From the Department of Public Health Sciences
Karolinska Institutet, Stockholm, Sweden

NOVEL METHODS FOR DOSE-RESPONSE META-ANALYSIS

Alessio Crippa

Karolinska Institutet

NOVEL METHODS FOR DOSE-RESPONSE META-ANALYSIS

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Alessio Crippa

Lecture hall Inghesalen, Widerströmska huset, Karolinska Institutet, Solna
Friday the 13th of April, 2018, at 10:00

Principal supervisor:
Associate Professor Nicola Orsini
Karolinska Institutet
Department of Public Health Sciences

Co-supervisor:
Professor Alicja Wolk
Karolinska Institutet
Institute of Environmental Medicine

Professor Matteo Bottai
Karolinska Institutet
Institute of Environmental Medicine

Professor Donna Spiegelman
Harvard T.H. Chan School of Public Health
Department of Epidemiology

Opponent:
Professor Christopher H. Schmid
Brown University
Center for Evidence Based Medicine

Examination board:
Associate Professor Nele Brusselaers
Karolinska Institutet
Department of Microbiology, Tumor and Cell Biology

Associate Professor Antonio Gasparrini
London School of Hygiene and Tropical Medicine
Department of Social & Environmental Health Research

Professor Paul Lambert
University of Leicester
Department of Health Sciences

"The function of the expert reviewer is not to be more right than other people,
but to be wrong for more sophisticated reasons."
—Iain Chalmers and Douglas G. Altman
Systematic Reviews, 1995

Abstract

Dose–response meta-analysis is a statistical procedure for combining and contrasting the evidence on the association between a continuous exposure and the risk of a health outcome. Different papers refined selected aspects of the methodology, such as implementation of flexible strategies and extensions to multivariate meta-analysis. However, there were still several relevant questions that needed to be addressed. This thesis aims to address these issues by developing and implementing new strategies and ad-hoc measures (Paper I), including tools for evaluating the goodness-of-fit (Paper II), a new measure for quantifying the impact of heterogeneity (Paper III), a strategy to deal with differences in the exposure range across studies (Paper IV), and a one-stage approach to estimate complex models without excluding relevant studies (Paper V).

In Paper I, we described the implementation of the main aspects of the methodology in the **dosresmeta** R package available on CRAN. Dedicated functions were written to facilitate specific tasks such as definition of the design matrix and prediction of the pooled results. We illustrated how to estimate both linear and non-linear curves, conduct test of hypotheses, and present the results in a tabular and graphical format reanalyzing published aggregated dose–response data.

In Paper II, we discussed how to evaluate the goodness-of-fit. The proposed solutions consist of descriptive measures to summarize the agreement between fitted and observed data (the deviance and the coefficient of determination), and graphical tools to visualize the fit of the model (decorrelated residuals-versus-exposure plot). A reanalysis of two published meta-analyses exemplified how these tools can improve the practice of quantitative synthesis of aggregated dose–response data.

In Paper III, we proposed and characterized a new measure, $\hat{R}_b$, to quantify the proportion of the variance of the pooled estimate attributable to the between-study heterogeneity. Contrary to the available measures of heterogeneity, $\hat{R}_b$ does not make any assumption about the distribution of the within-study error variances, nor does it require specification of a typical value for these quantities. The performance of the proposed measure was evaluated in an extensive simulation study. We demonstrated how to present and interpret the $\hat{R}_b$ re-analyzing three published meta-analyses.

In Paper IV, we extended a point-wise approach to dose–response meta-analysis of aggregated data. The strategy consists of combining predicted relative risks for a fine grid of exposure values based on potentially different dose–response models. A point-wise approach can improve the flexibility in modeling the study-specific curves and may limit the impact of extrapolation by predicting the study-specific relative risks based on the observe exposure range. We illustrated the methodology using both individual and aggregated participant data.

In Paper V, we formalized a one-stage approach for dose–response meta-analysis in terms of a linear mixed model. We explained the main aspects of the methodology and how to address the same questions frequently answered in a two-stage analysis. Using both hypothetical and real data, we showed how the one-stage approach can facilitate the assessment of heterogeneity over the exposure range, model comparison, and prediction of individual dose–response associations. The main advantage is that flexible curves can be estimated regardless of the number of data-points in the individual analyses.

In conclusion, the methods presented in this thesis enrich the set of tools available for applying dose–response meta-analyses and for addressing specific questions including goodness-of-fit evaluation (Paper II) and quantification of heterogeneity (Paper III). In addition, we presented alternative models for pooling results in case of heterogeneous exposure range (Paper IV) and for estimating complex models without excluding relevant studies (Paper V). The proposed methods have been illustrated using real data and implemented in user-friendly R packages available on CRAN (Paper I).

Sammanfattning

Dos-respons metaanalys är ett statistiskt förfarande för att kombinera och jämföra resultat från epidemiologiska studier där sambandet mellan en kontinuerlig variabel och en hälsorisk har undersökts. Tidigare studier har förfinat delar av metoden, till exempel genom införande av flexibla strategier och utökning till multivariata modeller; men trots detta kvarstår flera relevanta metodologiska frågor. Denna avhandling syftar till att besvara ett antal av dessa frågor genom att utveckla och presentera: nya strategier och ad hoc-modeller (Artikel 1); nya metoder för att bedöma goodness-of-fit (Artikel 2); ett nytt mått för att kvantifiera påverkan av heterogenitet (Artikel 3); en ny strategi för att hantera storleksskillnader i exponering mellan studier (Artikel 4); och en ny enstegsmetod för att estimera komplexa modeller utan att exkludera relevanta studier (Artikel 5).

Artikel 1 beskriver hur huvuddelarna av ovanstående metoder har införts i R-paketet **dosresmeta**, tillgängligt via CRAN. De nya funktioner implementerades för att förenkla vissa uppgifter, så som att definiera en designmatris och prediktera sammanslagna resultat. Artikeln illustrerar även beräkningen av linjära och icke-linjära kurvor och utförandet av hypotestester. Utöver detta presenteras resultat från återanalyserade metaanalyser med sammanslagen dos-responsdata.

Artikel 2 utvärderar bedömningen av goodness-of-fit-testet. Den föreslagna metoden består dels av beräkning av deskriptiva mätvärden för att summera likheter och skillnader mellan predikterade och observerade data (avvikelse- och bestämningskoefficient), och dels av grafiska verktyg för att visualisera den predikterade modellen. En åter analys av publicerade metaanalyser exemplifierar hur metoden kan användas för att förbättra kvantitativ syntes av sammanslagen dos-responsdata.

Artikel 3 presenterar ett nytt mått ($\hat{R}_b$) för att kvantifiera andelen varians i den sammanslagna skattningen som kan förklaras av heterogeniteten mellan olika studier. I motsats till tidigare mått på heterogenitet kräver $\hat{R}_b$ varken något antagande om fördelningen av inom-studiefelvariationer eller någon specifikation av deras värden. Resultatet av det föreslagna måttet har utvärderats i en omfattande simuleringsstudie samt genom återanalys av publicerade metaanalyser.

Artikel 4 beskriver hur vi har vidareutvecklat ett punktvist tillvägagångssätt för metaanalys med sammanslagen dos-responsdata. Metoden kombinerar predikterade relativa risker för finfördelade exponeringsvärden baserat på potentiellt olika dos-responsmodeller. Ett punktvist tillvägagångssätt kan förbättra flexibiliteten i modelleringen av studiespecifika kurvor, och minska påverkan från extrapolering, genom att prediktera studiespecifika relativa risker baserat på observerad exponeringsstorlek.

Artikel 5 presenterar en enstegsmetod för att estimera dos-respons metaanalys för linjära mixade modeller, vilket vanligtvis utförs som en tvåstegsmetod. Fördelarna med en enstegsmetod är många, så som underlättad bedömning av exponeringsheterogenitet och modellskillnader samt förbättrad prediktion av individuella dos-responssamband.

Sammanfattningsvis utökar och förbättrar metoderna i denna avhandling de tillgängliga verktygen för dos-respons metaanalys. Metoderna har illustrerats med hjälp av befintlig data och är implementerade i det lättillgängliga R-paket.

List of publications

I. Alessio Crippa, and Nicola Orsini
   Multivariate dose–response meta-analysis: the dosresmeta R Package
   Journal of Statistical Software 2016; 72(1), 1–15

II. Andrea Discacciati, Alessio Crippa, and Nicola Orsini
   Goodness of fit tools for dose–response meta-analysis of binary outcomes
   Research Synthesis Methods 2017 Jun; 8(2):149.

III. Alessio Crippa, Polyna Khudyakov, Molin Wang, Nicola Orsini, and Donna Spiegelman
   A new measure of between-studies heterogeneity in meta-analysis
   Statistics in Medicine 2016; 35(21), 3661–75

IV. Alessio Crippa, Ilias Thomas, and Nicola Orsini
   A pointwise approach to dose-response meta-analysis of aggregated data
   Manuscript 2018

V. Alessio Crippa, Andrea Discacciati, Matteo Bottai, Donna Spiegelman, and Nicola Orsini
   One-stage dose–response meta-analysis for aggregated data
   Manuscript 2018

The articles will be referred to in the text by their Roman numerals, and are reproduced in full at the end of the thesis.

Related publications

- Ehimen C. Aneni, Alessio Crippa, Chukwuemeka U. Osondu, Javier Valero-Elizondo, Adnan Younus, Khurram Nasir, and Emir Veledar
  Estimates of Mortality Benefit From Ideal Cardiovascular Health Metrics: A Dose Response Meta-Analysis
  Journal of the American Heart Association 2017 Dec 1; 6(12):e006904.

- Alessio Crippa, Susanna C. Larsson, Andrea Discacciati, Alicja Wolk, and Nicola Orsini
  Red and processed meat consumption and risk of bladder cancer: a dose–response meta-analysis of epidemiological studies
  European Journal of Nutrition 2016 Dec; 22:1-3.

- Andrea D. Smith, Alessio Crippa, James Woodcock, and Søren Brage
  Physical activity and incident type 2 diabetes mellitus: a systematic review and dose–response meta-analysis of prospective cohort studies
  Diabetologia 2016;59(12):2527-45.

- Marco Vinceti, Tommaso Filippini, Alessio Crippa, Agnès de Sesmaisons, Lauren A. Wise, and Nicola Orsini
  Meta-Analysis of Potassium Intake and the Risk of Stroke
  Journal of the American Heart Association 2016; 5(10), e004210

- Alessio Crippa, and Nicola Orsini
  Dose–response meta-analysis of differences in means
  BMC Medical Research Methodology 2016 Dec, 16(1):91.

- Emir Veledar, Alessio Crippa, Chukwuemeka U. Osondu, Adnan Younus, and Khurram Nasir
  Letter to Editor: Ideal cardiovascular health metrics and risk of cardiovascular disease or mortality
  International Journal of Cardiology 2016 Nov 1; 222:737

- Alessio Crippa, Andrea Discacciati, Nicola Orsini, and Viktor Oskarsson
  Letter: coffee consumption and gallstone disease—a cautionary note on the assignment of exposure values in dose–response meta-analyses
  Alimentary Pharmacology & Therapeutics 2016; 43(1), 166-167

- Susanna C. Larsson, Alessio Crippa, Nicola Orsini, Alicja Wolk, and Karl Michaëlsson
  Milk consumption and mortality from all causes, cardiovascular disease, and cancer: a systematic review and meta-analysis
  Nutrients 2016; 7(9), 7749-7763

- Daniela Di Giuseppe, Alessio Crippa, Nicola Orsini, and Alicja Wolk
  Fish consumption and risk of rheumatoid arthritis: a dose-response meta-analysis
  Arthritis Research & Therapy 2014; 16(5), 446

- Alessio Crippa, Andrea Discacciati, Susanna C. Larsson, Alicja Wolk, and Nicola Orsini
  Coffee consumption and mortality from all causes, cardiovascular disease, and cancer: a dose–response meta-analysis
  American Journal of Epidemiology 2014; 180(8), 763-775

Contents

List of abbreviations

AIC	Akaike Information Criterion
BLUP	Best Linear Unbiased Prediction
CI	Confidence Intervals
CRAN	Comprehensive R Archive Network
CS	Cubic Splines
df	Degrees of Freedom
GLS	Generalized Least Squares
GRSS	Generalized Residual Sum of Squares
GTSS	Generalized Total Sum of Squares
FP2	Second-degree Fractional Polynomials
ICC	Intraclass Correlation Coefficient
IPD	Individual Participant Data
logRR	log–Relative Risk
LOWESS	Locally Weighted Scatterplot Smoothing
ML	Maximum Likelihood
RCS	Restricted Cubic Splines
REML	Restricted Maximum Likelihood
R^2	Coefficient of Determination
RR	Relative Risk
SEER	Surveillance, Epidemiology, and End Results
VPC	Variance Partition Coefficient
WLS	Weighted Least Squares

Chapter 1

Introduction

A single experiment can hardly provide a definitive answer to a scientific question. Science is oftentimes referred to as a cumulative process where results from many studies, aiming to address a common question of interest, contribute to create and update the scientific evidence. In the cumulative paradigm, meta-analysis is the statistical methodology to combine and compare the current evidence in the field. This process lies at the heart of the concept of evidence-based medicine and plays a major role in policy and decision making.

Epidemiological studies often assess whether the occurrence of a health outcome (e.g. mortality, incidence of a disease) varies according to a quantitative exposure (e.g. amount of physical activity, alcohol intake). The quantitative exposure is frequently divided in intervals and the results are expressed in a tabular format as relative risks for different exposure groups. A high-versus-low meta-analysis contrasts the outcome risk in the highest exposure category relative to the lowest. This common approach, however, discards the results for intermediate categories and thus provides only a partial picture. The rich information of the quantitative exposure is lost and the contrasts may be related to different exposure intervals.

A dose–response meta-analysis, instead, has the potential to be more informative and powerful since it uses the whole available information to estimate the dose–response association. Because the estimates are computed using a common reference group, it might not be appropriate to regress the relative risks on the assigned dose using ordinary least squares. Greenland and Longnecker (1992) described in their seminal paper how to reconstruct the correlation within set of relative risks and incorporate it in the dose–response analysis using generalized least squares regression. Since then, the number of published dose–response meta-analyses has rapidly increased in many fields of application including oncology, public health, environmental sciences, nutrition, endocrinology, and internal medicine. Additional papers refined selected aspects of the proposed methodology, mainly focusing on the implementation of flexible strategies in modeling non-linear associations and incorporating the advancements of multivariate meta-analysis. However, there were still several relevant questions that needed to be addressed including how to assess the goodness-of-fit, how to quantify the impact of heterogeneity, how to deal with differences in the exposure range across studies, and how to estimate complex models without excluding relevant studies.

This thesis aims to address these issues by developing and implementing new strategies and ad-hoc measures. The proposed methodologies are demonstrated reanalyzing published meta-analyses and are implemented in user friendly packages written in the free and open source R language, in order to bridge the gap between theory and application.

Chapter 2

Background

2.1 Meta-analysis

Relevant research questions are typically addressed by independent investigators in multiple studies. Sampling error and possibly differences in the investigations will inevitably produce diverse results, sometimes even conflicting. Evidence-based medicine requires a synthesis of the available evidence to optimize the decision-making process (Haidich, 2010).

Meta-analysis, or more generally quantitative review synthesis, is the statistical methodology for integrating and synthetizing the information arising from multiple studies (Borenstein *et al.*, 2009). Using appropriate statistical models, quantitative reviews contrast and pool results in the hope of identifying similarities and explaining differences across study findings. Meta-analysis represents the state of the art for systematically reviewing the evidence, as indicated by the increasing number of published meta-analyses over the last 40 years (Figure 2.1).

The classical approach for meta-analysis consists of an inverse variance weighted average of the study-specific results or estimates. A fixed-effect model for meta-analysis assumes that all the studies estimate a single common parameter (Rice *et al.*, 2017). The hypothesis of homogeneity of the estimates is rarely applicable in biomedical and social sciences where studies typically differ in terms of design, disease classification, exposure measurement, and implemented statistical analyses (Colditz *et al.*, 1995). In such cases, heterogeneity across estimates is expected and should be considered in the analysis (Higgins, 2008). If the parameters estimated in the studies are not identical but related, a random-effects models can be used to identify those similarities or to explain the observed heterogeneity (Higgins *et al.*, 2009).

2.1.1 Random-effects meta-analysis

In a meta-analysis of I studies indexed by $i = 1, \ldots, I$, we denote $\hat{\beta}_i$ the estimate of an effect of interest (effect size) in the i-th study. A random-effects model for meta-analysis can be written as

$$\hat{\beta}_i = \beta + b_i + \varepsilon_i \tag{2.1}$$

where β is the underlying mean effect, oftentimes the main parameter of interest. The random-
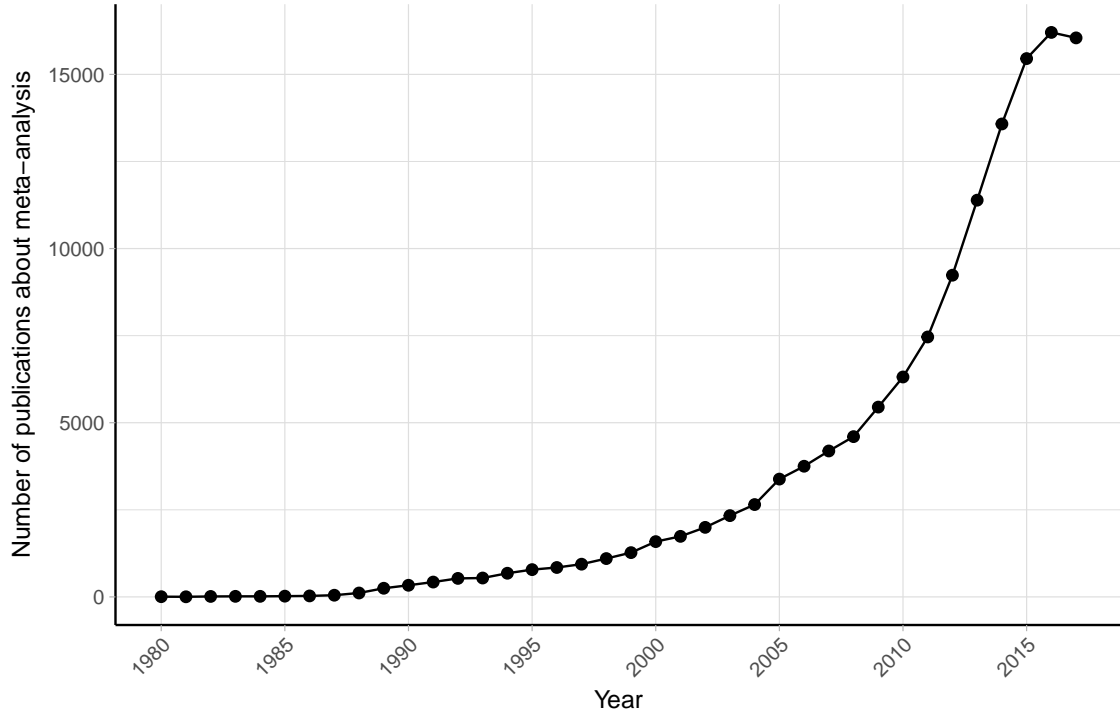
Figure 2.1: Number of publications about meta-analysis (results from Medline search using text "meta-analysis" until January 2018).

effects b_i represent the study-specific deviations from the mean effect β and is assumed to follow a generic distribution f with mean 0 and variance equal to τ^2, the between-studies heterogeneity. The within-study error components ε_i have also mean 0 and variance equal to $\hat{v}_i$, an estimate of the sampling variance of $\hat{\beta}_i$. Because the sample size in the individual investigations is often large, the uncertainty around the estimates of the sampling variance is negligible. Therefore, $\hat{v}_i$ can be considered fixed and denoted as v_i. In addition, for the central limit theorem, $\varepsilon_i \sim \mathcal{N}(0, v_i)$, or alternatively, $\hat{\beta}_i | b_i \sim \mathcal{N}(\beta + b_i, v_i)$.

An inverse variance-weighted approach for meta-analysis estimates the mean effect β as a weighted average of the study-specific effects $\hat{\beta}_i$ (Whitehead and Whitehead, 1991; DerSimonian and Laird, 1986)

$$\hat{\beta} = \frac{\sum_{i=1}^{I} \hat{\beta}_i \hat{w}_i}{\sum_{i=1}^{I} \hat{w}_i} \tag{2.2}$$

$$\widehat{\mathrm{Var}}(\hat{\beta}) = \left(\sum_{i=1}^{I} \hat{w}_i \right)^{-1} \tag{2.3}$$

with weights $\hat{w}_i = \left(v_i + \hat{\tau}^2 \right)^{-1}$ and $\hat{\tau}^2$ being an estimate of the between-study heterogeneity.

2.1.2 Test and estimates of heterogeneity

A second parameter of interest, often overlooked, is the between-study heterogeneity, τ^2. Focusing on the mean effect alone may provide only a limited piece of information, especially in case of heterogeneous effects (Borenstein *et al.*, 2010). Indeed, an evaluation of the extent of heterogeneity is a crucial step in determining the appropriateness of presenting a summary measure of the observed effect sizes.

Presence of heterogeneity is frequently defined as the excess in the variability of $\hat{\beta}_i$ above that expected alone by chance. A summary measure of the observed variability is represented by the Q statistic

$$Q = \sum_{i=1}^{I} \left(\hat{\beta}_i - \hat{\beta}_{\text{fe}}\right)^2 v_i^{-1} \tag{2.4}$$

where $\hat{\beta}_{\text{fe}} = \sum_{i=1}^{I} \hat{\beta}_i v_i^{-1} / \sum_{i=1}^{I} v_i^{-1}$ is the estimate of β in a fixed-effect model. Based on this statistic, Cochran (1954) developed a test for assessing the hypothesis of homogeneity of the study-specific estimates. Under the null hypothesis of no heterogeneity ($H_0 : \tau^2 = 0$) the Q statistic follows a χ^2 distribution with $I - 1$ degrees of freedom (df). A p value less than 0.10 is often used as a cut point for testing presence of between-studies variability. It is known, however, that the test is sensitive to the number of studies, failing to reject the null hypothesis even for high value of τ^2 when I is small and rejecting H_0 for negligible between-studies variation when I is big (Higgins and Thompson, 2002; Takkouche *et al.*, 1999). Therefore, failing to reject the null hypothesis does not provide evidence supporting homogeneity in the effect sizes (Biggerstaff and Tweedie, 1997). In addition, the dichotomization heterogeneous/homogeneous is not very informative, especially because heterogeneity is almost always present (Higgins, 2008).

An estimate of τ^2, instead, directly provides information about the amount of heterogeneity and is thus the most natural measure of between-studies variability. Based on the expectation of Q, DerSimonian and Laird (1986) proposed the following estimator for τ^2 using the method of moments

$$\hat{\tau}^2_{\text{DL}} = \max\left\{0, \frac{Q - (I - 1)}{\sum_{i=1}^{I} v_i^{-1} - \sum_{i=1}^{I} v_i^{-2} / \sum_{i=1}^{I} v_i^{-1}}\right\} \tag{2.5}$$

The moment-based estimator is one of the most popular estimators of τ^2 because it has a simple non-iterative formulation and does not require any distributional assumption for the random-effects. This estimator only assumes a finite first order moment. Other common non-iterative alternatives include estimators based on the variance components (Hedges, 1983) and on methods for estimating the error variance in weighted linear models (Sidik and Jonkman, 2005). Iterative methods based on maximizing the likelihood or restricted likelihood can also be used by specifying a distributional form for the random-effects. The more conventional choice is typically a normal distribution $b_i \sim \mathcal{N}\left(0, \tau^2\right)$, which implies $\beta_i \sim \mathcal{N}\left(\beta, \tau^2\right)$ and $\hat{\beta}_i \sim \mathcal{N}\left(\beta + b_i, \tau^2 + v_i\right)$.

Although τ^2 is the more natural and appropriate measure of between-study variability, the actual value is difficult to interpret because it depends on type of effect size (e.g. log relative

risk, standardized mean difference) and has no upper limit. Therefore, both evaluation of the degree (or levels) and the comparison of heterogeneity in different meta-analyses can hardly be based on the estimate of τ^2 alone.

2.1.3 Measures of heterogeneity

To complement the test-based approach and the information provided by $\hat{\tau}^2$, measures that quantify the impact of heterogeneity have been proposed (Higgins and Thompson, 2002). Higgins and Thompson (2002) presented several possibilities in the simpler case where all the sampling variances v_i are equal to a fixed and known value σ^2.

Two measures aim to estimate the ratio $\sigma^2/(\sigma^2 + \tau^2)$, namely the $H^2 = Q/(I-1)$ that represents the excess in Q statistic relative to its degrees of freedom, and $R^2 = \mathrm{Var}\left(\hat{\beta}\right)/\mathrm{Var}\left(\hat{\beta}_{\mathrm{fe}}\right)$ which describes the inflation in the variability of the mean effect in a random-effects model compared with a fixed-effect analysis. Other measures, instead, relate the between-studies heterogeneity, τ^2, to the marginal or unconditional variability $\tau^2 + v_i$, which is defined by the sum of within- and between-study components. These measures can be more easily interpreted as the percentage of the total variability due to heterogeneity, similar to the Intraclass Correlation Coefficient (ICC) defined for random intercept linear models. These measures directly involve the within-terms v_i that generally varies across the studies. The most popular measures, namely the $\hat{R}_I$ (Takkouche et $al.$, 1999) and the I^2 (Higgins and Thompson, 2002), replaced v_i with a statistic in order to summarize the observed distribution.

Takkouche et $al.$ (1999) chose

$$s_1^2 = \frac{I}{\sum_{i=1}^{I} v_i^{-1}} \tag{2.6}$$

that is the harmonic mean of the inverse of the sampling variances. Higgins and Thompson (2002), instead, described the "typical" within-study variance as

$$s_2^2 = \frac{(I-1)\sum_{i=1}^{I} v_i^{-1}}{\left(\sum_{i=1}^{I} v_i^{-1}\right)^2 - \sum_{i=1}^{I} v_i^{-2}} \tag{2.7}$$

that provided a direct relationship with the Q statistic when τ^2 is estimated using the method of moments: $I^2 = (Q-(I-1))/Q$. Both statistics can be expressed as a percentage where 0% corresponds to no heterogeneity and increasing values imply higher levels of heterogeneity. It is known that these measures depend on the precision of the study-specific estimates and tend to increase to 100% when the v_i are much smaller than the estimated τ^2 (Takkouche et $al.$, 1999; Higgins and Thompson, 2002). A complementary measure is the between-studies coefficient of variation, defined as $\tau^2/|\hat{\beta}|$, that does not directly depend on the within-study variances. However, it increases quickly as $\hat{\beta}$ becomes smaller, and is not defined for $\hat{\beta} = 0$.

2.2 Categorical models for dose–response analysis

Epidemiological studies often assess the strength and direction of the association between exposures and the occurrence of a health outcome. When the exposure of interest is measured on a continuous scale, the additional information on the shape of the relationship is mostly of interest. Investigating how the outcome risk varies throughout the exposure range can provide insights on the causal mechanism (Hill, 1965). Different patterns can be identified such as an increased/decreased outcome risk for increasing values of the exposure or a threshold effect.

A common approach in epidemiology is to include the continuous variable as covariate in an appropriate statistical model. By doing so, the outcome risk is assumed to linearly depend on the covariate. A frequent alternative is to divide the quantitative exposure in categories. Possible advantages of such a categorical approach is that it relaxes the linearity assumption and facilitates the interpretation of the estimated regression coefficients. In addition, the results can be easily presented in a tabular format (Orsini *et al.*, 2011a).

A recent survey among top medical and epidemiological journals estimated that categorization occurred 86% of the times while a linear trend was reported 56% of the times (Turner *et al.*, 2010).

2.2.1 Aggregated dose–response data

In a categorical approach the quantitative exposure is divided in $J + 1$ categories. The corresponding indicator or dummy variables index by $j = 1, \ldots, J$ are included in the model in place of the exposure variable. The results from such a categorical dose–response analysis are expressed as relative measures of association using one category (corresponding to the omitted dummy variable) as referent. Depending on the study-design and on the statistical model, the results consist of estimated odds ratios, rate ratios, or risk ratios (generally referred to as relative risks (RRs)) for the different exposure categories, possibly adjusted for potential confounders. The corresponding 95% confidence intervals (CI) $\widehat{RR}_L, \widehat{RR}_U$ provide information on the uncertainty related to the estimated regression coefficients. Additional information about the assigned dose (mean or median within exposure intervals), the number of cases and the total number of subjects or person-time usually complements the reported results. The general structure and notation for aggregated or summarized dose–response data are presented for a generic i-th study in Table 2.1. The subscript i in J_i highlights that independent studies may categorize the continuous exposure using different number of categories.

The effect sizes considered in a meta-analysis of multiple aggregated dose–response data consist of the estimated $\log \widehat{RR}$s and the corresponding standard errors that can be easily derived from the data available in Table 2.1

$$\widehat{SE}\left(\log \widehat{RR}\right) = \frac{\log\left(\widehat{RR}_U\right) - \log\left(\widehat{RR}_L\right)}{2\, z_{1-\alpha/2}} \tag{2.8}$$

where $z_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of a standard normal distribution, usually approximated

Table 2.1: Aggregated results from a categorical dose–response analysis.

Exposure level	Assigned dose	Cases	n^a	$\widehat{RR}$	95% CI
0	x_{i0}	c_{i0}	n_{i0}	1	—
1	x_{i1}	c_{i1}	n_{i1}	$\widehat{RR}_{i1}$	$\widehat{RR}_{Li1}, \widehat{RR}_{Ui1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
J_i	x_{iJ_i}	c_{iJ_i}	J_{iJ_i}	$\widehat{RR}_{iJ_i}$	$\widehat{RR}_{LiJ_i}, \widehat{RR}_{UiJ_i}$

[a] Depending on the study design, this column reports either total number of subjects or amount of person-time.

to 1.96 for the common $\alpha = 5\%$ level.

A distinctive feature of aggregated dose–response data is the correlation among the (log) $\widehat{RR}$s, which arises from the fact that they are estimated using a common reference group. Each $\widehat{RR}$ has the same baseline risk as denominator that works as comparator. If the observed baseline risk happens to be high or low just by chance, the estimated $\widehat{RR}$s will be also higher or lower than expected (Schmid *et al.*, 1998). This adds complexity in evaluating a trend from a categorical dose–response analysis or in directly comparing results based on different baseline categories.

