Karolinska Institutet

Workshop on Dose–Response Meta–Analysis

Day 2

Nicola Orsini, Alessio Crippa

18th April 2024

Main learning points of Day 2

- Estimate and present a non-linear dose–response model
- Use degree-2 polynomials and restricted cubic splines to model the dose
- Conduct and present statistical inference under common effect and random-effects meta-analysis
- Examine the statistical heterogeneity across studies

Vitamin D and Colorectal Cancer Risk

The *Journal of Clinical Oncology (JCO)* serves its readers as the single most credible, authoritative resource for disseminating significant clinical oncology research.

Impact Factor = 45

Abstract

A B S T R A C T

Purpose
To conduct a systematic review of prospective studies assessing the association of vitamin D intake or blood levels of 25-hydroxyvitamin D [25(OH)D] with the risk of colorectal cancer using meta-analysis.

Methods
Relevant studies were identified by a search of MEDLINE and EMBASE databases before October 2010 with no restrictions. We included prospective studies that reported relative risk (RR) estimates with 95% CIs for the association between vitamin D intake or blood 25(OH)D levels and the risk of colorectal, colon, or rectal cancer. Approximately 1,000,000 participants from several countries were included in this analysis.

Results
Nine studies on vitamin D intake and nine studies on blood 25(OH)D levels were included in the meta-analysis. The pooled RRs of colorectal cancer for the highest versus lowest categories of vitamin D intake and blood 25(OH)D levels were 0.88 (95% CI, 0.80 to 0.96) and 0.67 (95% CI, 0.54 to 0.80), respectively. There was no heterogeneity among studies of vitamin D intake ($P = .19$) or among studies of blood 25(OH)D levels ($P = .96$). A 10 ng/mL increment in blood 25(OH)D level conferred an RR of 0.74 (95% CI, 0.63 to 0.89).

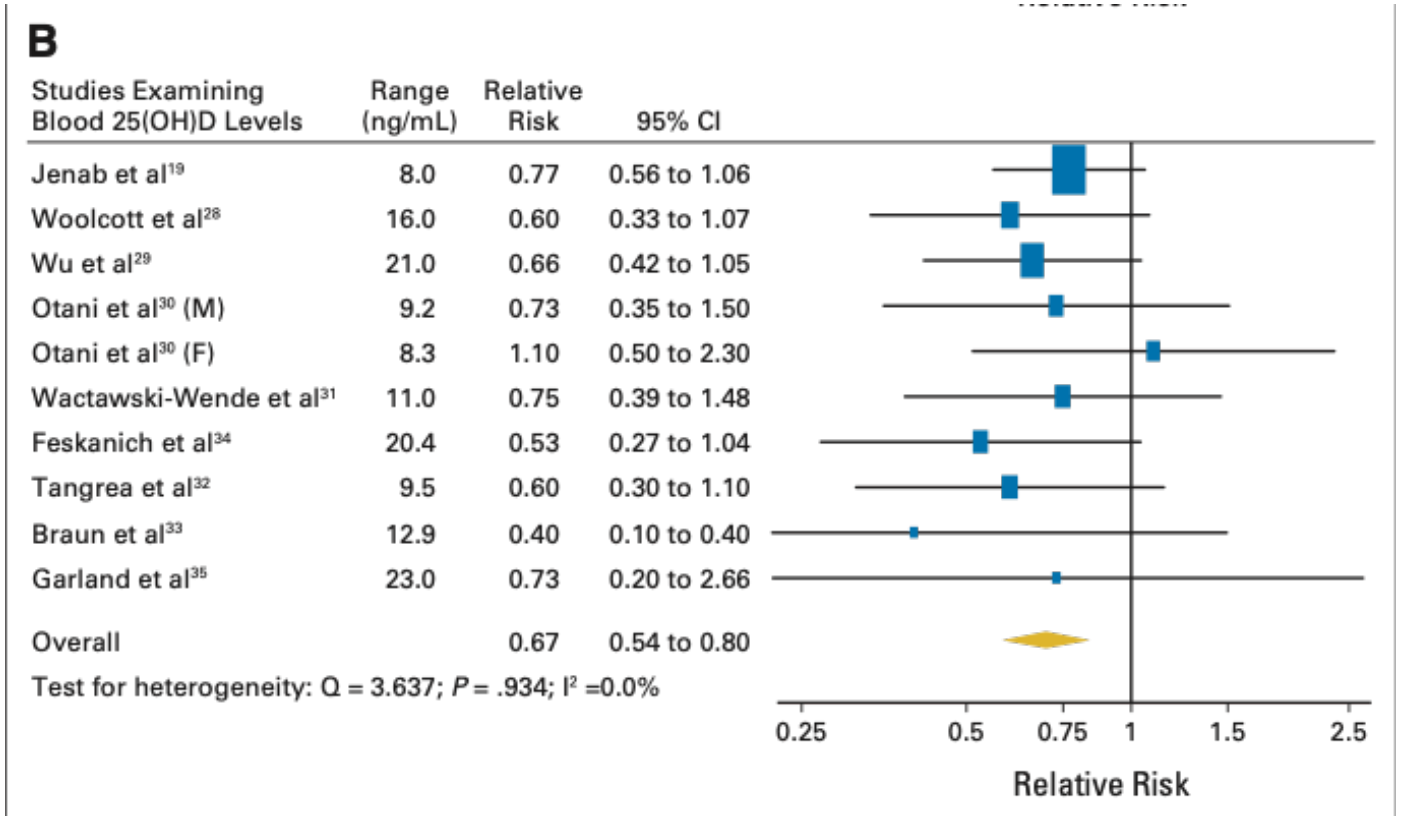
Conclusion
Vitamin D intake and blood 25(OH)D levels were inversely associated with the risk of colorectal cancer in this meta-analysis.

High vs low approach

High Versus Low Vitamin D or 25(OH)D Levels

The multivariable-adjusted RRs for each study and combination of all studies for the highest versus lowest categories of vitamin D intake or blood 25(OH)D levels are shown in Figure 2. Results

Forest plot of top vs bottom approach



B

Studies Examining Blood 25(OH)D Levels	Range (ng/mL)	Relative Risk	95% CI
Jenab et al[19]	8.0	0.77	0.56 to 1.06
Woolcott et al[28]	16.0	0.60	0.33 to 1.07
Wu et al[29]	21.0	0.66	0.42 to 1.05
Otani et al[30] (M)	9.2	0.73	0.35 to 1.50
Otani et al[30] (F)	8.3	1.10	0.50 to 2.30
Wactawski-Wende et al[31]	11.0	0.75	0.39 to 1.48
Feskanich et al[34]	20.4	0.53	0.27 to 1.04
Tangrea et al[32]	9.5	0.60	0.30 to 1.10
Braun et al[33]	12.9	0.40	0.10 to 0.40
Garland et al[35]	23.0	0.73	0.20 to 2.66
Overall		0.67	0.54 to 0.80

Test for heterogeneity: Q = 3.637; P = .934; I^2 =0.0%

Methods

each corresponding RR estimate. We examined a potential nonlinear dose-response relationship between 25(OH)D levels and colorectal cancer by modeling 25(OH)D levels using restricted cubic splines with three knots at percentiles 25%, 50%, and 75% of the distribution. A *P* value for nonlinearity was calculated by testing the null hypothesis that the coefficient of the second spline is equal to 0.[15-16]

Results

Dose-Response Meta-Analysis

Next, we assessed the dose-response relationship between blood 25(OH)D levels and the risk of colorectal cancer. We found obvious evidence of statistically significant departure from linearity ($P < .001$). A 10 ng/mL increment in blood 25(OH)D level conferred an RR of 0.74 (95% CI, 0.63 to 0.89; Fig 3).

