Novel methods for dose–response meta-analysis

Alessio Crippa

Department of Public Health Sciences (PHS)
Karolinska Institutet

April 13, 2018

Acknowledgements

Main supervisor

- Nicola Orsini

Co-supevisor

- Alicja Wolk
- Matteo Bottai
- Donna Spiegelman

Co-authors

Opponent

- Christopher H. Schmid

Examination board

- Nele Brusselaers
- Antonio Gasparrini
- Paul Lambert

The audience

Dose–response meta–analysis

Summarize and contrast results on the relation between a quantitative exposure and the occurrence of a health outcome.

Dose–response meta–analysis

Summarize and contrast results on the relation between a quantitative exposure and the occurrence of a health outcome.

Research questions based on multiple studies:

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

Dose–response meta–analysis

Summarize and contrast results on the relation between a quantitative exposure and the occurrence of a health outcome.

Research questions based on multiple studies:

- 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?

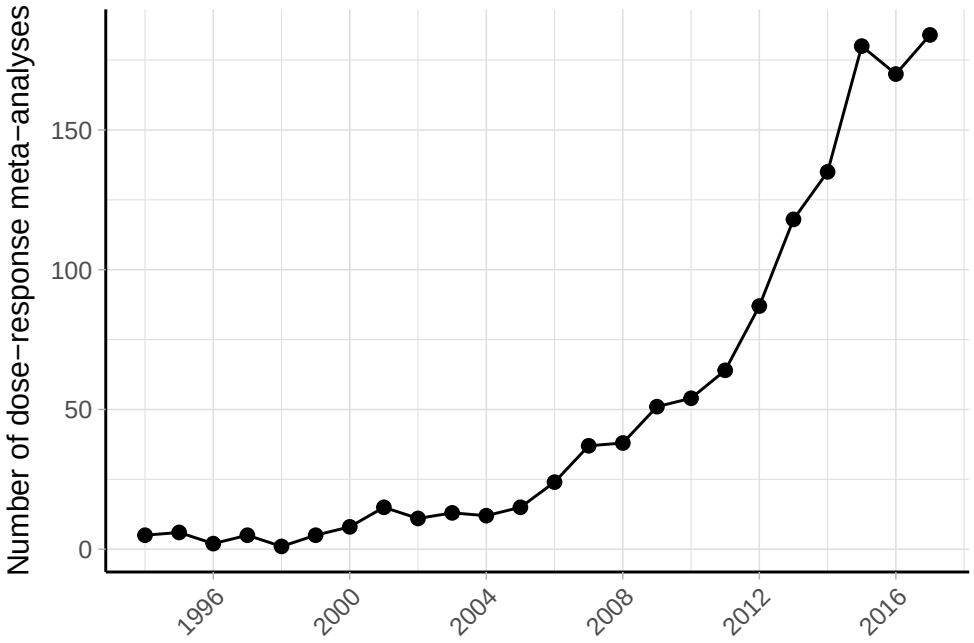
Dose–response meta–analysis

Summarize and contrast results on the relation between a quantitative exposure and the occurrence of a health outcome.

Research questions based on multiple studies:

- 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? Which factors can explain the observed heterogeneity?

Increasing number of dose–response meta-analyses



Data source: Google scholar

- Several research areas
- Many leading medical and epidemiological journals
- International health organizations and academic institutions
- Measures of public health impact

Aggregated dose–response data

An example from a prospective study on coffee consumption (cups/day) and all-cause mortality (Crippa et al., *Am. J. Epidemiol*, 2014)

Exposure category	Dose	Cases	n	$\widehat{\mathrm{RR}}$	95% CI
0-1	0.5	57	249	1.00	—
2-3	2.5	136	655	0.75	0.57, 0.99
4-5	4.5	144	619	0.84	0.64, 1.10
6+	6.5	115	387	1.09	0.83, 1.43

The $\widehat{\mathrm{RR}}$s are not independent

The predicted relative risk for reference category is 1

Two stage dose–response meta-analysis

First stage

Define and estimate a common dose–response model in each study ($i = 1, \ldots, I$)

Two stage dose–response meta-analysis

First stage

Define and estimate a common dose–response model in each study ($i = 1, \ldots, I$)

Second stage

Combine study-specific regression coefficients using meta-analysis

Challenges

- Lack of free and open source software (*Paper I*)

Challenges

- Lack of free and open source software (*Paper I*)
- Assessment of goodness of fit of dose–response meta–analytic models has not yet been discussed (*Paper II*)

Challenges

- Lack of free and open source software (*Paper I*)
- Assessment of goodness of fit of dose–response meta–analytic models has not yet been discussed (*Paper II*)
- Little emphasis is placed on the assumptions underlying the common measures of heterogeneity (*Paper III*)

Challenges

- Lack of free and open source software (*Paper I*)
- Assessment of goodness of fit of dose–response meta–analytic models has not yet been discussed (*Paper II*)
- Little emphasis is placed on the assumptions underlying the common measures of heterogeneity (*Paper III*)
- The effect of differential shape and exposure distribution is hard to be addressed in a two-stage approach (*Paper IV*)

Challenges

- Lack of free and open source software (*Paper I*)
- Assessment of goodness of fit of dose–response meta–analytic models has not yet been discussed (*Paper II*)
- Little emphasis is placed on the assumptions underlying the common measures of heterogeneity (*Paper III*)
- The effect of differential shape and exposure distribution is hard to be addressed in a two-stage approach (*Paper IV*)
- Dose–response and meta-regression models may be affected by small number of data points in some of the studies (*Paper V*)