2.2.2 High vs. low and categorical meta-analysis

A common approach for synthetizing the information from multiple aggregated dose–response data is to limit the analysis to a subset of the available results. In particular, a high- versus-low meta-analysis focuses on the results for the extreme exposure categories (Yu *et al.*, 2013). By selecting only the last row of the aggregated dose–response data, the meta-analytic models discussed in section 2.1.1 are used for combining and contrasting the results, with $\hat{\beta}_i = \log \widehat{RR}_{iJ_i}$.

The major limitation of a high- versus-low approach is that only a subset of the data is analyzed, while the remaining information about intermediate exposure categories is excluded from the analysis. As a consequence, much of the information about the shape of the dose–response is lost and the power of detecting an association may dramatically decrease. For example, in cases where only moderate exposure values have a lower or higher outcome risk, e.g. U-shaped associations, a high- versus-low approach would wrongly conclude that there is no relationship between the exposure and the health outcome.

In addition, in a high- versus-low analysis, the highest and the lowest category may be associated to a different exposure value in the studies included in the meta-analysis. To limit the impact of heterogeneous category definitions, practitioners should carefully plan the analysis by selecting the $\widehat{RR}$s for exposure categories whose definition is more consistent across studies. If the choice of baseline category also substantially differs, the $\widehat{RR}$s can be re-expressed using an alternative reference category implementing dedicated methodologies (Hamling *et al.*, 2008).

An alternative remedy, although less common, is to conduct a categorical meta-analysis, which consists of separate univariate meta-analyses pooling the results from comparable expo-

sure categories. A dose–response association is then deducted from observing the combined $\widehat{RR}$s for increasing dose levels. Apart from evident difficulties in identifying $\widehat{RR}$s for homogeneous exposure intervals in applied works, this approach does not take into account the correlations across set of log $\widehat{RR}$s and suffers from the same problem of guessing a trend from a categorical dose–response analysis.

2.3 Dose–response meta-analysis

The aim of a dose–response meta-analysis is to make inference about the shape of the association from multiple aggregated dose–response data. As compared to the previous strategies, it has the advantages of using all the information available and the potential to be more informative. By describing the variation of the outcome over the entire exposure range, a dose–response meta-analysis allows one to answer the following questions:

- Is there any association between increasing dose levels and the outcome? If so, what is the shape of the relationship?

- Which exposure values are associated with the minimum or maximum response?

- Is there any difference in the study-specific dose–response associations across studies? Which factors can explain the observed heterogeneity?

The statistical problems for modelling sets of correlated relative risks in a dose–response analysis were first presented by Greenland and Longnecker (1992). Their seminal paper is now a standard reference for applied works. The number of published dose–response meta-analyses increased exponentially from 9 in 2000 to 172 in 2016 (Figure 2.2). The most popular research fields of application include oncology, environmental and public health, nutrition epidemiology, and general internal medicine. Dose–response meta-analyses are published in many leading medical and epidemiological journals, including *JAMA, Lancet, Stroke, Gastroenterology, American J of Medicine, American J of Clinical Nutrition, American J Epidemiology, International J Epidemiology, Journal National Cancer Institute, International J of Cancer, Statistics in Medicine* and many others. The method is also routinely used by international organizations such as the World Cancer Research Fund and American Institute for Cancer Research for reviewing the evidence on the relations between life-style factors (e.g. diet and physical activity) and cancer. Guidelines based on these quantitative reviews are central to promote the overall health and prevent many chronic diseases.

The common approach for dose–response meta-analysis consists of a two-stage procedure, where the regression coefficients for the study-specific trends are first estimated separately within each study, and then combined using meta-analysis. In the next sections I cover the main methodological aspects related to each stage of the analysis.
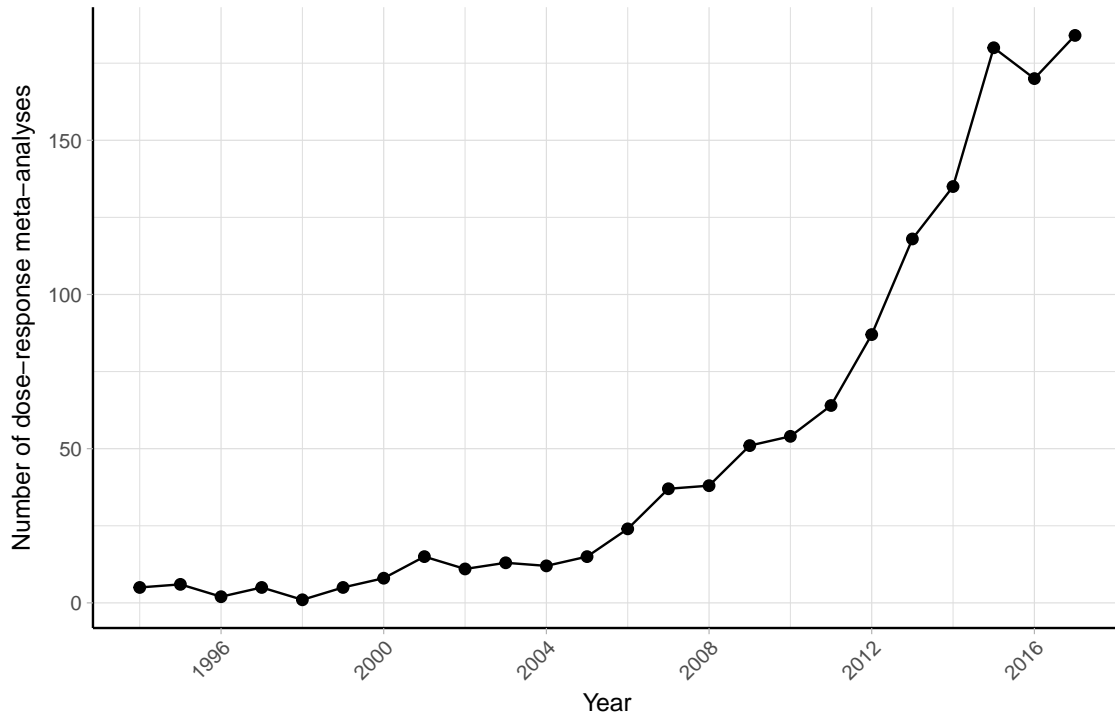
Figure 2.2: Number of citations of the paper by Greenland and Longnecker (1992) obtained from Google Scholar 1992-2017 (until January 2018).

2.3.1 First stage: study-specific trends

If we had access to the individual participant data (IPD), the dose–response model for a simple linear trend could be written as

$$\log(\lambda(x, \mathbf{c})) = \beta_0 + \beta_1 x + \boldsymbol{\gamma}^\top \mathbf{c} \tag{2.9}$$

where x is the quantitative exposure and $\mathbf{c}$ the set of possible confounders. The outcome variable is the log transformation of the mean outcome (e.g. odds, risk, or rate). Transformations of the exposure variable can be included to relax the linearity assumption, such as a quadratic term

$$\log(\lambda(x, \mathbf{c})) = \beta_0 + \beta_1 x + \beta_2 x^2 + \boldsymbol{\gamma}^\top \mathbf{c} \tag{2.10}$$

This thesis focuses on methods for estimating a dose–response relationship from a summary of the initial individual participant data. In particular, aggregated data from a categorical analysis can be often retrieved from published articles. The aim of the first stage of a dose–response meta-analysis is to estimate the β coefficients in Equation 2.9 and 2.10 using aggregated dose–response data. We consider the notation presented in Table 2.1 with $i = 1, \ldots, I$ indexing the studies and $j = 1, \ldots, J_i$ the non-referent dose levels of a generic i-th study. The corresponding two models can be written as

$$\log\left(\widehat{\mathrm{RR}}_{ij}\right) = \log\left(\hat{\lambda}\left(x = x_{ij}\right)\right) - \log\left(\hat{\lambda}\left(x = x_{i0}\right)\right) = \beta_1\left(x_{ij} - x_{i0}\right) \tag{2.11}$$

$$\log\left(\widehat{RR}_{ij}\right) = \log\left(\hat{\lambda}\left(x = x_{ij}\right)\right) - \log\left(\hat{\lambda}\left(x = x_{i0}\right)\right) = \beta_1\left(x_{ij} - x_{i0}\right) + \beta_2\left(x_{ij}^2 - x_{i0}^2\right) \quad (2.12)$$

More generally, the i-th dose–response model is defined as

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i \qquad (2.13)$$

The outcome $\mathbf{y}_i$ is the J_i length vector of the non-referent log $\widehat{RR}$s while $\mathbf{X}_i$ the $J_i \times p$ design matrix containing the p transformations of the assigned dose used to model the dose–response association

$$\mathbf{X}_i = \begin{bmatrix} g_1(x_{i1}) - g_1(x_{i0}) & \cdots & g_p(x_{ip}) - g_p(x_{i0}) \\ \vdots & & \vdots \\ g_1(x_{iJ_i}) - g_1(x_{i0}) & \cdots & g_p(x_{iJ_i}) - g_p(x_{i0}) \end{bmatrix} \qquad (2.14)$$

In the linear trend analysis (model 2.11), $\mathbf{X}_i$ includes only the dose levels, $p = 1$, $g_1(x) = x$ (identity function)

$$\mathbf{X}_i = \begin{bmatrix} x_{i1} - x_{i0} \\ \vdots \\ x_{iJ_i} - x_{i0} \end{bmatrix}$$

while $p = 2$ columns are needed in the quadratic model 2.12: $g_1(x) = x$ and $g_2(x) = x^2$

$$\mathbf{X}_i = \begin{bmatrix} x_{i1} - x_{i0} & x_{i1}^2 - x_{i0}^2 \\ \vdots & \vdots \\ x_{iJ_i} - x_{i0} & x_{iJ_i}^2 - x_{i0}^2 \end{bmatrix}$$

A distinctive feature of the models 2.13 is the absence of the intercept term. The reference row in Table 2.1 is not actually used for the estimation of the regression coefficients but introduces the constraint on the predicted log $\widehat{RR}$, which needs to be 0 ($\widehat{RR} = 1$) for the reference dose value x_{i0}, as is explicit in models 2.11 and 2.12.

Approximation of the covariance between log relative risks

A particular characteristic of summarized dose–response data is that the log $\widehat{RR}$s are reported with different precision and are constructed using the same baseline group. Thus, the error terms $\boldsymbol{\varepsilon}_i$ in Equation 2.13 are heterogeneous and correlated, with a covariance matrix structured as

$$\text{Cov}(\boldsymbol{\epsilon}_i) = \mathbf{S}_i = \begin{bmatrix} \sigma_{i11} & & & & \\ \vdots & \ddots & & & \\ \sigma_{i1j} & & \sigma_{ijj} & & \\ \vdots & & & \ddots & \\ \sigma_{i1J_i} & \cdots & \sigma_{iJ_ij} & \cdots & \sigma_{iJ_iJ_i} \end{bmatrix} \qquad (2.15)$$

with the variance of the log $\widehat{RR}$s on the diagonal (σ_{ijj}) and the pairwise covariances as non-diagonal elements ($\sigma_{ijj'}$).

Two methods have been proposed to approximate the covariances $\sigma_{ijj'}$ (Greenland and Long-necker, 1992; Hamling *et al.*, 2008). Greenland and Longnecker described an algorithm to construct a table of pseudo or effective counts (number of cases and participants or person-time) that would produce the adjusted log $\widehat{RR}$s as those published. A unique solution for the algorithm is ensured by keeping the margins of the pseudo-counts equal to the observed ones. Alternatively, Hamling et al. modified the previous algorithm in such a way that the pseudo-counts would also match the standard errors of the log $\widehat{RR}$s.

Estimation

The dose–response coefficients $\boldsymbol{\beta}_i$ can be efficiently estimated using generalized least squares estimator (GLS), which minimizes the quadratic loss function $\left(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_i \right)^\top \mathbf{S}_i^{-1} \left(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_i \right)$ with respect to $\boldsymbol{\beta}_i$ assuming the covariance matrix $\mathbf{S}_i$ known

$$\hat{\boldsymbol{\beta}}_i = (\mathbf{X}_i^\top \mathbf{S}_i^{-1} \mathbf{X}_i)^{-1} \mathbf{X}_i^\top \mathbf{S}_i^{-1} \mathbf{y}_i$$
$$\widehat{\mathrm{Var}}\left(\hat{\boldsymbol{\beta}}_i\right) = (\mathbf{X}_i^\top \mathbf{S}_i^{-1} \mathbf{X}_i)^{-1} \tag{2.16}$$

The GLS estimates in Equation 2.16 do not require any distributional assumption for the error terms. However, for the central limit theory, the error terms follow approximately a normal distribution $\epsilon_i \sim \mathcal{N}\left(\mathbf{0}, \mathbf{S}_i\right)$. Using this additional assumption, the log-likelihood of model 2.13 is

$$\ell\left(\boldsymbol{\beta}_i\right) = -\frac{J_i}{2} \log(2\pi) - \frac{1}{2} \log|\mathbf{S}_i| - \frac{1}{2}\left[\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}_i\right)^\top \mathbf{S}_i^{-1}\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}_i\right)\right] \tag{2.17}$$

Interestingly, the maximum likelihood (ML) estimates that maximize the log-likelihood 2.17 coincides with the GLS estimates in 2.16. Introducing the normality distribution for the random errors facilitates the inference, i.e. test of hypothesis and confidence intervals, on the $\boldsymbol{\beta}_i$ coefficients. The estimates in 2.16 are a linear combination of normal distributions, $\mathbf{y}_i \sim \mathcal{N}\left(\mathbf{X}_i\boldsymbol{\beta}_i, \mathbf{S}_i\right)$, and therefore are also normally distributed $\hat{\boldsymbol{\beta}}_i \sim \mathcal{N}\left(\boldsymbol{\beta}_i, \mathrm{Var}\left(\hat{\boldsymbol{\beta}}_i\right)\right)$.

The ML and GLS estimators give unbiased estimates of $\boldsymbol{\beta}_i$ regardless of the specification of $\mathbf{S}_i$ (Orsini *et al.*, 2006). As a consequence, a weighted least squares estimator (WLS) that assumes independence of the log $\widehat{RR}$s will also produce unbiased estimates. However, taking into account the correlation will improve the efficiency of the estimator. I investigated the differences between the GLS and WLS estimators using a simulation study of 5000 aggregated dose–response data where the true trends were linear ($\beta_{\mathrm{TRUE}} = -0.014$). As expected, both the estimators were unbiased but the empirical distribution of the GLS estimator was more concentrated around the true β value 2.3. The empirical distributions of the estimated standard errors were shifted, with the mean standard error for the WLS estimates being 10% lower than the corresponding GLS value. This had a direct effect on the inference for the estimated linear trend. For instance, it may be interesting to fit a quadratic curve as in 2.12 and test the hypothesis $H_0 : \beta_2 = 0$, i.e. departure from a linear trend. Using inference based on

WLS estimators the null hypothesis were wrongly rejected 3.96% of the time, lower than the nominal level $\alpha = 5\%$. The corresponding number for the GLS estimator was instead closer to the specified rejection rate (4.8%).

I also implemented simulations assuming a quadratic curve with the true coefficients $\boldsymbol{\beta}_{\text{TRUE}} = (-0.092, 0.003)$. The empirical bivariate distributions of $\hat{\boldsymbol{\beta}}_i$ were centered around the true parameter, with levels curve more concentrated for the GLS estimates (Figure 2.4). Similarly to the linear trend simulations, the distributions of the estimates of the standard errors for the two estimators where shifted, with the mean of $\widehat{\text{SE}}\big(\hat{\beta}_1\big)$ and $\widehat{\text{SE}}\big(\hat{\beta}_2\big)$ being 7 % lower and 6% higher, respectively, when comparing WLS to GLS estimates.
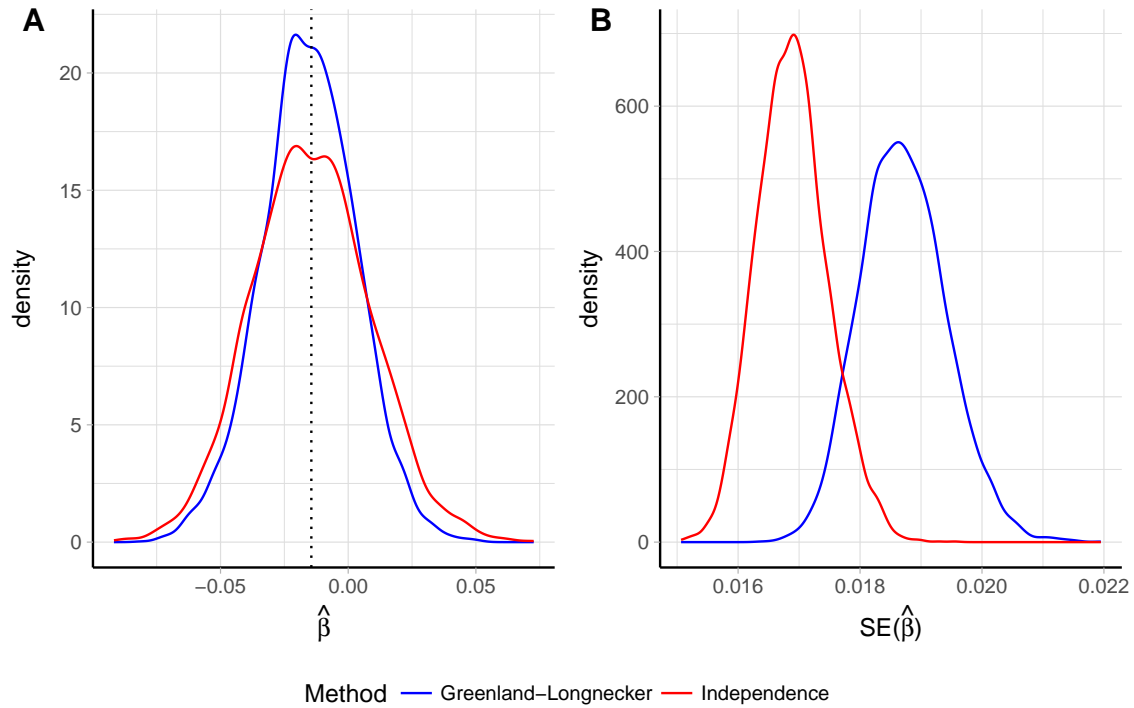
Figure 2.3: Empirical distribution of the $\hat{\beta}$ (panel A) and $\widehat{\text{SE}}\big(\hat{\beta}_i\big)$ (panel B) for a linear trend assuming independence of the log $\widehat{\text{RR}}$ and reconstructing the covariances using the Greenland and Longnecker's method. Results are based on simulations with 5000 replications and a true linear trend $\beta = -0.014$.

2.3.2 Second stage: multivariate meta-analysis

The study-specific dose–response curves are defined by the p transformations, $g_1(x), \ldots, g_p(x)$, and the estimated regression coefficients $\hat{\boldsymbol{\beta}}_i$. A pooled dose–response can be obtained by combining the $\hat{\boldsymbol{\beta}}_i$ coefficients. For that purpose, the same functional relationship needs to be defined across the studies. Therefore, the transformations of the exposure were not subscripted by the study index i.

The p length vector of the $\hat{\boldsymbol{\beta}}_i$ parameters and the accompanying $p \times p$ covariances matrices $\widehat{\text{Var}}\big(\hat{\boldsymbol{\beta}}_i\big)$ serve as outcome in the meta-analytic model. We consider the setting with $p \geq 2$ and relate the univariate case as a simpler instance of the more general multivariate case. Since the
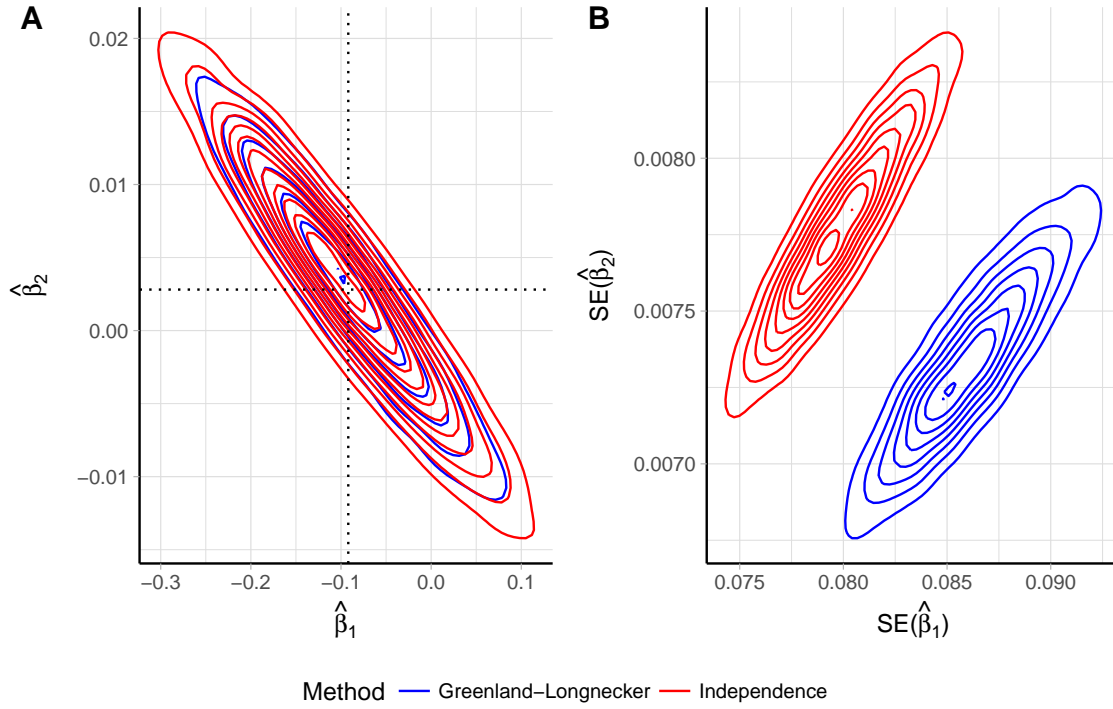
Figure 2.4: Empirical bivariate distribution of the beta coefficients (panel A) and their standard errors (panel B) for a quadratic trend assuming independence of the log $\widehat{\text{RR}}$ and reconstructing the covariances using the Greenland and Longnecker's method. Results are based on simulations with 5000 replications and a true quadratic trend $\beta_1 = -0.092, \beta_2 = 0.003$.

dimension of the outcome is no longer univariate, extensions of models 2.1 to the multivariate settings can be implemented for accommodating the synthesis of correlated estimates (Berkey *et al.*, 1998; Gasparrini *et al.*, 2012; Ritz *et al.*, 2008).

Model definition

A multivariate random-effects model has a similar formulation as in the univariate case

$$\hat{\boldsymbol{\beta}}_i = \boldsymbol{\beta} + \mathbf{b}_i + \boldsymbol{\varepsilon}_i \tag{2.18}$$

The unobserved random effects $\mathbf{b}_i$ are now of dimension p, still representing study-specific deviation from the mean $\boldsymbol{\beta}$ parameter. As before, $\text{E}[\mathbf{b}_i] = \mathbf{0}$ and $\text{Var}[\mathbf{b}_i] = \boldsymbol{\Psi}$, the $p \times p$ between-study variance matrix. Specification of a parametric distribution for the random-effects may facilitate the inference (especially confidence intervals) and improve the prediction of marginal and conditional dose–response associations. Typically a multivariate normal distribution is assumed $\mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi})$. Hence, we can write the marginal model of 2.18 as

$$\hat{\boldsymbol{\beta}}_i \sim \mathcal{N}(\boldsymbol{\beta}, \boldsymbol{\Sigma}_i) \tag{2.19}$$

where the marginal variance $\boldsymbol{\Sigma}_i = \widehat{\text{Var}}(\hat{\boldsymbol{\beta}}_i) + \boldsymbol{\Psi}$ is defined by the sum of the within-study and between-studies variance components. The model 2.19 implies a two-stage sampling procedure

where the study-specific $\boldsymbol{\beta}_i$ parameters are assumed to be sampled from a multivariate normal distribution centered around the population average parameter $\boldsymbol{\beta}$. The study-specific estimates $\hat{\boldsymbol{\beta}}_i$ are themselves sampled from a multivariate distribution with zero mean and error variance. The multivariate random-effects model 2.19 can be extended to meta-regression models by including study-levels covariates that might change the shape of the dose–response relationship. The dose–response coefficients are then modeled as a linear combination of the m study-level covariates $\mathbf{u}_i = (u_{i1}, \ldots, u_{im})$, with $u_{i1} = 1$ representing the intercept term

$$\hat{\boldsymbol{\beta}}_i \sim \mathcal{N}\left(\widetilde{\mathbf{X}}_i \boldsymbol{\beta}, \boldsymbol{\Sigma}_i\right) \tag{2.20}$$

The $p \times pm$ design matrix $\widetilde{\mathbf{X}}_i$ is constructed taking the Kronecker product between the $\mathbf{u}_i$ and the identity matrix of dimension p, $\mathbf{I}_{(p)}$

$$\widetilde{\mathbf{X}}_i = \mathbf{I}_{(p)} \otimes \mathbf{u}_i^\top = \begin{bmatrix} 1 & u_{i2} & \cdots & u_{im} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & & & \ddots & & & \\ 0 & 0 & \cdots & 0 & \cdots & 1 & u_{i2} & \cdots & u_{im} \end{bmatrix} \tag{2.21}$$

For example, the $\widetilde{\mathbf{X}}_i$ matrix relating the effect of a binary variable u_i to the dose–response coefficients for a quadratic trend is

$$\widetilde{\mathbf{X}}_i = \mathbf{I}_{(2)} \otimes \mathbf{u}_i^\top = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes (1, u_i) = \begin{bmatrix} 1 & u_i & 0 & 0 \\ 0 & 0 & 1 & u_i \end{bmatrix}$$

The dimension of $\hat{\boldsymbol{\beta}}$ is now $p \times m$. The coefficients related to the intercept terms are interpreted as the mean dose–response coefficients when all the study-level covariates $\mathbf{u}$ are equal to zero. The remaining coefficients indicate how the mean dose–response association varies with respect to the corresponding study-level covariate.

Estimation

Several methods are available for estimating the parameters of interest, namely the $p \times m$ dose–response coefficients in $\boldsymbol{\beta}$ and the $p(p+1)/2$ length vector $\boldsymbol{\xi}$ containing the elements on or above the diagonal of the between-studies covariance $\boldsymbol{\Psi}$. There is generally no reason to assume a specific covariance structure (White *et al.*, 2011). We consider here likelihood-based estimators (Verbeke, 1997; Pinheiro and Bates, 2010). In particular, ML estimators estimate simultaneously $\boldsymbol{\beta}$ and $\boldsymbol{\xi}$ by maximizing the log-likelihood of the marginal model 2.20

$$\ell\left(\boldsymbol{\beta}, \xi\right) = -\frac{1}{2} I p \log(\pi) - \frac{1}{2} \sum_{i=1}^{I} \log |\boldsymbol{\Sigma}_i| - \frac{1}{2} \sum_{i=1}^{I} \left[\left(\hat{\boldsymbol{\beta}}_i - \widetilde{\mathbf{X}}_i \boldsymbol{\beta}\right)^\top \boldsymbol{\Sigma}_i^{-1} \left(\hat{\boldsymbol{\beta}}_i - \widetilde{\mathbf{X}}_i \boldsymbol{\beta}\right)\right] \tag{2.22}$$

ML estimators, however, don't take into account the loss of degrees of freedom due to the $\boldsymbol{\beta}$ estimation (Harville, 1977). Alternatively, restricted maximum likelihood methods (REML)

maximizes a set of contrasts defined as a function of the only covariance parameters

$$
\begin{aligned}
\ell_R(\xi) = & -\frac{1}{2}(Ip - pm) - \frac{1}{2}\sum_{i=1}^{I}\log|\Sigma_i| - \frac{1}{2}\sum_{i=1}^{I}\log\left|\widetilde{\mathbf{X}}_i^\top \Sigma_i \widetilde{\mathbf{X}}_i\right| + \\
& -\frac{1}{2}\sum_{i=1}^{I}\left[(\hat{\boldsymbol{\beta}}_i - \widetilde{\mathbf{X}}_i\boldsymbol{\beta})^\top \Sigma_i^{-1}(\hat{\boldsymbol{\beta}}_i - \widetilde{\mathbf{X}}_i\boldsymbol{\beta})\right]
\end{aligned}
\tag{2.23}
$$

Both estimation methods require iterative algorithms, where conditional estimates of $\hat{\boldsymbol{\beta}}$ are plugged into either 2.22 or 2.23, regarded as function of ξ only, until convergence. More details on the implementation of iterative methods for maximizing Equations 2.22 and 2.23 are described by Gasparrini *et al.* (2012).

Hypothesis testing, heterogeneity, and model comparison

There are two main domains of interest for making inference that relate either to the fixed-effects $\boldsymbol{\beta}$ or the variance components in $\boldsymbol{\Psi}$. Using the normality assumption for the random-effects, inference is based on the approximated normal distribution for $\hat{\boldsymbol{\beta}}$, with mean and covariance matrix defined similarly as in Equation 2.16.

Since the mean dose–response association is defined by the $\boldsymbol{\beta}$, the hypothesis of no association can be evaluated by testing $H_0 : \boldsymbol{\beta} = \mathbf{0}$. Alternatively, a subset or linear combinations of the elements in $\boldsymbol{\beta}$ may be of interest. For example, in a quadratic trend the non-linearity is introduced by the quadratic term x^2. Thus, testing $H_0 : \beta_2 = 0$ is a possible way for evaluating departure from a linear dose–response relationship.

As previously presented in section 2.1.2, the coefficients defining $\boldsymbol{\Psi}$ are not nuisance parameters rather they are useful for quantifying the variation of the study-specific associations $\boldsymbol{\beta}_i$. Similar measures for testing and quantifying the impact of heterogeneity have been extended to the multivariate setting (Berkey *et al.*, 1996). In particular, the Q statistic

$$
Q = \sum_{i=1}^{I}(\hat{\boldsymbol{\beta}}_i - \widetilde{\mathbf{X}}_i\hat{\boldsymbol{\beta}}_{\text{fe}})^\top \widehat{\text{Var}}(\hat{\boldsymbol{\beta}}_i)^{-1}(\hat{\boldsymbol{\beta}}_i - \widetilde{\mathbf{X}}_i\hat{\boldsymbol{\beta}}_{\text{fe}})
\tag{2.24}
$$

with $\hat{\boldsymbol{\beta}}_{\text{fe}}$ estimated under a fixed-effect model is used to test $H_0 : \boldsymbol{\Psi} = \mathbf{0}$. Under the null hypothesis, the Q statistic follow a χ^2 distribution with $Ip - pm$ degrees of freedom. When $p = 1$ the formulations 2.4 and 2.24 coincide. The multivariate extension of the I^2 was derived relating the Q statistics to its degrees of freedom $I^2 = \max\left\{0, \frac{Q-(Ip-pm)}{Ip-pm}\right\}$ (Jackson *et al.*, 2012).

