



Workshop on Dose–Response Meta–Analysis

Day 1

Nicola Orsini, Alessio Crippa

17th April 2024

Schedule

- ▶ **April 17th: Inghesalen**
 - ▶ Session 1: 9:00 - 12:00
 - ▶ Session 2: 13:00 - 16:00
- ▶ **April 18th: Karolina**
 - ▶ Session 3: 9:00 - 12:00
 - ▶ Session 4 13:00 - 16:00
- ▶ **April 19th: Karolina**
 - ▶ Session 5: 9:00 - 10:15
 - ▶ Group work: 10:15 - 12:00
 - ▶ Conclusions 13:00 - 14:30

Each session will be structured as:

- ▶ ~ 1.5 h theoretical lecture
- ▶ ~ 15 min coffee/fika break
- ▶ ~ 1.5 h hands-on computer sessions (in groups)

All material will be available on [Canvas](#)

Computer sessions

You will work in groups on a common set of questions, with each group analyzing a different problem and data set.

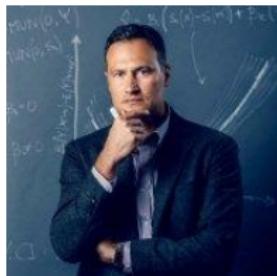
Questions, and Stata and R code for answering the questions will be available on [Canvas](#) at the start of the hands-on computer sessions.

For R, you can also find the scripts for the solution in a dedicated [Posit Cloud](#), where the code can be executed without need of setting up R and the required packages.

Main objectives

- ▶ Understand concepts and methodologies for conducting dose-response meta-analyses
- ▶ Be able to dose transformations such as splines
- ▶ Interpret different meta-analytic models (common effect, random-effects)
- ▶ Assess and quantify statistical heterogeneity
- ▶ Develop practical implementation skills

About the teachers



Nicola Orsini

- ▶ Highly Cited Researcher for the work on dose-response meta-analysis
- ▶ Wrote two book chapters and co-authored many dose-response meta-analyses.
- ▶ Author of the `glst` and `drmeta` Stata commands.



Alessio Crippa

- ▶ Earned his PhD on developments for dose-response meta-analysis.
- ▶ Author of the `dosresmeta` R package.
- ▶ Published and co-authored dose-response meta-analyses in different fields.

Main learning points of Day 1

- ▶ Conduct a dose-response meta-analysis based on individual and aggregated data
- ▶ Estimate and interpret linear dose-response models with different study designs and outcomes
- ▶ Conduct and present statistical inference (test, confidence interval) under common effect and random-effects meta-analysis
- ▶ Present graphically the estimated linear dose-response model using a chosen referent

Blood Pressure Effects of Sodium Reduction: Dose–Response Meta-Analysis of Experimental Studies. *Circulation* 2021

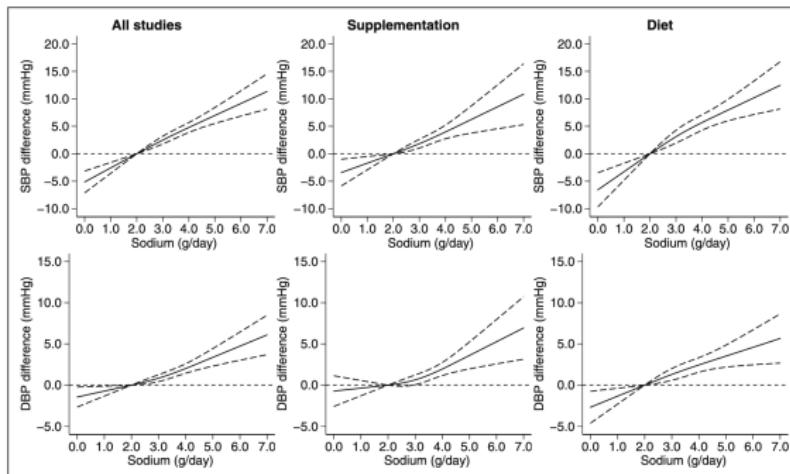
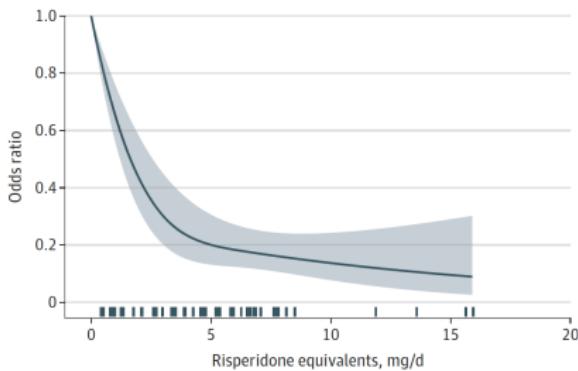


Figure 2. Dose-response meta-analysis of changes in SBP and DBP levels (mm Hg) according to achieved sodium excretion in the treatment and control groups at the end of the trials (all studies) and by type of intervention (supplementation or diet).

The average curve (solid line) with 95% confidence limits (dashed lines) was estimated with a 1-stage random-effects restricted cubic spline model, using 2 g/d as referent. DBP indicates diastolic blood pressure; and SBP, systolic blood pressure.