Dose–response plot

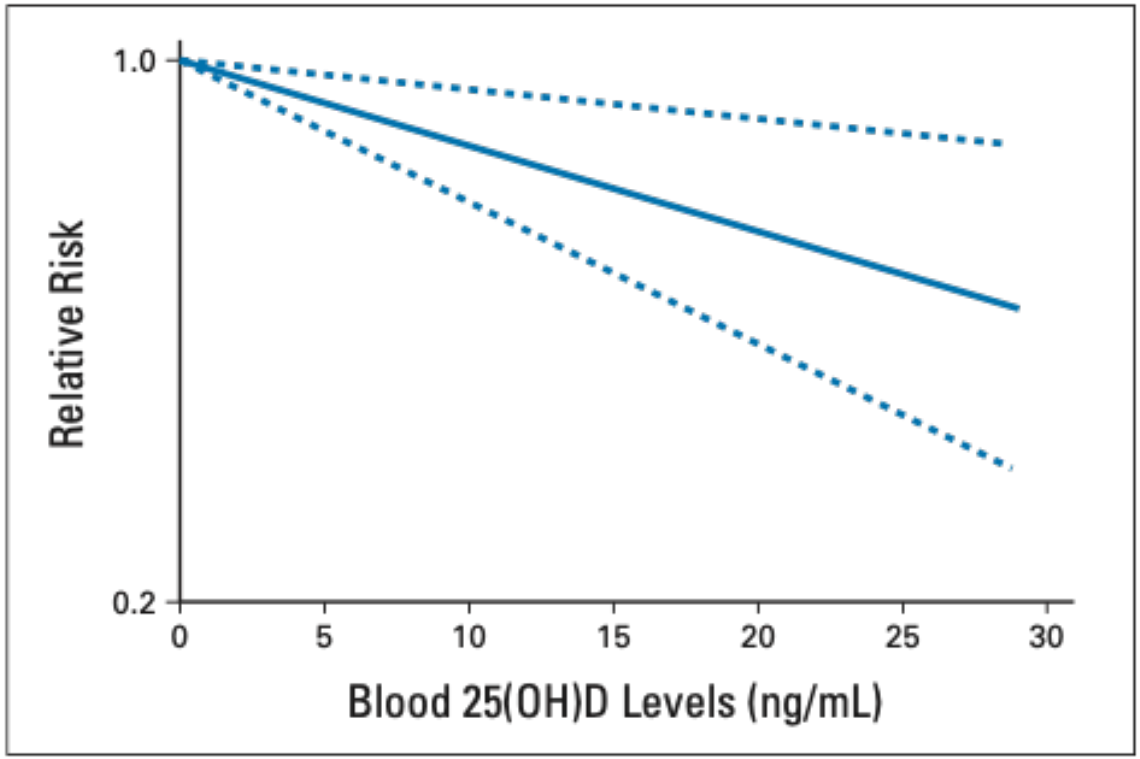


Fig 3. Dose-response relationship between blood 25-hydroxyvitamin D [25(OH)D] levels and the relative risk of colorectal cancer. Adjusted relative risks and 95% CIs (dashed lines) are reported. Blood 25(OH)D levels were modeled with a linear trend in a random-effects meta-regression model. The median value of the lowest reference interval (9.8 ng/mL) was used to estimate all relative risks. The vertical axis is on a log scale.

Questions

- What is the relative risk (95% CI) contrasting the exposure extremes (saying 30 vs 0 ng/mL)?
- How the plot reconcile with 33% lower colorectal cancer risk comparing highest vs lowest category of blood vitamin D?
- What is the rationale for using 0 ng/mL as referent?
- Would it be possible to re-do the plot (RR, 95% CI) using another exposure value as comparison?

Flexible dose-response models

In a dose–response meta–analysis, as in any regression model, one can move beyond linearity for quantitative predictors. Some examples are

- U-shape
- J-shape
- V-shape
- Threshold
- Spike at zero
- Steps

Non–linear dose–response model

A non-linear dose–response meta–analysis is conducted using some transformations of the dose x_i.

$$\boldsymbol{y}_i = g\left(x_i, \beta_i\right) + \varepsilon_i, \quad \varepsilon_i \sim N\left(\mathbf{0}, \boldsymbol{\Sigma}_i\right), \quad i = 1, \ldots, I$$

where β_i is the $p \times 1$ vector of regression coefficients to be estimated within each study.

Two transformations of the dose

Using just two transformations, the regression function can be written as

$$g(x_i, \boldsymbol{\beta}_i) = [g_1(x_{ij}) - g_1(x_{i0})]\beta_{1i} + [g_2(x_{ij}) - g_2(x_{i0})]\beta_{2i}$$

where, for example, $g_1(x_{ij})$ is the value of the first transformation for the j-th dose value level in the i-th study, and $g_1(x_{i0})$ is the value of the first transformation at the referent in the i-th study.

Quadratic function

A degree 2-polynomial or simply quadratic function is probably the easiest approach to detect strong indications against a simpler linear trend.

$$g_1(x_{ij}) = x_{ij}$$
$$g_2(x_{ij}) = x_{ij}^2$$

With a regression function that simplifies to

$$g(x_i, \boldsymbol{\beta}_i) = (x_{ij} - x_{i0})\beta_{1i} + (x_{ij}^2 - x_{i0}^2)\beta_{2i}$$

Spline function

A single spline transformation of the exposure x is defined by the location of a knot k and a certain degree d as follows

$I(x > k)(x - k)^d$ with $d = 0, 1, 2, 3$

The indicator function $I(x > k)$ takes value 1 if $x > k$ and 0 if $x \leq k$. A more compact way of writing a spline is $(x - k)_+^d$. According to the degree of the polynomial and the number of knots we can estimate piecewise constant, linear, quadratic, and cubic splines.

Restricted cubic splines

A restricted cubic spline model with three knots $\boldsymbol{k} = (k_1, k_2, k_3)$ is defined only in terms of 2 regression coefficients. The two spline transformations can be defined as follows (Harrell F. 2001):