The fit of alternative non-nested meta-analytical models can be compared using information criteria indices such the Akaike information criterion (AIC), which is defined as AIC = $-2\ell(\boldsymbol{\beta}, \xi) + 2k$, a descriptive measure depending on the maximum log-likelihood and k, the number of estimated parameters. It is worth to note that the AIC can be used for comparing the fit of different analyses such as alternative meta-regression models. However, it is not clear if these indices can be used for comparing different dose–response models, such as linear vs non-linear.

Prediction

Interpretation of a single regression coefficient of $\hat{\boldsymbol{\beta}}$ may be difficult. The dose–response findings are best communicated in a graphical form as predicted (log) relative risk for selected exposure levels using one value as referent. Obtaining predictions and presenting them either in a graphical or tabular form is an important step following the estimation of the model. Based on the model 2.18, the predicted $\log RR$ for a dose level x_v relative to the level x_{ref} can be calculated as

$$\log\widehat{RR}(x_v \text{ vs } x_{\text{ref}}) = \mathbf{X}_v\hat{\boldsymbol{\beta}} \tag{2.25}$$

$$\text{Var}\left(\log\widehat{RR}(x_v \text{ vs } x_{\text{ref}})\right) = \mathbf{X}_v\widehat{\text{Var}}\left(\hat{\boldsymbol{\beta}}\right)\mathbf{X}_v^\top \tag{2.26}$$

where $\mathbf{X}_v$ is the design matrix evaluated in x_v, as defined in the first-stage analysis (Equation 2.14). For example, the predicted $\log RR$ for the quadratic model 2.12 comparing x_v versus x_{ref} is

$$\log\widehat{RR}(x_v \text{ vs } x_{\text{ref}}) = \hat{\beta}_1\left(x_v - x_{\text{ref}}\right) + \hat{\beta}_2\left(x_v^2 - x_{\text{ref}}^2\right)$$

Of note, the referent dose x_{ref} is an arbitrary value and thus does not need to correspond to any of the study-specific reference values x_{i0}.

A confidence interval for the predicted $\log\widehat{RR}(x = x_v)$ is based on the normal distribution of $\hat{\boldsymbol{\beta}}$

$$\log\widehat{RR}(x_v \text{ vs } x_{\text{ref}}) \mp z_{1-\alpha/2}\,\text{Var}\left(\log\widehat{RR}(x_v \text{ vs } x_{\text{ref}})\right)^{\frac{1}{2}}$$

Formulas 2.25 and 2.26 can be extended to meta-regression models. The predicted $\log RR$ conditional on a specific study-level covariate pattern $\mathbf{u} = \mathbf{u}_v$ is

$$\log\widehat{RR}\left(x_v \text{ vs } x_{\text{ref}}, \mathbf{u} = \mathbf{u}_v\right) = \mathbf{X}_v\left(\mathbf{I}_{(p)} \otimes \mathbf{U}_v^\top\right)\hat{\boldsymbol{\beta}} \tag{2.27}$$

$$\text{Var}\left(\log\widehat{RR}\left(x_v \text{ vs } x_{\text{ref}}, \mathbf{u} = \mathbf{u}_v\right)\right) = (\mathbf{X}_v\mathbf{U}_v)\widehat{\text{Var}}\left(\hat{\boldsymbol{\beta}}\right)(\mathbf{X}_v\mathbf{U}_v)^\top \tag{2.28}$$

Inference on study-specific dose–response associations can be enhanced by exploiting the information from the multivariate distribution for the random-effects. The best linear unbiased prediction (BLUP) for the study-specific regression coefficients $\mathbf{b}_i$ is defined as

$$\hat{\boldsymbol{b}}_i = \boldsymbol{\Psi}\boldsymbol{\Sigma}_i^{-1}\left(\hat{\boldsymbol{\beta}}_i - \widetilde{\mathbf{X}}_i\hat{\boldsymbol{\beta}}\right) \tag{2.29}$$

Study-specific predicted log relative risks can be obtained as in Equations 2.25 and 2.27 by replacing the mean parameter $\hat{\boldsymbol{\beta}}$ with the individual dose–response coefficients $\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{b}}_i$

2.3.3 History of previous methodological research

Dose–response meta-analysis has received attention not only in applied works but also in theoretical articles that covered different aspects of the methodology. Greenland and Longnecker (1992) originally presented the two-stage approach for efficiently estimating a linear trend

in a fixed-effect analysis. An alternative model for estimating curvilinear model, referred to as "pool first", was also presented. The technique consists of a one-stage approach where the aggregated data are considered altogether and a single model as in Equation 2.13 is fitted. By first combining the data, more flexible curve parametrized by multiple parameters ($p \geq 2$) can be estimated. The study-specific dose–response analyses are limited by the minimum number of non-referent log RRs across the studies. For example, if the aggregated data for a study consists of only one non referent log RR, only a univariate model with $p = 1$ can be estimated. The authors refined the methodology by extending the two-stage approach to allow for heterogeneity limited to a linear trend analysis (Berlin *et al.*, 1993).

Flexible dose–response models

The primary interest of the methodological research was in presenting alternative strategies for estimating non-linear curves. Bagnardi *et al.* (2004) described the use of fractional polynomials and restricted cubic splines using aggregated dose–response data. Based on a practical example on the association between alcohol consumption and all-cause mortality, the authors showed how implementation of these flexible techniques may prevent misleading results from conventional polynomials (e.g. quadratic) curves. Second-degree fractional polynomials (FP2) consist of a large family of curves defined in the general form of

$$\text{FP2}(x) = \beta_1 x^{p_1} + \beta_2 x^{p_2} \tag{2.30}$$

where p_1 and p_2 are chosen in the set of power coefficients $\{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$ (Royston and Altman, 1994; Royston, 2000). When $p = 0$, x^p becomes $\log(x)$, while if $p_1 = p_2$ the second transformation of x becomes $x^{p_2} \log(x)$. The advantage of FP2 models is that different shapes, including U- and J-shapes, can be estimated by only two coefficients chosen using different combinations for the power terms (p_1, p_2). Typically, the best fitting fractional polynomial is chosen in such a way that the (p_1, p_2) corresponds to the model with the highest likelihood, or equivalently, lowest deviance (Royston, 2001).

A popular alternative to flexibly model the dose–response association is represented by the use of splines (De Boor *et al.*, 1978), largely presented by Orsini *et al.* (2011b) using data from the Pooling Project of Prospective Studies of Diet and Cancer (`http://www.hsph.harvard.edu/poolingproject`). Spline functions consist of consecutive polynomials connected at specific points of the exposure range called knots. Choosing $\mathbf{k} = (k_1, \ldots, k_K)$ knots and third order polynomials, the model, also known as cubic splines (CS), is defined as

$$\text{CS}(x) = \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \sum_{l=1}^{K-1} \beta_{l+3}(x - k_l)_+^3 \tag{2.31}$$

where the '+' notation has been used ($u_+ = u$ if $u \geq 0$ and $u_+ = 0$ otherwise). To avoid strange behaviors at the extremes of the exposure range, the model 2.31 is constrained to be linear before and after the first and last knots, respectively. For example, using the minimum number

of three knots a restricted cubic spline model (RCS) can be specified in terms of two coefficients

$$\text{RCS}(x) = \beta_1 x + \beta_2 \left[(x - k_1)_+^3 - \frac{k_3 - k_1}{k_3 - k_2} (x - k_2)_+^3 + \frac{k_2 - k_1}{k_3 - k_2} (x - k_3)_+^3 \right] \qquad (2.32)$$

The second transformation is generally divided by $(k_3 - k_1)^2$ to improve the numerical behavior and to put the spline transformations on the same scale (Harrell Jr, 2015) (see Appendix A for the derivation of a RCS model).

The previous strategies have been often presented for exposures where 0 was the natural reference categories (e.g. alcohol consumption). Liu *et al.* (2009) extended the methodology for handling non-zero reference categories, as in the case of Body Mass Index. In particular, they first described how to construct the design matrix in terms of contrasts, as clarified in Equation 2.14. Misspecification of the design matrix (e.g. $\log \text{RR} = \beta_1(x_1 - x_0) + \beta_2(x_1 - x_0)^2$ instead of $\log \text{RR} = \beta_1(x_1 - x_0) + \beta_2(x_1^2 - x_0^2)$) may increase the risk of generating artifacts and misleading conclusions. Note that for zero exposure categories the problem is generally not relevant since many functions return zero for $x = 0$ ($g(0) = 0$).

Multivariate meta-analysis

The major contribution for estimating non-linear curves in a random-effects analysis came with the extension of univariate meta-analytic models to the multivariate case. The formalization and implementation of multivariate meta-analysis enabled the extension of a two-stage dose–response meta-analysis to the more complex case of multiple parameters association (Gasparrini *et al.*, 2012). The multivariate framework can accommodate the synthesis of correlated outcomes or estimates, as those derived in the first stage of a dose–response meta-analyses. The application of the strategies presented in 2.30 and 2.32 in a random-effects setting has been easily facilitated by the implementation of dedicated packages for multivariate meta-analysis (White *et al.*, 2011; Jackson *et al.*, 2011). This is probably the reason why a one-stage approach for dose–response meta-analysis was no further extended.

The methodological advancements for meta-analysis were diverse and numerous (see Sutton and Higgins (2008) for an overview). Important improvements that directly affected how results are presented related mainly to the quantification and assessment of heterogeneity, with the definition of the measures presented in section 2.1.2. In addition, many other articles enriched the set of tools for pooling study-specific effects, with a direct application to the second stage of a dose–response meta-analysis. Among the many, it is worth to mention the implementation of several estimation methods for the between-study variability (see Langan *et al.* (2017) for a comparison based on simulation studies); advancement in performing meta-regression (Van Houwelingen *et al.*, 2002); proposal of sequential approaches (Pogue and Yusuf, 1997) and statistical power (Sutton *et al.*, 2007); and introduction of Bayesian methods (Sutton and Abrams, 2001).

Covariance and sensitivity analysis

Orsini *et al.* (2006) refined the initial formulas presented by Greenland and Longnecker for approximating the study-specific covariance matrices depending on the study-design. Berrington and Cox (2003) described an alternative method that avoids the reconstruction of the covariance matrices. Instead, upper and lower bounds for the covariance matrix are used in a sensitivity analysis of the dose–response coefficients, adopting a range of plausible values for the covariances. While the alternative algorithm proposed by Hamling *et al.* (2008) was presented in section 2.3.1, Easton *et al.* (1991) proposed the implementation of the floating absolute risks where the parameters and their standard errors can be estimated without specifying a baseline group and thus can be regarded as independent. Using individual participant data from the Pooling Project of Prospective Studies of Diet and Cancer, negligible differences in the reconstructed covariance matrices were found comparing the three approaches (Orsini *et al.*, 2011b). Of note, none of the methods would be needed if the authors of the original articles provided the covariance matrix along with the estimated coefficients, as it is usually done in consortia projects.

Berlin *et al.* (1993) presented alternatives for assigning the dose levels within exposure categories and illustrated the use of meta-regression models for investigating the possible effect of study-level characteristics on the estimated linear trends. Shi and Copas (2004) further discussed the issue of dose assignment in grouped measures allowing for arbitrary dose levels. In addition, they investigated the effect of heterogeneity and publication bias by means of sensitivity analyses. A similar problem of dose assignment was addressed by using a likelihood approach limited to a linear trend analysis (Takahashi and Tango, 2010). This idea has been further extended to the case of restricted cubic splines (Takahashi *et al.*, 2013).

2.3.4 Description of current practice

There were still many open research questions that need to be addressed to improve the synthesis of aggregated dose–response data. In order to identify the most relevant questions, I began by observing current practice in applied works. I searched the PubMed database for articles published between January 1, 2013 and April 1, 2013 using the research query ("meta-analysis" [Title] and "dose–response" [Title]) and, after excluding irrelevant articles, I found a total of 42 applied dose–response meta-analyses. The authors of the select articles conducted a linear trend analysis most of the times (25 times, 60%) while only 17 (40%) articles considered non-linear associations by means of restricted cubic splines (15) and fractional polynomials (2). All the papers modelling non-linear curves reported a graphical presentation of the pooled dose–response association.

Interestingly, none of the retrieved articles evaluated the goodness-of-fit of the selected dose–response model. The assessment of how the estimated curve fits the aggregated data should be a natural and important step in a dose–response analysis. We address this important issue by presenting relevant measures and graphical tools to help the assessment of goodness-of-fit.

The majority of the screened papers (39, 93%) quantify the impact of heterogeneity by reporting results for the Q test and the value of I^2. While the limitations of the Q test approach are widely known, little emphasis is placed on the assumptions underneath the definition of the established measures of heterogeneity, i.e. all the estimates being reported with the same precision, which is unlikely to be met in almost all the applications. A measure of the impact of heterogeneity that does not require such an assumption would be desirable. We overcome this limitation by proposing an alternative measure of heterogeneity and comparing the performance of the new and available measures.

None of the surveyed meta-analyses discussed the sensitivity of the overall dose–response relationship to differences in study-specific exposure distributions. This analysis can be relevant in case of studies reporting results for heterogeneous exposure range. Application of a point-wise average approach originally presented by Sauerbrei and Royston (2011) in the context of individual participant data may represent an interesting alternative to the averaging of regression coefficients.

In all the meta-analyses assessing departure from linearity, the authors excluded those studies reporting less than two non-referent RRs. Indeed, a two-stage dose–response meta-analysis requires that all the models in the dose–response analysis are identifiable ($p \leq \min(J_i)$). A one-stage approach would avoid that requirement. Such an approach is conceptually easier to understand, and more elegant from a statistical point of view. In addition, it allows investigation of much more flexible dose–response curves, that are not possible within the context of a traditional two-stage analysis.

2.4 Software

Dissemination of new statistical methodologies is certainly facilitated by the development and implementation of statistical software components. Many theoretical papers have not been considered in applied works because of lack of user-friendly software.

Orsini *et al.* (2006) described the `glst` command in Stata, the first publicly available procedure dedicated for dose–response meta-analysis. The command implements both the one- and two-stage approaches limited, in case of a random-effects analysis, to a linear trend. A two-stage random-effects meta-analysis of non-linear relationships can be performed with the aid of the `mvmeta` command for multivariate meta-analysis (White *et al.*, 2011). Several worked examples and codes are available at `http://stats4life.se/drm`. Later on, Li and Spiegelman (2010) wrote `%metadose`, a two-steps macro for dose–response meta-analysis, where estimation of non-linear relationships is restricted to a fixed-effect analysis.

The majority of the applied meta-analyses retrieved in our survey were performed using the `glst` procedure in Stata (36, 87%), while 2 used the `metadose` macro in SAS, and 2 studies used functions in RevMan. No dedicated package was available in the free software programming language R (R Core Team, 2017).

Chapter 3

Aims of the thesis

The overall aims of this thesis were to develop and implement new methods for dose–response meta-analysis, in order to deal with the methodological aspects that have not yet been addressed.

More specifically, the aims were:

- To develop, maintain, and share a package for dose–response meta-analysis in the open source and free R software.

- To present and discuss relevant measures and graphical tools to assess the goodness-of-fit in dose–response meta-analytical models, which is often neglected.

- To develop a new measure of between-study heterogeneity in the broader context of meta-analysis that does not make any assumption about the distribution of the within-study error variances.

- To move beyond the specification of a unique model across the studies exploring possible advantages of a point-wise approach, especially in case of dose–response meta-analysis where the exposure range varies substantially across the studies.

- To avoid exclusion of studies in order to fit more complex and informative models in an alternative one-stage approach for dose–response meta-analysis.

Chapter 4

Methods

4.1 The dosresmeta R package

The first version of the **dosresmeta** R package was released on the Comprehensive R Archive Network (CRAN) on September 2013. It is listed in the CRAN task view Meta-Analysis (`https://CRAN.R-project.org/view=MetaAnalysis`), a guide that covers the vast collection of R packages for facilitating meta-analysis of summary statistics.

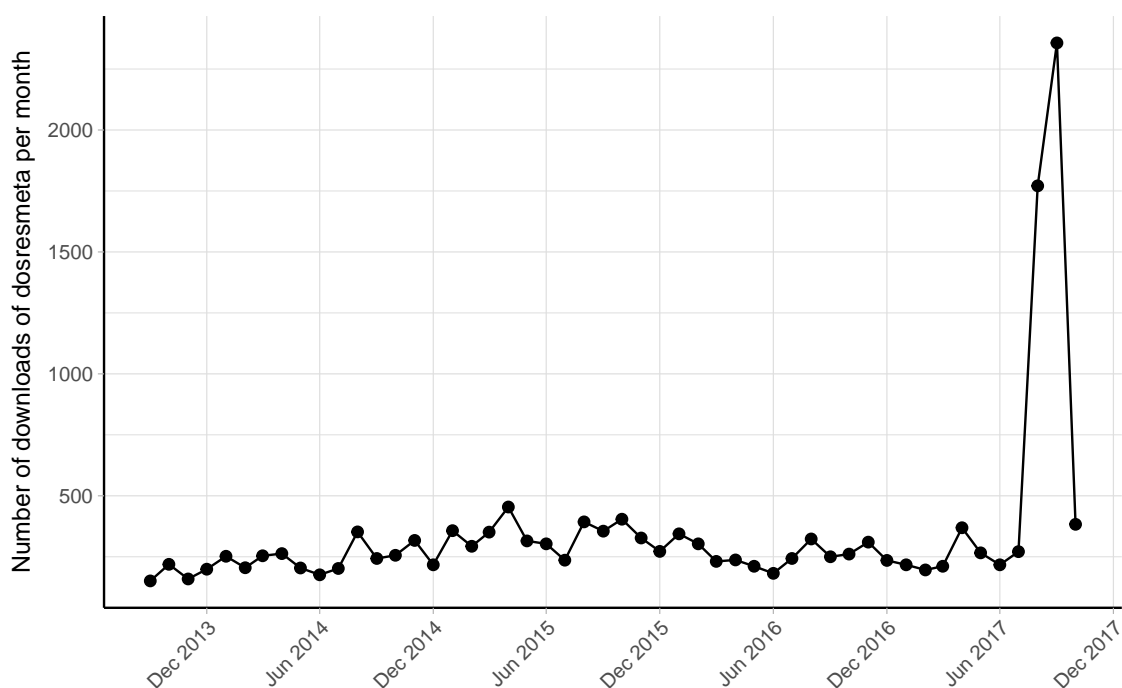Figure 4.1: Monthly number of downloads of the **dosresmeta** R package from the RStudio CRAN mirror September 2013 - December 2017.

The **dosresmeta** package is now available in the updated version 2.0.1 and new features are being implemented in the version under development on GitHub (`https://github.com/alecri/dosresmeta`). Currently, the **dosresmeta** package is downloaded and used world-

wide, with a median number of 260 downloads/month (Figure 4.1). The countries where it has been downloaded most are Great Britain (4005), United States (3905), and China (1605) (Figure 4.2). Working examples, codes, and data are available at `http://alecri.github.io/software` to fully reproduce figures and numbers presented in both applied and theoretical papers.
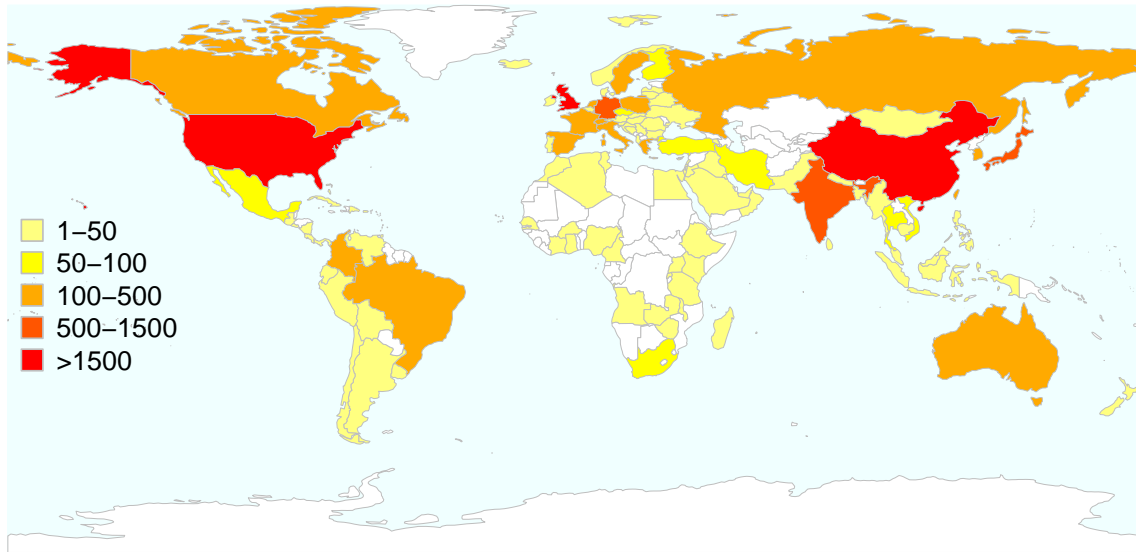

Figure 4.2: Total number of downloads of the **dosresmeta** R package worldwide from the RStudio CRAN mirror September 2013 - December 2017.

The implementation of the package is presented in Paper I, which is also offered as a free guide for the package in a vignette accessible by typing `browseVignettes("dosresmeta")` from the R console.

4.1.1 Architecture and design of the package

The initial version 1.0 of the **dosresmeta** R package implemented the two-stage approach for dose–response meta-analysis described in Sections 2.3.1 and 2.3.2. The package included some facilities for efficiently estimating the dose–response associations across the included studies and used the **mvmeta** package for combining the study-specific regression coefficients (Gasparrini *et al.*, 2012). The main novelty of the version 1.0 was the implementation of `gl` and `hamling` functions for reconstructing the covariance matrices among sets of log relative risks using the methods developed by Greenland and Longnecker (1992) and Hamling *et al.* (2008) (Figure 4.3). In version 1.3, dedicated functions were written for summarizing and displaying results, and for predicting the pooled dose–response association as described in Section 2.3.2. Compared to other routines, the `predict` function offers the possibility of

deriving the predicted curve also for dose levels that are not observed in the analyzed data. The same applies for the choice of the reference dose value. The main advantage is that publication quality curves and tables (i.e. combined results for desired dose values) can be easily obtained with a few lines of code. Practical examples are available at `https://alecri.github.io/software/dosresmeta.html`. Furthermore, the `center` argument was added in the main function for centering the design matrix as described by Liu *et al.* (2009). The argument has been set to TRUE by default for preventing possible errors when modelling non-linear curves, especially in case of non-zero exposure reference categories. Finally, additional arguments were introduced for allowing the specification of a list of covariance matrices directly by the user. This can be useful in pooling projects where the principle investigators share the results of harmonized analyses.

New capabilities and functions were written in the version under development available on GitHub and were finally included in the major release version 2.0 on CRAN. The **dosresmeta** package was largely redesigned in the internal functions but kept unchanged the external form and arguments for backward compatibility. Three main features were introduced: the extension of the two-stage approach for dose–response meta-analysis of differences in means (rather than log relative risks) (Crippa and Orsini, 2016a), the possibility of fitting meta-regression models and the implementation of an alternative one-stage approach. The first was achieved by extending the choices of the `covariance` argument for results presented in terms of mean and standardized mean differences, which related to the `covar.smd` function. The alternative `covariance == "indep"` can be specified for assuming independence of the log relative risks or differences in means. This is particularly useful when the information for reconstructing the covariances is not available (see additional (useful) code section on the referenced site for examples).
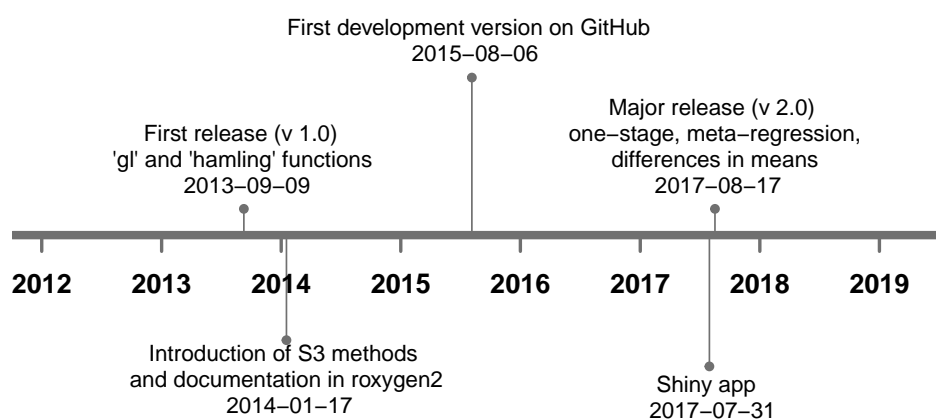

Figure 4.3: Development of the **dosresmeta** R package over time.

The implementation of the one-stage approach and related functions is discussed in more detail

in the methods for Paper V. The updated **dosresmeta** package also implements functions to facilitate specific aspects of a dose–response meta-analysis. These includes the assessment of goodness-of-fit discussed in Paper II, tests for fixed- and random-effects coefficients, conditional and marginal predictions, and the use of fractional polynomials.

Based on the last version of the **dosresmeta** package, an interactive interface is also available at `http://alessiocrippa.com/shiny/dosresmeta`. The web-app can be useful for introducing the concepts of dose–response meta-analysis to those researchers who are not familiar with the R software.

4.1.2 Description of the package

The **dosresmeta** package can be downloaded from CRAN by typing directly in R

```
R> install.packages("dosresmeta")
```

The version under development is instead available from GitHub

```
R> install.packages("devtools")
R> devtools::install_github("alecri/dosresmeta")
```

The package consists of a main function `dosresmeta` with the following arguments

```
R> str(dosresmeta)
function (formula, id, v, type, cases, n, sd, data, mod = ~1, intercept = F,
    center = T, se, lb, ub, covariance = "gl", method = "reml", proc = "2stage",
    Slist, method.smd = "cohen", control = list())
```

The dose–response model is specified in the `formula` argument in a symbolic representation. For example, if logrr and dose are the variable names for the log relative risk and assigned doses, a linear trend is specified as `logrr ~ dose` while a quadratic curve as `logrr ~ dose + I(dose^2)`. The variables are defined in a `data.frame` whose name is specified in the `data` argument. By default `intercept = FALSE` does not include the intercept term in the covariance matrix, which is constructed in terms of contrasts unless `center = FALSE`. The `id` argument specifies the name for the study id variable (can be omitted for single study analysis). The standard errors for the log relative risks are specified in the `se` argument, or alternatively, either the variances (`v`) or the lower (`lb`) and upper bounds (`ub`) of the relative risks need to be specified. The additional information about the study-design (`type`), the number of cases (`cases`), and participants or amount of person-time (`n`) is used for reconstructing the covariance of the log relative risk (or mean differences) using the method specified in the `covariance` argument (default is the Greenland and Longnecker's method). A list of covariance matrices can be passed to the `Slist` argument when `covariance = "user"`. A two-stage procedure with REML estimation method is the default. A one-stage procedure (`proc = "1stage"`) and either ML estimator `method = "ml"` or a fixed-effect analysis `method = "fixed"` can be adopted. Residual heterogeneity can be modeled in a meta-regression analysis

by specifying covariates in the `mods` argument. For example, a different curve depending on the study design can be specified with `mods = ~ type`. Finally, a list of parameters can be passed to the `control` argument to control the fitting process.

The `dosresmeta` function returns an object of class "`dosresmeta`" with the information from the dose–response meta-analytic model. The `print` and `summary` methods display and produce a summary of the content of a `dosresmeta` object. The `predict` method facilitates the presentation of the results of a dose–response meta-analysis

```
R> str(dosresmeta:::predict.dosresmeta)
function (object, newdata, xref, expo = FALSE, xref_vec, ci.incl = TRUE,
    se.incl = FALSE, xref_pos = 1, delta, order = FALSE, ci.level = 0.95,
    ...)
```

where object contains the results of the `dosresmeta` function. A new `data.frame` with the desired doses can be passed to the `newdata` argument for obtaining the corresponding predicted log relative risks. If not provided, the predictions will be calculated for the assigned dose values available from the studies. The `expo` argument can be set to TRUE to predict log relative risk and confidence intervals (unless `ci.incl = FALSE`) on the exponential scale. The reference value can be specified with the `xref` argument, or better, specifying the line of the newdata which serves as referent (`xref_pos` argument). For non-linear models, a vector needs to be provided in `xref_vec` instead of `xref`. The `delta` argument is useful for predicting the linear increase in the outcome for a delta increase in the exposure, and is thus only appropriate in a linear trend analysis. In the updated version of the **dosresmeta** package, a `blup` method has also been implemented for predicting the study-specific random-effects and hence the conditional curves.

Additional functions can be listed

```
R> ls("package:dosresmeta")
 [1] "covar.logrr"       "covar.smd"        "dosresmeta"
 [4] "dosresmeta.control" "dosresmeta.fit"    "fpgrid"
 [7] "fracpol"           "gof"              "grl"
[10] "hamling"           "vpc"              "waldtest"
```

Use `ls(getNamespace("dosresmeta"), all.names=TRUE)` for a complete list including hidden auxiliary functions.

4.1.3 Datasets

Several datasets from published dose–response meta-analyses and methodological articles have been included in the **dosresmeta** package. To get a list as in Table 4.1 with the names and description type

```
R> data(package = "dosresmeta")
```

Table 4.1: Data sets available in the **dosresmeta** R package.

