosresmeta` function is an object of class "`dosresmeta`". The corresponding `print` and `summary` methods can be used to display and inspect the elements of the object. The `predict` method allows the user to obtain predictions as described in Section 2.2. The predictions can be expressed on the RR scale by specifying the optional argument `expo` equal to `TRUE`. The increase in the log RR associated to a d unit increase in the exposure can be obtained with the optional argument `delta = d`.

4. Examples

Single study

Consider the case-control study used by Greenland and Longnecker (1992) on alcohol consumption (grams/day) and breast cancer risk. The data `cc_ex` is included in the `dosresmeta` package.

```
R> library("dosresmeta")
R> data("cc_ex")
R> print(cc_ex, row.names = F, digits = 2)

 gday dose case control   n crudeor adjrr   lb  ub logrr
```

```
    Ref.     0  165      172 337     1.00     1.0 1.00 1.0  0.00
    <2.5     2   74       93 167     0.83     0.8 0.51 1.3 -0.22
 2.5-9.3     6   90       96 186     0.98     1.2 0.73 1.9  0.15
    >9.3    11  122       90 212     1.41     1.6 0.99 2.5  0.45
```

Assuming a log-linear dose-response association between alcohol consumption and breast cancer risk, we estimate the model using the following code

```
R> mod.cc <- dosresmeta ( formula = logrr ~ dose, type = "cc", cases = case,
+                n = n, lb = lb, ub = ub, data = cc_ex)
R> summary(mod.cc)

Call:  dosresmeta(formula = logrr ~ dose, type = "cc", cases = case,
    n = n, data = cc_ex, lb = lb, ub = ub)

One-stage fixed-effects meta-analysis
Covariance approximation: Greenland & Longnecker

Chi2 model: X2 = 4.8333 (df = 1), p-value = 0.0279

Fixed-effects coefficients
      Estimate  Std. Error       z  Pr(>|z|)  95%ci.lb  95%ci.ub
dose    0.0454      0.0207  2.1985    0.0279    0.0049    0.0859  *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

1 study 3 values, 1 fixed and 0 random-effects parameters
 logLik      AIC      BIC
 0.7835   0.4330  -0.4684
```

The change in the log relative risk of breast cancer corresponding to 1 gram/day increase in alcohol consumption was 0.0454. On the exponential scale, every 1 gram/day increase of alcohol consumption was associated with a 4.6 % ($\exp(0.0454) = 1.046$) higher breast cancer risk. The `predict` function allows the user to express the log-linear trend for any different amount. For example, by setting `delta = 11`

```
R> predict(mod.cc, delta = 11, exp = TRUE)

 delta     pred    ci.lb    ci.ub
    11 1.648255 1.055709 2.573384
```

Every 11 grams/day increment in alcohol consumption was associated with a significant 65% (95% CI = 1.06 , 2.57) higher breast cancer risk.

Multiple studies

We now perform a dose-response meta-analysis of 8 prospective cohort studies participating in the Pooling Project of Prospective Studies of Diet and Cancer used by Orsini *et al.* (2012).

There are 6 exposure intervals (from 0 grams/day to 45 grams/day) for each study. A total of 3,646 cases and 2,511,424 person-years were included in the dose-response analysis. All relative risks were adjusted for smoking status, smoking duration for past and current smokers (years), number of cigarettes smoked daily for current smokers, educational level, body mass index, and energy intake (kcal/day).

The data `ex_alcohol_crc` is included in the `dosresmeta` package. Below is a snapshot of the data set for the first two studies.

```
R> data("alcohol_crc")
R> print(alcohol_crc[1:12 ,], row.names = F, digits = 2)

  id type dose cases peryears  logrr   se
 atm   ir  0.0    28    22186  0.000   NA
 atm   ir  1.8    38    43031 -0.417 0.25
 atm   ir  9.2    43    53089 -0.396 0.25
 atm   ir 22.9    32    45348 -0.488 0.26
 atm   ir 35.7    16    19791 -0.279 0.32
 atm   ir 58.4    27    19920  0.202 0.29
 hpm   ir  0.0   100   103002  0.000   NA
 hpm   ir  2.1    65   106826 -0.416 0.16
 hpm   ir  9.5   104   119846 -0.099 0.14
 hpm   ir 18.8    63    58034  0.094 0.17
 hpm   ir 36.7    46    33081  0.205 0.18
 hpm   ir 59.4    30    18455  0.343 0.22
```