$$g_1(x_{ij}) = x_{ij}$$

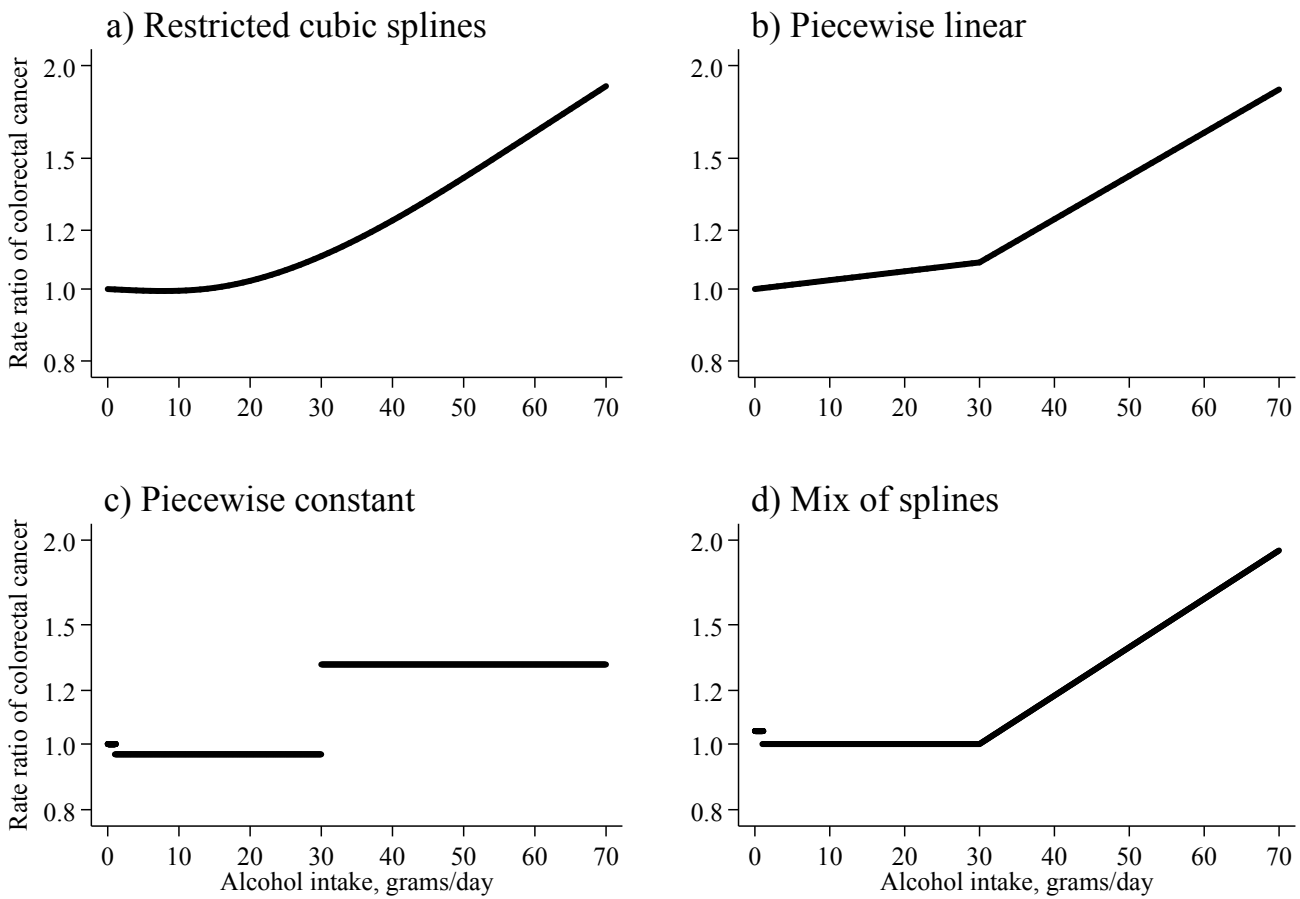
$$g_2(x_{ij}) = \frac{(x_{ij} - k_1)_+^3 - \frac{k_3 - k_1}{k_3 - k_2}(x_{ij} - k_2)_+^3 + \frac{k_2 - k_1}{k_3 - k_2}(x_{ij} - k_3)_+^3}{(k_3 - k_1)^2}$$

Restricted cubic splines

The dose-response function is constrained to be linear before the first knot k_1 where by definition $g_2(x_{ij})$ takes on value zero. It can be shown that the dose-response function is also linear beyond the last knot k_3.

The location of the knots is usually derived from fixed percentiles of the overall distribution of the exposure. The parameters β_{1i} and β_{2i} jointly define the shape of the dose-response relationship.

Comparison of different spline functions



Parameters of the random-effects multivariate meta-analysis

$$\begin{pmatrix} \beta_{1i} \\ \beta_{2i} \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \begin{bmatrix} \xi_1 & \\ \xi_3 & \xi_2 \end{bmatrix} \right)$$

The model is in terms of 2 fixed-effects plus 2 variances and 1 covariance of the 2 random-effects.

Estimation of study-specific non-linear curves

Generalized least square estimation.

$$\hat{\beta}_i = (X_i^\top \Sigma_i^{-1} X_i)^{-1} X_i^\top \Sigma_i^{-1} y_i$$
$$S_i = \mathrm{Cov}(\beta_i) = (X_i^\top \Sigma_i^{-1} X_i)^{-1}$$

X_i is the design matrix as defined in the non-linear model.

Note: $\hat{\beta}_i$ is now a vector of regression coefficients, with S_i (co)variance matrix.

Multivariate random-effects meta-analysis

$$\hat{\beta}_i \sim \mathcal{N}_p\left(\hat{\beta}, S_i + \Psi\right)$$

$$\hat{\beta} = \left(\sum_{i=1}^{I} (S_i + \Psi)^{-1}\right)^{-1} \left(\sum_{i=1}^{I} (S_i + \Psi)^{-1} \hat{\beta}_i\right)$$

$$\mathrm{Var}\left(\hat{\beta}\right) = \left(\sum_{i=1}^{I} (S_i + \Psi)^{-1}\right)^{-1}$$

Testing hypotheses: Overall effect

The shape of the summary dose-response curve depends on the vector of estimated coefficients $\hat{\beta}$ under either a common effect or a random effects model.

The hypothesis of overall no effect can be addressed by testing $H_0 : \beta = \mathbf{0}$. Depending on the functional form g specified, testing one specific regression coefficient may detect specific characteristics of the shape (i.e. non-linearity, shift in level, change in slope).

Testing hypotheses: Departure from linearity

This can be done by testing $H_0^* : \boldsymbol{\beta}^* = \mathbf{0}$, where $\boldsymbol{\beta}^*$ refers to the part of the vector of coefficients defining those characteristics.

For example, if the function g consists of restricted cubic spline models with 3 knots (2 splines), a p-value detecting departure from linearity can be obtained by testing the coefficient of the second spline equal to zero.

The Wald-type tests of hypothesis for combined parameters $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^*$ can be based on the estimated $\hat{\boldsymbol{\beta}}$ and its variance/covariance matrix $\mathbf{V}(\hat{\boldsymbol{\beta}})$.

Quantiles marginal and conditional dose-response

Marginal

$$Q_p^M = (\boldsymbol{X}^* - \boldsymbol{x}_0^*)\hat{\boldsymbol{\beta}} + \phi^{-1}(p)\mathrm{diag}[(\boldsymbol{X}^* - \boldsymbol{x}_0^*)V(\hat{\boldsymbol{\beta}})(\boldsymbol{X}^* - \boldsymbol{x}_0^*)']^{1/2}$$

Conditional

$$Q_p^C = (\boldsymbol{X}^* - \boldsymbol{x}_0^*)\hat{\boldsymbol{\beta}} + \phi^{-1}(p)\mathrm{diag}[(\boldsymbol{X}^* - \boldsymbol{x}_0^*)(V(\hat{\boldsymbol{\beta}}) + \hat{\boldsymbol{\Psi}})(\boldsymbol{X}^* - \boldsymbol{x}_0^*)']^{1/2}$$