Name	Description
alcohol_crc	Eight published studies on the relation between alcohol intake and colorectal cancer (Orsini *et al.*, 2011b)
alcohol_cvd	Six published studies on the relation between alcohol intake and cardiovascular disease risk (Liu *et al.*, 2009)
alcohol_esoph	Fourteen case-control studies on the relation between alcohol consumption and esophageal cancer (Rota *et al.*, 2010)
alcohol_lc	Four published studies on the relation between alcohol intake and lung cancer (Orsini *et al.*, 2011b)
ari	Five clinical trials on the relation between aripiprazole and schizophrenia (Crippa and Orsini, 2016a)
bmi_rc	Four case-control studies on the relation between Body Mass Index and renal cell cancer (Liu *et al.*, 2009)
cc_ex	Case-control data on alcohol and breast cancer risk (Greenland and Longnecker, 1992)
ci_ex	Cumulative incidence data on high-fat dairy food and colorectal cancer risk (Orsini *et al.*, 2006)
coffee_cancer	Eight prospective studies on the relation between coffee consumption and cancer mortality (Crippa *et al.*, 2014)
coffee_cvd	Thirteen prospective studies on the relation between coffee consumption and cardiovascular mortality (Crippa *et al.*, 2014)
coffee_mort	Twenty-one prospective studies on the relation between coffee consumption and all-cause mortality (Crippa *et al.*, 2014)
coffee_mort_add	Additional two prospective studies on the relation between coffee consumption and all-cause mortality (Nilsson *et al.*, 2012)
coffee_stroke	Eleven prospective studies on the relation between coffee consumption and stroke risk (Larsson and Orsini, 2011)
fish_ra	Six studies on the relation between fish consumption and rheumatoid arthritis risk (Di Giuseppe *et al.*, 2014)
ir_ex	Incidence-rate data on fiber intake and coronary heart disease risk (Orsini *et al.*, 2006)
milk_mort	Eleven prospective studies on the relation between milk consumption and all-cause mortality (Larsson *et al.*, 2015)
milk_ov	Nine studies on the relation between milk consumption and ovarian cancer (Larsson *et al.*, 2006)
oc_breast	Twenty-two case-control studies on the relation between oral contraceptives use and breast cancer (Berlin *et al.*, 1993)
process_bc	Ten studies on the relation between processed meat and bladder cancer (Crippa *et al.*, 2016b)
red_bc	Twelve studies on the relation between red meat and bladder cancer (Crippa *et al.*, 2016b)
sim_os	Simulated data for one-stage dose-response meta-analysis (Crippa *et al.*, 2018a)

4.2 Goodness-of-fit

The aim of Paper II was to address how to evaluate the goodness-of-fit in dose–response meta-analysis. The distance between observed and predicted data is often at the heart of model checking. Although we are working with linear regression models, features of aggregated dose–response data can complicate this comparison. The predicted curve is presented using a referent, x_{ref}, which may differ from the study-specific reference values. Thus the predicted and observed log relative risks can no longer be directly compared because their baseline group is not the same. Even in the case where the reference values are all the same and equal to x_{ref}, the covariance among the study-specific RRs and the heterogeneity across the studies adds further difficulties in evaluating if the chosen model provides a good summary of the observed data. The proposed tools for assessing the goodness-of-fit are presented in a fixed-effect analysis where meta-regression models as in 2.20 are typically employed to explain the observed heterogeneity. We consider relevant measures, tests, and graphical tools that take into account the correlation of the observed data.

4.2.1 Deviance

The first natural measure of goodness-of-fit is a comparison between the predicted and observed data points, i.e. the non-referent log relative risks. This can be done by analyzing the residual term errors, which are defined by the difference between the observed log RRs, $\mathbf{y}_i$, and the marginal prediction

$$\hat{\boldsymbol{e}}_{\boldsymbol{i}} = \mathbf{y}_i - \mathbf{X}_i \widetilde{\mathbf{X}}_i \hat{\boldsymbol{\beta}} \tag{4.1}$$

A summary for the error terms is the deviance

$$D = \sum_{i=1}^{I} \left(\mathbf{y}_i - \mathbf{X}_i \widetilde{\mathbf{X}}_i \hat{\boldsymbol{\beta}} \right)^{\top} \mathbf{S}_i^{-1} \left(\mathbf{y}_i - \mathbf{X}_i \widetilde{\mathbf{X}}_i \hat{\boldsymbol{\beta}} \right) = \sum_{i=1}^{I} \hat{\boldsymbol{e}}_{\boldsymbol{i}}^{\top} \mathbf{S}_i^{-1} \hat{\boldsymbol{e}}_{\boldsymbol{i}} \tag{4.2}$$

The deviance D measures the total squared deviation between observed and predicted data taking into account the covariance matrices $\mathbf{S}_i$ of the error terms. It is usually referred to as generalized residual sum of squares (GRSS) (Draper and Smith, 2014). Decreasing values of D indicate a better agreement between reported and fitted log RRs, with 0 being the lower bound that corresponds to perfect agreement (saturated model).

The D statistic can be used as a test for model specification. Under the null hypothesis that the model is correctly specified, D follows a χ^2 distribution with df $= \sum_{i=1}^{I} J_i - pm$. This is equivalent to test if the residual variance, corrected for the correlation of the error terms, is larger than expected assuming that the dose–response model is correct. A small p value can be interpreted as evidence that the fitted model fails in accounting the observed variation in the log relative risks. As for the Q test, the deviance has no upper bound and is thus difficult to interpret the absolute value of D.

4.2.2 Coefficient of determination

To complement the D statistic (and the corresponding test), the coefficient of determination, R^2, can be used as a descriptive measure of goodness-of-fit (Hagquist and Stenbeck, 1998; Kvålseth, 1985). The R^2 is the complement to the unit of the ratio between the GRSS and the generalized sum of squares, GTSS $= \sum_{i=1}^{I} \mathbf{y}_i^\top \mathbf{S}_i^{-1} \mathbf{y}_i$. Because dose–response models do not have the intercept term, the R^2 can be defined as in (Buse, 1973; Theil, 1958)

$$R^2 = 1 - \frac{\text{GRSS}}{\text{GTSS}} = 1 - \frac{\sum_{i=1}^{I} \left(\mathbf{y}_i - \mathbf{X}_i \widetilde{\mathbf{X}}_i \hat{\boldsymbol{\beta}} \right)^\top \mathbf{S}_i^{-1} \left(\mathbf{y}_i - \mathbf{X}_i \widetilde{\mathbf{X}}_i \hat{\boldsymbol{\beta}} \right)}{\sum_{i=1}^{I} \mathbf{y}_i^\top \mathbf{S}_i^{-1} \mathbf{y}_i} \tag{4.3}$$

R^2 is a dimensionless number that is bounded between 0 and 1 and can be generally interpreted as the proportion of the generalized total sum of squares explained by the dose–response model and study-level covariates. The lower bound 0 corresponds to the case where all the β coefficients are equal to 0 and therefore the model is not able to explain the variability in the observed log relative risks. Contrary, the upper bound 1 indicates a perfect agreement between reported and fitted data.

By construction the R^2 can not decrease as the number of regression coefficients increases. A penalized, or adjusted, version that takes into account this behavior can be defined

$$R^2_{\text{adj}} = 1 - \frac{\sum_{i=1}^{I} J_i}{\sum_{i=1}^{I} J_i - m} \left(1 - R^2 \right) \tag{4.4}$$

Possibly, a low R^2 may indicate that more flexible transformations of the exposure are needed, or that there is considerable residual variability that might be explained by study-level covariates.

4.2.3 Visual tools

The visual assessment of the goodness-of-fit may reveal specific patterns in the data that might otherwise go undetected by only looking at summary measures such as the deviance and the coefficient of variation (Kvålseth, 1985). The graphical comparison can sometimes be cumbersome because different factors may affect such analysis. The aggregated dose–response data are presented using one group as comparison. This feature has two important consequences when deducing an overall trend from the reported (log) relative risks. The first is that different parameterizations, or models, can be graphically compared only if the predicted risk for the reference category is similar. The second involves the correlation of the modelled data points, which makes it difficult to evaluate if a specific model fits the reported data.

To illustrate both problems, I consider IPD from one of the registers involved in the Surveillance, Epidemiology, and End Results (SEER) Program (`https://seer.cancer.gov`), and focus on the association between age at diagnosis and breast cancer mortality. Using 35, 60, 70 years as cut points, I modelled age using 5 categories in a Poisson model adjusting for potential confounders and produced aggregated dose–response results in Table 4.2, where the

first category served as referent.

Table 4.2: Aggregated dose–response data on the adjusted association between age and breast cancer mortality based on one registry from the SEER program using the first age category as referent.

Age category	Mean age	Cases	Person-years	IRR	95% CI	log IRR	SE
[20,35]	32.4	56	2099.3	1.00	—	0.00	0.00
(35,60]	49.7	451	37585.3	0.68	(0.51, 0.90)	-0.39	0.14
(60,70]	65.6	220	20590.7	0.71	(0.53, 0.96)	-0.34	0.15
(70,97]	77.0	215	17896.5	0.77	(0.57, 1.04)	-0.27	0.15

Using the methodology presented in Section 2.3.1, I estimated a linear trend and graphically compared the predicted curve and the modelled data (left panel of Figure 4.4). While the empirical log RRs may suggest an inverse association between age and breast cancer mortality, the estimated linear trend indicates opposite conclusions (a 0.17% increase in mortality for every 5-year increase of age at diagnosis). The linear trend does not seem to properly fit the data since the fitted line doesn't event pass through the reported log RRs. The problem is that there are few cases in the reference age category and so the log RRs are typically larger. By changing the group comparison to the second age category, this issue is partially solved (right panel of Figure 4.4). To address the further issue of the covariance, we proposed graphical tools based on the analysis of the decorrelated residuals.
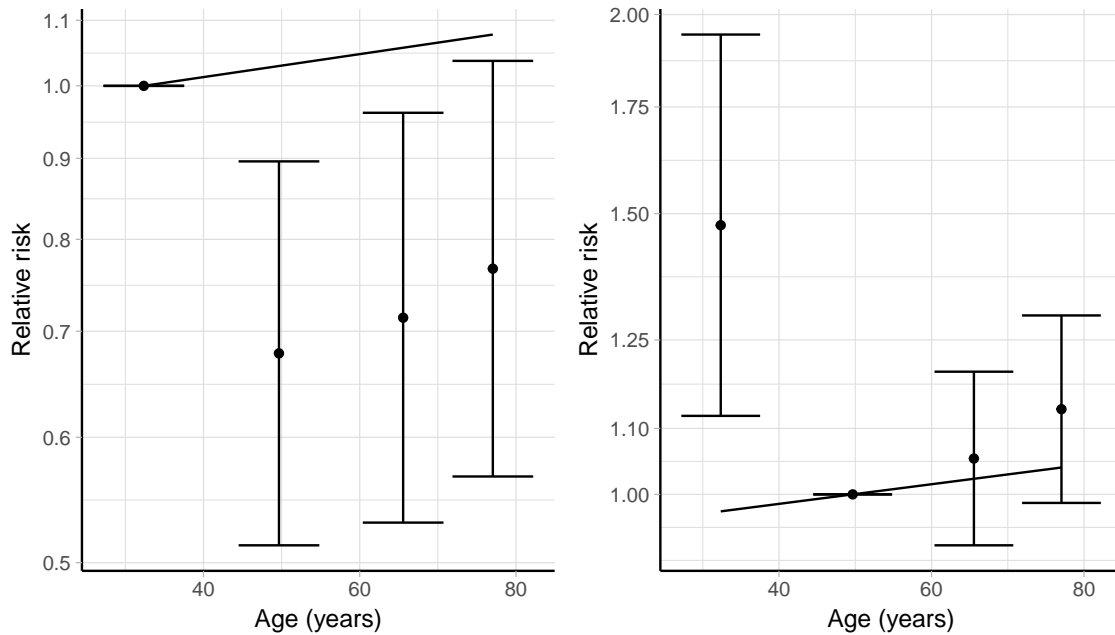
Figure 4.4: Comparison between observed and fitted data, changing the comparison group from the first age category (left panel) to the second one (right panel), based on the aggregated data on the association between age and breast cancer mortality presented in Table 4.2.

The residuals can be decorrelated using the Cholesky factorization of the covariance matrices $\mathbf{S}_i = \mathbf{C}_i\mathbf{C}_i^\top$, with $\mathbf{C}_i$ being a lower $J_i \times J_i$ triangular matrix. The decorrelated residuals $\hat{\boldsymbol{e}}_i^*$

are then calculated as

$$\hat{e}_i^* = \mathbf{C}_i^{-1}\left(\mathbf{y}_i - \mathbf{X}_i\widetilde{\mathbf{X}}_i\hat{\boldsymbol{\beta}}\right) = \mathbf{C}_i^{-1}\hat{e}_i \tag{4.5}$$

and plotted against the exposure. Such a graphical analysis is the equivalent to a residual vs. predictors/predicted outcome plot after a model estimated through ordinary least squares. Since the variables in the plot are transformed, the actual value of the decorrelated residuals is not directly interpretable. However, it is still possible to detect specific patterns for the residuals over the exposure range. In such a case, the plot can indicate lack of fit of the chosen model for select exposure values. Overlaying a locally weighted scatterplot smoothing (LOWESS) may help to visualize possible patterns, while a different shape or color for the residuals can distinguish them according to study-level covariates in case of meta-regression models.

4.3 A new measure of heterogeneity

One of the possible reasons for large values of D or similarly small values of R^2 is the presence of heterogeneity in the dose–response associations. The measures of heterogeneity presented in Section 2.1.3, namely $\hat{R}_I$ and I^2 were derived based on the assumption of homogeneity for the error variance terms v_i, which is unlikely to be met in applications. To illustrate this point, let us consider the two hypothetical distributions for the within-study error terms in Table 4.3. The distribution of v_i in the first example (Analysis A) is much more homogeneous compared to the second scenario (Analysis B). The coefficient of variation for the v_i in the second example is 20 times higher than the corresponding number in the first scenario. While it seems reasonable to assume homogeneity of the v_i for the Analysis A, in the second case this hypothesis is not appropriate. Nonetheless, both $\hat{R}_I$ and I^2 would summarize the two distributions with a single common number.

Table 4.3: Hypothetical distributions of within-study errors in two meta-analyses of 5 studies.

Analysis	$v_1,\dots,v_5$	CV_{v_i}	s_1^2	s_2^2
A	5, 5.2, 4.9, 5.3, 4.8	0.04	5.0	5.0
B	4, 17, 15, 2, 3.8	0.84	5.0	4.4

4.3.1 Definition and properties

The available measures relate the between-study heterogeneity to the overall variability, whose definition also varies across studies. A different approach consists of measuring the impact of heterogeneity in determining the variance of the combined effect in a random-effects analysis. This quantity depends on the τ^2 and on the variance of the mean $\hat{\beta}$, quantities that don't require the assumption of homogeneity for the within-study error terms.

To determine how τ^2 contributes to $\widehat{\mathrm{Var}}\left(\hat{\beta}\right)$, we consider the hypothetical case where all the estimates $\hat{\beta}_i$ are reported with no error, i.e. $v_i = 0\ \forall i = 1,\dots,I$. The weights used in the meta-analytic model depends only on τ^2, and following Equation 2.3 the variance of the combined effect is $\widehat{\mathrm{Var}}\left(\hat{\beta}\right) = \left(\sum_{i=1}^I \hat{\tau}^2\right)^{-1} = \hat{\tau}^2/I$. In the more realistic case where $v_i > 0$,

the variance of $\hat{\beta}$ will increase to incorporate the uncertainty in the reported estimates. The contribution of the heterogeneity is still defined by $\hat{\tau}^2/I$. The new measure of heterogeneity, $\hat{R}_b$, can be written as

$$\hat{R}_b = \frac{\hat{\tau}^2}{I\widehat{\text{Var}}(\hat{\beta})} = \frac{\hat{\tau}^2}{I/\left(\sum_{i=1}^{I}\frac{1}{v_i+\hat{\tau}^2}\right)} = \frac{1}{I}\sum_{i=1}^{I}\frac{\hat{\tau}^2}{\hat{\tau}^2+v_i} \tag{4.6}$$

The $\hat{R}_b$ is a dimensionless number that can be expressed as a percentage, ranging from 0% (corresponding to $\hat{\tau}^2 = 0$, i.e. no observed heterogeneity) to 100%, for the hypothetical case where all the effects are estimated with no error. The proposed measure is defined as a function of the estimated heterogeneity ($\hat{\tau}^2$), the number of studies (I), and the within-study error terms (v_i). It satisfies the criteria required for a measure of heterogeneity (Higgins and Thompson, 2002): it is a non-decreasing function of $\hat{\tau}^2$, it is invariant to scale transformation of the estimates, and is not intrinsically affected by the number of studies included in the analysis. Similarly to $\hat{R}_I$ and I^2, the proposed measure is a function on the within-error terms v_i and so it tends to 1 in case of meta-analysis of very precise estimates, even for relative small values of $\hat{\tau}^2$. As already mentioned in Section 2.1.3, the use of the between-studies coefficient of variation can complement these measures (Takkouche et al., 1999). The right hand of Equation 4.6 expresses $\hat{R}_b$ as an average of the ratios of the $\hat{\tau}^2$ to the study-specific overall variance $v_i + \hat{\tau}^2$. It is easy to derive that $\hat{R}_b$ coincides with the definition of $\hat{R}_I$ and I^2 in case of homogeneity of the v_i. When the within-errors vary across study, the difference between $\hat{R}_b$ and I^2 will depend on the actual values of v_i, while it can be proven that $\hat{R}_b \leq \hat{R}_I$.

An estimate of the between-study heterogeneity is required for the computation of $\hat{R}_b$ as well for the alternative measures $\hat{R}_I$ and I^2. The moment-based estimator presented in Equation 2.5 is a standard choice in many applied meta-analyses and it has the advantage of having a closed formulation. In addition, the estimator is consistent (Jackson et al., 2010). Using this estimation method, $\hat{R}_b$ can be expressed as a function of the Q statistic

$$\hat{R}_b = \frac{1}{I}\sum_{i=1}^{I}\frac{Q-(I-1)}{Q+a_i-(I-1)} \tag{4.7}$$

with $a_i = v_i\left(\sum_{i=1}^{I}w_i - \sum_{i=1}^{I}w_i^2/\sum_{i=1}^{I}w_i\right)$. The asymptotic variance for $\hat{R}_b$ can be derived by using the delta method on the relation 4.7

$$\widehat{\text{Var}}(\hat{R}_b) \approx \left(\frac{1}{I}\sum_{i=1}^{I}\frac{a_i}{(Q+a_i-(I-1))^2}\right)^2 \text{Var}(Q) \tag{4.8}$$

The formula for $\text{Var}(Q) = 2(I-1)+4\left(S_1-\frac{S_2}{S_1}\right)\tau^2+2\left(S_2-2\frac{S_3}{S_1}+\frac{S_2^2}{S_1^2}\right)$, with $S_r = \sum_{i=1}^{I}w_i^r$, was first presented by Biggerstaff and Tweedie (1997).

When the number of studies is large, $\hat{R}_b$ is asymptotically normally distributed and thus Wald-type confidence interval can be constructed $\hat{R}_b \mp z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{R}_b)}$.

4.4 A point-wise approach

Measures of heterogeneity are frequently employed in applied works. In a dose–response meta-analysis, however, there are two aspects of heterogeneity that can not be captured using standard methods. More specifically, in a two-stage analysis the same functional transformations $g_1, \ldots, g_p$ need to be defined for all the studies so that the regression coefficients can be properly combined by meta-analysis. This may have important consequences in case the chosen dose–response model adequately fits only a subset of the analyzed studies. To illustrate this aspect, I considered a subset of 4 studies on the association between milk consumption (ml/day) and all-cause mortality (Larsson *et al.*, 2015). I modelled the individual curves using FP2 and selected in each study the combination of power terms which maximized the log likelihood (or equivalently with the minimum AIC). The chosen power terms in the individual analyses differ although the predicted curves describe a similar association (Figure 4.5). Forcing a unique set of power terms may decrease the fit of some of the study-specific analyses and thus produce more unstable estimates for the dose–response parameters.

Another difficulty may be encountered in uniformly defining the *g* transformations across the studies. For instance, we might decide to model the previous association between milk consumption and mortality using RCS. A percentile approach is commonly adopted for choosing where to place the **k** in Equation 2.32 (Harrell, 2013). The first and third knots are located at fixed percentiles, generally at the 10-th and 90-th percentiles of the exposure distribution, while the median exposure value is chosen for the remaining knot. If the exposure distributions largely vary (especially in terms of range definition), as it is in this case, it may not be possible to equally locate the knots in all the studies. Indeed, in our example only one study considered dose levels higher than the upper knot (433 ml/day). As a consequence, the second spline transformation for the other studies will be equal 0 and the model is not estimable (the design matrix is not invertible).

A related problem directly affects the prediction of the pooled log relative risks. The maximum dose levels varies largely across studies, with values 715, 441, 369, 146 ml/day. The predictions from the second stage analysis, however, are obtained using the combined $\hat{\boldsymbol{\beta}}$, which disregards the information about the exposure. All the studies contribute in predicting the log relative risks, even if some of them only reported results for low exposure values. In such a way, the combined curve may be severely affected by these extrapolations.

A point-wise approach for meta-analysis of aggregated dose–response data may properly address the described issues. A point-wise approach was first presented by Sauerbrei and Royston (2011) for meta-analysis of continuous covariates based on IPD. The strategy consists of separately estimating the study-specific curves and, based on them, calculating the predicted log relative risks for selected exposure values. The combined curve is obtained by averaging the predicted relative risks instead of calculating them from the combined regression coefficients. The described methodology has the advantage of fitting potential diverse curves across the studies, and limiting the study-specific predicted log relative risks to the observed exposure range.
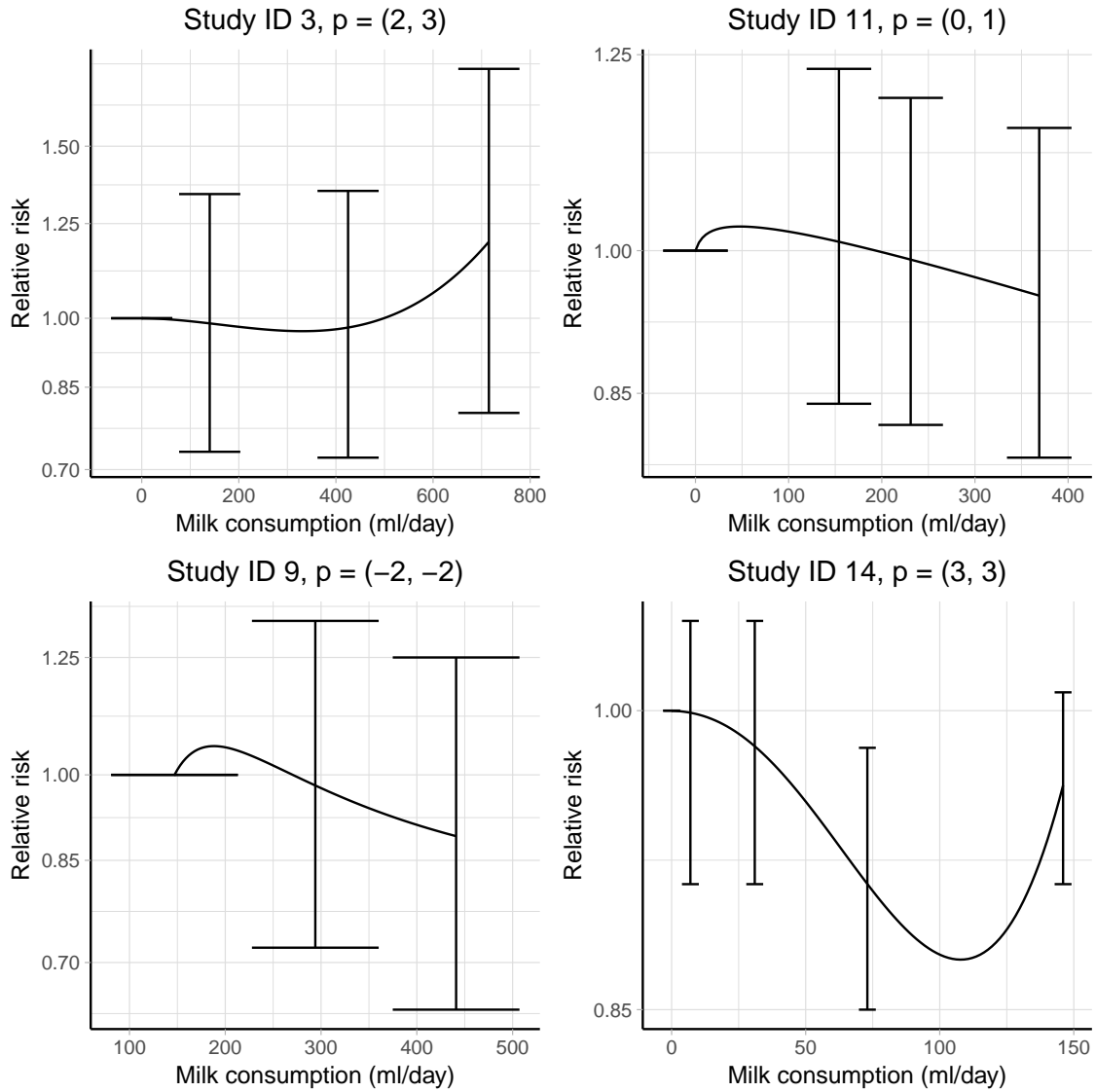
Figure 4.5: Dose–response associations between milk consumption (ml/day) and all-cause mortality in 4 studies. The curves are modelled using fractional polynomials with the sets of power terms p (reported in the title) chosen by maximizing the study-specific likelihoods. The results are presented on the log scale using the observed reference values as comparators.

4.4.1 Estimation and prediction of study-specific curves

The first step of a point-wise approach is similar to the methodology presented in Section 2.3.1 with the difference that the g transformations in the design matrices 2.14 are subscripted by the study index i, to highlight that they may differ across the studies, also in terms of number (p_i)

$$\mathbf{X}_i = \begin{bmatrix} g_{i1}(x_{i1}) - g_{i1}(x_{i0}) & \cdots & g_{ip_i}(x_{ip}) - g_{ip_i}(x_{i0}) \\ \vdots & & \vdots \\ g_{i1}(x_{iJ_i}) - g_{i1}(x_{i0}) & \cdots & g_{ip_i}(x_{iJ_i}) - g_{ip_i}(x_{i0}) \end{bmatrix} \tag{4.9}$$

For example, different power terms for FP2 can be chosen or, for a RCS analysis, the knots can be located separately in each study. Potentially, a mixture of curves could be also estimated across the study in order to improve the fit of the individual dose–response analyses. The regression coefficients are then estimated using GLS estimators presented in Equation 2.16.

Once the curves have been estimated, the results of the first step analysis are presented in terms of predicted log relative risks for a set of n_i exposure values $\mathbf{x}_{n_i}$ using a suitable common value x_{ref} as comparator. The index i is used to highlight that the chosen dose values may differ across the studies. In particular, the predictions in each study can be limited to the observed exposure range: $\max\left(\mathbf{x}_{n_i}\right) \leq \max(\mathbf{x_i})$.

$$\hat{\mathbf{y}}_i = \mathbf{X}_{n_i}\hat{\boldsymbol{\beta}}_i \tag{4.10}$$

$$\hat{\mathbf{v}}_i = \text{diag}\left(\mathbf{X}_{n_i}\widehat{\text{Var}}\left(\hat{\boldsymbol{\beta}}_i\right)\mathbf{X}_{n_i}^{\top}\right) \tag{4.11}$$

with formulas similar to those presented in Equation 2.25 and 2.26, with the notable difference that the predictions are based on the study-specific $\hat{\boldsymbol{\beta}}_i$ rather than the mean $\hat{\boldsymbol{\beta}}$ coefficients.

4.4.2 Averaging of dose–response predictions

In a point-wise strategy, the combined dose–response curve is derived by pooling the study-specific predicted log RRs derived in Equation 4.10. The second step consists of $n = \max(n_i)$ univariate meta-analyses where the effect sizes are the elements in $\hat{\mathbf{y}}_i$ with the within-error variances $\hat{\mathbf{v}}_i$. The combined predicted (log) relative risks are estimated using the weighted average presented in Equation 2.2.

As a final result, the combined dose–response curve is graphically presented as a smooth function of the combined predicted relative risks for the chosen n dose levels. In addition, all the results of the univariate meta-analyses can be pointwisely presented, such as estimates of the heterogeneity τ^2 and related measures ($\hat{R}_b$, I^2, and $\hat{R}_i$), the Q statistic, and the study-specific weights, with the potential of a much richer set of results.

4.5 A one-stage model

The study-specific dose–response analyses in a standard two-stage or alternative point-wise approach suffer from the limited number of data points. Most of the individual studies present results for 2 to 4 non-referent exposure categories, but cases where dichotomization of the exposure occurred are not rare (Turner *et al.*, 2010). In the latter case, only dose–response models parameterized by $p = 1$ coefficient can be estimated (e.g. linear trend). Indeed, a standard requirement for meta-analysis of non-linear curves is that the studies provide at least two non-referent relative risks.

The extension of the one-stage approach to a random-effects meta-analysis may overcome the exclusions of studies with limited number of individual data points. A one-stage approach is conceptually easier since the entire analysis can be formulated in a single statistical model.

Flexible curves characterized by multiple parameters ($p > 3$) can also be easily accommodated, so that more elaborate research questions can be answered, without losing the information from those studies usually excluded in a two-stage analysis.

4.5.1 Model definition

A one-stage dose–response meta-analysis can be written in the form of a linear mixed-effects model

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i \tag{4.12}$$

where the terms $\mathbf{y}_i$, $\mathbf{X}_i$, and $\boldsymbol{\varepsilon}_i$ are defined exactly as those presented in the first step of a two-stage analysis (Equation 2.13). The quantities $\boldsymbol{\beta}$ and $\mathbf{b}_i$, instead, are the population average parameter and unobserved random-effects as defined in the second step (Equation 2.18). The additional $\mathbf{Z}_i$ is the $n_i \times p$ design matrix for the random-effects and coincides with $\mathbf{X}_i$. Assuming a multivariate distribution for the random-effects terms, the marginal model for a one-stage analysis is

$$\mathbf{y}_i \sim \mathcal{N}\left(\mathbf{X}_i \boldsymbol{\beta}, \mathbf{Z}_i \boldsymbol{\Psi} \mathbf{Z}_i^\top + \mathbf{S}_i\right) \tag{4.13}$$

Similarly, the conditional model is defined as

$$\mathbf{y}_i \mid \boldsymbol{b}_i \sim \mathcal{N}\left(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \boldsymbol{b}_i, \mathbf{S}_i\right) \tag{4.14}$$

Note that the conditional and marginal models are now defined for the log relative risks rather than the dose–response coefficients. In particular, the definition of the marginal variance $\boldsymbol{\Sigma}_i = \mathbf{Z}_i \boldsymbol{\Psi} \mathbf{Z}_i^\top + \mathbf{S}_i$ is quite different. It depends not only on the within and between components but also on the corresponding dose value x_{ij} (or equivalently z_{ij}) associated with y_{ij}.