Paper I

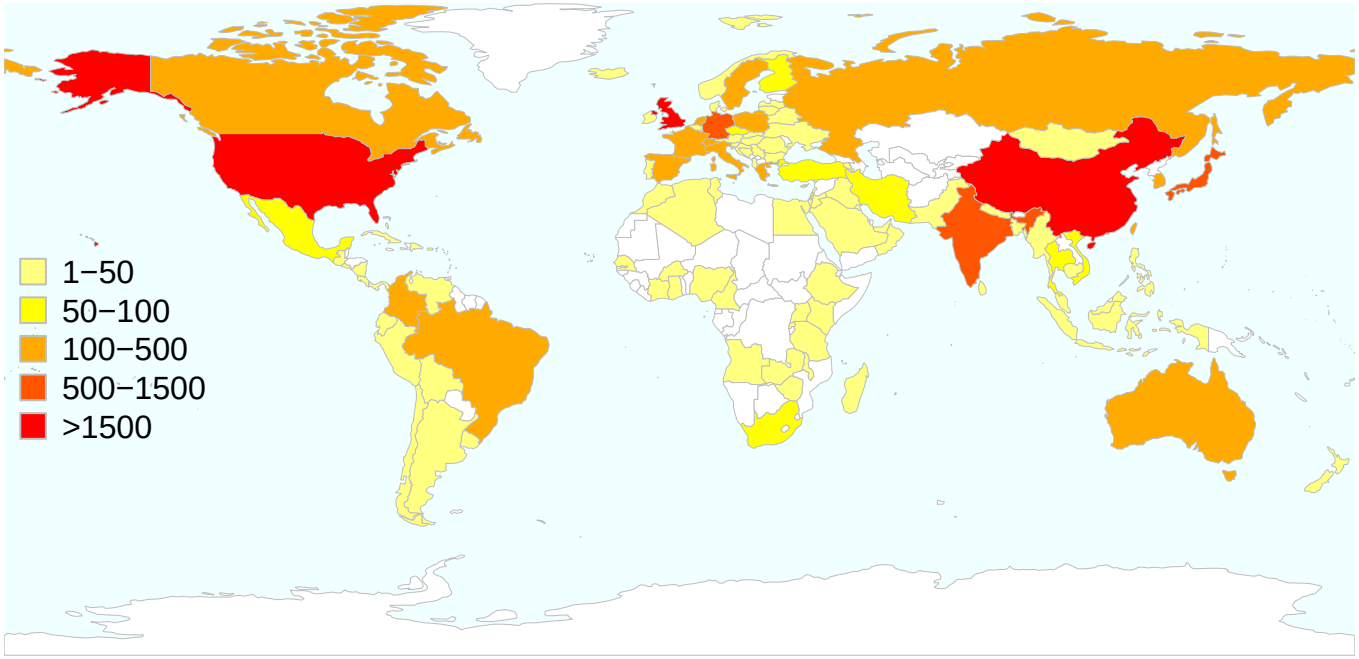
Multivariate dose–response meta-analysis: the dosresmeta R Package. *J. Stat. Softw*, 2016

Specific aim

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

The dosresmeta R package

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



Codes and examples and at

`https://alecri.github.io/software/dosresmeta.html`

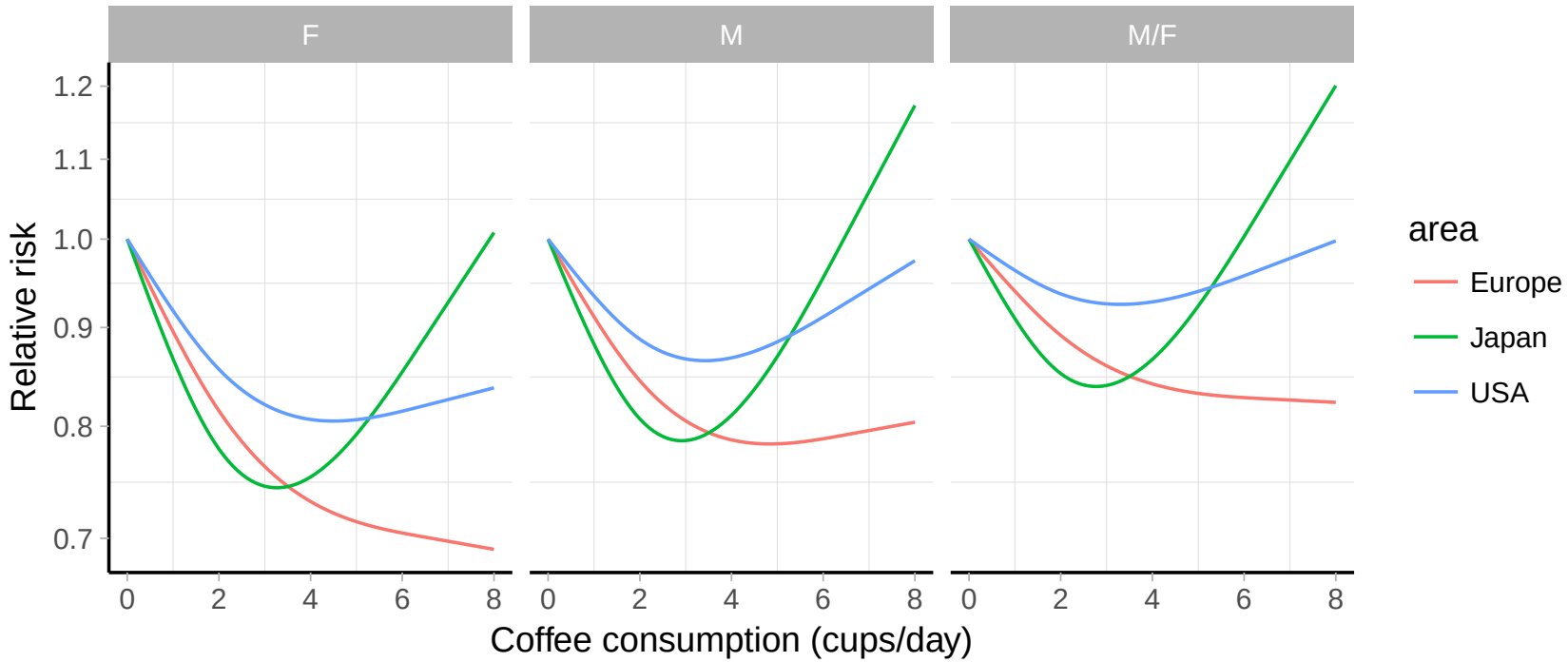
Package Description

- Two-stage dose–response meta-analysis
- Greenland and Longnecker, and Hamling method
- `print` and `summary` function
- Meta-regression models
- Dedicated `predict` function
- Methodologies presented in the thesis