Examination of Dosing of Antipsychotic Drugs for Relapse Prevention in Patients With Stable Schizophrenia. *JAMA Psychiatry* 2021

Figure 1. Relapse



The dose-response curve for the primary outcome relapse after pooling all drugs using the primary scientific dose-equivalence method (the maximum effective dose method). The marks on the x-axis indicate for which doses data from study arms were available. A total of 26 studies with 71 individual dose comparisons including 1742 patients included 14 publications (Table 2).

Daily steps and all-cause mortality: a meta-analysis of 15 international cohorts. *Lancet Public Health* 2022

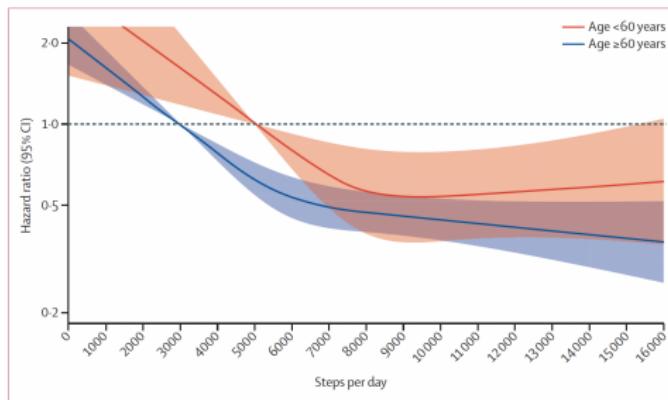


Figure 3: Dose-response association between steps per day and all-cause mortality, by age group
 Thick lines indicate hazard ratio estimates, with shaded areas showing 95% CIs. Reference set at the median of the means in the lowest quartile group (age ≥60 years = 3000 steps per day and <60 years = 5000 steps per day). Model is adjusted for age, accelerometer wear time, race and ethnicity (if applicable), sex (if applicable), education or income, body-mass index, and study-specific variables for lifestyle, chronic conditions or risk factors, and general health status. $p_{\text{nominal}} = 0.012$ by age group. 14 studies included in spline analysis, excluded Baltimore Longitudinal Study of Aging.³³ The y-axis is on a log scale.

The world's most comprehensive analysis of cancer prevention and survival research



World
Cancer
Research
Fund International



Analysing research on cancer prevention and survival



CONTINUOUS UPDATE PROJECT

Dose-response analysis in health risk assessment



- ▶ Consortium of researchers from 4 EU Member States (Italy, Sweden, Greece, Portugal)
- ▶ Example: Potassium intake in relation to blood pressure levels in adult population

What's in common in these examples?

- ▶ There is a quantitative factor measured in either experimental or observational studies
- ▶ Effect sizes can be of any type (mean difference, odds ratios, hazard ratios)
- ▶ Research questions are about the shape of the dose-response relationship or some specific less known aspects of it
- ▶ Design of the meta-analysis can be either retrospective (previously published) or prospective (pooling projects)
- ▶ Meta-analysis models are specified and estimated to learn from multiple studies

Example of study protocol (17 pages)

Environment International 157 (2021) 106828



ELSEVIER

Contents lists available at [ScienceDirect](#)

Environment International

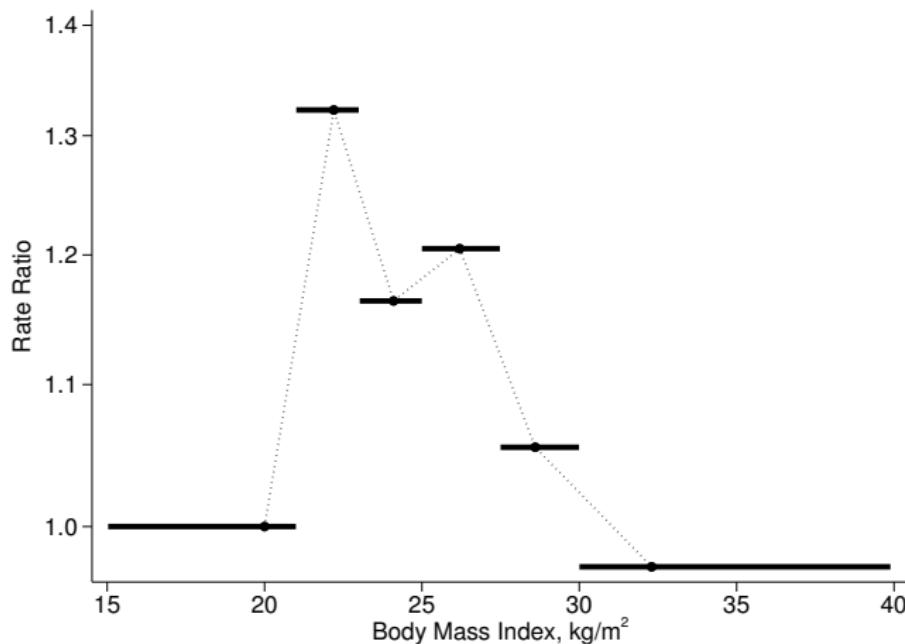
journal homepage: www.elsevier.com/locate/envint