First we assume a log-linear relation between alcohol consumption and colorectal cancer risk using a random-effect model. The estimation is carried out by running the following line

```
R> lin <- dosresmeta(formula = logrr ~ dose, id = id, type = type, se = se,
+                      cases = cases, n = peryears, data = alcohol_crc)
R> summary(lin)

Call:  dosresmeta(formula = logrr ~ dose, id = id, type = type, cases = cases,
    n = peryears, data = alcohol_crc, se = se)

Two-stage random-effects meta-analysis
Estimation method: REML
Covariance approximation: Greenland & Longnecker

Chi2 model: X2 = 22.6704 (df = 1), p-value = 0.0000

Fixed-effects coefficients
             Estimate  Std. Error       z  Pr(>|z|)  95%ci.lb
(Intercept)    0.0064      0.0014  4.7613    0.0000    0.0038
             95%ci.ub
(Intercept)    0.0091  ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Between-study random-effects (co)variance components
  Std. Dev
    0.0000

Univariate Cochran Q-test for residual heterogeneity:
Q = 4.7797 (df = 7), p-value = 0.6868
I-square statistic = 0.0%

8 studies, 8 values, 1 fixed and 1 random-effects parameters
  logLik       AIC       BIC
 29.6176  -55.2352  -55.3434
```

We found a significant log-linear dose-response association between alcohol consumption and colorectal cancer risk ($p < 0.001$) and no evidence of heterogeneity across studies ($Q = 4.78$, p value $= 0.6868$).

The change in colorectal cancer risk associated with every 12 grams/day (standard drink) can be obtained with the `predict` function.

```
R> predict(lin, delta = 12, exp = TRUE)

 delta     pred  ci.lb    ci.ub
    12 1.080313 1.0465 1.115218
```

Every 12 grams/day increase in alcohol consumption was associated with a significant 8% (95% CI = 4.6, 11.5) increased risk of colorectal cancer.

The log-linear assumption between alcohol consumption and colorectal cancer risk can be relaxed by using regression splines. A possibility is to use restricted cubic spline model as described by Orsini *et al.* (2012): 4 knots at the 5th, 35th, 65th and 95th percentiles of the aggregated exposure distribution. The **rms** package (Harrell 2013) provides three $(4 - 1)$ variables (the initial exposure and two splines transformations) to be included in the dose-response model. Statistical heterogeneity across studies can be taken into account by using a random-effects approach.

```
R> library("rms")
R> knots <- quantile(alcohol_crc$dose, c(.05, .35, .65, .95))
R> spl <- dosresmeta(formula = logrr ~ rcs(dose, knots), type = type, id = id,
+                    se = se, cases = cases, n = peryears,  data = alcohol_crc)
R> summary(spl)

Call:  dosresmeta(formula = logrr ~ rcs(dose, knots), id = id, type = type,
    cases = cases, n = peryears, data = alcohol_crc, se = se)

Two-stage random-effects meta-analysis
```

```
Estimation method: REML
Covariance approximation: Greenland & Longnecker

Chi2 model: X2 = 28.3632 (df = 3), p-value = 0.0000

Fixed-effects coefficients
                        Estimate  Std. Error        z  Pr(>|z|)
rcs(dose, knots)dose     -0.0097      0.0082  -1.1937    0.2326
rcs(dose, knots)dose'     0.1274      0.0808   1.5757    0.1151
rcs(dose, knots)dose''   -0.2144      0.1488  -1.4407    0.1497
                        95%ci.lb  95%ci.ub
rcs(dose, knots)dose     -0.0257    0.0062
rcs(dose, knots)dose'    -0.0311    0.2858
rcs(dose, knots)dose''   -0.5061    0.0773
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Between-study random-effects (co)variance components
                        Std. Dev                  Corr
rcs(dose, knots)dose      0.0112  rcs(dose, knots)dose
rcs(dose, knots)dose'     0.0777                    -1
rcs(dose, knots)dose''    0.1274                     1

rcs(dose, knots)dose    rcs(dose, knots)dose'
rcs(dose, knots)dose'
rcs(dose, knots)dose''                     -1

Univariate Cochran Q-test for residual heterogeneity:
Q = 20.5624 (df = 21), p-value = 0.4859
I-square statistic = 0.0%

8 studies, 24 values, 3 fixed and 6 random-effects parameters
  logLik       AIC       BIC
 43.0662  -68.1324  -58.7317
```

The risk of colorectal cancer is significantly varying according to alcohol consumption. We reject the null hypothesis of overall no association (all three regression coefficients simultaneously equal to zero) between alcohol consumption and colorectal cancer risk ($\chi^2 = 28.36$, $p < 0.001$). The simpler log-linear dose-response model can be obtained from the restricted cubic spline model by constraining the regression coefficients for the second and the third spline, `rcs(dose, knots).dose'` and `rcs(dose, knots).dose"`, equal to zero. Therefore, a Wald-type test for the hypothesis of deviation from log-linearity can be carried out as follows

```
R> waldtest(b = coef(spl), Sigma = vcov(spl), Terms = 2:3)

Wald test:
```

```
----------

Chi-squared test:
X2 = 5.1, df = 2, P(> X2) = 0.078
```