Meta-regression models can be estimated by including in $\mathbf{X}_i$ the interaction terms between the p transformations of the exposure variable and study-level covariates, while leaving unchanged the definition of the $\mathbf{Z}_i$ matrix for the random-effects. Assuming a quadratic curve and a binary study-level covariate u_i, for example, the design matrices for the fixed- and random-effects can be written as

$$\mathbf{X}_i = \begin{bmatrix} x_{i1} - x_{i0} & x_{i1}^2 - x_{i0}^2 & (x_{i1} - x_{i0})u_i & \left(x_{i1}^2 - x_{i0}^2\right)u_i \\ \vdots & \vdots & \vdots & \vdots \\ x_{iJ_i} - x_{i0} & x_{iJ_i}^2 - x_{i0}^2 & \left(x_{iJ_i} - x_{i0}\right)u_i & \left(x_{iJ_i}^2 - x_{i0}^2\right)u_i \end{bmatrix}$$

$$\mathbf{Z}_i = \begin{bmatrix} x_{i1} - x_{i0} & x_{i1}^2 - x_{i0}^2 \\ \vdots & \vdots \\ x_{iJ_i} - x_{i0} & x_{iJ_i}^2 - x_{i0}^2 \end{bmatrix}$$

4.5.2 Estimation, hypothesis testing, and model comparison

The parameters to be estimated are the $m \times p$ fixed effects and the $p \times (p+1)/2$ elements of the between-study variance matrix. We consider likelihood-based estimators that maximize either the log-likelihood of model 4.14

$$\ell\left(\boldsymbol{\beta}, \boldsymbol{\xi}\right) = -\frac{1}{2}n\log(2\pi) - \frac{1}{2}\sum_{i=1}^{I}\log|\boldsymbol{\Sigma}_i\left(\boldsymbol{\xi}\right)| - \frac{1}{2}\sum_{i=1}^{I}\left[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^{\top}\left(\boldsymbol{\Sigma}_i\left(\boldsymbol{\xi}\right)\right)^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})\right] \quad (4.15)$$

or the corresponding restricted version

$$\begin{aligned}\ell_R\left(\boldsymbol{\xi}\right) = &-\frac{1}{2}(n-p)\log(2\pi) - \frac{1}{2}\sum_{i=1}^{I}\log|\boldsymbol{\Sigma}_i\left(\boldsymbol{\xi}\right)| - \frac{1}{2}\log\left|\sum_{i=1}^{I}\mathbf{X}_i^{\top}\left(\boldsymbol{\Sigma}_i\left(\boldsymbol{\xi}\right)\right)^{-1}\mathbf{X}_i\right| + \\ &-\frac{1}{2}\sum_{i=1}^{I}\left[\left(\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}\right)^{\top}\left(\boldsymbol{\Sigma}_i\left(\boldsymbol{\xi}\right)\right)^{-1}\left(\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}\right)\right]\end{aligned} \quad (4.16)$$

As in case of multivariate meta-analysis, likelihood estimators requires iterative algorithms to optimize the functions 4.15 and 4.16. In particular, the Nelder-Mead method can be employed to find the maximum for the parameters in the multidimensional domain. For ease of computation, both likelihoods are expressed as function of only the $\boldsymbol{\xi}$ parameters, which corresponds to the elements of the lower triangular Cholesky decomposition of $\boldsymbol{\Psi}$. The algorithms start with an initial guess for $\boldsymbol{\Psi}$, obtain an estimate for $\boldsymbol{\beta}$ using GLS estimators and maximize the objective function in terms of $\boldsymbol{\xi}$. The steps are iterated until convergence is achieved.

Test of hypothesis and confidence intervals are based on the established theory for mixed models. Inference on the fixed-effects coefficients is conducted in a similar way as presented in Section 2.3.2, using the asymptotic normal distribution for the estimator of $\boldsymbol{\beta}$. Tests for the variance components, instead, generally require a mixture of χ^2 because the coefficients to be tested can only be positive. When $p \geq 2$, however, the distribution of the test statistic is difficult to implement, so that alternative measures can be instead applied.

Following the idea behind the definition of the ICC, the marginal variance in 4.13 can also be decomposed in the within-study and between-study components. The dose–response model 4.12 is a mixed model with random-effects for the slope terms and with no intercept. Thus, the between-study variance is a quadratic function of the assigned dose values. For this setting Goldstein *et al.* (2002) defined the Variance Partition Coefficient (VPC) as the ratio of the between-studies component by the total residual variability

$$\text{VPC}_{ij} = \frac{\mathbf{z}_{ij}\boldsymbol{\Psi}\mathbf{z}_{ij}^{\top}}{\mathbf{z}_{ij}\boldsymbol{\Psi}\mathbf{z}_{ij}^{\top} + s_{ij}^2} \quad (4.17)$$

The VPC is indexed by both i and j because it depends both on the observed dose value z_{ij} and the variance for the log relative risk s_{ij}^2. Values for VPC can be expressed as percentage to quantify the proportion of residual variance attributable to heterogeneity. Because the VPC will typically vary for different doses, overlaying a LOWESS smoother in a scatter plot VPC versus

dose levels may help to examine the impact of heterogeneity over the exposure range.

Information criteria based on the log-likelihood 4.15 can be used to compared the fit of alternative models. As opposed to the corresponding measure from a two-stage analysis, the AIC from a one-stage model is not limited to the comparison of meta-regression models assuming the same functional relationship. Instead, the fit of different dose–response analyses can be properly compared using these fit indices since the log-likelihood is conditional on the modelled log relative risks rather than the study-specific regression coefficients.

4.5.3 Prediction

Predictions for the combined, or marginal, curve are obtained as in Equation 2.25, where now the $\hat{\boldsymbol{\beta}}$ coefficients were estimated using the log relative risks as outcome variables rather than the study-specific $\hat{\boldsymbol{\beta}}_i$.

Predictions are also available for study-specific curves. Using the normality distribution for the random-effects, Henderson *et al.* (1959) computed the asymptotic BLUP of the random-effects $\boldsymbol{b}$ as

$$\hat{\mathbf{b}}_i = \hat{\boldsymbol{\Psi}} \mathbf{Z}_i^\top \hat{\boldsymbol{\Sigma}}_i^{-1} \left(\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}} \right) \tag{4.18}$$

The conditional dose–response coefficients are defined as $\mathbf{X}_i \hat{\boldsymbol{\beta}} + \hat{\mathbf{b}}_i$. Of note, individual curves defined by p parameters can be predicted for those studies reporting $J_i < p$ non-referent relative risks. The BLUP employ the information of the entire distribution of the random-effects to make the best possible prediction.

4.5.4 Comparison with two-stage analysis

Previous methodological articles have oftentimes implemented new methods using the two-stage approach, mainly because of computation reasons. The alternative one-stage approach has been frequently referred to as equivalent and was not further investigated. The tools to assess the goodness-of-fit in Paper II were established using the one-stage framework. We proved in the supplementary material of the original paper that the two techniques give identical results in a fixed-effects analysis of non-linear curves, not only for the simpler case of a linear trend. In the appendix of Paper V, we extended the equivalence to the setting of non-linear curves for a random-effects model. In order to provide the same point and interval estimates, the study-specific models in the two-stage analysis need to be estimable, i.e. $p \leq \min(J_i)$. In practical examples, small discrepancies in $\hat{\boldsymbol{\beta}}$ and $\widehat{\mathrm{Var}}\left(\hat{\boldsymbol{\beta}}\right)$ may be related to differences in the optimization methods for the objective functions of the two techniques.

Chapter 5

Results

We illustrated the usage of the developed methodologies and measures by reanalyzing data from published dose–response meta-analyses.

5.1 Paper I

We demonstrated the main aspects of the methodology reanalyzing aggregated dose–response data from 21 prospective studies on the association between coffee consumption (cups/day) and all-cause mortality (Crippa *et al.*, 2014). The data set `coffee_mort` is included in the package, with the first six lines printed below.

```
R> library(dosresmeta)
R> data("coffee_mort")
R> head(coffee_mort)
  id            author year type dose cases    n      logrr        se gender   area
1  1   LeGrady et al. 1987   ci  0.5    57  249  0.0000000 0.0000000      M    USA
2  1   LeGrady et al. 1987   ci  2.5   136  655 -0.2876821 0.1391187      M    USA
3  1   LeGrady et al. 1987   ci  4.5   144  619 -0.1743534 0.1373198      M    USA
4  1   LeGrady et al. 1987   ci  6.5   115  387  0.0861777 0.1401409      M    USA
5  2 Rosengren et al. 1991   ci  0.0    17  192  0.0000000 0.0000000      M Europe
6  2 Rosengren et al. 1991   ci  1.5    88 1121 -0.1203576 0.2531537      M Europe
```

5.1.1 Single study analysis

We showed how to estimate the dose–response association in a single study. For that purpose, we selected the first study ID 1 consisting of 3 non-referent log relative risks (Legrady *et al.*, 1987). One relevant feature of aggregated dose–response data is the correlation arising from having a common comparator. The `covar.logrr` function can be used to reconstruct the covariance matrix accordingly to the method by Greenland and Longnecker (1992).

```
R> legrady <- subset(coffee_mort, id == 1)
R> covar.logrr(cases = cases, n = n, y = logrr, v = se^2, type = type,
+             data = legrady)
```

```
          [,1]        [,2]        [,3]
[1,] 0.01935402 0.01279718 0.01259433
[2,] 0.01279718 0.01885673 0.01254856
[3,] 0.01259433 0.01254856 0.01963947
```

The alternative method by Hamling *et al.* (2008) can be used by specifying `covariance = "h"` in the `covar.logrr` function. The reconstructed covariance matrix is used to efficiently estimate the dose–response association. For example, a linear trend $y_{1j} = \beta_1(x_{1j} - x_{10}) + \varepsilon_{1j}$ can be estimated with

```
R> lin_le <- dosresmeta(logrr ~ dose, se = se, type = type, cases = cases, n = n,
+                   data = legrady)
R> lin_le
Call:  dosresmeta(formula = logrr ~ dose, type = type, cases = cases,
    n = n, data = legrady, se = se)

Fixed-effects coefficients:
  dose
0.0328

1 study 3 values, 1 fixed and 0 random-effects parameters
 logLik     AIC       BIC
-0.7914   3.5827    2.6813
```

The change in the log relative risk of all-cause mortality associated with a 1 cup/day increase in coffee consumption was 0.033. That is, the increment of 1 cup/day of coffee was associated with a 3.4% (exp(0.033) = 1.034) higher mortality risk. The `predict` method facilitates the computation for the predicted linear increase for any arbitrary amount of coffee consumption. For example, setting `delta = 3`

```
R> predict(lin_le, delta = 3, expo = TRUE)
 delta      pred     ci.lb     ci.ub
     3 1.103507 0.9744976 1.249596
```

In the study by Legrady *et al.* (1987), 3 cups/day increase in coffee consumption was associate with a 10% (95% CI 0.97, 1.25) higher mortality risk.

Alternative curves can be specified in the formula argument. For example, a quadratic trend $y_{1j} = \beta_1(x_{1j} - x_{10}) + \beta_2(x_{1j}^2 - x_{10}^2) + \varepsilon_{1j}$ can be estimated.

```
R> quadr_le <- dosresmeta(logrr ~ dose + I(dose^2), se = se, type = type,
+                   cases = cases, n = n, data = legrady)
R> quadr_le
Call:  dosresmeta(formula = logrr ~ dose + I(dose^2), type = type, cases = cases,
    n = n, data = legrady, se = se)

Fixed-effects coefficients:
    dose  I(dose^2)
```

```
  -0.2135      0.0334


1 study 3 values, 2 fixed and 0 random-effects parameters
 logLik      AIC       BIC
 3.5738   -3.1475   -4.9503
```

The coefficients of the quadratic model are not directly interpretable. The results can be instead presented in terms of predicted relative risks for selected values of coffee consumption.

```
R> predict(quadr_le, newdata = data.frame(dose = 0:6), expo = TRUE)
  dose I(dose^2)      pred      ci.lb      ci.ub
1    0          0 1.0000000 1.0000000 1.0000000
2    1          1 0.8352219 0.7209175 0.9676497
3    2          4 0.7458585 0.5796705 0.9596916
4    3          9 0.7121374 0.5191844 0.9768005
5    4         16 0.7269823 0.5166453 1.0229519
6    5         25 0.7934812 0.5680037 1.1084654
7    6         36 0.9259812 0.6805480 1.2599276
```

The quadratic model suggested a U-shaped inverse association, with the maximum risk reduction, 29 % (95% CI 0.52, 0.98), observed for 3 cups/day of coffee compared.

5.1.2 Multiple studies

The chosen curves can be estimated also for the other studies. One possibility is to use the **tidyverse** package (Wickham, 2017) in order to obtain the linear trends and corresponding variances.

```
R> library(tidyverse)
R> lin_i <- coffee_mort %>%
+   split(.$id) %>%
+   map(~ dosresmeta(logrr ~ dose, se = se, type = type,
+                    cases = cases, n = n, data = .x))
R> lin_bi <- map_dbl(lin_i, ~ coef(.x))
R> lin_vi <- map_dbl(lin_i, ~ vcov(.x))
R> head(cbind(bi = lin_bi, vi = lin_vi))
          bi           vi
1  0.03283124 4.470869e-04
2 -0.02360445 3.561441e-04
3 -0.01430817 5.833618e-05
4 -0.04777017 6.194806e-04
5 -0.04736154 1.219069e-03
6 -0.02027627 1.127937e-04
```

The mean linear trend can be calculated with standard packages for meta-analysis such as the **mvmeta** package (Gasparrini *et al.*, 2012).

```
R> mvmeta(lin_bi, lin_vi)
Call:  mvmeta(formula = lin_bi ~ 1, S = lin_vi)

Fixed-effects coefficients:
(Intercept)
    -0.0326

22 studies, 22 observations, 1 fixed and 1 random-effects parameters
  logLik       AIC       BIC
 36.0221  -68.0442  -65.9551
```

Alternatively, the two steps of a two-stage analysis (dose–response and pooling) are unified and simplified in the dosresmeta function. In a single call, the study-specific linear trends are estimated and combined with the results being stored in the lin object.

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

The summary method displays the measures and tests of interest.

```
R> summary(lin)
Call:  dosresmeta(formula = logrr ~ dose, id = id, type = type, cases = cases,
    n = n, data = coffee_mort, se = se)

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

Chi2 model: X2 = 41.8314 (df = 1), p-value = 0.0000

Fixed-effects coefficients
             Estimate  Std. Error         z  Pr(>|z|)  95%ci.lb  95%ci.ub
(Intercept)   -0.0326      0.0050   -6.4677    0.0000   -0.0424   -0.0227


Between-study random-effects (co)variance components
  Std. Dev
    0.0172

Univariate Cochran Q-test for residual heterogeneity:
Q = 77.0088 (df = 21), p-value = 0.0000
I-square statistic = 72.7%

22 studies, 22 values, 1 fixed and 1 random-effects parameters
  logLik       AIC       BIC
 36.0221  -68.0442  -65.9551
```

There was an inverse association between increasing levels of coffee consumption and all-cause mortality risk, with a mean relative risk of $\exp(-0.03) = 0.97$ for a 1 cup/day increase. Similarly to the single study analysis, the predict function returns the combined result for any amount of coffee consumption.

```
R> predict(lin, delta = 3, expo = T)
 delta      pred     ci.lb     ci.ub
     3 0.9069412 0.8804891 0.9341879
```

The linear trend appeared to be heterogeneous as indicated by both the Q test ($Q = 77$, p value
< 0.01) and $I^2 = 73\%$. A possible alternative for both reducing the observed heterogeneity
and relaxing the linearity assumption is to model the dose–response as a quadratic curve.

```
R> quadr <- dosresmeta(logrr ~ dose + I(dose^2), id = id, se = se, type = type,
+                      cases = cases, n = n, data = coffee_mort)
R> summary(quadr)
Call:  dosresmeta(formula = logrr ~ dose + I(dose^2), id = id, type = type,
    cases = cases, n = n, data = coffee_mort, se = se)

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

Chi2 model: X2 = 75.9675 (df = 2), p-value = 0.0000

Fixed-effects coefficients
                      Estimate  Std. Error        z  Pr(>|z|)  95%ci.lb  95%ci.ub
dose.(Intercept)       -0.0847      0.0138  -6.1315    0.0000   -0.1118   -0.0576
I(dose^2).(Intercept)   0.0095      0.0023   4.1751    0.0000    0.0050    0.0139


Between-study random-effects (co)variance components
          Std. Dev     Corr
dose        0.0491     dose
I(dose^2)   0.0081  -0.9811

Univariate Cochran Q-test for residual heterogeneity:
Q = 113.4826 (df = 42), p-value = 0.0000
I-square statistic = 63.0%

22 studies, 44 values, 2 fixed and 3 random-effects parameters
   logLik       AIC        BIC
  96.2340 -182.4680  -173.7796
```

The overall test $H_0 : \beta_1 = \beta_2 = 0$ (Chi2 model) indicated that the mortality risk significantly
varied according to coffee consumption levels ($\chi^2_2 = 76$, p value < 0.001). The quadratic
model reduces to the simpler linear trend analysis when $\beta_2 = 0$. Thus, the univariate test for
β_2 ($z = 4.18$, p value < 0.001) suggested that the all-cause mortality risk was related in a
non-linear fashion with coffee consumption. The heterogeneity in the study-specific regression
coefficients was reduced but its impact was still important, with the multivariate $I^2 = 63\%$.

The predicted log relative risks from the linear and quadratic analyses can be presented in a
graphical format using $x_{\text{ref}} = 0$ cups/day as referent (Figure 5.1).

```
R> xref <- 0
R> pred <- data.frame(dose = c(xref, seq(0, 8, .1))) %>%
+    predict(quadr, newdata = ., expo = T) %>%
+    cbind(lin = predict(lin, newdata = ., expo = T))
R> ggplot(pred, aes(dose, pred, ymin = ci.lb, ymax = ci.ub)) +
+    geom_line() + geom_ribbon(alpha = .1) +
+    geom_line(aes(y = lin.pred), linetype = "dashed") +
+    scale_y_continuous(trans = "log", breaks = scales::pretty_breaks()) +
+    labs(x = "Coffee consumption (cups/day)", y = "Relative Risk")
```
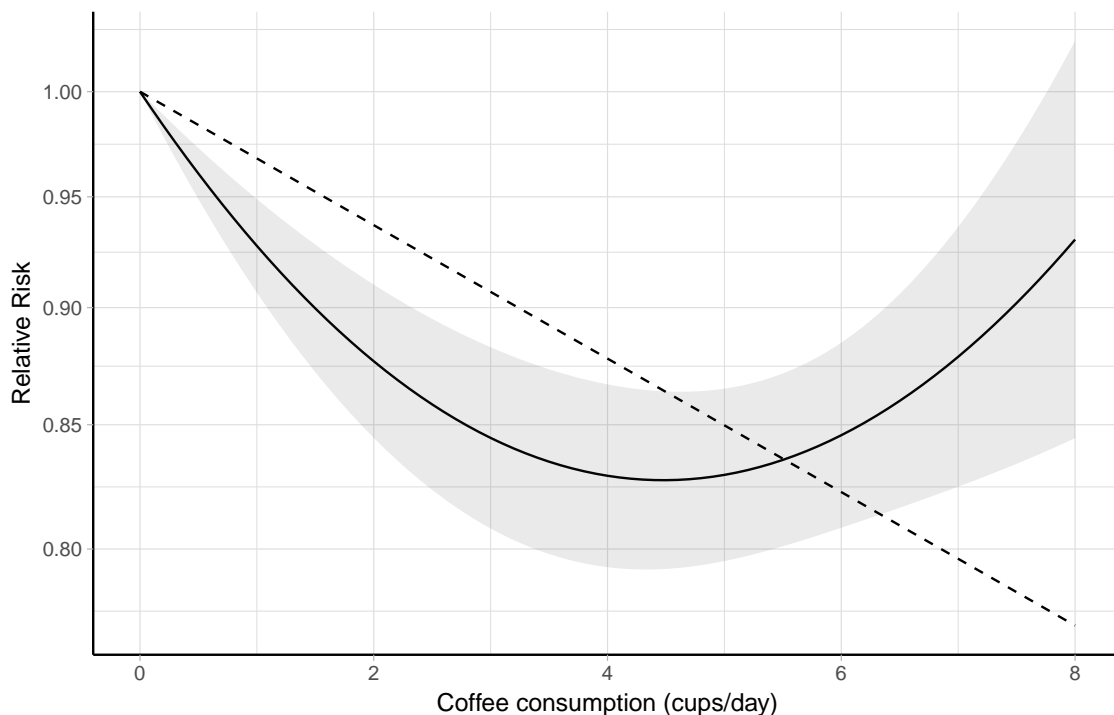

Figure 5.1: Combined dose–response association between coffee consumption and all-cause mortality (solid line) with 95% confidence intervals (shaded area). Coffee consumption was modelled with a quadratic curve in a two-stage random-effects meta-analysis. The dashed line represents the combined linear trend. The value 0 cups/day served as referent. The relative risks are plotted on the log scale.

Alternatively, the predicted relative risks for desired exposure values, say from 0 to 5 cups/day can be presented in a table.

```
R> filter(pred, dose %in% 0:5) %>%
+    select(-`I(dose^2)`, -lin.dose) %>%
+    unique() %>% round(3)
  dose  pred ci.lb ci.ub lin.pred lin.ci.lb lin.ci.ub
1    0 1.000 1.000 1.000    1.000     1.000     1.000
3    1 0.928 0.907 0.949    0.968     0.958     0.978
4    2 0.877 0.845 0.910    0.937     0.919     0.956
5    3 0.845 0.808 0.883    0.907     0.880     0.934
6    4 0.829 0.793 0.867    0.878     0.844     0.913
7    5 0.829 0.795 0.865    0.850     0.809     0.893
```

The previous predictions can be easily re-expressed using a different exposure value as referent, e.g. 1.5 cup/day, by changing `xref <- 1.5` and rerunning the previous code to produce a new figure or table.

The study-specific dose–response coefficients and related covariance matrices are stored in the results and can be easily accessed. For example, the $\hat{\boldsymbol{\beta}}_i$ and $\widehat{\mathrm{Var}}\left(\hat{\boldsymbol{\beta}}_i\right)$ for the quadratic model in the first two studies are

```
R> quadr$bi[1:2, ]
          dose      I(dose^2)
1 -0.213506072  0.033448197
2 -0.008546787 -0.001605319
R> quadr$Si[1:2]
[[1]]
               [,1]           [,2]
[1,]   0.0073978648 -0.0009437912
[2,]  -0.0009437912  0.0001281500


[[2]]
               [,1]           [,2]
[1,]   0.0043603384 -4.268928e-04
[2,]  -0.0004268928  4.551165e-05
```

and can be used to plot the individual curves in Figure B.1.

```
R> newd <- data.frame(dose = c(xref, seq(0, 6, .1)))
R> p_indiv <- cbind(newd, map(array_branch(quadr$bi, 1),
+                            ~ exp(.x[1]*newd$dose + .x[2]*newd$dose^2))) %>%
+   gather(study, pred, -dose) %>%
+   ggplot(aes(dose, pred, group = study)) + geom_line() +
+   scale_y_continuous("Relative Risk", trans = "log", breaks = c(.5, 1, 2, 5, 10),
+                   limits = c(.25, 10)) + labs(x = "Coffee consumption (cups/day)")
```

The previous analyses suggested that coffee consumption may be inversely associated with all-cause mortality in a non-linear fashion. The highest risk reduction, 17% (95% CI 0.79, 0.87), was observed for 4 cups/day of coffee consumption.

5.2 Paper II

We used the tools presented in Section 4.2 to evaluate the goodness-of-fit of the previous analyses. The `gof` function returns a display of the quantities of interest, namely the deviance test and the adjusted and unadjusted R^2. We started with the simpler linear trend estimated in a fixed-effect analysis.

```
R> gof(lin, fixed = TRUE)
Goodness-of-fit statistics:

Deviance test:
```

```
D = 225.244 (df = 78), p-value = 0.000
```

```
Coefficient of determination R-squared: 0.488
Adjusted R-squared: 0.482
```

The fit for the model assuming a linear relationship was poor (Analysis A in Table 5.1). In particular, the deviance test rejected the null hypothesis that the model was properly specified ($D = 225.2$, p value < 0.001) while the percentage of the accounted variation by the analysis was 49%. In addition, the decorrelated residuals vs exposure plot in panel A of Figure 5.2 indicated a specific pattern with negative residuals for low values of the exposure (before 5 cups/day) and positive ones for high levels of coffee consumption.

A possible solution for addressing the lack of fit of the previous analysis is to consider a non-linear example such as a quadratic curve (Analysis B). The fit slightly improved as indicated by R^2 which increased to 65% ($R^2_{adj} = 0.64$). The deviance test, however, still showed evidence of lack of fit ($D = 155.1$, p value < 0.001). Furthermore, even if the variability of the residuals reduced (panel B of Figure 5.2), the LOWESS smoother showed a tendency for the residuals to be more likely negative for low and very high exposure values.
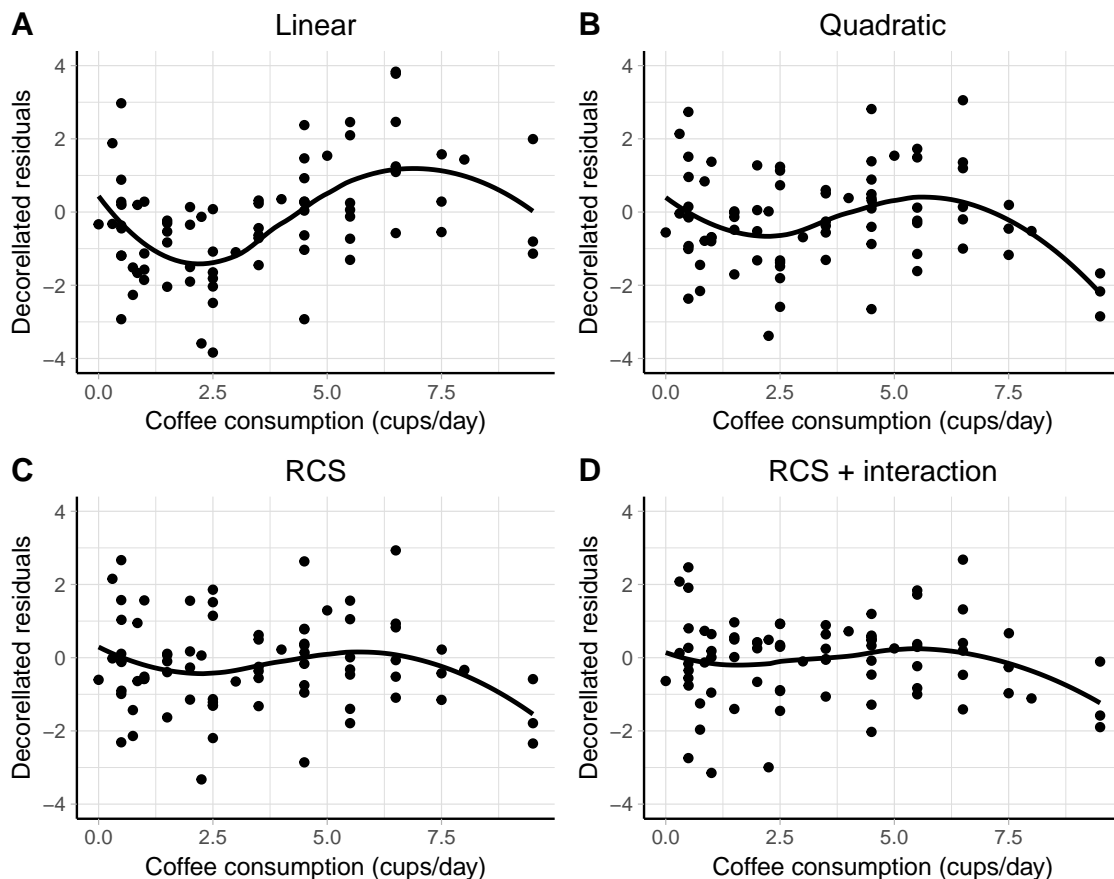

Figure 5.2: Decorrelated residuals versus exposure plots with LOWESS smother for different modelling strategies in a dose–response meta-analysis between coffee consumption and all-cause mortality (Crippa *et al.*, 2014).

An alternative strategy (Analysis C) was to model coffee consumption using RCS with 3 knots located at the fixed quantiles 0.10, 0.5, and 0.9 corresponding to 0, 2, 6.5 cups/day. Since the number of parameters is the same as in the Analysis B ($p = 2$), the R^2 can be used to compare the fit of the two strategies. In particular, the RCS analysis had a better fit, as indicated by the higher $R^2 = 0.68$ ($R^2_{adj} = 0.67$). The pattern in the residuals also leveled off around zero with the exception for the 3 highest dose levels. The deviance test, however, rejected again the null hypothesis for model specification, possibly because of the heterogeneity in the individual curves.

Coffee consumption varies substantially across countries both in terms of coffee powder, methods of preparation, and amount of cup size. In addition, the effect of coffee on all-cause mortality may have a different impact depending on the sex of participants. To address this variability, we employed meta-regression models to explain differences across the studies (Analysis D). We included two study-level covariates indicating the geographical area where the study was conducted (Europa, USA, and Japan) and the sex of the participants (only men, only women, and both sexes). Both the decorrelated vs exposure plot and $R^2_{adj} = 0.74$ indicated an improvement in the overall fit of the analysis even if the p value for the deviance test was below the nominal value ($D = 100.4$, p value $= 0.008$).

Table 5.1: Goodness-of-fit tests and measures for dose–response meta-analysis of coffee consumption and all-cause mortality (Crippa *et al.*, 2014)

Analysis	Model	Deviance	df	p value	R^2	R^2_{adj}
A	Linear	225.244	78	0.000	0.488	0.482
B	Quadratic	155.093	77	0.000	0.648	0.638
C	RCS [a]	141.332	77	0.000	0.679	0.671
D	RCS + interaction [b]	100.372	69	0.008	0.772	0.739

[a] 3 knots located at the 10th, 50th, and 90th percentiles of the distribution of coffee.
[b] As in c) + interaction with gender of participants (only men, only women, both sexes) and geographical area (Europe, USA, Japan) included as categorical study-level covariates.

5.3 Paper III

We illustrated how to employ the new measure of heterogeneity, $\hat{R}_b$, reanalyzing aggregated dose–response data on the association between meat and bladder cancer (Crippa *et al.*, 2016b). Five cohort and 8 case-control studies reported results for either red or processed meat and bladder cancer risk including a total of 10,271 cases and 1,066,027 participants. The data for both the associations are also included in the `dosresmeta` package.

5.3.1 Processed meat and bladder cancer

We started by considering a linear trend analysis for the association between processed meat and bladder cancer risk. The effect sizes for the meta-analysis in the second part of a two-stage dose–response meta-analysis were the linear trends for a 50 g per day increase in processed

meat consumption (Figure 5.3). The combined relative risk was 1.2 (95% CI 1.06, 1.37). The Q test detected statistical heterogeneity ($Q = 17.2$, p value $= 0.07$). The $\hat{R}_b = 38\%$ (95% CI 37, 40) indicated that the contribution of heterogeneity in the pooled analysis was limited (Table 5.2), as also suggested by the alternative measures $I^2 = 42\%$ (95% CI 40, 43) and $R_I = 43\%$ (95% CI 41, 44). Because the within-study variances were substantially homogenous, the differences between the alternative measures were negligible.