where

$\boldsymbol{X}^*$ indicates a matrix of user specified transformations

$\boldsymbol{x}_0^*$ indicates a matrix of reference values

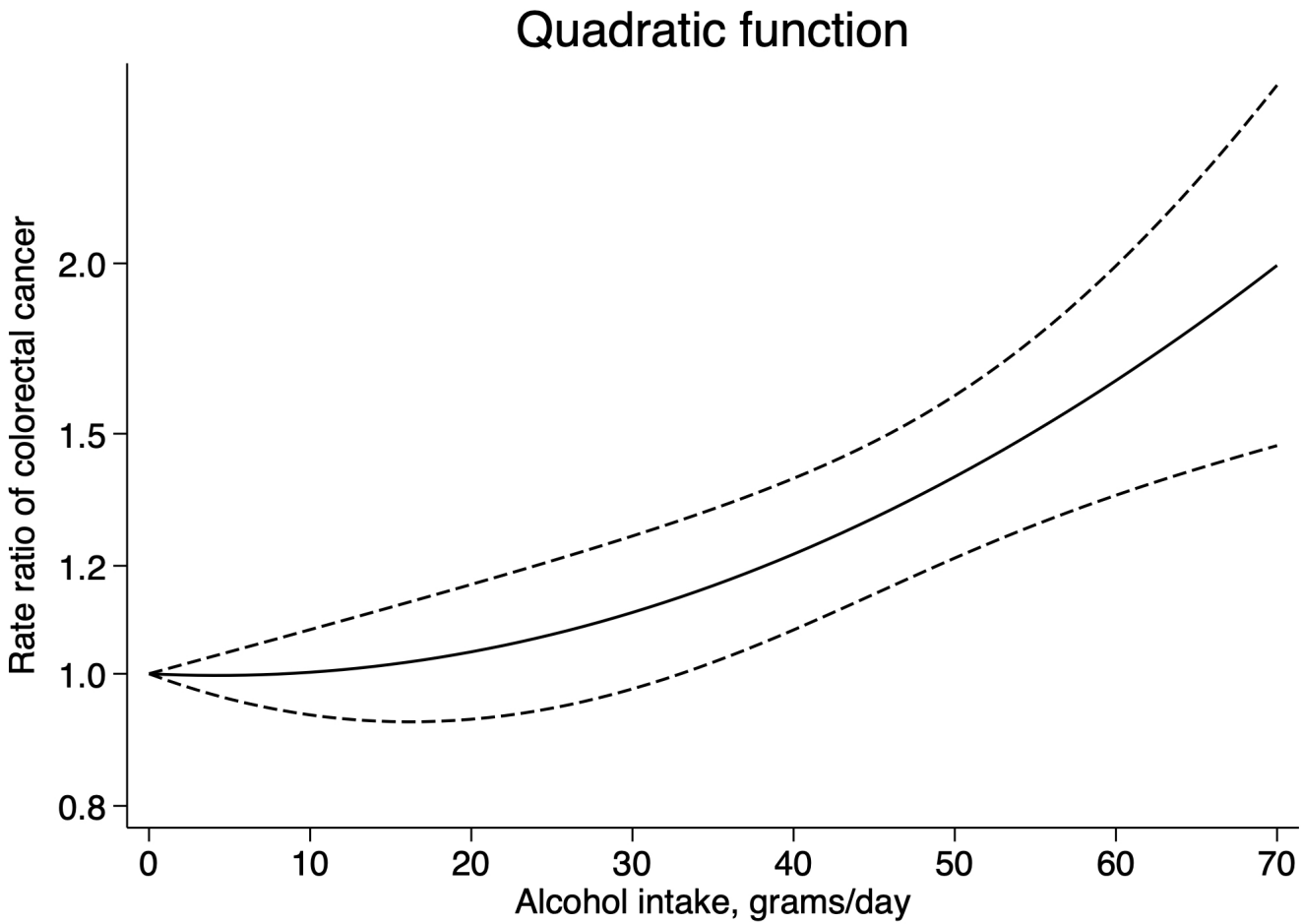
Alcohol intake and colorectal cancer risk

We combine the dose-response relation between alcohol intake and colorectal cancer rate arising from 8 prospective cohort studies including 489,979 women and men participating in the Pooling Project of Prospective Studies of Diet and Cancer. A total of 3,646 cases and 2,511,424 person-years are included in this analysis (AJE, 2012).

Quadratic function for alcohol intake

```
Two-stage random-effect dose-response model      Number of studies =         8
Optimization   = reml                               Number of obs =        40
           AIC = -140.06                            Model chi2(2) =     25.91
Log likelihood = 75.030097                            Prob > chi2 =    0.0000
------------------------------------------------------------------------------
       logrr | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
        dose |  -.0013479    .004476    -0.30   0.763    -.0101206    .0074248
       dose2 |     .00016   .0000853     1.88   0.061    -7.20e-06    .0003272
------------------------------------------------------------------------------
------------------------------------------
  Random-effects parameters  |   Estimate
-----------------------------+------------
var(dose,dose)               |    .0000268
var(dose2,dose2)             |    8.71e-09
cov(dose,dose2)              |   -4.83e-07
------------------------------------------
LR test vs. no random-effects model = .15184004      Prob >= chi2(3) = 0.9850
```

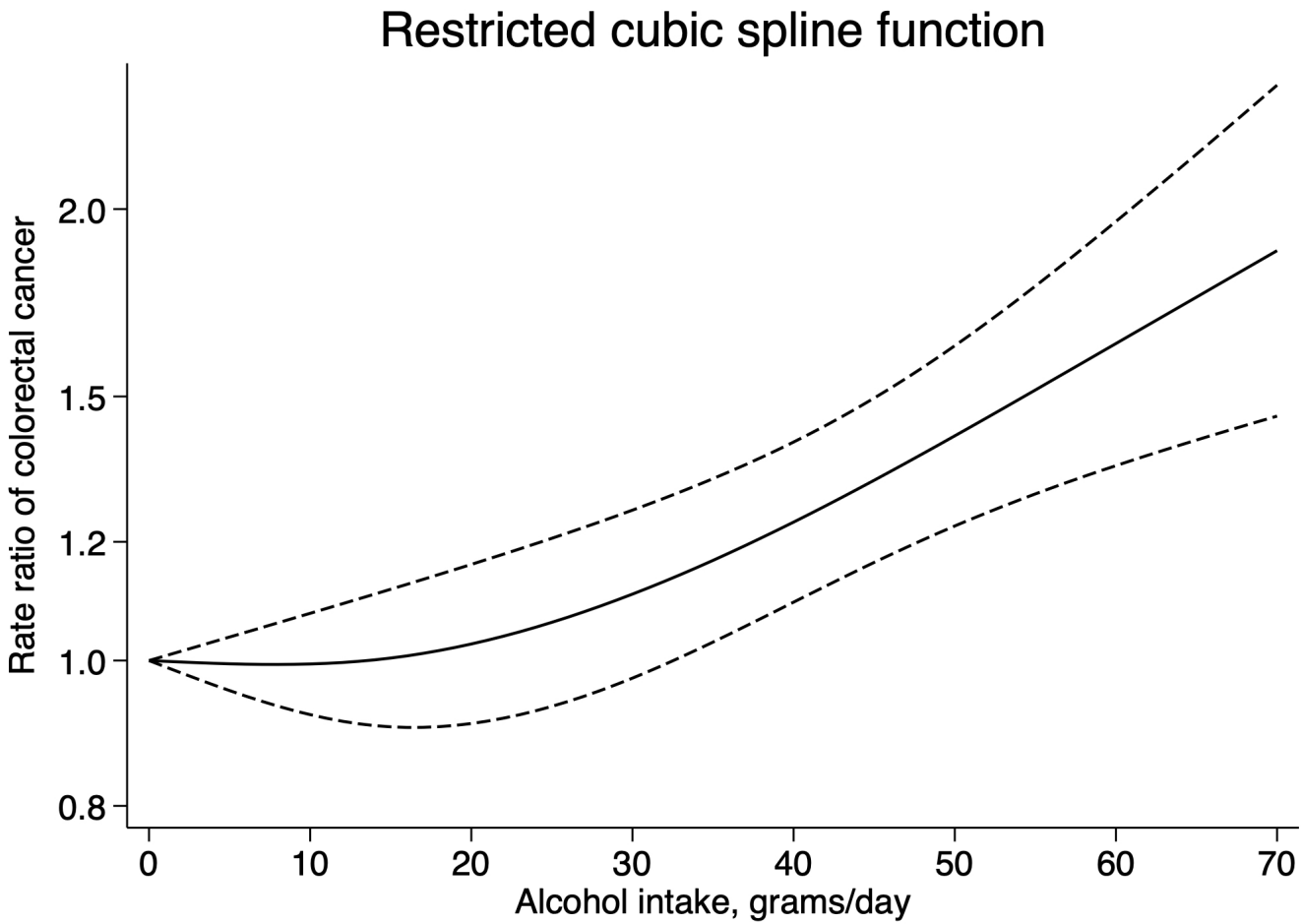
Graph the predicted summary dose-response function