The marginally significant p value ($\chi^2 = 5.1$, $p = 0.078$) suggested that the risk of colorectal cancer may not vary in a log-linear fashion with alcohol consumption. We presented the predicted relative risks arising from the log-linear and restricted cubic spline models (Figure 1) using 0 grams/day as referent ($x_{\text{ref}} = 0$).

```
R> newdata <- data.frame(dose <- seq(0, 60, 1))
R> xref <- 0
R> with(predict(spl, newdata, xref, exp = TRUE),{
+    plot(get("rcs(dose, knots)dose"), pred, type = "l", ylim = c(.8, 1.8),
+         ylab = "Relative risk", xlab = "Alcohol consumption, grams/day",
+         log = "y", bty = "l", las = 1)
+    matlines(get("rcs(dose, knots)dose"), cbind(ci.ub, ci.lb),
+             col = 1, lty = "dashed")
+  })
R> points(dose, predict(lin, newdata, xref)$pred, type = "l", lty = 3)
R> rug(alcohol_crc$dose)
```

A tabular presentation of predicted point and interval estimates of the relative risks for selected values of the exposure of the two fitted models is greatly facilitated by the `predict` function. For example, a table of pooled relative risks of colorectal cancer risk for a range of alcohol consumption between 0 and 60 grams/day (with step by 12 grams/day) is obtained as follow

```
R> dataTab <- data.frame(dose = seq(0, 60, 12))
R> predLin <- predict(lin, dataTab, exp = TRUE)
R> predSpl <- predict(spl, dataTab, exp = TRUE)
R> round(cbind(lin = predLin, spl = predSpl[4:6]), 2)

  lin.dose lin.pred lin.ci.lb lin.ci.ub spl.pred spl.ci.lb spl.ci.ub
1        0     1.00      1.00      1.00     1.00      1.00      1.00
2       12     1.08      1.05      1.12     0.95      0.84      1.07
3       24     1.17      1.10      1.24     1.07      0.94      1.21
4       36     1.26      1.15      1.39     1.22      1.07      1.39
5       48     1.36      1.20      1.55     1.37      1.20      1.56
6       60     1.47      1.26      1.73     1.53      1.25      1.88
```

The reference exposure used to obtain predicted relative risks can be easily modified by setting x_{ref} to a different value. For example, one could use $x_{\text{ref}} = 12$ grams/day (corresponding to 1 standard drink) as graphically shown in Figure 2.