No evidence of non-linearity was observed using either a quadratic curve (p value $= 0.96$) or a RCS model with 3 knots located at fixed percentiles (p value $= 0.92$).

**processed meat and bladder cancer
for every 50 g per day increment**

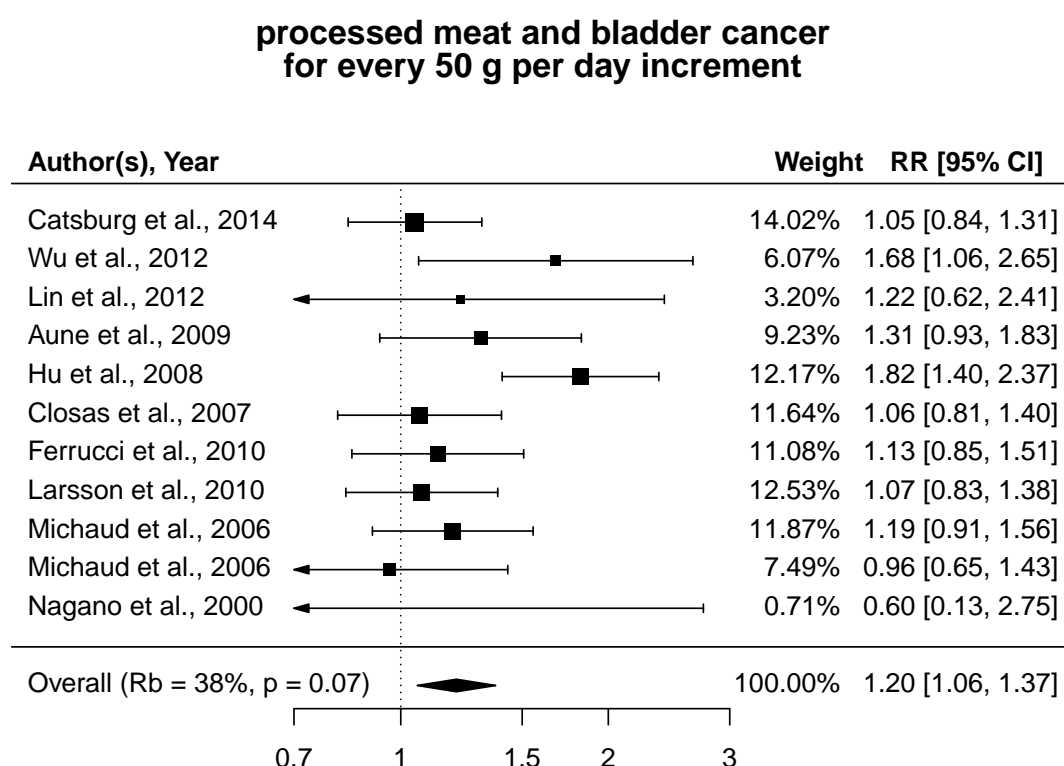| Author(s), Year	Weight	RR [95% CI]
Catsburg et al., 2014	14.02%	1.05 [0.84, 1.31]
Wu et al., 2012	6.07%	1.68 [1.06, 2.65]
Lin et al., 2012	3.20%	1.22 [0.62, 2.41]
Aune et al., 2009	9.23%	1.31 [0.93, 1.83]
Hu et al., 2008	12.17%	1.82 [1.40, 2.37]
Closas et al., 2007	11.64%	1.06 [0.81, 1.40]
Ferrucci et al., 2010	11.08%	1.13 [0.85, 1.51]
Larsson et al., 2010	12.53%	1.07 [0.83, 1.38]
Michaud et al., 2006	11.87%	1.19 [0.91, 1.56]
Michaud et al., 2006	7.49%	0.96 [0.65, 1.43]
Nagano et al., 2000	0.71%	0.60 [0.13, 2.75]
Overall (Rb = 38%, p = 0.07)	100.00%	1.20 [1.06, 1.37]

Figure 5.3: Relative risks of bladder cancer for every 50 g increase per day in processed meat consumption.

5.3.2 Red meat and bladder cancer

An analogous analysis was conducted for the association between red meat consumption and bladder cancer risk, with the estimated linear trends expressed for a 100 g per day increase (Crippa *et al.*, 2016b). The pooled relative risk was 1.22 (95% CI 1.05, 1.41). The study-specific associations in Figure 5.4, however, appeared to be heterogeneous ($Q = 60$, p value < 0.01). In particular, the contribution of the heterogeneity in determining the variance of the combined effect was $\hat{R}_b = 67\%$, indicating a moderate impact. Given that the distribution of within-error terms was higher as compared to the previous analysis (Figure B.2), the differences with the

alternative measures were higher: $I^2 = 80\%$ (95% CI 79, 81) and $R_I = 89\%$ (95% CI 88, 89), which suggested a larger heterogeneity across the study specific trends.

Similarly to the case of processed meat, no evidence of non-linearity was found (p value equal to 0.73 and 0.74 for β_2 in the quadratic and RCS model.)
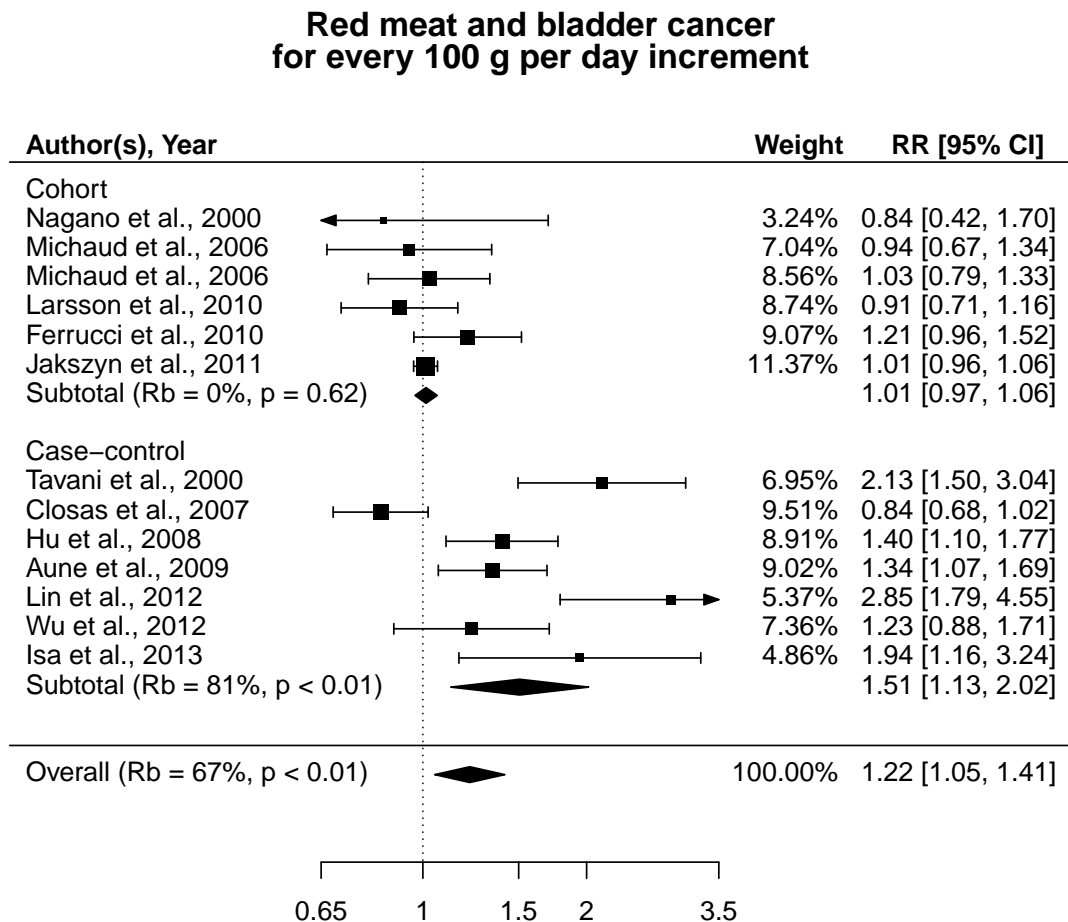

Figure 5.4: Relative risks of bladder cancer for every 100 g increase per day in red meat consumption, separately for cohort and case-control studies.

The variability across the linear trends could be related to differences in the study design. Figure 5.4 reports also the results separately for cohort and case-control studies. No association was observed in the subset of the prospective studies (combined RR $= 1.01$ (95% CI 0.97, 1.06)). The individual associations were homogenous ($Q = 3.5$, p value $= 0.62$), with $\hat{R}_b = 0\%$. The results for case-control studies, instead, largely varied across studies $Q = 39.8$, p value < 0.01 and $\hat{R}_b = 81\%$. A summary of the results for the alternative measures of heterogeneity in the different analyses is presented in Table 5.2.

Table 5.2: Measures of heterogeneity for dose–response meta-analysis between processed and red meat and bladder cancer risk (Crippa *et al.*, 2016b).

Analysis	$\hat{\beta}$ (95% CI)	Q test, p values	CV_{v_i}	$\hat{R}_b$ (95% CI)	I^2 (95% CI)	$\hat{R}_I$ (95% CI)
Processed meat	1.2 (1.06, 1.37)	17, 0.07	0.33	38 (37, 40)	42 (40, 43)	43 (41, 44)
Red meat	1.22 (1.05, 1.41)	60, < 0.01	5.94	67 (66, 68)	80 (79, 81)	89 (88, 89)
Red meat, Prospective	1.01 (0.97, 1.06)	4, 0.6	3.51	0 (0, 4)	0 (0, 8)	0 (0, 100)
Red meat, Case-control	1.51 (1.13, 2.02)	40, < 0.01	0.36	81 (80, 82)	85 (84, 86)	86 (85, 86)

5.4 Paper IV

The proposed and alternative measures of between-studies variability cannot capture the heterogeneity related to differences in the exposure range distribution. We presented the implementation of a point-wise strategy as a possible remedy, reanalyzing the data between red meat consumption and bladder cancer risk described in the previous section. The data consists of 13 independent studies where the exposure varied across studies not only in terms of definition and measurement but also in terms of range of assessment. For example, the minimum red meat consumption in one study (85.5 g per day in study ID 7) was higher then the maximum exposure value in other studies (71.7 g per day in study ID 3). Furthermore, only 3 out 13 studies reported relative risks for meat consumption greater than 150 grams/day. Descriptive measures of red meat consumption across studies are reported in Table C.1 and graphically presented in Figure B.3.

We decided to model the association using FP2. To allow for differential dose–response relationships, we fitted the FP2 models for different combination of power terms separately in each studies and selected the study-specific best fitting polynomials as the model with the lowest AIC. As expected, the best study-specific power terms varied across studies (Table C.2), with the most common polynomial defined by $p = (-2, -2)$ (8 out of 13 studies). Interestingly, the corresponding analysis using the AIC from the a one-stage model selected $p = (-1, -0.5)$ as the best FP2, a combination of power terms that was not preferred in any of the individual dose–response analyses. Figure 5.5 displays the predicted curves based on the best FP2 for each study, limited to the observed exposure range (blue solid lines). Extrapolation are represented by dashed lines while the red curves represent the predicted association from the one-stage model, which imposed the common power terms $p = (-1, -0.5)$ across the studies. While the alternative strategies gave similar results in the observed exposure range, larger differences were observed for high levels of red meat consumption. Taking study ID 5 as an example, the predicted relative risks for 150 g compared to 15 g of red meat consumption per day were 2.14 (95% CI 0.81, 5.67) and 1.13 (95% CI 0.84, 1.53) for the FP2 with the study specific $p_5 = (3, 3)$ and the common power terms $p = (-1, -0.5)$, respectively.
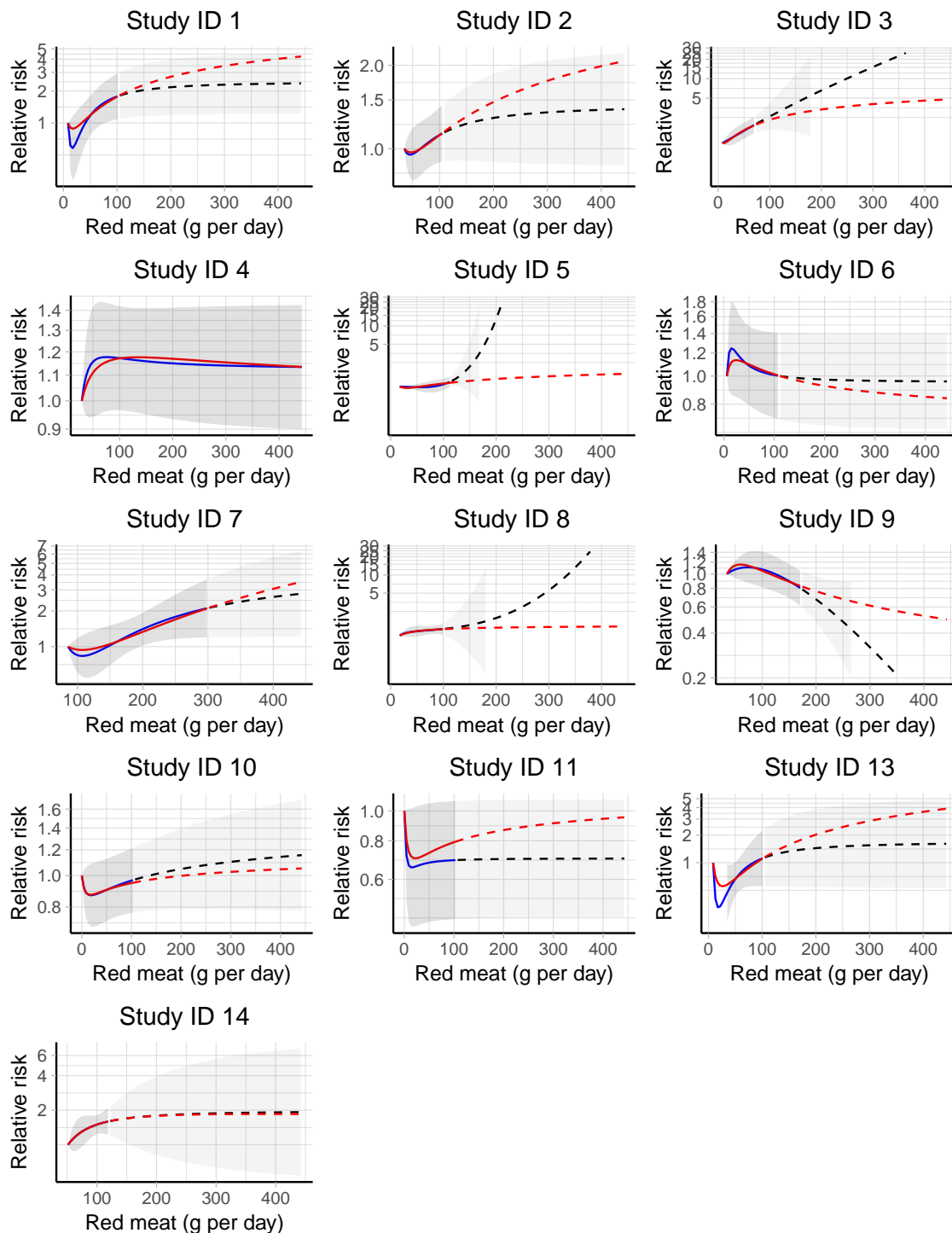
Figure 5.5: Study-specific dose–response curve between red meat consumption and bladder cancer risk. Blue and red solid lines are the predicted relative risks based, respectively, on the best fitting fractional polynomial (FP2) and on a common FP2 with p = (-1, -.5) limited to the observed exposure range. Dashed lines indicate extrapolated predicted relative risks. Confidence intervals for best fitting FP2 are represented by shaded areas (lighter colors correspond to extrapolations). All predicted relative risks are presented on the log scale using the study-specific reference values as comparators.

In a point-wise dose–response meta-analysis, we chose to predict the study specific log relative risks for a fine grid of exposure values ranging from 5 g to 300 g with step by 5 g, using 85 g/day (median reference dose value) as referent. The study-specific predictions were limited to the observed exposure range so that study ID 3, for example, predicted log RRs for dose values up to 70 g/days. Thus, the number of study-specific predicted log RRs varies across the selected red meat consumption levels.
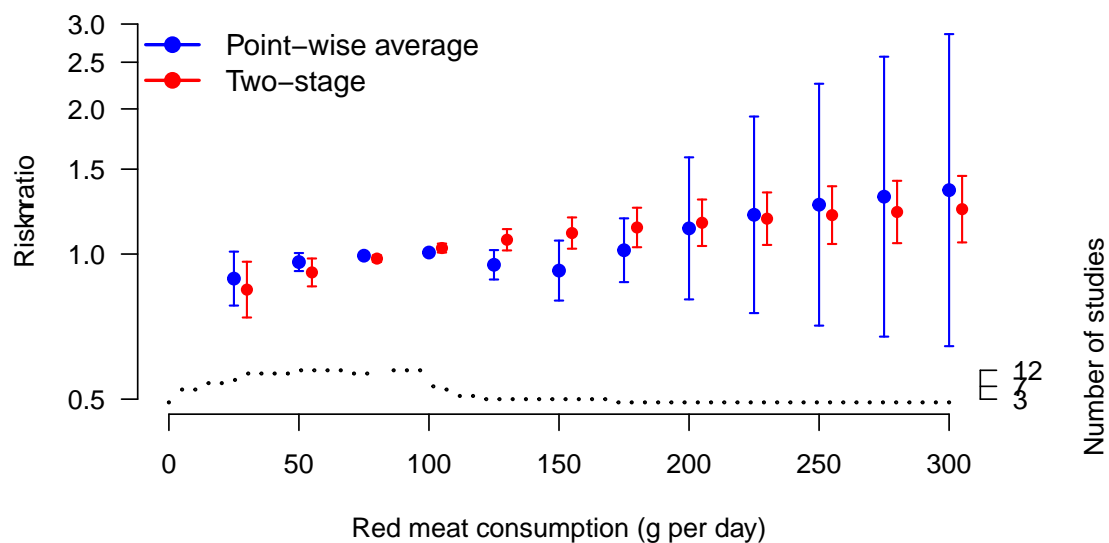

Figure 5.6: Comparison between pointwise and one-stage predicted relative risks for the association between red meat consumption (g per day) and bladder cancer risk. The step function at the bottom indicates the number of studies contributing to the prediction in the pointwise analysis. The relative risks are presented on the log scale using 85 g per day as referent.

The combined results from the 59 meta-analyses can be presented pointwisely. Figure 5.6 plots the combined predicted relative risk from a point-wise strategy and compares them with the corresponding one-stage analysis. In the reanalyzed data, the two strategies provided similar point estimates. However, the confidence intervals from the point-wise analysis were larger as compared to the corresponding one-stage model, reflecting the higher uncertainty in the predictions. For example, for 100 g the predicted RRs were 0.92 (95% CI 0.80, 1.07) for the point-wise strategy and 1.11 (95% CI 1.03, 1.19) for the one-stage model. Similarly, for 250 g the two predictions were 1.27 (95% CI 0.71, 2.26) and 1.20 (95% CI 1.05, 1.38).

One advantage of a point-wise strategy is that additional statistics from the meta-analytic models can be presented pointwisely (Figure 5.7). For example, the estimates for the between-study heterogeneity, $\hat{\tau}^2$, were lower than 0.02 for red meat consumption between 20 g and 150 g per day, and rapidly increased for higher values (panel A). The weights of the individual studies changed over the range of the exposure values. Panel B displays the standardized
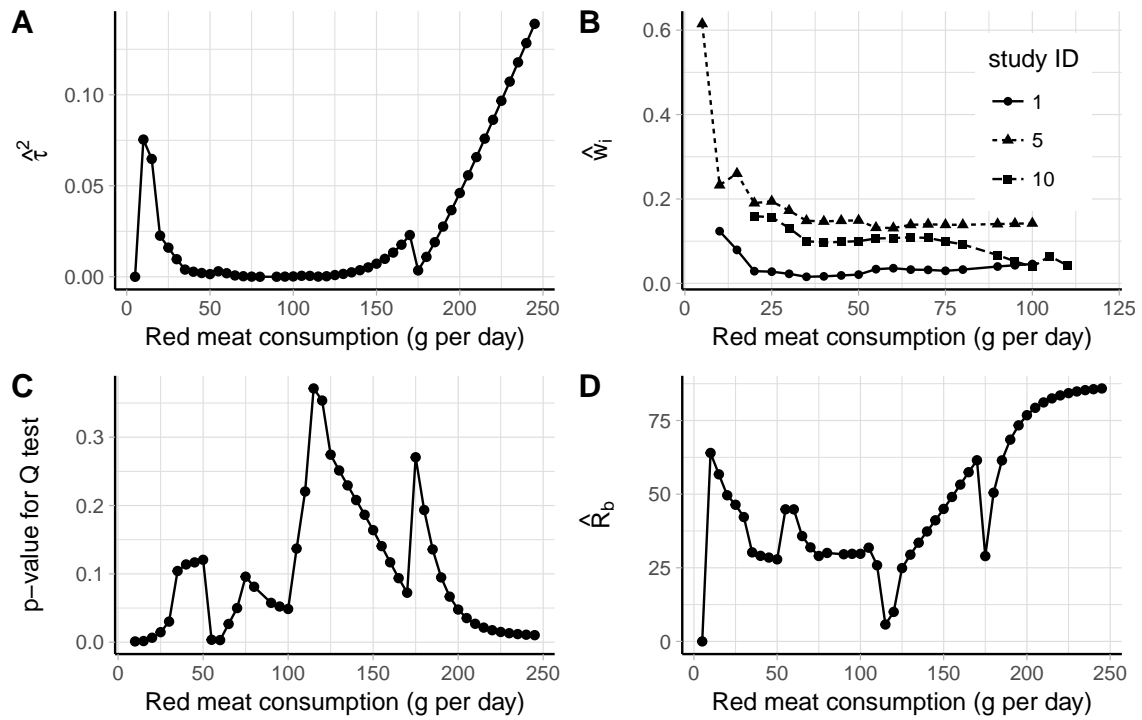
Figure 5.7: Point-wise results for a meta-analysis between red meat consumption (g per day) and bladder cancer risk: estimates of $\hat{\tau}^2$ (A), random-effects weights for three studies (B), p value for ther Q test (C), and $\hat{R}_b$ (D).

weights for three of the included studies. The percentage weight for study ID 5 was very large for red meat consumption lower than 30 g, while it stabilized around 15% after that. For study ID 10, the weight declined from 20% to 5% as the dose value increased, while it remained constant around 3% for study ID 1. The results from the Q test in panel C indicated presence of statistical heterogeneity for red meat consumption betwee 100 and 200 g. The impact of heterogeneity quantified by the $\hat{R}_b$ varied between 25% and 60% for exposure values below 75 g per day, while it reached 85% for higher dose levels.

5.5 Paper V

A common limitation of a two-stage and point-wise approach is that the dose–response analyses are limited by the low number of data points in each study, usually $J_i \leq 3$. We demonstrated an alternative one-stage model for dose–response meta-analysis using a subset of the data on coffee consumption and all-cause mortality presented in Sections 5.1 and 5.2. The data consists of the results for 12 independent populations including 5508 cases and 750959 participants (Crippa *et al.*, 2014). The results for two cohorts (Nilsson *et al.*, 2012) were excluded in the initial analysis of a non-linear trend because they only reported 1 non-referent risk. We performed a one-stage random-effects meta-analysis assuming a quadratic relationship on the whole data set and compared the results with the corresponding two-stage analysis which excluded the results by Nilsson *et al.* (2012).

Although both the analyses suggested a U-shaped association between increasing levels of coffee consumption and mortality risk, the predicted curve from a two-stage model provided lower relative risk estimates (Figure 5.8). Using 0 cups/day, the predicted RRs for 2 cups/day were 0.86 (95% CI 0.81, 0.91) for the two-stage analysis, and 0.89 (95% CI 0.83, 0.96) for the one-stage model. Similarly, for 5 cups/day the corresponding numbers were 0.78 (95% CI 0.72, 0.84) and 0.82 (95% CI 0.76, 0.89). Interestingly, the standard errors for the beta coefficients were lower in the two-stage analysis $\left(\mathrm{SE}\left(\hat{\beta}_1 \right) = 0.021, \mathrm{SE}\left(\hat{\beta}_2 \right) = 0.003 \right)$, even if the corresponding estimates from the one-stage analysis were based on a larger number of data points $\left(\mathrm{SE}\left(\hat{\beta}_1 \right) = 0.027, \mathrm{SE}\left(\hat{\beta}_2 \right) = 0.004 \right)$. As a consequence, the confidence intervals for the reported predicted relative risks were larger in the one-stage analysis, while the p value for non-linearity $H_0 : \beta_2 = 0$ was lower in a two-stage analysis: $p = 0.008$ compared to $p = 0.154$ from the alternative one-stage model.
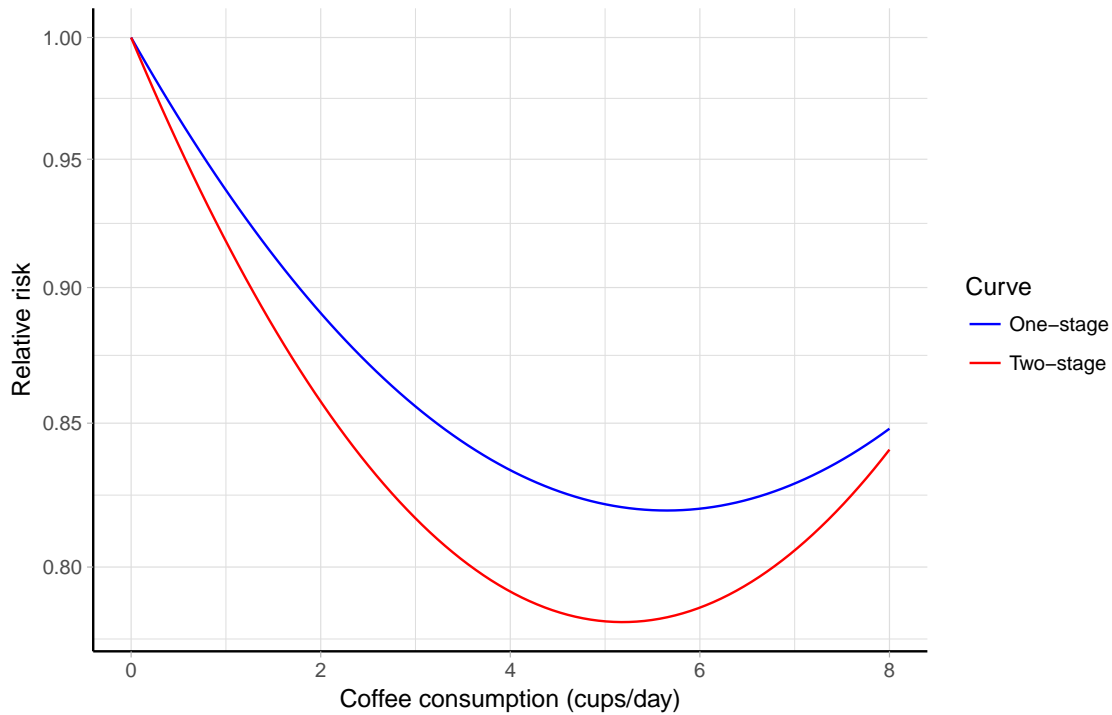

Figure 5.8: Combined quadratic association between coffee consumption (cups/day) and all-cause mortality estimated using a one-stage (blue line) and two-stage (red line) approach. The predicted relative risks are plotted on the log scale using 0 cups/day as referent.

The exclusion of 2 studies affected also the estimation of the variance components in $\mathbf{\Psi}$. For example, the variances of the random-effects based on 10 out of the 12 studies (two-stage) were 0.0013 for b_1 and 5×10^{-5} for b_2, while they were 0.00544 and 1.5×10^{-4}, respectively, in the alternative model. The $I^2 = 45\%$ from the multivariate two-stage analysis indicated a moderate impact of the heterogeneity. The same question can be addressed in a one-stage analysis by looking at the variance partition coefficients for the observed dose values. The VPC plot suggested a limited impact of heterogeneity between 20 and 40% for coffee consumption below 7 cups/day, while it become substantial for higher exposure values (Figure B.5).

Oftentimes it can be informative to provide a graphical presentation not only of the marginal association but also of the study-specific or conditional curves. The multivariate normal distribution for the random-effects is used to predict the study-specific regression coefficients, exploiting the information from the between-studies heterogeneity matrix (Table C.3). It is worthwhile to notice that the non-linear curves in the one-stage approach can be predicted also for those studies providing only one non-referent log RR (study ID 27 and 28), which were instead excluded in the traditional approach. The predicted conditional curves are graphically presented in Figure 5.9.