Coffee consumption and all-cause mortality

```
R> data("coffee_mort")
R> # linear model
R> lin <- dosresmeta(logrr ~ dose, id = id, se = se, type = type,
+                    cases = cases, n = n, data = coffee_mort)
R> # restricted cubic spline model
R> k <- quantile(coffee_mort$dose, c(.1, .5, .9))
R> spl <- dosresmeta(logrr ~ rcs(dose, k), id = id, se = se, type = type,
+                    cases = cases, n = n, data = coffee_mort)
R> # restricted cubic spline meta-regression model
R> spl_reg <- dosresmeta(logrr ~ rcs(dose, k), id = id, se = se,
+                        cases = cases, n = n, type = type, data = coffee_mort,
+                        mod = ~ gender + area)
```

```
R> expand.grid(dose = seq(0, 8, .1), gender = levels(coffee_mort$gender),
+                area = levels(coffee_mort$area)) %>%
+   cbind(predict(spl_reg, newdata = ., expo = T)) %>%
+   ggplot(aes(dose, pred, col = area)) + geom_line() + facet_grid(~ gender) +
+   scale_y_continuous(trans = "log", breaks = pretty_breaks()) +
+   labs(x = "Coffee consumption (cups/day)", y = "Relative risk")
```



Paper II

Goodness of fit tools for dose–response meta-analysis of binary outcomes
Res Synth Meth, 2017

Specific aim

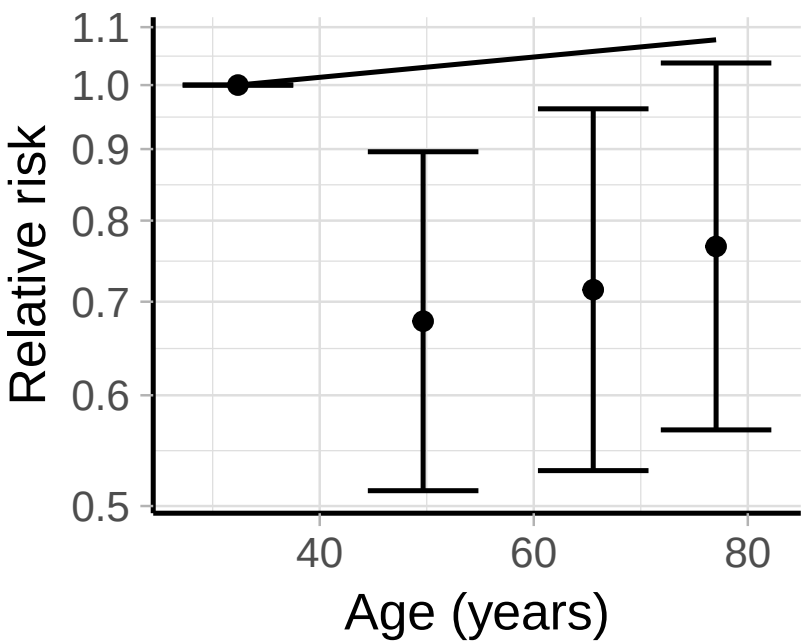
- To present and discuss relevant measures and graphical tools to assess the goodness-of-fit in dose–response meta-analytic models

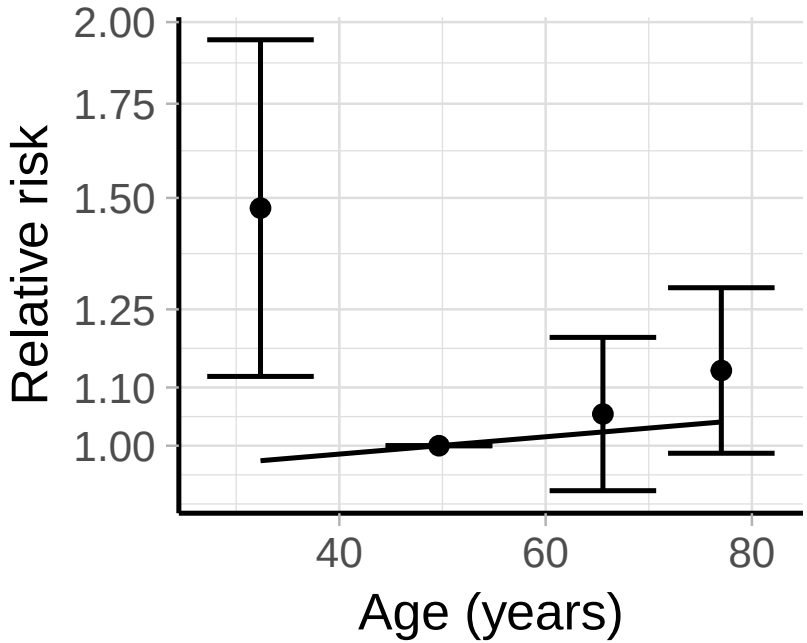
Goodness-of-fit

Does the pooled curve adequately summarize the aggregate data?

This question is typically ignored in published meta-analyses

A graphical comparison may be not be appropriate





Proposed tools

Deviance (D)

- Total absolute distance between fitted and reported (log) RRs
- Test for model specification

Proposed tools

Deviance (D)

- Total absolute distance between fitted and reported (log) RRs
- Test for model specification

Coefficient of determination (R^2)

- Descriptive measure of agreement
- Dimensionless measure bounded between 0 and 1

Proposed tools

Deviance (D)

- Total absolute distance between fitted and reported (log) RRs
- Test for model specification

Coefficient of determination (R^2)