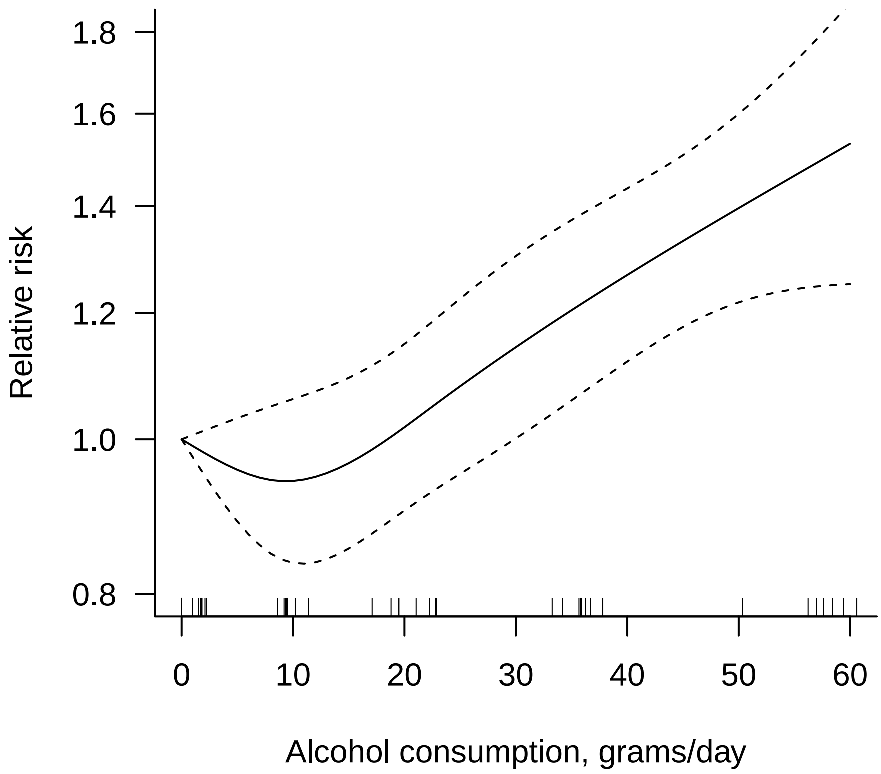

Figure 1: Pooled dose-response association between alcohol consumption and colorectal cancer risk (solid line). Alcohol consumption was modeled with restricted cubic splines in a multivariate random-effects dose-response model. Dash lines represent the 95% confidence intervals for the spline model. The dotted line represents the linear trend. Tick marks below the curve represent the positions of the study-specific relative risks. The value of 0 grams/day served as referent. The relative risks are plotted on the log scale.

5. Conclusion

We presented the key elements involved in a dose-response meta-analysis of epidemiological data. These elements includes reconstructing the variance/covariance matrix of published relative risks, testing hypothesis, predictions, and graphical presentation of the pooled trend. We described a two-stage approach to combine either linear or non-linear dose-response associations. Applications of the novel R package **dosresmeta** were illustrated through several worked examples.

One strength of this paper is to provide an accessible introduction to multivariate dose-response meta-analysis. In particular, we showed how to flexibly model a quantitative exposure using spline transformations and how to present graphically the predicted pooled relative risks using different reference values. Although dose-response meta-analyses are increasingly

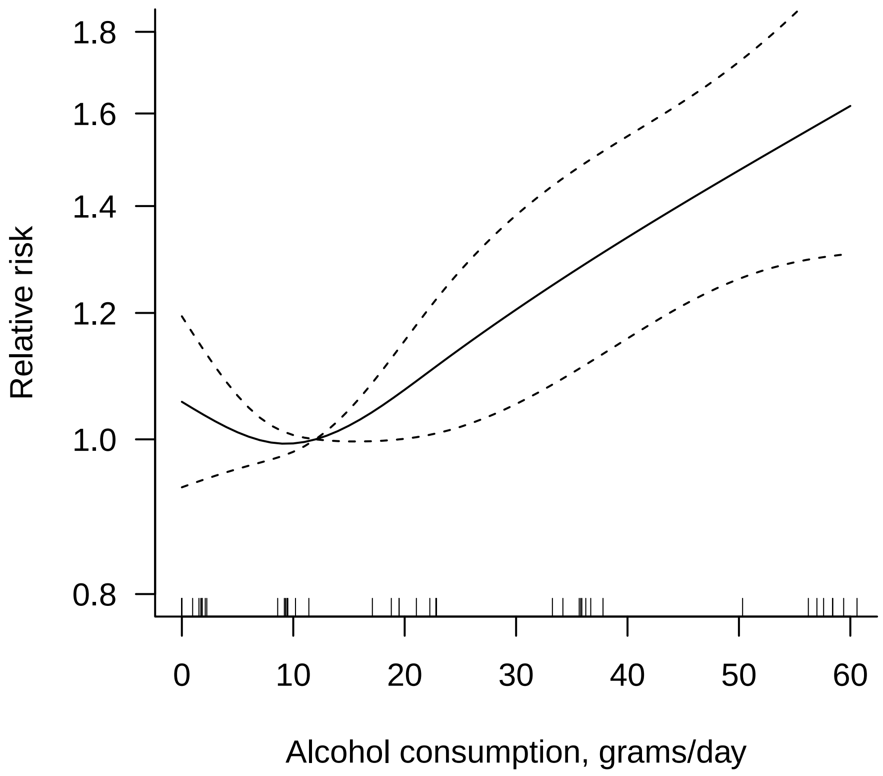

Figure 2: Pooled dose-response association between alcohol consumption and colorectal cancer risk (solid line). Alcohol consumption was modeled with restricted cubic splines in a multivariate random-effects dose-response model. Dash lines represent the 95% confidence intervals for the spline model. The dotted line represents the linear trend. Tick marks below the curve represent the positions of the study-specific relative risks. The value of 12 grams/day served as referent. The relative risks are plotted on the log scale.

popular and published in the medical literature using commercial statistical software, no procedures were available for the free software programming language of R. Therefore, another strength of this paper is to present the first release of the **dosresmeta** package written for R.

In conclusion, this paper described the main steps involved in a dose-response meta-analysis. The **dosresmeta** package as well as worked examples can be useful to introduce researchers to the application of this increasingly popular method.

References

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." *American Journal of Epidemiology*, **159**(11), 1077–1086.