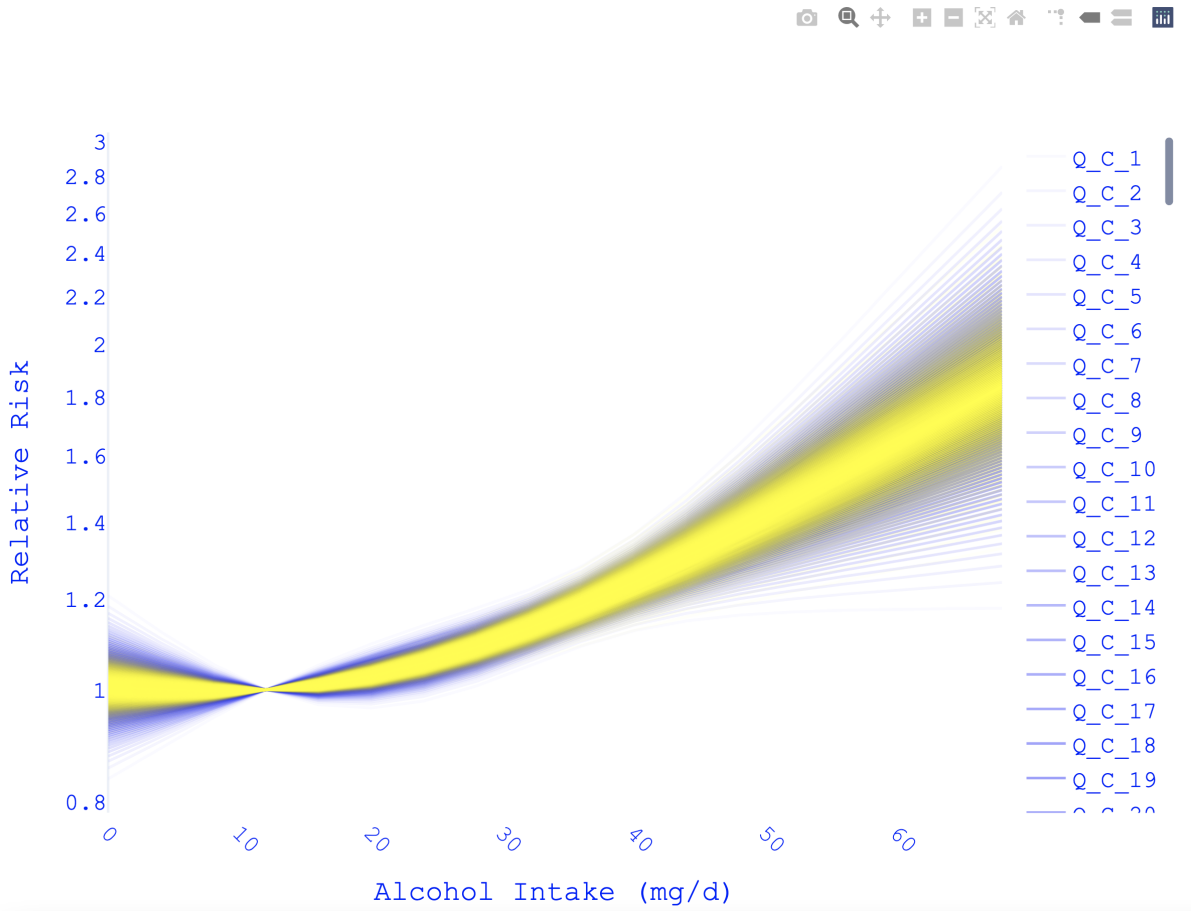
Restricted cubic spline function for alcohol intake

```
Two-stage random-effect dose-response model      Number of studies =         8
Optimization   = reml                               Number of obs =        40
           AIC = -72.82                             Model chi2(2) =     26.17
Log likelihood = 41.411693                            Prob > chi2 =    0.0000
------------------------------------------------------------------------------
       logrr | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
      doses1 |  -.0011522   .0042726    -0.27   0.787    -.0095264     .007222
      doses2 |   .0207752    .010731     1.94   0.053    -.0002571    .0418075
------------------------------------------------------------------------------
-----------------------------------------
  Random-effects parameters  |   Estimate
-----------------------------+-----------
var(doses1,doses1)           |    .0000323
var(doses2,doses2)           |    .0001858
cov(doses1,doses2)           |   -.0000775
-----------------------------------------
LR test vs. no random-effects model = .30925274      Prob >= chi2(3) = 0.9583
```

Graph the predicted summary dose-response function



Marginal and conditional quantiles



Click Here

What comes next

- Revise some concepts on heterogeneity.
- Better characterize what we mean by heterogeneity in the dose–response associations.
- How we deal with, and measure statistical heterogeneity.
- Subgroup analyses, and meta-regression models.

Relevant research questions

The aim of a dose–response meta–analysis is not only to provide a summary association, but also to make sense of the pattern of the study-specific curves.

- Is there evidence of heterogeneity in the study-specific curves?
- What is the variability of the study-specific curves?
- What are the implications of the observed heterogeneity?
- What proportion of the observed dispersion can be attributed to differences among studies?
- Can we explain (part of) the observed heterogeneity?

Heterogeneity across studies

Differences between studies might be due to:

- Design and follow-up.
- Populations, participants and patients.
- Treatment or exposure definition, type of intervention.
- Outcome definition.
- ...

Quantifying heterogeneity

Heterogeneity refers to the (excess of) variability in the effect sizes across studies.

Intuition from a forest plot:

- *Consistency in effect sizes*: if the effect sizes are similar and tightly clustered around the overall effect estimate (low heterogeneity).
- *Divergence in effect sizes*: wide variation with confidence intervals that do not overlap (high heterogeneity).

Subjective interpretation and black/white description.

We want to better describe and quantify this variation.

Better understanding statistical heterogeneity

Let's start from the simpler linear trend.

A set of alternative measures and statistics are available which tackle different aspect of the heterogeneity:

- Q statistic and the Q test.
- Between-studies variance (T^2) and standard deviation (T).
- Ratio of true heterogeneity to total observed variation (I^2).
- Prediction intervals.

The Q statistic

Q quantifies the total amount of observed variation in the study-specific linear trend.

$$Q = \sum_{i=1}^{I} W_i(\hat{\beta}_i - \hat{\beta})^2 = \sum_{i=1}^{I} \left(\frac{\hat{\beta}_i - \hat{\beta}}{\text{SE}\left(\hat{\beta}_i\right)} \right)^2$$

where W_i and $\hat{\beta}$ are the weights and summary measure from a fixed-effect model.