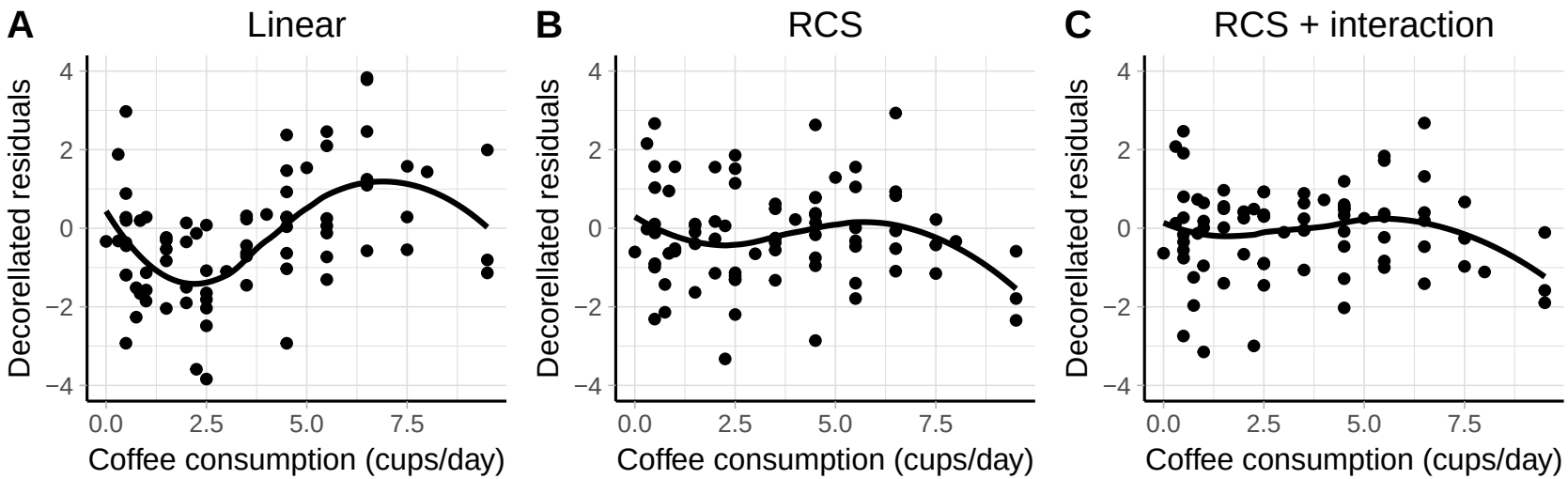
- Descriptive measure of agreement
- Dimensionless measure bounded between 0 and 1

Plot of decorrelated residuals versus exposure

- Visual assessment of the goodness of fit
- Evaluate how the pooled dose–response curve fits the data by exposure levels

Coffee consumption and all-cause mortality

Analysis	Model	Deviance	df	p value	R^2	$\mathrm{R}^2_{\mathrm{adj}}$
A	Linear	225.244	78	0.000	0.488	0.482
B	RCS	141.332	77	0.000	0.679	0.671
C	RCS + interaction	100.372	69	0.008	0.772	0.739



Paper III

A new measure of between-studies heterogeneity in meta-analysis. *Stat. Med.*, 2016

Specific aim

- To develop a new measure of between-study heterogeneity in the broader context of meta-analysis

Measures of heterogeneity

Heterogeneity measures, I^2 and R_I, relate the heterogeneity, τ^2, to the total variance, $\tau^2 + \sigma^2$

σ^2 is a summary measure of the observed within-study variance, v_i

Homogeneity of within-studies variances is unlikely to hold

Analysis	$v_1, \ldots, v_5$	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

R_b a new measure of heterogeneity

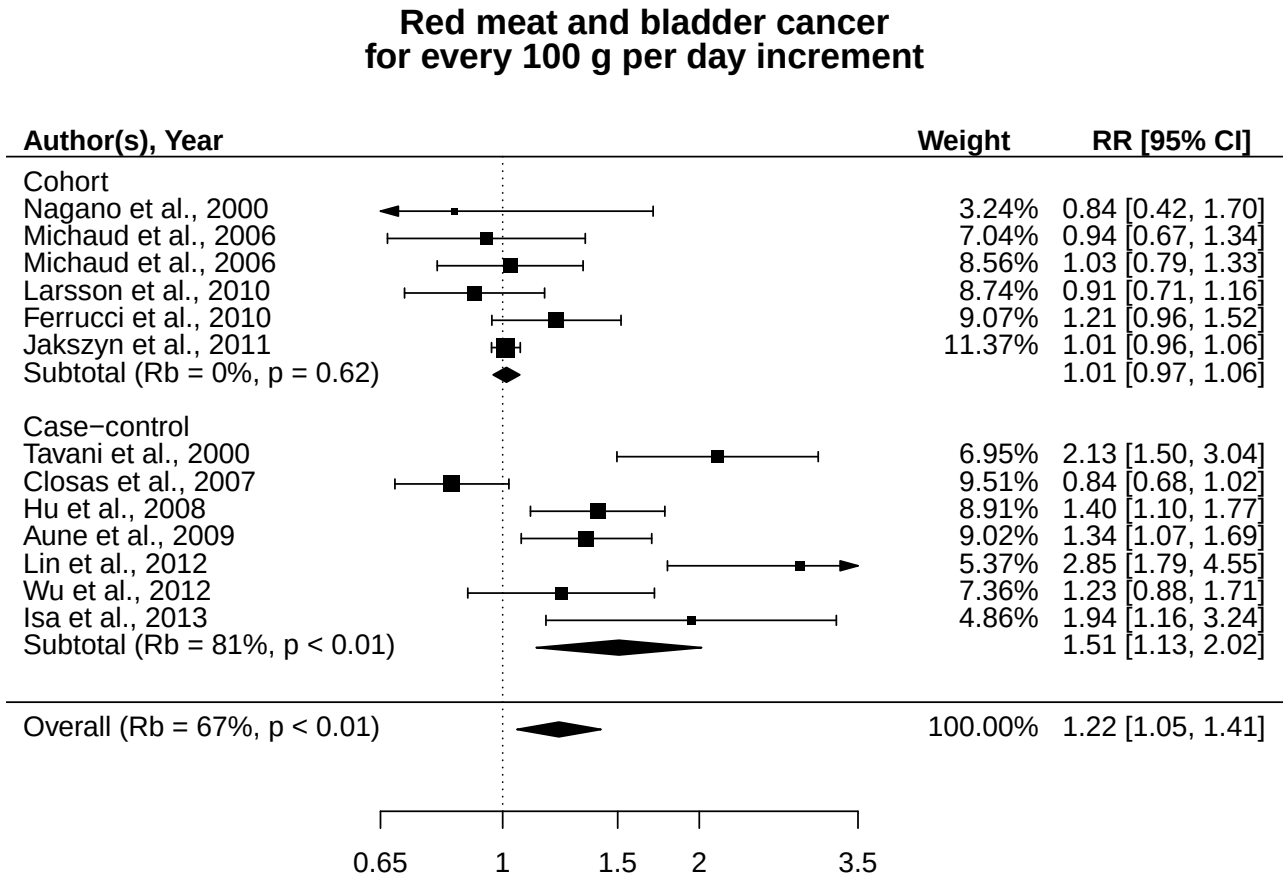
The new measure quantifies the contribution of τ^2 relative to the variance of the pooled random effects estimate