Different strategies can be chosen to model the dose–response association of interest. A one-stage approach, in particular, gives the opportunity to estimate a complex curve defined by multiple parameters in order to answer more elaborate research questions. A model with a spike at zero treats the low or unexposed participants as a separate group, and estimates the dose–response association only for the exposed groups. For example, we can think of the participants drinking a low amount coffee (less than 1 cup/day) as a separate group. The rest of the association can be described by two lines with a different slope before and after 4 cups/day. The mixed of linear splines model can be specified as

$$\mathbf{y}_i = (\beta_1 + b_{i1})(\mathbf{I}(\mathbf{x}_i < 1) - \mathbf{I}(x_{i0} < 1)) + (\beta_2 + b_{i2})((\mathbf{x}_i - 1)\mathbf{I}(\mathbf{x}_i \geq 1) - (x_{i0} - 1)\mathbf{I}(x_{i0} \geq 1)) +$$
$$+ (\beta_3 + b_{i3})((\mathbf{x}_i - 4)\mathbf{I}(\mathbf{x}_i \geq 4) - (x_{i0} - 4)\mathbf{I}(x_{i0} \geq 4)) + \epsilon_i$$

where the indicator function $\mathbf{I}$ takes on value 1 if the condition in the parenthesis is met, 0 otherwise. Another possibility consists of considering homogeneous groups of exposure in which the mortality risk can be considered constant. For instance, coffee consumption could be divided in intervals < 1, $[1, 3)$, $[3, 5)$, $[5, 7)$, and > 7 cups/day. The dose–response model is specified by including 4 dummy variables (x_1, x_2, x_3, and x_4) and using one level (e.g. $[1, 3)$) as referent

$$\mathbf{y}_i = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})(\mathbf{x}_{1i} - x_{1i0}) + (\beta_2 + b_{i2})(\mathbf{x}_{2i} - x_{2i0})(\beta_3 + b_{i3})(\mathbf{x}_{3i} - x_{3i0}) +$$
$$+ (\beta_4 + b_{i4})(\mathbf{x}_{4i} - x_{4i0}) + \epsilon_i$$

The two models are parameterized, respectively, by $p = 3$ and $p = 4$ parameters. It is unlikely to have enough data points (at least 4 non referent log RRs) for estimating the previous models in a two-stage analysis. Using a one-stage approach, instead, the predicted curve from those alternative analyses are presented graphically in Figure B.4 together with the quadratic model described before. In the spike at 0 model, the predicted mortality risk for the unexposed participants was 14% higher (95% CI 1.02, 1.28) as compared to participants drinking 1 cup/day. Every one cup per day increase in coffee consumption was associated with 2% (95% CI 0.95, 1.01) reduction in all-risk mortality, whereas the same number was less pronounced ($\beta_3 = 0.01$) for coffee consumption higher than 4 cups/day, with a decreased risk of 1% (95% CI 0.96, 1.02). The categorical model provided similar results with a 16% (95% CI 1.04, 1.28) higher mortality risk for never drinkers compared to the category $[1, 3)$ of coffee consumption.
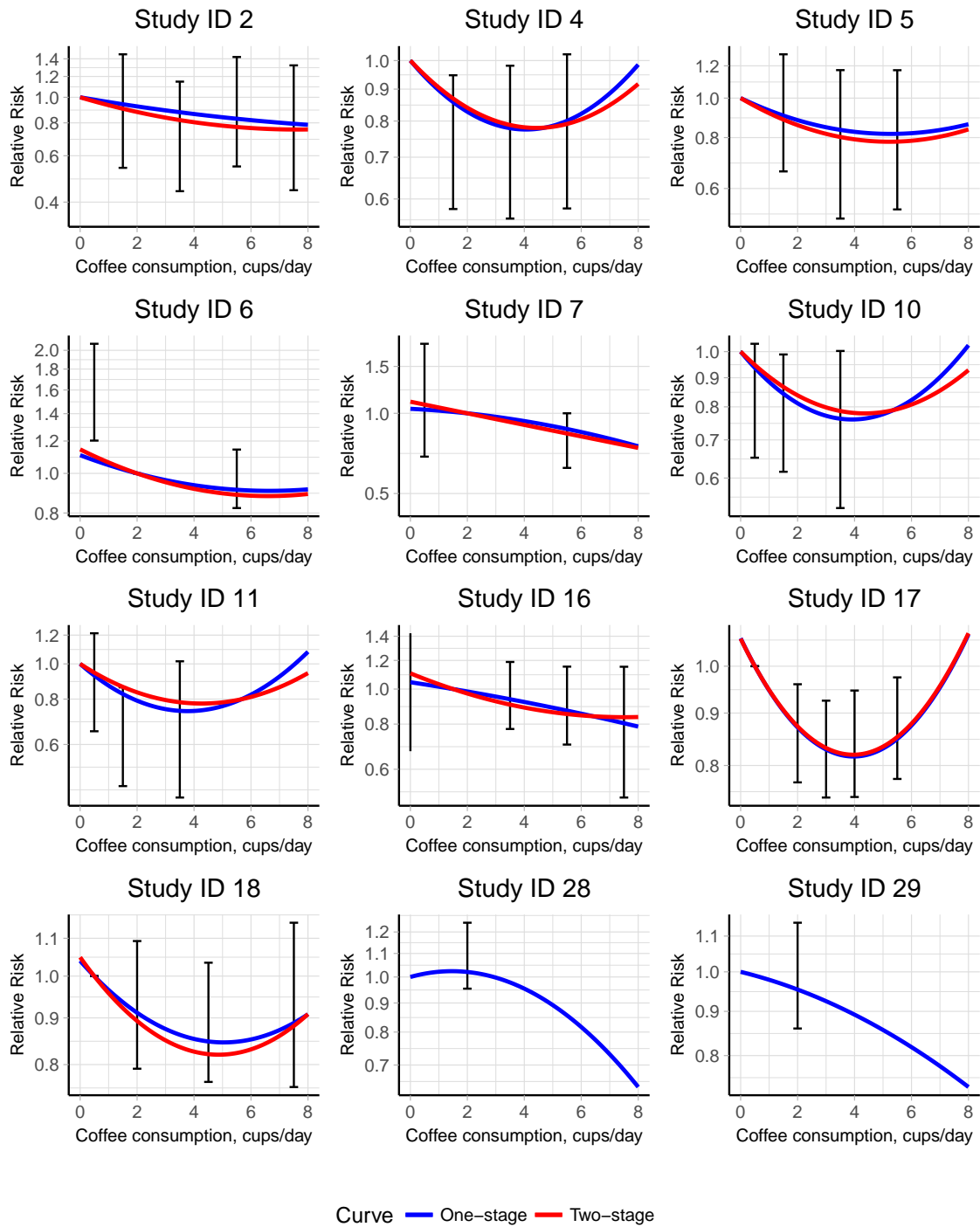
Figure 5.9: Conditional predicted quadratic curves for the association between coffee consumption and all-cause mortality in a one-stage (blue lines) and two-stage (red lines) approach. The relative risks are presented on the log scale using the study-specific reference categories as comparators.

Categories for coffee consumption greater or equal to 3 cup/day indicated a decreased mortality risk. We can evaluate if this further decline is significant by testing $H_0 : \beta_2 = \beta_3 = \beta_4 = 0$. The multivariate Wald test did not provide enough evidence to reject the null hypothesis ($\chi_3^2 =$

5.08, p value $= 0.17$). The fit of the alternative analyses can be compared by looking at the AIC. We selected the quadratic curve as the best fitting model since it had the lowest AIC (-29), as compared to the spike at 0 (-28) and the categorical model (-11).

After choosing the quadratic curve as the basis for modelling the dose–response association, we might try to relate the heterogeneity observed in the VPC plot to study-level covariates. For example, we can investigate if the sex of the study participants (3 levels: only men, only women, both sexes) substantially alter the combined quadratic association or partially explain the observed heterogeneity. This can be done by fitting a meta-regression model which includes in the fixed-effect matrix the interactions between the quadratic transformations (x and x^2) and the two dummy variables (u_1 and u_2) for the sex of the participants

$$
\begin{aligned}
\mathbf{y}_i = (\beta_1 + b_{i1})(\mathbf{x}_i - x_{i0}) + (\beta_2 + b_{i2})(\mathbf{x}_i^2 - x_{i0}^2) + \\
+ \beta_3(\mathbf{x}_i - x_{i0}) \times \mathbf{u}_{1i} + \beta_4(\mathbf{x}_i^2 - x_{i0}^2) \times \mathbf{u}_{1i} + \\
+ \beta_5(\mathbf{x}_i - x_{i0}) \times \mathbf{u}_{2i} + \beta_6(\mathbf{x}_i^2 - x_{i0}^2) \times \mathbf{u}_{2i} + \boldsymbol{\epsilon}_i
\end{aligned}
$$

The meta-regression model simplifies to the quadratic model when the coefficients for the interaction terms are equal to 0. A test for $H_0 : \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$ is equivalent to test if the dose–response associations differ according to sex of the participants. The multivariate Wald test ($\chi_3^2 = 2.4$, p value $= 0.66$) did not reject the null hypothesis. Furthermore, the meta-regression model was not able to explain the observed residual heterogeneity (Figure B.5).

Chapter 6

Discussion

We have proposed and developed new strategies and ad-hoc measures for dose–response meta-analysis, including tools for evaluating the goodness-of-fit, a new measure for quantifying the impact of heterogeneity, a strategy to deal with differences in the exposure range across studies, and a one-stage approach to estimate complex models without excluding relevant studies. The developed methodologies have been implemented in user-friendly R packages freely available on CRAN. Several codes for reproducing the results of this thesis and of the corresponding papers can be found on my website `https://alecri.github.io/software` and on GitHub at `https://github.com/alecri`.

6.1 Goodness-of-fit

An evaluation of the goodness-of-fit should be a natural step in a dose–response meta-analysis. In Paper II we discussed the relevant issue of how to evaluate the goodness-of-fit in a dose–response meta-analysis. Flexible parametric curves are estimated in order to summarize and represent the aggregated data in a synthetic format. It is important to check if the fitted meta-analytical model actually provides an adequate description of the data at hand.

The evaluation of the goodness-of-fit is usually carried out in practice by measuring the degree of agreement between the fitted and observed data. We have presented and discussed three tools (deviance, coefficients of determination, and decorrelated residuals versus exposure plot) specifically designed for assessing the goodness-of-fit in meta-analysis of aggregated dose–response data. In particular, the deviance can be employed for testing if the chosen meta-analytic model is properly specified, while the R^2 can be useful for quantifying from a descriptive point of view the proportion of variability accounted by the dose–response model. The fit of the dose–response analysis can be visually checked by inspecting the scatter plot of the decorrelated residuals versus the quantitative exposure.

The practical examples in Paper II and Section 5.2 illustrated the use of the proposed tools in evaluating the fit of the candidate dose–response models. In particular, we have shown how they can be useful for identifying specific dose–response patterns, investigating possible sources of heterogeneity, and generally evaluating if the combined dose–response association

can be an adequate summary of the observed data. Implementation of the proposed tools in applied works can strengthen the results or, on the contrary, raise doubts about the ability of the selected model in summarizing the available evidence.

As in the general case for the use of summary measures, one should be aware of the possible limitations of the developed tools. We have already seen that while a small p value for the deviance test for model specification is an indication that the posited model failed in accounting for the observed variation in the log relative risks, a large p value can not be interpreted as evidence that the model adequately explains the observed variability. In addition, a test based approach is generally unsatisfactory because it does not provide information about the actual fit of the analysis and suffers from low power due to the typically small number of data points in meta-analyses. Lastly, the p values for the global test of goodness-of-fit are not valid when the meta-analytical dose–response models are estimated driven by the observed data.

Possible explanations for a low value of the R^2 may be multiple. In fact, an R^2 close to zero may indicate that the selected model poorly fits the data, but also that there is no association between the quantitative exposure and the relative risk for the health outcome, or again that the model is correctly specified but the residual variability is still close to the overall variability. Finally, the visual inspection of the goodness-of-fit can reveal dose–response patterns in the modeled data but its judgment can be quite subjective. In case of sparse data, almost any patterns can be detected in the decorrelated residuals-versus-exposure plot.

More generally, the tools have been presented in a fixed-effect framework. The decorrelated residuals-versus-exposure plot can be directly extended to the case of a random-effects analysis by including the covariance matrix of the random-effects in the Cholesky decomposition. The other two measures do not have an explicit extension. Their usage as diagnostic tools, however, should be independent from the inclusion of the random-effects in the final model.

6.2 A new measure of heterogeneity

Another relevant aspect in a quantitative review, which is also related the assessment of goodness-of-fit, is the evaluation of the impact of heterogeneity. Indeed, a high variability in the reported effect sizes may undermine the appropriateness of presenting the combined effect as a summary measure. The common measures of heterogeneity have been developed under the unrealistic assumption of constant error variances. In Paper III we have proposed a new measure of heterogeneity, $\hat{R}_b$, that overcomes the limitation of the previous measures.

The $\hat{R}_b$ quantifies the impact of heterogeneity as the proportion of the variance of the combined effect due to the between-study variability. We have shown how $\hat{R}_b$ satisfies the properties required for a measure of heterogeneity without making any assumptions about the distribution of the within-study error terms. It can be expressed as the average of the study-specific intraclass correlation terms, i.e. the ratios of the τ^2 to the overall study-specific variance $\tau^2 + v_i$. Like I^2 and $\hat{R}_I$, the proposed measure tends to its upper limit 1 in case of meta-analysis of very precise estimates (small v_i). The between-study coefficient of variation can give additional information about the magnitude of heterogeneity compensating the shortcoming

of the available measures. The proposed measure of heterogeneity requires an estimate of τ^2, the between-study variability. Thus, confidence intervals should accompany the point estimates of $\hat{R}_b$ to reflect the uncertainty in the sample. We have proposed Wald type confidence intervals using the delta methods based on the relation between $\hat{R}_b$ and Q. The performances of the confidence intervals were tested throughout an extensive simulation study presented in Paper III.

We have shown how to present and interpret the new measure of heterogeneity by reanalyzing both univariate meta-analyses (in the illustrative examples of Paper III) and, more specifically, a dose–response meta-analysis (in Section 5.3). As expected, the $\hat{R}_b$ provided similar results as compared to both I^2 and $\hat{R}_I$ in case of effect sizes with homogeneous distribution for the within-error terms. On the contrary, differences were more evident as the variability of the v_i increased, with values of $\hat{R}_b$ generally lower than the corresponding I^2 and $\hat{R}_I$.

6.3 A point-wise approach

In Paper IV we have extended a point-wise approach originally presented for meta-analysis of individual patient data to the case of meta-analysis of aggregated dose–response data. The proposed strategy consists of combining the predicted log relative risks for a fine grid of exposure values arising from different study-specific dose–response analyses instead of combining the regression coefficients for a common dose–response model.

A point-wise approach has the potential advantage of improving the individual dose–response analyses since the study-specific models can be defined separately across the studies. Although the aim of a dose–response meta-analysis should be to estimate a common curve that uniformly fits the study-specific results, estimation of a single functional form may lower the fit of some individual analyses. We have illustrated in Section 4.4 and 5.4 the case of second degree fractional polynomials. In a two-stage approach, a single couple of power terms needs to be defined for all the studies so that the pooled dose–response curve can be derived by pooling the study-specific regression coefficients. In a point-wise approach, each study can choose a possibly different combination of power terms to better fit the observed data. In such a way, the predicted log relative risks will be closer to the observed ones. The combined curve can be then derived by pooling the individual predicted log relative risks pointwisely.

Another important advantage relates to the meta-analysis of heterogeneous exposure distributions where the quantitative exposure may differ not only in the definition and measurement but also in the range. The solution in a two-stage analysis could be to limit the prediction of the pooled curve to a subset of the observed exposure values. Depending of the extent of the diversity of the exposure ranges this might not be sufficient. We have illustrated this feature reanalyzing aggregated data on the association between milk consumption and all-cause mortality in the results of Paper IV and between red meat consumption and bladder cancer in Section 5.4. In the point-wise strategy, the predicted log relative risk can be limited to the observed exposure range. The combined curve is thus obtained by combining pointwisely a potential different number of log relative risks. Neglecting this type of heterogeneity may

have important consequences both in terms of point and interval estimates for the combined dose–response association. We have seen in Section 5.4 how the results based on a two-stage analysis may provide overconfident results for moderate to high values of red meat consumption, whereas a point-wise strategy limited the number of studies participating in the corresponding prediction and thus produced wider confident intervals, reflecting the uncertainty associated with the lower number of results. Finally, additional results from the univariate meta-analytic models can also be presented pointwisely, providing a richer description of the quantities of interest over the exposure range.

A possible limitation of the proposed approach is that the combined curve is obtained by means of separate univariate meta-analyses which are based on a set of common study-specifics analyses. As a consequence, the standard errors and confidence intervals may no longer be valid. A potential remedy would be to incorporate the covariance matrix for the study-specific predictions in the multivariate meta-analytic model. However, the number and the nature of the multivariate predictions are typically too high for the estimation algorithms to converge.

6.4 A one-stage model

In Paper V we have formalized and presented a one-stage model for meta-analysis of heterogeneous non-linear curves. The two steps of a two-stage approach, dose–response and pooling, can be written as a single procedure in terms of a linear mixed-effects model. The mixed-effects framework is particularly suitable for inferential procedures, marginal and conditional predictions, quantification of heterogeneity, goodness-of-fit and model comparison. The same questions frequently answered in a two-stage approach can be similarly addressed using a one-stage methodology.

The technique was initially presented in a fixed-effect analysis as a more flexible alternative of the two-stage methodology. Extensions to random-effects meta-analysis of non-linear curves have been typically framed into a two-stage framework because of the developments related to multivariate meta-analysis and for simplicity in the implementation using common statistical software. A one-stage model has oftentimes been regarded as equivalent. Even if we proved that a one-stage and two-stage approach give the same point estimates and inference, the one-stage methodology is more flexible and allows one to answer more elaborate research questions. Flexible curves can also be estimated based on the results from studies reporting a limited number of relative risks. In a two-stage meta-analysis, on the other hand, a typical requirement is that each study provides enough data for the individual dose–response analyses. For example, using either second order fractional polynomials or restricted cubic splines with 3 knots, $p = 2$ transformations are required for modeling non-linear associations. As a consequence, only studies providing at least 2 non-referent relative risks can be included in the non-linear analysis. The case where studies reported the results after dichotomizing the quantitative exposure are not rare. The data for these studies will be excluded in a two-stage meta-analysis. One important objective of a quantitative review, however, is to consider and analyze the whole body of evidence for a research question of interest. Systematic exclusion of studies because of

insufficient number of data points will necessarily discard useful information and thus provide only a partial summary. Furthermore, the assessment and investigation of between-studies variability will be also distorted, so that residual heterogeneity might be undetected.

Another advantage of a one-stage model is that many methodological aspects are greatly facilitated by using a single linear mixed-effects model. The tools presented in Paper II, for instance, were developed using the equivalence between the one- and two-stage approach in a fixed-effects analysis. The comparison of the fit in different dose–response analyses is also greatly facilitated by using information criteria such as the AIC, which are based on a common comparable likelihood.

Multiple routines implement linear mixed-effects models in different statistical packages. However, several aspects are specific to dose–response meta-analysis and it may be cumbersome to specify them using general commands for mixed-effects model. Therefore, we have implemented the one-stage methodology in the updated version of the **dosresmeta** package. Several example data sets and codes are available in order to facilitate applications of the proposed methodology.

Chapter 7

Conclusions

The methods presented in this thesis enrich the set of tools available for applying dose–response meta-analyses and for addressing specific questions, including how to evaluate the goodness-of-fit and how to measure the impact of the between-studies heterogeneity. Furthermore, this thesis describes alternative models for pooling results in case of heterogeneous exposure range and for estimating complex models without excluding relevant studies. The proposed methods have been illustrated using real data from published meta-analyses and implemented in user-friendly R packages available on CRAN.

More specifically we conclude the following:

- The `dosresmeta` R package has been widely used throughout the world and applied by practitioners in conducting dose–response meta-analyses. More recent developments are available to apply the methods presented in this thesis. Dedicated functions have been useful to avoid pitfalls frequently encountered in published meta-analyses, such as definition of the design matrix and prediction of the pooled results (Paper I).

- The proposed tools consist of descriptive measures to summarize the agreement between fitted and observed data (the deviance and the coefficient of determination), and graphical tools to visualize the fit of the model (decorrelated residuals-versus-exposure plot). These tools can be employed to identify systematic dose–response patterns and possible sources of heterogeneity, and to support the conclusions. Goodness-of-fit should be regularly evaluated in applied dose–response meta-analyses (Paper II).

- The new measure of heterogeneity, $\hat{R}_b$, quantifies the proportion of the variance of the pooled estimate attributable to the between-study heterogeneity. Contrary to the available measures of heterogeneity, it does not require specification of a typical value for these quantities. Therefore, we recommend the use of the $\hat{R}_b$ as a preferred measure for quantifying the impact of heterogeneity (Paper III).

- A point-wise strategy for dose–response meta-analysis does not require the specification of a unique model as in the traditional approaches, and therefore allows for more flexibility in modeling the individual curves. In addition, the extent of extrapolation is limited

by predicting the study-specific relative risk based on the observe exposure range. The use of the described strategy may improve the robustness of the results, especially in case of heterogeneous exposure range (Paper IV).

- The proposed one-stage approach for dose–response meta-analysis consists of a linear mixed-effects model, offering useful tools for describing the impact of heterogeneity over the exposure range, for comparing the fit of different models, and for predicting individual dose–response associations. The main advantage is that flexible curves can be estimated regardless of the number of data-points in the individual analyses (Paper V).

Chapter 8

Future research

Based on the conclusions presented in this thesis, future research includes:

- Implementing additive models as a smoother for dose–response meta-analysis. The non-parametric regression models can be use to investigate and identify the shape of the dose–response relationship. Formulae for estimation of additive models need to be extended to take into account the correlation of the error terms and the lack of intercept term.

- A limited set of tools is available for evaluating possible sources of bias for dose–response meta-analysis. In particular, a set of tools including descriptive measures, tests, and plots would be desirable for examining the likelihood of publication bias. Following this direction, a similar application of the trim and fill method could provide some aid in performing such a sensitivity analysis.

- Random-effects models for dose–response meta-analysis focus on estimating the population average risk-exposure association. Methods for evaluating the influence of specific data points and the effect of possible outliers are not available. A possibility could be to switch the focus from the mean to selected percentiles such as the median, which is generally less sensitive to extreme observations.

- Bayesian methods for dose–response meta-analysis have not yet been presented. A Bayesian perspective has the advantages of incorporating pertinent information that can be available from external sources. In addition, the uncertainty for all the parameters can be directly specified in the model. More generally, communication of the results can be enhanced by making probability statements about the quantities of interest.

- More generally, study selection is a frequent issue in meta-analyses of aggregated data. On the other hand, sharing of individual participant data is oftentimes difficult because of privacy agreements and costs involved in the data collection. A solution could be the implementation of a platform where practitioners are allowed to upload aggregated data without the need to have them published.

Appendix A

Restricted cubic splines

A Restricted Cubic Splines (RCS) model with 3 knots $\mathbf{k} = (k_1, \ldots, k_3)$ can be derived from a corresponding Cubic Splines (CS) model by forcing the curve to be linear at the extremes of the exposure distribution.

The CS model with 3 knots $\mathbf{k}$ is defined as

$$\text{CS}(x) = \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 (x - k_1)^3_+ + \beta_5 (x - k_2)^3_+ + \beta_6 (x - k_3)^3_+ \qquad \text{(A.1)}$$

where the '+' notation has been used ($u_+ = u$ if $u \geq 0$ and $u_+ = 0$ otherwise).

A RCS model restricts the CS function in equation A.1 to be linear before the first knot (k_1) and after the last knot (k_3). The first linearity constraint requires the model A.1 to be linear for $x \leq k_1$

$$\text{CS}(x) = \beta_1 x + \beta_2 x^2 + \beta_3 x^3$$

Hence, $\beta_2 = 0 \wedge \beta_3 = 0$.

The second linearity constraint requires the model A.1 to be linear for $x \geq k_3$

$$
\begin{aligned}
\text{CS}(x) =& \beta_1 x + \beta_4 \left(x^3 - 3x^2 k_1 + 3x k_1^2 - k_1^3 \right) + \beta_5 \left(x^3 - 3x^2 k_2 + 3x k_2^2 - k_2^3 \right) + \\
&+ \beta_6 \left(x^3 - 3x^2 k_3 + 3x k_3^2 - k_3^3 \right) = \\
=& -\left(\beta_4 k_1^3 + \beta_5 k_2^3 + \beta_6 k_3^3 \right) + \left(\beta_1 + 3\beta_4 k_1^2 + 3\beta_5 k_2^2 + 3\beta_6 k_3^2 \right) x + \\
&- 3 \left(\beta_4 k_1 + \beta_5 k_2 + \beta_6 k_3 \right) x^2 - \left(\beta_4 + \beta_5 + \beta_6 \right) x^3
\end{aligned}
$$

$$
\begin{cases}
\beta_4 k_1 + \beta_5 k_2 + \beta_6 k_3 = 0 \\
\beta_4 + \beta_5 + \beta_6 = 0
\end{cases}
\quad
\begin{cases}
\beta_4 k_1 + \beta_5 k_2 - \beta_4 k_3 - \beta_5 k_3 = 0 \\
\beta_6 = -\beta_4 - \beta_5
\end{cases}
\quad
\begin{cases}
\beta_5 = -\beta_4 \frac{k_3 - k_1}{k_3 - k_2} \\
\beta_6 = -\beta_4 + \beta_4 \frac{k_3 - k_2}{k_2 - k_1}
\end{cases}
$$

$$
\begin{cases}
\beta_5 = -\beta_4 \frac{k_3 - k_1}{k_3 - k_2} \\
\beta_6 = -\beta_4 + \beta_4 \frac{k_3 - k_1}{k_3 - k_2}
\end{cases}
\quad
\begin{cases}
\beta_5 = -\beta_4 \frac{k_3 - k_1}{k_3 - k_2} \\
\beta_6 = \beta_4 \frac{k_2 - k_1}{k_3 - k_2}
\end{cases}
\qquad \text{(A.2)}
$$

We can rewrite equation A.1 with $\beta_2 = 0 \wedge \beta_3 = 0$ and equations A.2

$$\text{RCS}(x) = \beta_1 x + \beta_4 \left[(x-k_1)_+^3 - \frac{k_3-k_1}{k_3-k_2}(x-k_2)_+^3 + \frac{k_2-k_1}{k_3-k_2}(x-k_3)_+^3 \right] \qquad \text{(A.3)}$$

that is a function of two variables: the quantitative exposure x and a transformation of x.
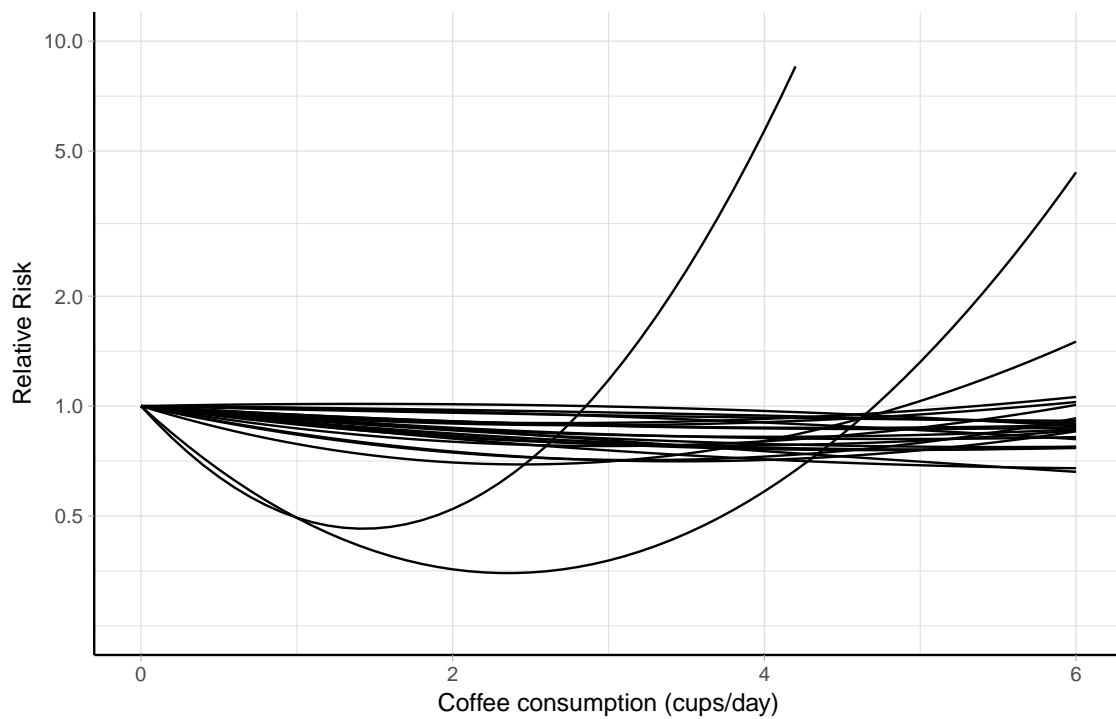
Appendix B

Supplementary figures

Figure B.1: Study-specific quadratic associations between coffee consumption and all-cause mortality. The relative risks are presented on a log scale using 0 cups/day as referent (Crippa *et al.*, 2016b).
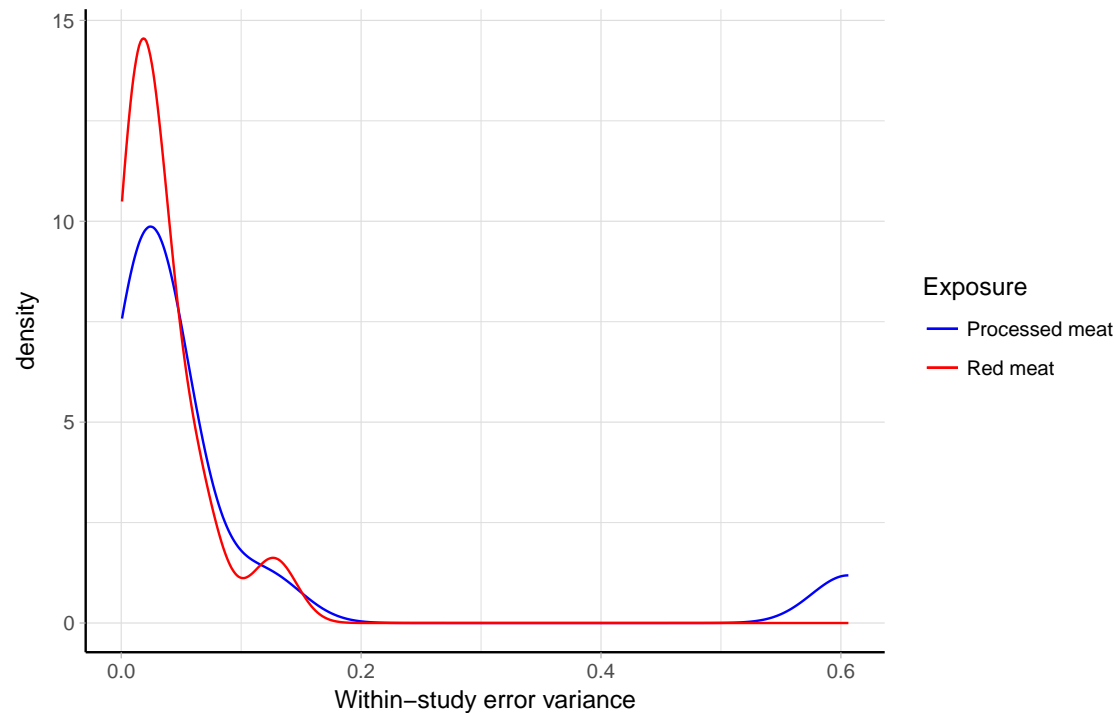
Figure B.2: Empirical distributions for within-error terms for the study-specific linear trend in a dose–response meta-analysis between processed and red meat and bladder cancer risk (Crippa *et al.*, 2016b).
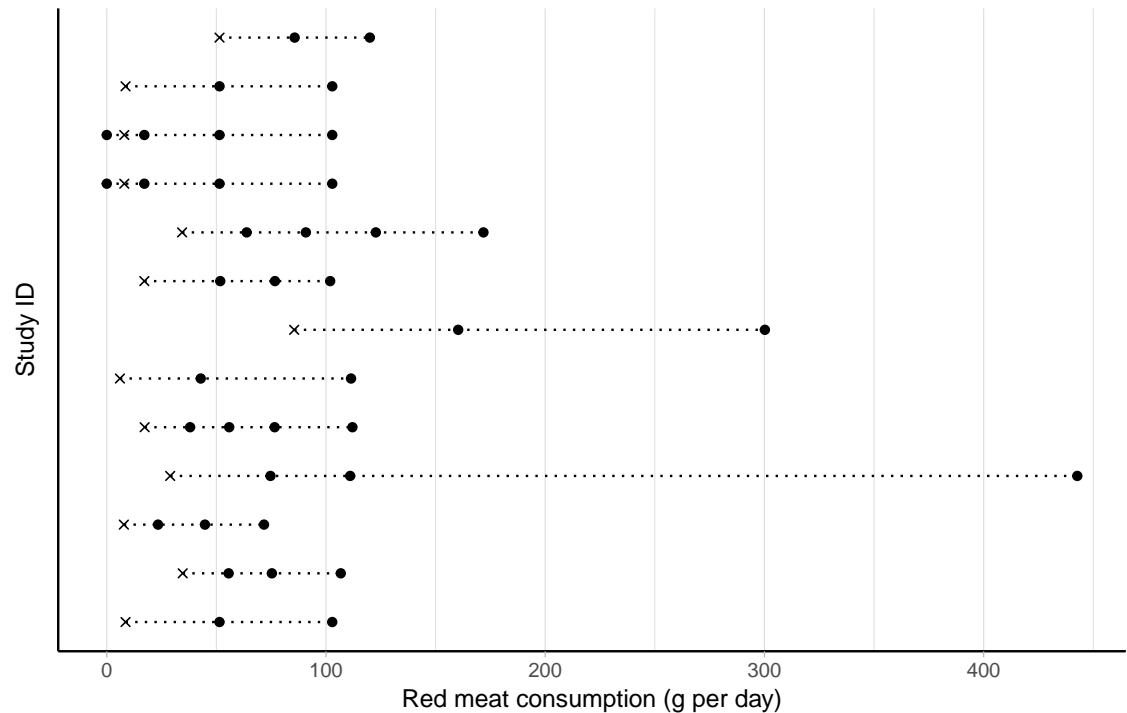