Under the fixed-effect assumption: $E[Q] = \text{I} - 1$.

$Q - \text{I} - 1$ is the excess in the variation in the study-specific linear trend.

The Q test

$H_0 : \tau^2 = 0$, i.e. that the underlying true effect size is the same for all studies.

Under the null hypothesis, $Q \sim \chi^2 (\mathrm{df}) = I - 1$

p-value $= \Pr (Q \geq q | H_0)$

Failing to reject the null hypothesis does not imply that studies are homogeneous. It is known that the test for heterogeneity:

- has low power when only a few studies;
- is possibly oversensitive when many studies.

Estimating τ^2 and τ

τ^2 and τ are the variance and standard deviation of the true effect sizes. They are on the same metric as the effect and reflects the absolute amount of variation.

Many different estimators exist, with then restricted maximum likelihood method being the most popular.

They are difficult to interpret but can be used to describe the distribution of study-specific linear trend about the mean effect (2-sigma rule).

$$\Pr\left(\hat{\beta} - 2\hat{\tau} \leq \hat{\beta}_i \leq \hat{\beta} + 2\hat{\tau}\right) \approx 95.45\%$$

A measure of Inconsistency (I^2)

I^2 which quantifies the amount of total variability attributed to between study heterogeneity.

$$I^2 = \frac{Q - \mathrm{df}}{Q} \times 100\% = \frac{\mathrm{Variance}_{\mathrm{between}}}{\mathrm{Variance}_{\mathrm{total}}} \times 100\%$$

- Intuitive interpretation (percentage).
- It is easy to calculate from previously published meta-analysis.
- It reflects the extent of overlap of confidence intervals.

It is convenient to interpret as a measure of inconsistency, and not as a measure of the real variation.

Prediction intervals

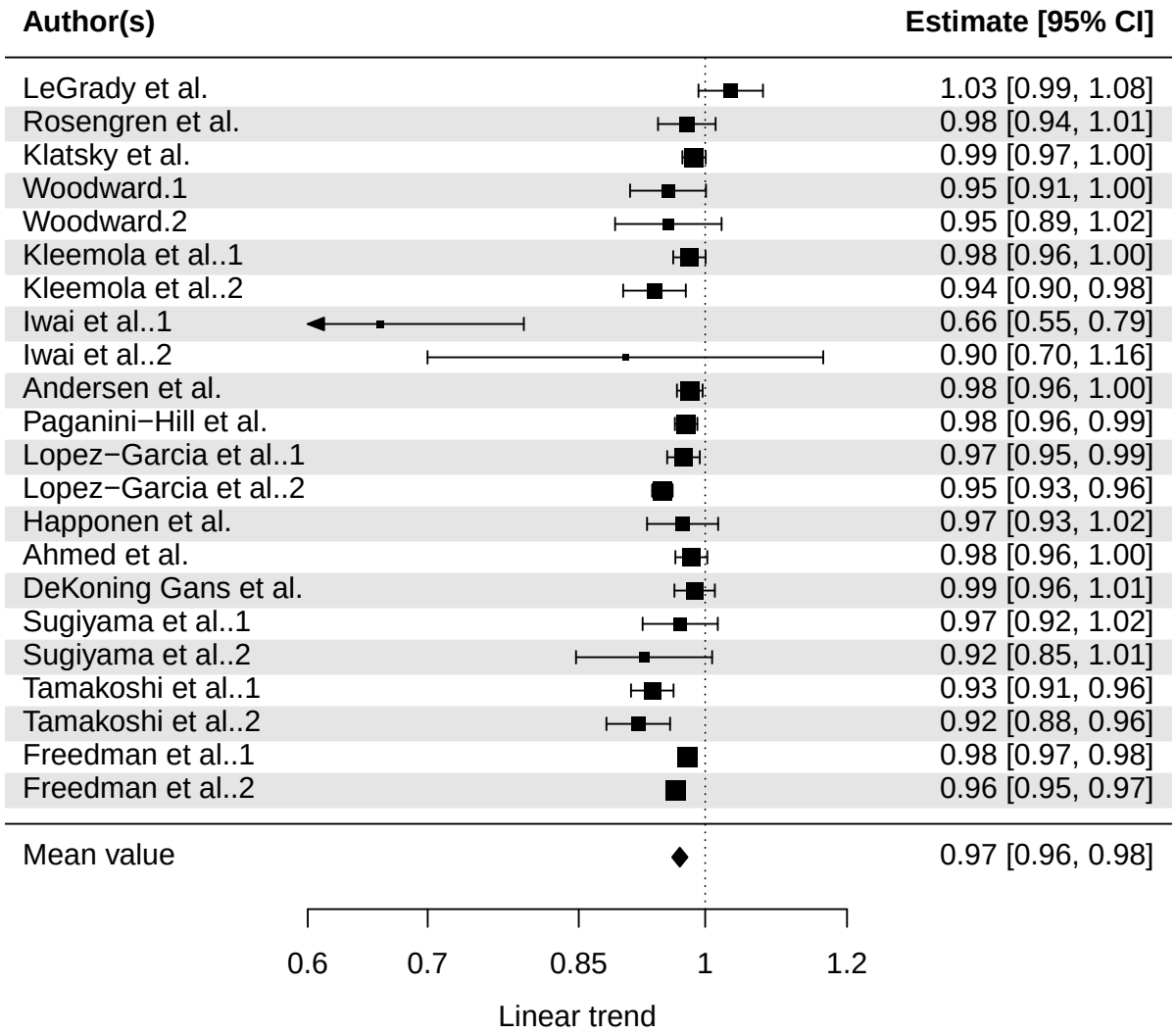
Prediction intervals in a meta-analysis estimate the range within which the true effect size of a new study is expected to fall.

$$\hat{\beta} \mp t_{I-2} \cdot \sqrt{\hat{\tau}^2 + \widehat{\mathrm{SE}}\left(\hat{\beta}\right)}$$

Clinicians can use prediction intervals to assess the potential range of treatment effects in individual patients, considering the heterogeneity observed across studies.

Decision-making should consider both the point estimate and the width of the prediction interval to account for uncertainty due to heterogeneity.

Coffee consumption and all-cause mortality (linear trend)