$$R_b = \frac{\tau^2}{I\mathrm{Var}\left(\hat{\beta}_{re}\right)} = \frac{1}{I}\sum_{i=1}^{I}\frac{\tau^2}{v_i + \tau^2} \quad (1)$$

R_b satisifies the properties for a measure of heterogeneity

R_b is a consistent and asymptotically normal distributed estimator (Wald-type confidence intervals)

It coincides with I^2 and R_I when $v_i = \sigma^2 \; \forall i = 1, \ldots, I$



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)
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)

Paper IV

A pointwise approach to dose–response meta-analysis of aggregated data.

Specific aim

- To move beyond the specification of a unique model across the studies exploring possible advantages of a point-wise approach

Paper IV

Possible limitations of a two-stage approach

- Common study-specific functional relationship (1st stage)
- Information on study-specific exposure range is not considered (2nd stage)

Paper IV

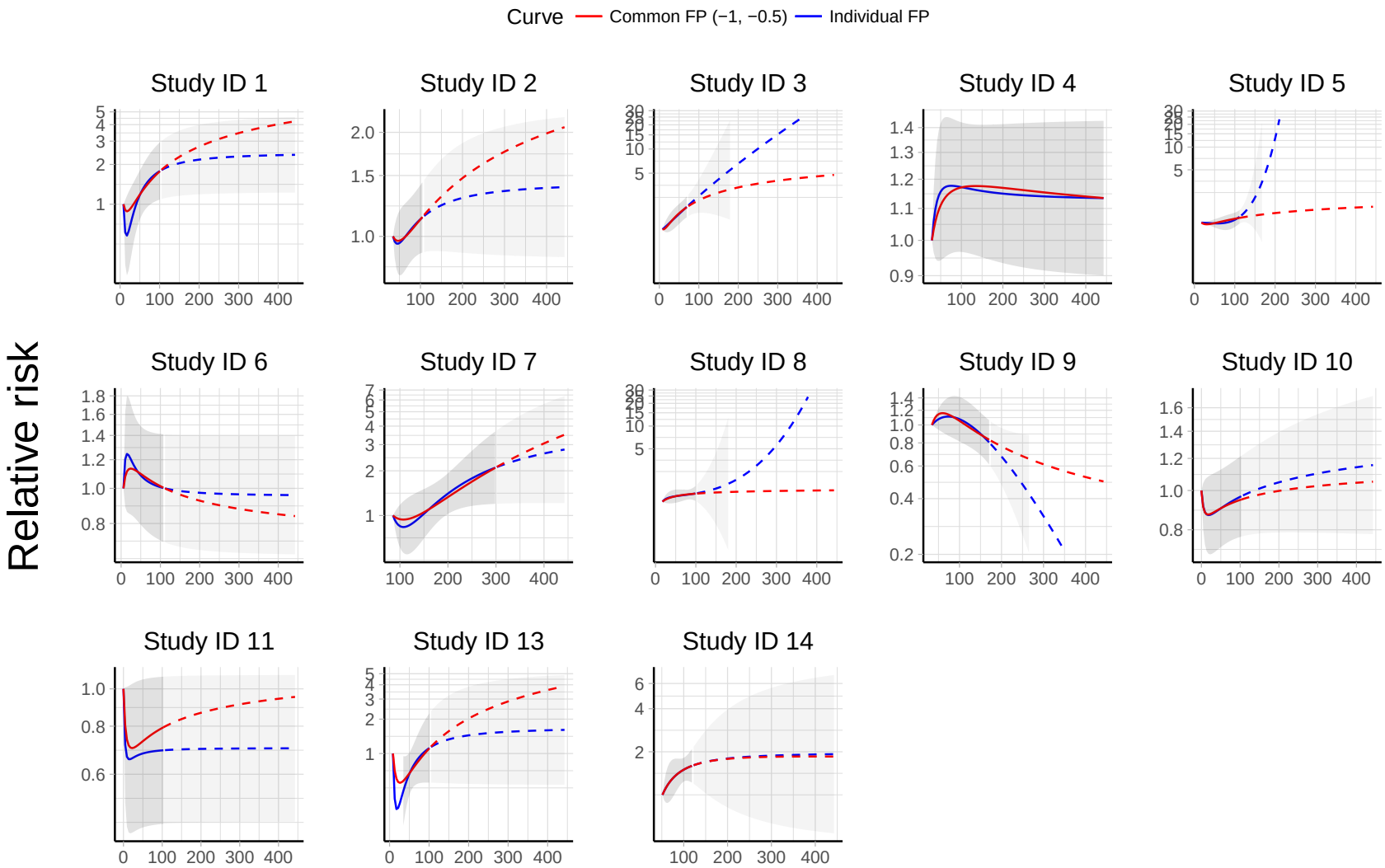
Possible limitations of a two-stage approach

- Common study-specific functional relationship (1st stage)
- Information on study-specific exposure range is not considered (2nd stage)

Consequences

- Poor fit in some of the study-specific dose–response analyses
- Risk of extrapolating predicted relative risks

Individual curves for 13 studies on red meat and bladder cancer risk



A point-wise average approach

It consists of

- Estimating study-specific dose–response curves
- Predicting study-specific effects (log RRs) for a grid of exposure values
- Combining study-specific effects