Figure B.3: Graphical visualization of the study-specific exposure distribution for 13 studies included in a dose–response meta-analysis between red meat consumption (g per day) and bladder cancer risk. The crosses and circles are, respectively, the referent and non-referent assigned doses of red meat consumption.
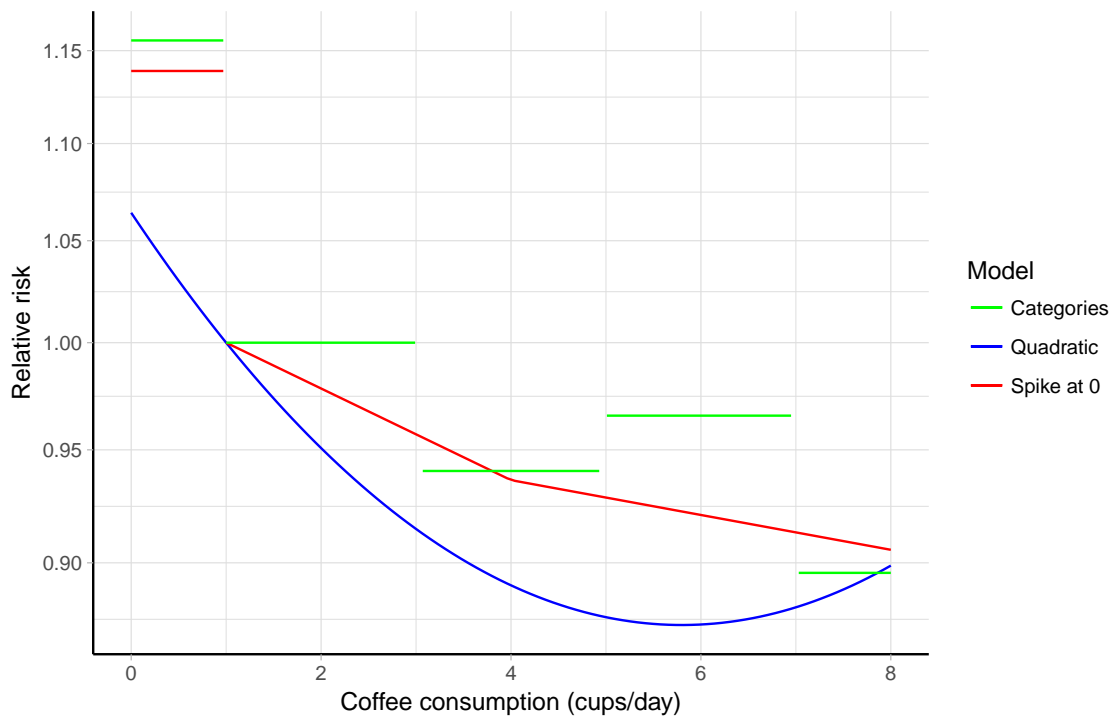
Figure B.4: Comparison of different strategies (quadratic, spike at 0, and categorical models) in a dose–response meta-analysis of coffee consumption (cups/day) and all-cause mortality. The relative risks are presented on the log scale using 1 cup/day as referent.

Figure B.5: Variance Partition Coefficient, VPC_{ij}, versus observed dose levels plot and LOWESS smoother for dose–response meta-analysis between coffee consumption (cusp/day) and all-cause mortality using a quadratic and meta-regression model.

Appendix C

Supplementary tables
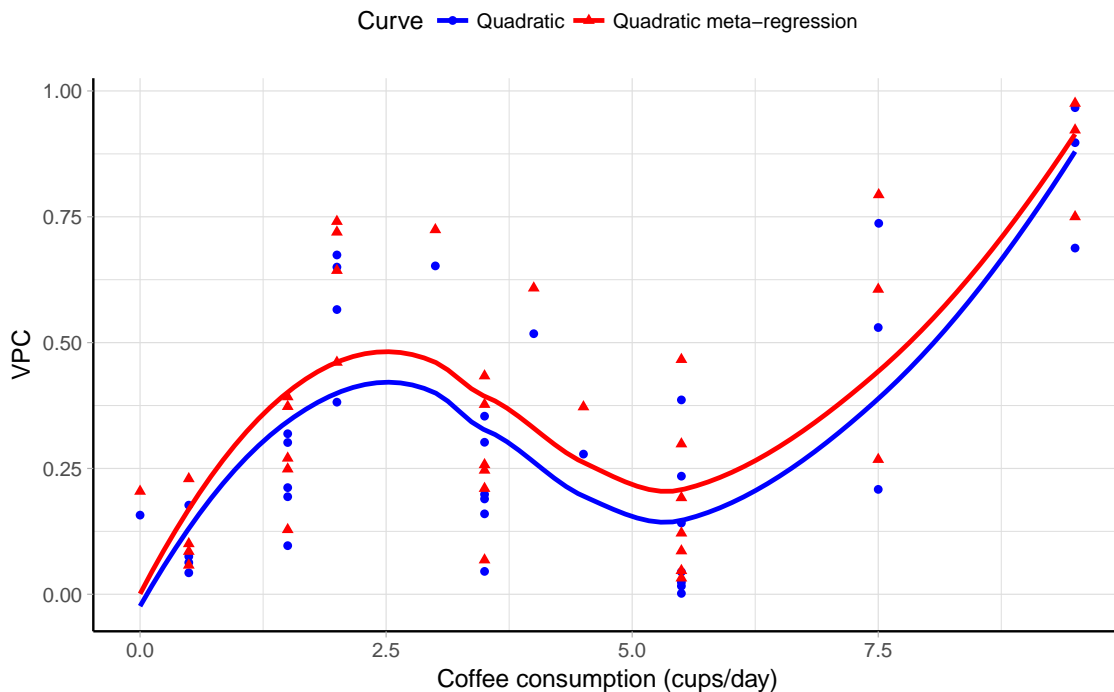
Table C.1: Descriptive statistics of the assigned dose levels for 13 studies included in a dose–response meta-analysis between red meat consumption (g per day) and bladder cancer risk.

ID	Referent	Min	P25	Median	P75	Max
1	8.6	8.6	30.0	51.4	77.1	102.9
2	34.6	34.6	50.3	65.5	83.2	106.7
3	7.8	7.8	19.5	34.1	51.5	71.7
4	28.9	28.9	63.2	92.8	193.9	442.7
5	17.3	17.3	38.0	55.8	76.6	112.1
6	6.0	6.0	24.4	42.9	77.1	111.4
7	85.5	85.5	122.9	160.3	230.2	300.2
8	17.1	17.1	43.1	64.2	83.0	101.9
9	34.4	34.4	63.8	90.8	122.7	171.8
10	8.0	0.0	8.0	17.1	51.4	102.9
11	8.0	0.0	8.0	17.1	51.4	102.9
13	8.6	8.6	30.0	51.4	77.1	102.9
14	51.4	51.4	68.6	85.7	102.9	120.0

Table C.2: AIC for the study-specific second-degree fractional polynomials with power terms specified by p in a dose–response meta-analysis between red meat consumption (g per day) and bladder cancer risk. The last row reports the power term corresponding to the lowest AIC.

p	Study ID												
	1	2	3	4	5	6	7	8	9	10	11	13	14
(-2, -2)	1.12	-3.66	-1.88	-4.04	-4.29	-0.71	1.76	-2.7	-3.43	-3.5	-1.8	2.38	-0.37
(-2, -1)	1.12	-3.66	-2.17	-3.98	-4.56	-0.71	1.76	-2.71	-3.8	-3.67	-1.43	2.38	-0.37
(-2, -0.5)	1.12	-3.65	-2.3	-3.94	-4.71	-0.71	1.76	-2.71	-3.97	-3.75	-1.2	2.38	-0.37
(-2, 0)	1.12	-3.65	-2.41	-3.91	-4.88	-0.71	1.76	-2.71	-4.14	-3.8	-0.95	2.38	-0.37
(-2, 0.5)	1.12	-3.64	-2.47	-3.88	-5.04	-0.71	1.76	-2.72	-4.28	-3.8	-0.7	2.38	-0.37
(-2, 1)	1.12	-3.62	-2.48	-3.85	-5.21	-0.71	1.76	-2.73	-4.39	-3.73	-0.47	2.38	-0.37
(-2, 2)	1.12	-3.6	-2.39	-3.83	-5.5	-0.71	1.76	-2.74	-4.51	-3.39	-0.14	2.38	-0.37
(-2, 3)	1.12	-3.56	-2.19	-3.82	-5.74	-0.71	1.76	-2.75	-4.52	-3.1	0	2.38	-0.37
(-1, -1)	1.12	-3.65	-2.33	-3.89	-4.77	-0.71	1.76	-2.71	-4.07	-3.87	-0.82	2.38	-0.37
(-1, -0.5)	1.12	-3.64	-2.39	-3.84	-4.9	-0.71	1.76	-2.71	-4.19	-3.94	-0.45	2.38	-0.37
(-1, 0)	1.12	-3.63	-2.45	-3.79	-5.04	-0.71	1.76	-2.71	-4.31	-3.95	-0.06	2.38	-0.37
(-1, 0.5)	1.12	-3.62	-2.48	-3.74	-5.18	-0.71	1.76	-2.71	-4.41	-3.85	0.32	2.38	-0.37
(-1, 1)	1.12	-3.61	-2.48	-3.7	-5.32	-0.71	1.76	-2.71	-4.47	-3.63	0.63	2.38	-0.37
(-1, 2)	1.12	-3.58	-2.43	-3.65	-5.57	-0.71	1.76	-2.72	-4.53	-3.01	1	2.38	-0.37
(-1, 3)	1.12	-3.56	-2.32	-3.63	-5.78	-0.71	1.76	-2.73	-4.5	-2.6	1.13	2.38	-0.37
(-0.5, -0.5)	1.12	-3.63	-2.43	-3.78	-5.01	-0.71	1.76	-2.71	-4.3	-3.98	-0.01	2.38	-0.37
(-0.5, 0)	1.12	-3.63	-2.46	-3.72	-5.13	-0.71	1.76	-2.71	-4.39	-3.92	0.44	2.38	-0.37
(-0.5, 0.5)	1.12	-3.61	-2.48	-3.66	-5.26	-0.71	1.76	-2.71	-4.46	-3.72	0.82	2.38	-0.37
(-0.5, 1)	1.12	-3.6	-2.48	-3.6	-5.38	-0.71	1.76	-2.71	-4.51	-3.4	1.11	2.38	-0.37
(-0.5, 2)	1.12	-3.58	-2.45	-3.53	-5.61	-0.71	1.76	-2.71	-4.53	-2.69	1.39	2.38	-0.37
(-0.5, 3)	1.12	-3.55	-2.38	-3.5	-5.81	-0.71	1.76	-2.71	-4.48	-2.32	1.46	2.38	-0.37
(0, 0)	1.12	-3.62	-2.48	-3.64	-5.24	-0.71	1.76	-2.7	-4.45	-3.76	0.89	2.38	-0.37
(0, 0.5)	1.12	-3.61	-2.48	-3.56	-5.35	-0.71	1.76	-2.7	-4.5	-3.45	1.23	2.38	-0.37
(0, 1)	1.12	-3.59	-2.48	-3.49	-5.46	-0.71	1.76	-2.7	-4.53	-3.05	1.45	2.38	-0.37
(0, 2)	1.12	-3.57	-2.47	-3.39	-5.67	-0.71	1.76	-2.69	-4.52	-2.39	1.61	2.38	-0.37

Table C.2: AIC for the study-specific second-degree fractional polynomials with power terms specified by p in a dose–response meta-analysis between red meat consumption (g per day) and bladder cancer risk. The last row reports the power term corresponding to the lowest AIC. *(continued)*

p	Study ID												
---	1	2	3	4	5	6	7	8	9	10	11	13	14
(0, 3)	1.12	-3.54	-2.43	-3.34	-5.84	-0.71	1.76	-2.69	-4.45	-2.15	1.63	2.38	-0.37
(0.5, 0.5)	1.12	-3.6	-2.48	-3.46	-5.44	-0.71	1.76	-2.7	-4.53	-3.07	1.49	2.38	-0.37
(0.5, 1)	1.12	-3.58	-2.48	-3.37	-5.54	-0.71	1.76	-2.69	-4.54	-2.67	1.63	2.38	-0.37
(0.5, 2)	1.12	-3.56	-2.48	-3.23	-5.72	-0.71	1.76	-2.67	-4.5	-2.21	1.7	2.38	-0.37
(0.5, 3)	1.12	-3.54	-2.47	-3.16	-5.88	-0.71	1.76	-2.66	-4.42	-2.14	1.7	2.38	-0.37
(1, 1)	1.12	-3.57	-2.48	-3.25	-5.62	-0.71	1.76	-2.68	-4.53	-2.37	1.7	2.38	-0.37
(1, 2)	1.12	-3.55	-2.48	-3.07	-5.78	-0.71	1.76	-2.65	-4.47	-2.17	1.73	2.38	-0.37
(1, 3)	1.12	-3.53	-2.48	-2.97	-5.92	-0.71	1.76	-2.63	-4.38	-2.24	1.73	2.38	-0.37
(2, 2)	1.12	-3.54	-2.48	-2.8	-5.9	-0.71	1.76	-2.6	-4.38	-2.33	1.74	2.38	-0.37
(2, 3)	1.12	-3.52	-2.46	-2.62	-6.01	-0.71	1.76	-2.55	-4.28	-2.54	1.73	2.38	-0.37
(3, 3)	1.12	-3.52	-2.4	-2.38	-6.09	-0.71	1.76	-2.46	-4.19	-2.78	1.73	2.38	-0.37
Best p	-2, -2	-2, -2	0.5, 0.5	-2, -2	3, 3	-2, -2	-2, -2	-2, 3	0.5, 1	-0.5, -0.5	-2, -2	-2, -2	-2, -2

Table C.3: Conditional predicted coefficients for quadratic curves in 12 studies on the association between coffee consumption (cups/day) and all-cause mortality based on a one-stage (os) and two-stage (ts) approach.

ID	$\hat{\beta}_{1os}$	$\hat{\beta}_{2os}$	$\hat{\beta}_{1ts}$	$\hat{\beta}_{2ts}$
2	-0.043	0.002	-0.074	0.005
4	-0.125	0.015	-0.112	0.013
5	-0.077	0.007	-0.094	0.009
6	-0.060	0.005	-0.080	0.006
7	-0.012	-0.003	-0.050	0.000
10	-0.140	0.018	-0.115	0.013
11	-0.159	0.021	-0.118	0.014
16	-0.028	-0.001	-0.074	0.005
17	-0.134	0.017	-0.132	0.017
18	-0.082	0.008	-0.101	0.010
28	0.032	-0.011		
29	-0.019	-0.002		

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.

Berkey C, Hoaglin D, Antczak-Bouckoms A, Mosteller F, Colditz G (1998). "Meta-analysis of multiple outcomes by regression with random effects." *Statistics in medicine*, **17**(22), 2537–2550.

Berlin JA, Longnecker MP, Greenland S (1993). "Meta-analysis of epidemiologic dose-response data." *Epidemiology*, **4**(3), 218–228.

Berrington A, Cox D (2003). "Generalized least squares for the synthesis of correlated information." *Biostatistics*, **4**(3), 423–431.

Biggerstaff B, Tweedie R (1997). "Incorporating variability in estimates of heterogeneity in the random effects model in meta-analysis." *Statistics in medicine*, **16**(7), 753–768.

Borenstein M, Hedges LV, Higgins J, Rothstein HR (2009). *References - Introduction to Meta-Analysis*. Wiley Online Library.

Borenstein M, Hedges LV, Higgins J, Rothstein HR (2010). "A basic introduction to fixed-effect and random-effects models for meta-analysis." *Research synthesis methods*, **1**(2), 97–111.

Buse A (1973). "Goodness of fit in generalized least squares estimation." *The American Statistician*, **27**(3), 106–108.

Cochran WG (1954). "The combination of estimates from different experiments." *Biometrics*, **10**(1), 101–129.

Colditz GA, Burdick E, Mosteller F (1995). "Heterogeneity in meta-analysis of data from epidemiologic studies: a commentary." *American journal of epidemiology*, **142**(4), 371–382.

Crippa A, Discacciati A, Bottai M, Spiegelman D, Orsini N (2018a). "One-stage dose–response meta-analysis for aggregated data." *Manuscript*.

Crippa A, Discacciati A, Larsson SC, Wolk A, Orsini N (2014). "Coffee consumption and mortality from all causes, cardiovascular disease, and cancer: a dose-response meta-analysis." *American journal of epidemiology*, **180**(8), 763–775.

Crippa A, Khudyakov P, Wang M, Orsini N, Spiegelman D (2016a). "A new measure of between-studies heterogeneity in meta-analysis." *Statistics in medicine*, **35**(21), 3661–3675.

Crippa A, Larsson SC, Discacciati A, Wolk A, Orsini N (2016b). "Red and processed meat consumption and risk of bladder cancer: a dose–response meta-analysis of epidemiological studies." *European journal of nutrition*, pp. 1–13.

Crippa A, Orsini N (2016a). "Dose-response meta-analysis of differences in means." *BMC medical research methodology*, **16**(1), 91.

Crippa A, Orsini N (2016b). "Multivariate dose-response meta-analysis: The dosresmeta R Package." *Journal of statistical software, Code Snippets*, **72**(1), 1–15. doi:10.18637/jss.v072.c01.

Crippa A, Thomas I, Orsini N (2018b). "A pointwise approach to dose-response meta-analysis of aggregated data." *Submitted*.

De Boor C, De Boor C, Mathématicien EU, De Boor C, De Boor C (1978). *A practical guide to splines*, volume 27. Springer-Verlag New York.

DerSimonian R, Laird N (1986). "Meta-analysis in clinical trials." *Controlled clinical trials*, **7**(3), 177–188.

Di Giuseppe D, Crippa A, Orsini N, Wolk A (2014). "Fish consumption and risk of rheumatoid arthritis: a dose-response meta-analysis." *Arthritis research & therapy*, **16**(5), 446.

Discacciati A, Crippa A, Orsini N (2017). "Goodness of fit tools for dose–response meta-analysis of binary outcomes." *Research Synthesis Methods*, **8**(2), 149.

Draper NR, Smith H (2014). *Applied regression analysis*. John Wiley & Sons.

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 M (2012). "Multivariate meta-analysis for non-linear and other multi-parameter associations." *Statistics in medicine*, **31**(29), 3821–3839.

Goldstein H, Browne W, Rasbash J (2002). "Partitioning variation in multilevel models." *Understanding Statistics: Statistical Issues in Psychology, Education, and the Social Sciences*, **1**(4), 223–231.

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.

Hagquist C, Stenbeck M (1998). "Goodness of fit in regression analysis–R2 and G2 reconsidered." *Quality and Quantity*, **32**(3), 229–245.

Haidich AB (2010). "Meta-analysis in medical research." *Hippokratia*, **14**(Suppl 1), 29.

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 F (2013). "Regression modeling strategies." *as implemented in R package 'rms' version*, **3**(3).

Harrell Jr FE (2015). *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. Springer.

Harville DA (1977). "Maximum likelihood approaches to variance component estimation and to related problems." *Journal of the American Statistical Association*, **72**(358), 320–338.

Hedges LV (1983). "A random effects model for effect sizes." *Psychological Bulletin*, **93**(2), 388.

Henderson CR, Kempthorne O, Searle SR, Von Krosigk C (1959). "The estimation of environmental and genetic trends from records subject to culling." *Biometrics*, **15**(2), 192–218.

Higgins J, Thompson SG (2002). "Quantifying heterogeneity in a meta-analysis." *Statistics in medicine*, **21**(11), 1539–1558.

Higgins J, Thompson SG, Spiegelhalter DJ (2009). "A re-evaluation of random-effects meta-analysis." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **172**(1), 137–159.

Higgins JP (2008). "Commentary: Heterogeneity in meta-analysis should be expected and appropriately quantified." *International journal of epidemiology*, **37**(5), 1158–1160.

Hill AB (1965). "The environment and disease: association or causation?"

Jackson D, Riley R, White IR (2011). "Multivariate meta-analysis: Potential and promise." *Statistics in Medicine*, **30**(20), 2481–2498.

Jackson D, White IR, Riley RD (2012). "Quantifying the impact of between-study heterogeneity in multivariate meta-analyses." *Statistics in medicine*, **31**(29), 3805–3820.

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.

Kvålseth TO (1985). "Cautionary note about R 2." *The American Statistician*, **39**(4), 279–285.

Langan D, Higgins J, Simmonds M (2017). "Comparative performance of heterogeneity variance estimators in meta-analysis: a review of simulation studies." *Research synthesis methods*, **8**(2), 181–198.

Larsson SC, Crippa A, Orsini N, Wolk A, Michaëlsson K (2015). "Milk consumption and mortality from all causes, cardiovascular disease, and cancer: a systematic review and meta-analysis." *Nutrients*, **7**(9), 7749–7763.

Larsson SC, Orsini N (2011). "Coffee consumption and risk of stroke: a dose-response meta-analysis of prospective studies." *American Journal of Epidemiology*, **174**(9), 993–1001.

Larsson SC, Orsini N, Wolk A (2006). "Milk, milk products and lactose intake and ovarian cancer risk: a meta-analysis of epidemiological studies." *International journal of cancer*, **118**(2), 431–441.

Legrady D, Dyer AR, Shekelle RB, Stamler J, Liu K, Paul O, Lepper M, Shryock AM (1987). "Coffee consumption and mortality in the Chicago Western Electric Company study." *American journal of epidemiology*, **126**(5), 803–812.

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.

Nilsson LM, Winkvist A, Brustad M, Jansson JH, Johansson I, Lenner P, Lindahl B, Van Guelpen B (2012). "A traditional Sami diet score as a determinant of mortality in a general northern Swedish population." *International journal of circumpolar health*, **71**(1), 18537.

Orsini N, Bellocco R, Greenland S, *et al.* (2006). "Generalized least squares for trend estimation of summarized dose-response data." *Stata Journal*, **6**(1), 40.

Orsini N, Greenland S, *et al.* (2011a). "A procedure to tabulate and plot results after flexible modeling of a quantitative covariate." *Stata Journal*, **11**(1), 1.

Orsini N, Li R, Wolk A, Khudyakov P, Spiegelman D (2011b). "Meta-analysis for linear and nonlinear dose-response relations: examples, an evaluation of approximations, and software." *American journal of epidemiology*, **175**(1), 66–73.

Pinheiro J, Bates D (2010). *Mixed-Effects Models in S and S-PLUS*. Springer Science & Business Media. ISBN 978-1-4419-0318-1.

Pogue JM, Yusuf S (1997). "Cumulating evidence from randomized trials: utilizing sequential monitoring boundaries for cumulative meta-analysis." *Controlled clinical trials*, **18**(6), 580–593.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL `https://www.R-project.org/`.

Rice K, Higgins J, Lumley T (2017). "A re-evaluation of fixed effect (s) meta-analysis." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*.

Ritz J, Demidenko E, Spiegelman D (2008). "Multivariate meta-analysis for data consortia, individual patient meta-analysis, and pooling projects." *Journal of Statistical Planning and Inference*, **138**(7), 1919–1933.

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.

Royston P (2000). "A strategy for modelling the effect of a continuous covariate in medicine and epidemiology." *Statistics in medicine*, **19**(14), 1831–1847.

Royston P (2001). "Flexible parametric alternatives to the Cox model, and more." *Stata J*, **1**(1), 1–28.

Royston P, Altman DG (1994). "Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling." *Applied statistics*, pp. 429–467.

Sauerbrei W, Royston P (2011). "A new strategy for meta-analysis of continuous covariates in observational studies." *Statistics in medicine*, **30**(28), 3341–3360.

Schmid CH, Lau J, McIntosh MW, Cappelleri JC (1998). "An empirical study of the effect of the control rate as a predictor of treatment efficacy in meta-analysis of clinical trials." *Statistics in medicine*, **17**(17), 1923–1942.

Shi JQ, Copas J (2004). "Meta-analysis for trend estimation." *Statistics in medicine*, **23**(1), 3–19.

Sidik K, Jonkman JN (2005). "Simple heterogeneity variance estimation for meta-analysis." *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **54**(2), 367–384.

Sutton AJ, Abrams KR (2001). "Bayesian methods in meta-analysis and evidence synthesis." *Statistical methods in medical research*, **10**(4), 277–303.

Sutton AJ, Cooper NJ, Jones DR, Lambert PC, Thompson JR, Abrams KR (2007). "Evidence-based sample size calculations based upon updated meta-analysis." *Statistics in medicine*, **26**(12), 2479–2500.

Sutton AJ, Higgins J (2008). "Recent developments in meta-analysis." *Statistics in medicine*, **27**(5), 625–650.

Takahashi K, Nakao H, Hattori S (2013). "Cubic spline regression of J-shaped dose-response curves with likelihood-based assignments of grouped exposure levels." *J. Biom. Biostat*, **4**, 1–6.

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.

Takkouche B, Cadarso-Suárez C, Spiegelman D (1999). "Evaluation of old and new tests of heterogeneity in epidemiologic meta-analysis." *American journal of epidemiology*, **150**(2), 206–215.

Theil H (1958). "Economic forecasts and policy."

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.

Verbeke G (1997). *Linear mixed models for longitudinal data*. Springer.

White IR, *et al.* (2011). "Multivariate random-effects meta-regression: updates to mvmeta." *Stata Journal*, **11**(2), 255.

Whitehead A, Whitehead J (1991). "A general parametric approach to the meta-analysis of randomized clinical trials." *Statistics in medicine*, **10**(11), 1665–1677.

Wickham H (2017). *tidyverse: Easily Install and Load the 'Tidyverse'*. R package version 1.2.1, URL `https://CRAN.R-project.org/package=tidyverse`.

Yu WW, Schmid CH, Lichtenstein AH, Lau J, Trikalinos TA (2013). "Empirical evaluation of meta-analytic approaches for nutrient and health outcome dose-response data." *Research synthesis methods*, **4**(3), 256–268.

Acknowledgements

I would like to thanks all the many people who directly or indirectly contributed to this thesis for their support and encouragement during these years.

First of all, **Nicola Orsini**, my main supervisor. Thanks for giving me the opportunity of undertaking this doctoral studies and for your exceptional guidance. With your genuine and continuous enthusiasm, you pushed me outside my comfort zone and motivated me to do my best. Your advice and quotes have always been appreciated and will guide me in the important decisions inside and outside the academia.

My co-supervisors **Alicja Wolk**, **Donna Spiegelman**, and **Matteo Bottai**. Alicja, you welcomed me in your unit and introduced me to the challenging word of nutritional epidemiology. Thanks for your valuable comments and for forcing me to improve my Swedish. Donna, you gave me the opportunity to visit the Harvard School of Public Health and work under your supervision. We had very nice discussions and I learned a lot from your extensive experience. Thanks Matteo for always taking time to meet me and answer my questions. You helped me to think critically and look at things from a different perspective.

All the co-authors of the studies for your important contributions and for starting rewarding collaborations. A particular thanks to **Susanna Larsson** for your quick and frequent help, and to **Rino Bellocco** for putting me in contact with Nicola and for involving me in many activities. Meeting you on the train has probably changed my life. A warm thank you to **Rosaria Galanti** for engaging me in different collaborations and for being my personal life coach, especially for medical assistance.

Andrea Discacciati and **Andrea Bellavia**, for the nice and constructive discussions but more importantly for making our daily working life entertaining. Thanks to **Xingwu Zhou**, **Filip Andersson**, and **Fabio Castagna** for their support and aid, in particular during our frequent Ping-Pong games. **Daniela Di Giuseppe**, for our profitable collaborations and coffee breaks. **Renee Gardner**, for being a wonderful co-teacher and for proofreading this thesis.

Viktor Oskarsson and **Eleonor Säfsten**, with whom I shared office, concerns and laughs, and pleasant extra activities. Thanks for being patient with the too many Italians around you. An extra warm thanks goes to **Carolina Donat Vargas**, **Constance Boissin**, **Emilia Riggi**, **Nada Hana**, and **Paolo Frumento**. All my present and past colleagues at the department of Public Health Sciences: **Antonio Ponce De Leon, Ashley Mcallister, Asli Kulane, Bo Burström, Diana Corman, Dominika Seblova, Erika Saliba Gustafsson, Jad Shedrawy, Janne Ager-**