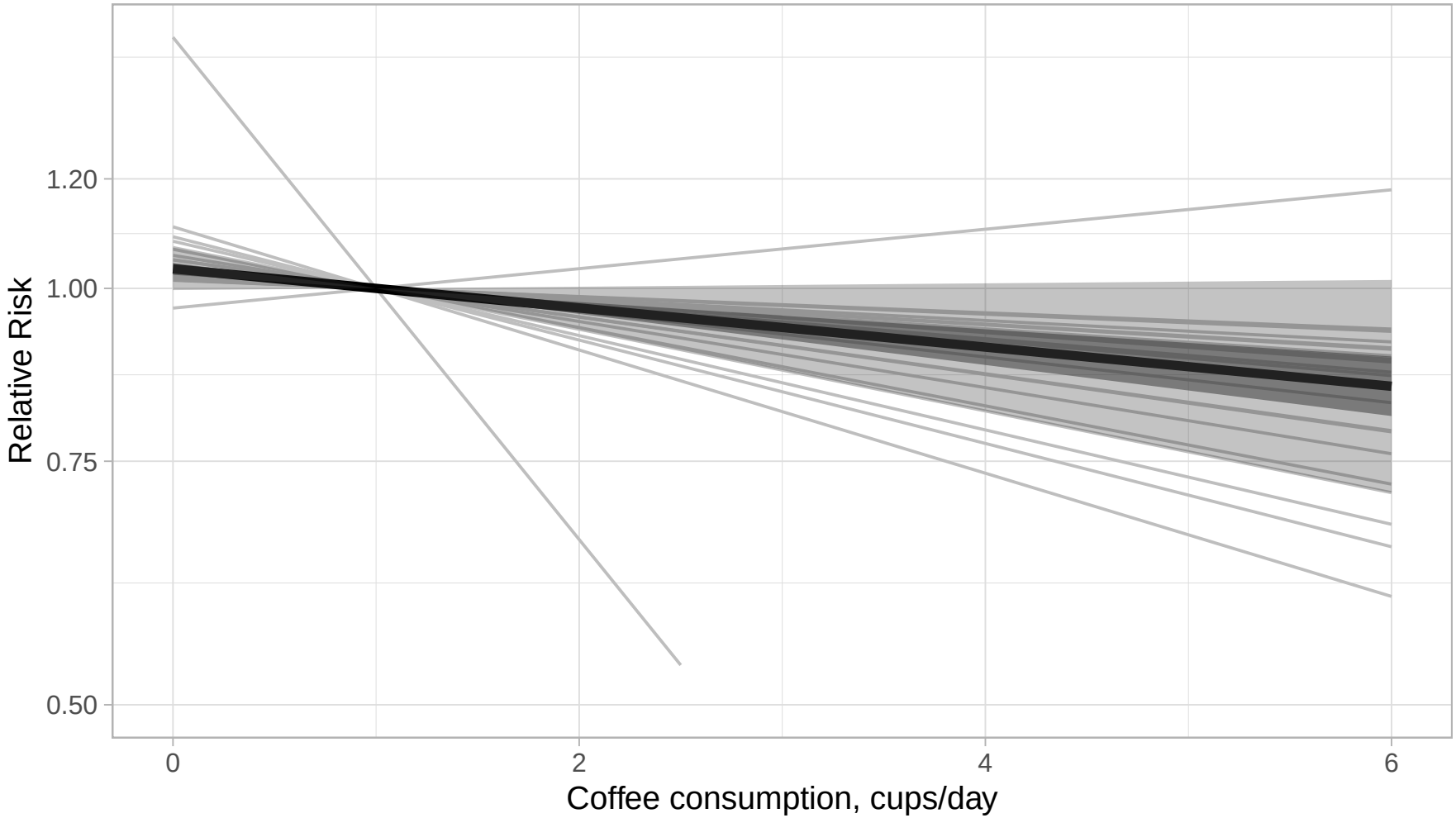
Coffee and all-cause mortality (linear trend)