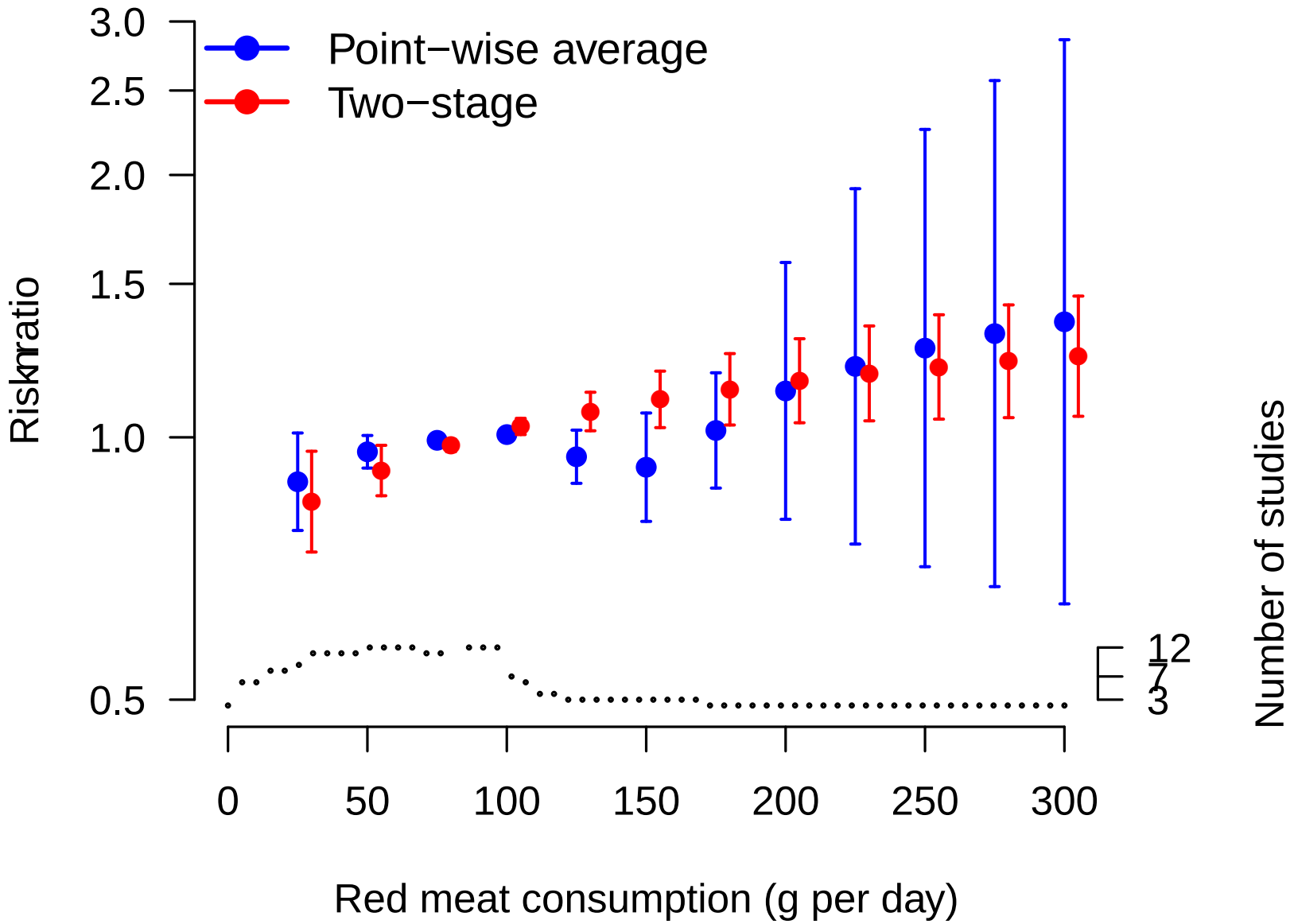
A point-wise average approach

It consists of

- Estimating study-specific dose–response curves
- Predicting study-specific effects (log RRs) for a grid of exposure values
- Combining study-specific effects

Advantages

- The dose–response analyses may vary across studies
- RR predictions can be limited to study-specific exposure ranges
- Results from univariate meta-analyses can be presented pointwisely


Point−wise average
Two−stage
Risk ratio
3.0
2.5
2.0
1.5
1.0
0.5
0
50
100
150
200
250
300
Red meat consumption (g per day)
12
7
3
Number of studies

Paper V

One-stage dose–response meta-analysis for aggregated data. *Stat. Methods Med. Res.*, 2018

Specific aim

- 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

Paper V

Study-specific dose–response analyses are often limited (1 to 3 log RRs)
Studies reporting one RR are excluded to model non-linear curves

A one-stage procedure for random-effects meta-analysis of non-linear curves

- Conceptually easier
- Fit more elaborate curves
- Avoid exclusion of studies with small observations