Berkey C, Anderson J, Hoaglin D (1996). "Multiple-Outcome Meta-Analysis of Clinical Trials." *Statistics in medicine*, **15**(5), 537–557.

Berrington A, Cox D (2003). "Generalized Least Squares for the Synthesis of Correlated Information." *Biostatistics*, **4**(3), 423–431.

Easton DF, Peto J, Babiker AG (1991). "Floating Absolute Risk: An Alternative to Relative Risk in Survival and Case-Control Analysis Avoiding an Arbitrary Reference Group." *Statistics in medicine*, **10**(7), 1025–1035.

Gasparrini A, Armstrong B, Kenward MG (2012). "Multivariate Meta-Analysis for Non-Linear and Other Multi-Parameter Associations." *Statistics in Medicine*, **31**(29), 3821–3839.

Greenland S, Longnecker MP (1992). "Methods for Trend Estimation from Summarized Dose-Response Data, with Applications to Meta-Analysis." *American journal of epidemiology*, **135**(11), 1301–1309.

Hamling J, Lee P, Weitkunat R, Ambühl M (2008). "Facilitating Meta-Analyses by Deriving Relative Effect and Precision Estimates for Alternative Comparisons from a Set of Estimates Presented by Exposure Level or Disease Category." *Statistics in medicine*, **27**(7), 954–970.

Harrell FE (2013). ***rms**: Regression Modeling Strategies*. R package version 4.1-0, URL http://CRAN.R-project.org/package=rms.

Harville DA (1977). "Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems." *Journal of the American Statistical Association*, **72**(358), 320–338.

Higgins J, Thompson SG (2002). "Quantifying Heterogeneity in a Meta-Analysis." *Statistics in medicine*, **21**(11), 1539–1558.

Jackson D, White IR, Riley RD (2013). "A Matrix-Based Method of Moments for Fitting the Multivariate Random Effects Model for Meta-Analysis and Meta-Regression." *Biometrical Journal*.

Jackson D, White IR, Thompson SG (2010). "Extending DerSimonian and Laird's Methodology to Perform Multivariate Random Effects Meta-Analyses." *Statistics in medicine*, **29**(12), 1282–1297.

Li R, Spiegelman D (2010). "The SAS% METADOSE Macro." URL http://www.hsph.harvard.edu/donna-spiegelman/software/metadose/.

Liu Q, Cook NR, Bergström A, Hsieh CC (2009). "A Two-Stage Hierarchical Regression Model for Meta-Analysis of Epidemiologic Nonlinear Dose-Response data." *Computational Statistics & Data Analysis*, **53**(12), 4157–4167.

Orsini N, Bellocco R, Greenland S (2006). "Generalized Least Squares for Trend Estimation of Summarized Dose-Response Data." *Stata Journal*, **6**(1), 40.

Orsini N, Li R, Wolk A, Khudyakov P, Spiegelman D (2012). "Meta-Analysis for Linear and Nonlinear Dose-Response Relations: Examples, an Evaluation of Approximations, and Software." *American journal of epidemiology*, **175**(1), 66–73.

Rota M, Bellocco R, Scotti L, Tramacere I, Jenab M, Corrao G, La Vecchia C, Boffetta P, Bagnardi V (2010). "Random-effects Meta-Regression Models for Studying Nonlinear Dose-Response Relationship, with an Application to Alcohol and Esophageal Squamous Cell Carcinoma." *Statistics in medicine*, **29**(26), 2679–2687.

Shi JQ, Copas J (2004). "Meta-Analysis for Trend Estimation." *Statistics in medicine*, **23**(1), 3–19.

Takahashi K, Tango T (2010). "Assignment of Grouped Exposure Levels for Trend Estimation in a Regression Analysis of Summarized Data." *Statistics in medicine*, **29**(25), 2605–2616.

Turner EL, Dobson JE, Pocock SJ (2010). "Categorisation of continuous risk factors in epidemiological publications: a survey of current practice." *Epidemiologic Perspectives & Innovations*, **7**(1), 9.

Van Houwelingen HC, Arends LR, Stijnen T (2002). "Advanced Methods in Meta-Analysis: Multivariate Approach and Meta-Regression." *Statistics in medicine*, **21**(4), 589–624.

White IR (2009). "Multivariate Random-Effects Meta-Analysis." *Stata Journal*, **9**(1), 40.

Affiliation:

Alessio Crippa
Unit of Nutritional Epidemiology
Unit of Biostatistics
Institute of Environmental Medicine, Karolinska Institutet
P.O. Box 210, SE-171 77, Stockholm, Sweden
E-mail: alessio.crippa@ki.se
URL: http://www.imm.ki.se/biostatistics