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

Chi2 model: X2 = 41.8249 (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  -6.4672    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
 37.5676  -71.1352  -69.0462
```

Interpretation of heterogeneity measures

- There was an excess of variability ($Q - df$) in the study-specific linear trend over the expected ($df = 21$)-
- The observed excess in expected variability was statistically significant ($p \leq 0.001$).
- The proportion of the total variation attributable to between-studies heterogeneity was 72.7 %.
- The between-study variance was 0.0172.
- The prediction interval for the linear trend of a new study is (0.93, 1.00)

Prediction intervals

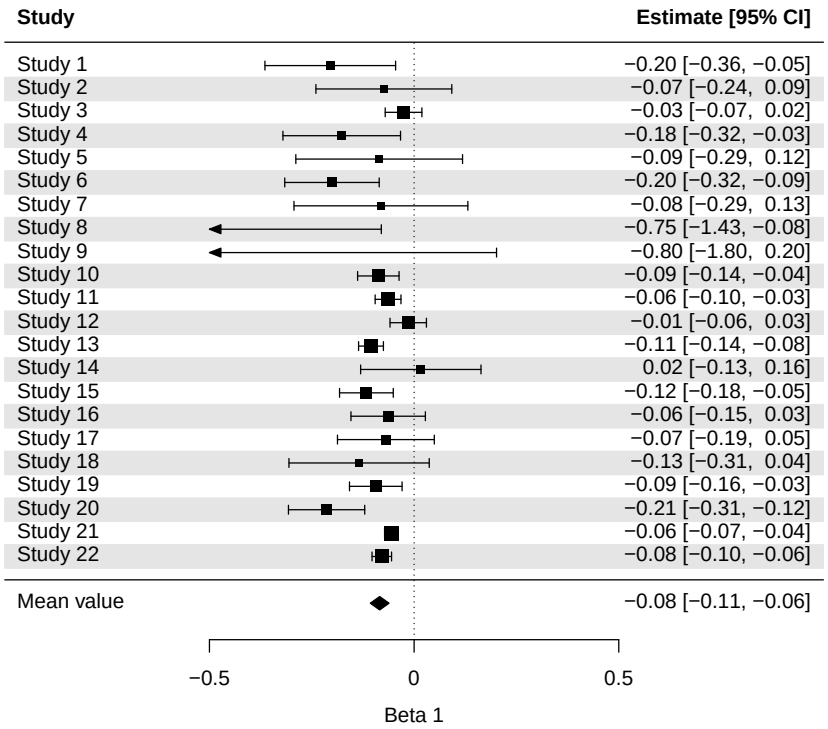


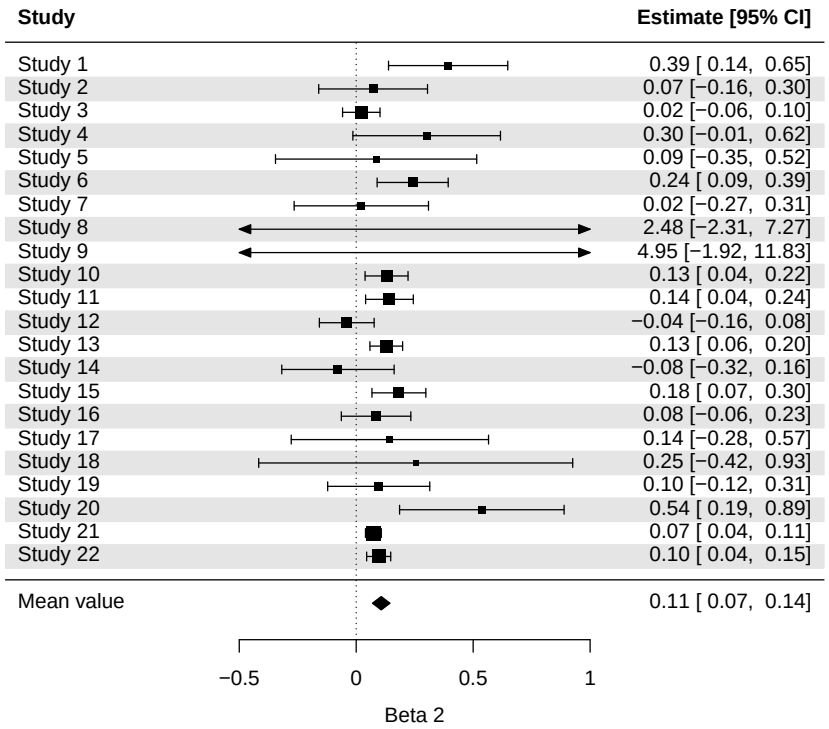
Extension to multivariate meta-analysis

Heterogeneity measures can be extended for non-linear dose–response ($p \geq 1$).

- $Q = \sum_{i=1}^{I} \left(\hat{\beta}_i - \hat{\beta}\right)^{\top} S_i^{-1} \left(\hat{\beta}_i - \hat{\beta}\right)$
- $Q \sim \chi^2(I - p)$
- Estimate of the (co)variance matrix Ψ
- $I^2 = \max\left\{\frac{Q-I-p}{Q}, 0\right\}$
- Prediction intervals.

Coffee consumption and all-cause mortality (splines)





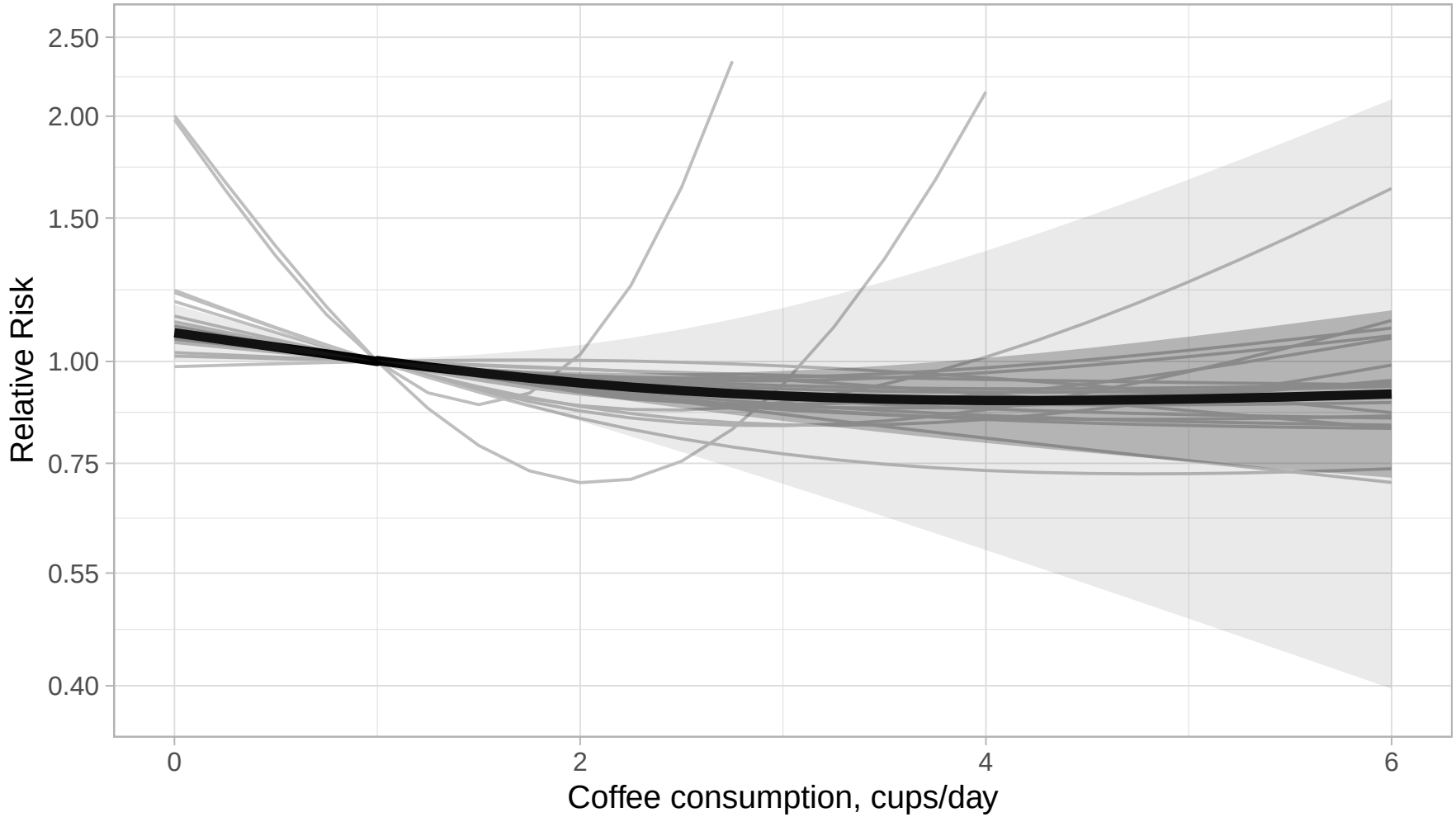
Coffee and all-cause mortality (splines)

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

Chi2 model: X2 = 74.3481 (df = 2), p-value = 0.0000

Fixed-effects coefficients
                   Estimate  Std. Error        z  Pr(>|z|)  95%ci.lb  95%ci.ub
rcs(dose, k)dose    -0.0811      0.0115  -7.0780    0.0000   -0.1035   -0.0586
rcs(dose, k)dose2    0.1072      0.0217   4.9451    0.0000    0.0647    0.1497

Between-study random-effects (co)variance components
                   Std. Dev              Corr
rcs(dose, k)dose     0.0372  rcs(dose, k)dose
rcs(dose, k)dose2    0.0707           -0.9601

Univariate Cochran Q-test for residual heterogeneity:
Q = 101.3912 (df = 42), p-value = 0.0000
I-square statistic = 58.6%

22 studies, 44 values, 2 fixed and 3 random-effects parameters
  logLik       AIC       BIC
 49.3872  -88.7745  -80.0861
```

Interpretation of heterogeneity measures

- There is still an excess of variability ($Q - df$) in the study-specific spline coefficients over the expected (df), which was statistically different from 0.
- The proportion of the total variation attributable to between-studies heterogeneity reduced to 58.6 %.
- The are now 3 (co)variance components.

Prediction intervals



What to do in case of substantial heterogeneity?

If the studies are highly heterogeneous you need to determine how to proceed.

- Don't combine results. Do a qualitative systematic review (evidence narratively)
- Combine with a random effects model.
- Combine only a subset of initial studies (those that are similar).
- Try to explain heterogeneity:
  - Subgroup analyses
  - Meta-regression models

Subgroup analyses

Explore whether the mean dose–response association varies across different subgroups of studies (effect modifier z at the study level).

Examples of subgroup variables include study-level characteristics (e.g. age, gender, baseline risk), study design, geographical area, ...

- Fit separate dose–response meta-analysis for the levels of the subgroup variable.
- Descriptively compare the dose–response coefficients and predicted curve.

How to compare curves

Is there any evidence of differential dose-response curve?

$$H_0 : \beta(z = 1) = \beta(z = 2) = \beta(z = 3)$$

Wald-test for multivariate correlated hypotheses.

Could I explain the observed heterogeneity?

Look for residual heterogeneity within stratified dose–response meta-analysis.

Possible limitations

It is important to pre-specify subgroup analyses to avoid bias and data-driven interpretations. Subgroup analyses are useful for one

categorical subgroup variable with a few levels. More subgroup

variables and/or continuous study-level predictors can be analyzed in a (multivariate) meta-regression model.

Meta-regression models

Formal statistical model for modeling the relationship between study characteristics (covariates) and dose–response coefficients

$$E\left[\hat{\beta}_i|Z\right] = \beta + \gamma z_i$$

z_i study-level variable in the i-th study.

Can theoretically be extended to continuous and multiple variable.

Advantages and limitations

- Statistical modelling of the differential dose–response curves.
- Formal hypothesis testing
  $H_0 : \beta(z = 1) = \beta(z = 2) = \beta(z = 3)$.
- Better quantification of the explained heterogeneity.

As for subgroup analyses, careful interpretation:

- Pre-specified comparisons.
- Limited number of studies.
- Possibility of ecological fallacy.
- Residual heterogeneity.

Coffee consumption and mortality