A one-stage approach

General form of a linear mixed model

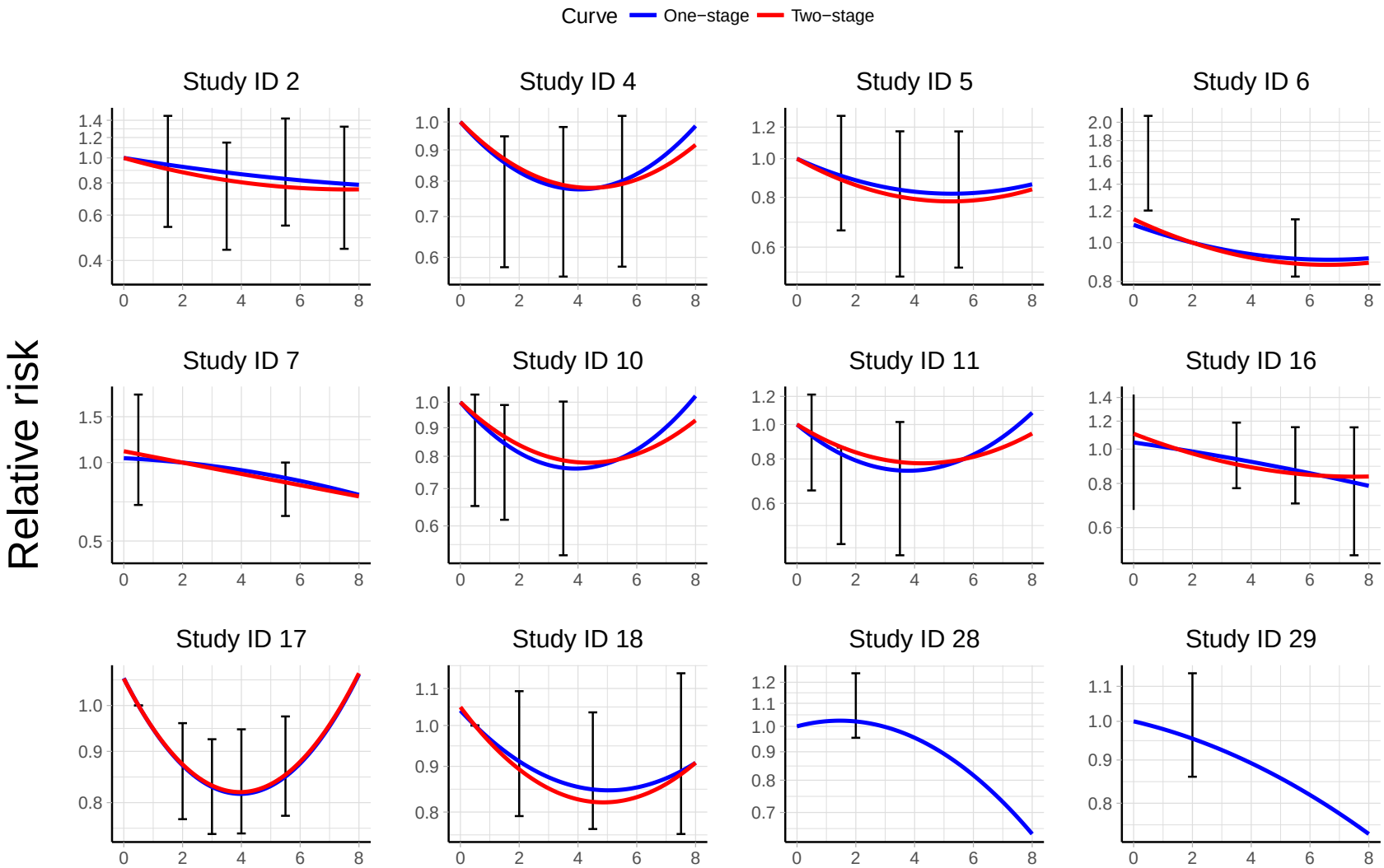
$$\mathbf{y}_i = \mathbf{X}_i\beta + \mathbf{Z}_i\mathbf{b}_i + \varepsilon_i \tag{2}$$

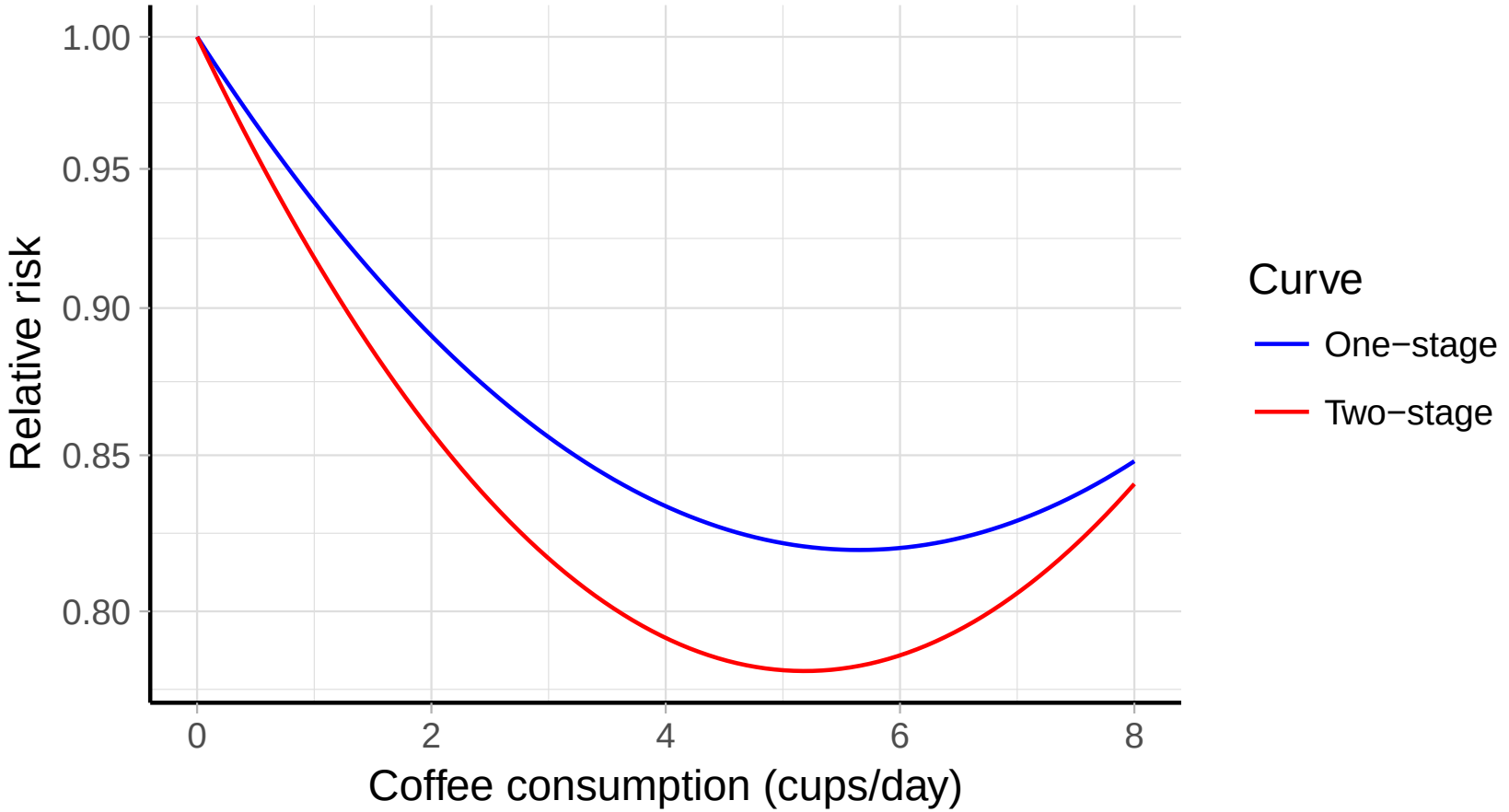
$\mathbf{Z}_i \equiv \mathbf{X}_i$ $\varepsilon_i \sim N(\mathbf{0}, \boldsymbol{S}_i)$ and $\mathbf{b}_i \sim N(\mathbf{0}, \boldsymbol{\Psi})$

Established theory for inference, heterogeneity assessment, and prediction

If the study-specific dose–response models are identifiable, the one- and two-stage approaches are equivalent

Individual curves for 12 studies on coffee and mortality




1.00
0.95
0.90
0.85
0.80
Relative risk
0
2
4
6
8
Coffee consumption (cups/day)
Curve
One−stage
Two−stage

Conclusions

Methodological advancements in dose–response meta-analysis

Conclusions

Methodological advancements in dose–response meta-analysis

Practice

- The `dosresmeta` R package greatly facilitates applications

Conclusions

Methodological advancements in dose–response meta-analysis

Practice

- The `dosresmeta` R package greatly facilitates applications

Interpretation

- The proposed tools can help to evaluate the goodness-of-fit
- The $\hat{R}_b$ quantifies the impact of heterogeneity without any assumption about the within-study error term

Conclusions

Methodological advancements in dose–response meta-analysis

Practice

- The `dosresmeta` R package greatly facilitates applications

Interpretation

- The proposed tools can help to evaluate the goodness-of-fit
- The $\hat{R}_b$ quantifies the impact of heterogeneity without any assumption about the within-study error term

Estimation

- A point-wise approach for evaluating heterogeneous curves and exposure distributions
- A one-stage meta-analysis for addressing more elaborated research questions based on all of the information available

Aggregated Data I

Category	A	B	RR	lb	ub
0	A_0	B_0	1	–	–
⋮	⋮	⋮	⋮	⋮	⋮
n	A_n	B_n	RR_n	lb_n	ub_n

$$RR_i = \frac{A_i B_0}{A_0 B_i} \qquad i = 1, \ldots, n$$

$$V_i = \begin{cases} \frac{1}{A_0} + \frac{1}{B_0} + \frac{1}{A_1} + \frac{1}{B_1} & \text{, if case-control} \\ \frac{1}{A_0} + \frac{1}{A_1} & \text{, if incidence rate} \\ \frac{1}{A_0} - \frac{1}{B_0} + \frac{1}{A_1} - \frac{1}{B_1} & \text{, if cumulative incidence} \end{cases} \tag{3}$$

Greenland and Longnecker's method

Newton's method optimization for the system of equation

$$\log(RR_i) + \log(A_0) + \log(N_i - A_i) - \log(A_i) - \log(N_0 - A_0)$$
$$i = 1, \ldots, n$$

Total numbers of cases and subjects is held constant

$N_i = A_i + B_i$ for case-control

Hamling's method

$2(n+1)$ system of equations

$$\begin{cases} p = B_0/B \\ z = B/A \\ RR_i = (A_i B_0)/(A_0 B_i) & i = 1, \ldots, n \\ V_i = \text{as defined in Equation 3} & i = 1, \ldots, n \end{cases}$$

$B = \sum_{i=0}^{n} B_i$ and $A = \sum_{i=0}^{n} A_i$

Fractional Polynomials

$$\mathrm{E}\left[\boldsymbol{y}_i|\boldsymbol{x}_i\right] = \beta_1 \boldsymbol{x}_i^{p_1} + \beta_2 \boldsymbol{x}_i^{p_2} \tag{4}$$

p_1 and p_2 in the set $\{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$

$\log(\boldsymbol{x}_i)$ for $p = 0$

$\beta_2 \boldsymbol{x}_i{}^{p_2} \log(\boldsymbol{x}_i)$ for $p_1 = p_2$

The best fitting model is the one with minimum AIC

Restricted cubic splines

with three knots (k_1, k_2, k_3)

$$\mathrm{E}\left[\boldsymbol{y}_i|\boldsymbol{x}_i\right] = \beta_1 f_1\left(\boldsymbol{x}_i\right) + \beta_2 f_2\left(\boldsymbol{x}_i\right) \tag{5}$$

$$f_1\left(\mathbf{x}_i\right) = \mathbf{x}_i$$

$$f_2\left(\mathbf{x}_i\right) = \frac{\left(\mathbf{x}_i - k_1\right)_+^3 - \frac{k_3 - k_1}{k_3 - k_2}\left(\mathbf{x}_i - k_2\right)_+^3 + \frac{k_2 - k_1}{k_3 - k_2}\left(\mathbf{x}_i - k_3\right)_+^3}{\left(k_3 - k_1\right)^2}$$

where $u_+ = u$ if $u \geq 0$ and $u_+ = 0$ otherwise

Goodness of fit measures

$$D = (\mathbf{y} - \mathbf{X}\beta)^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{X}\beta) \tag{6}$$

Under H_0 (correct model specification), $D \sim \chi^2_{n-p}$
If M_1 nested in M_2, $D(M_1) - D(M_2) \sim \chi^2_q$

$$R^2 = \frac{\sum_{i=1}^{K} (\mathbf{y}_i - \mathbf{X}_i\beta)^\top \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i\beta)}{\sum_{i=1}^{K} \mathbf{y}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{y}_i} \tag{7}$$

De-correlated residuals $\mathrm{e}_i^* = \mathbf{C}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i\beta)$, with $\mathbf{C}_i\mathbf{C}_i^\top = \boldsymbol{\Sigma}_i$