The effect of exposure to radiofrequency fields on cancer risk in the general and working population: A protocol for a systematic review of human observational studies

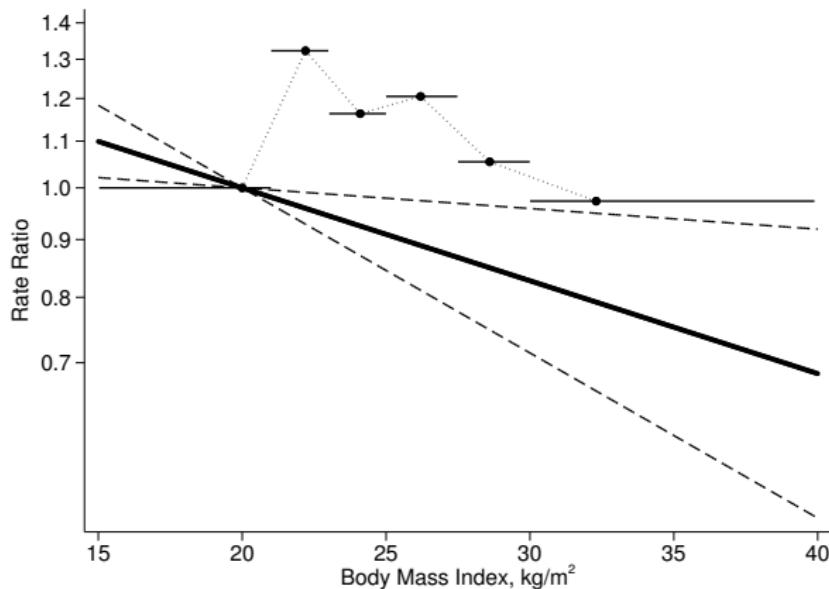
Example of section on dose-response meta-analysis

-
- c. We will perform dose-response *meta*-analyses of neoplasm risks per cumulative call time and total number of calls. We will use weighted mixed effects models suitable for table of correlated estimates ([Crippa et al. 2019](#); [Orsini 2021](#)). A single exposure value is assigned to each category based on what has been reported (mean, median, midpoint) within each study. In case the typical exposure value within each exposure interval is not available from the publication, it will be assigned according to its distribution. We will use regression splines of different degrees to answer specific questions about the dose-response relationships ([Orsini 2021](#); [Orsini and Spiegelman 2020](#)). The heterogeneity of dose-response gradients across studies is taken into account by using random-effects for the regression coefficients of the exposure transformations. The main target of statistical inference (test of hypothesis, confidence intervals) is the pointwise dose-response relationship for the average study. To examine the magnitude of heterogeneity across studies, the best linear unbiased predictions (BLUP) of the random effects will be used. A comparison of alternative candidate dose-response models will be done using the Akaike Information Criteria, balancing goodness of fit and overall number of parameters. Stratified analyses according to relevant design or scientific factors (e.g., gradient of susceptibility to systematic and differential exposure measurement errors) will be done using weighted mixed effects model.

Plot of the data for a single study



Alternative parametrizations of the exposure may not be graphically comparable



Data can be available in different forms

- ▶ **Individual Patient Data (IPD):**
 - ▶ Uses raw data from individual participants in each study
 - ▶ Requires collaboration with study authors or access to data repositories.
- ▶ **Aggregated Data (AD):**
 - ▶ Uses aggregate data from published reports
 - ▶ Uses aggregated data from standardized federated analysis

The key question is about the shape of the dose–response relationship in light of multiples studies

Two-stage approach

- ▶ Step 1. Specify and estimate a dose-response model within each study
- ▶ Step 2. Specify meta-analytical model based of study-specific estimated coefficients in Step 1

Step 1. Linear dose-response model

- ▶ A linear function is extremely popular
- ▶ Constant effect measured by one regression coefficient β_i
- ▶ Easy statistical inference
- ▶ Easy visualization

Minimal notation

- ▶ Studies indexed by $i = 1, \dots, I$
- ▶ y_i is the study-specific outcome variable (binary, continuous, time-to-event)
- ▶ x_i is the study-specific quantitative exposure

Popular statistical model

- ▶ Mean

$$\mu_i|x_i = \beta_{0i} + \beta_{1i}x_i$$

- ▶ Risk/Odds

$$\pi_i|x_i = \text{invlogit}(\beta_{0i} + \beta_{1i}x_i)$$

- ▶ Rate

$$\lambda_i|x_i = \exp(\beta_{0i} + \beta_{1i}x_i)$$

Step 2. Meta-analysis

- ▶ Common Effect. Same effect underlying all the studies

$$\beta_1 = \beta_2 = \dots = \beta_I = \beta$$

- ▶ Random effects. Distribution of effects

$$\beta_i \sim \mathcal{N}(\beta, \tau)$$

Target of inference in common and random-effects

- ▶ Test of hypothesis/Confidence about the **common** effect

$$H : \beta = 0 \text{ vs } \bar{H} : \beta \neq 0$$

$$C(\beta < \hat{Q}_p) = p$$

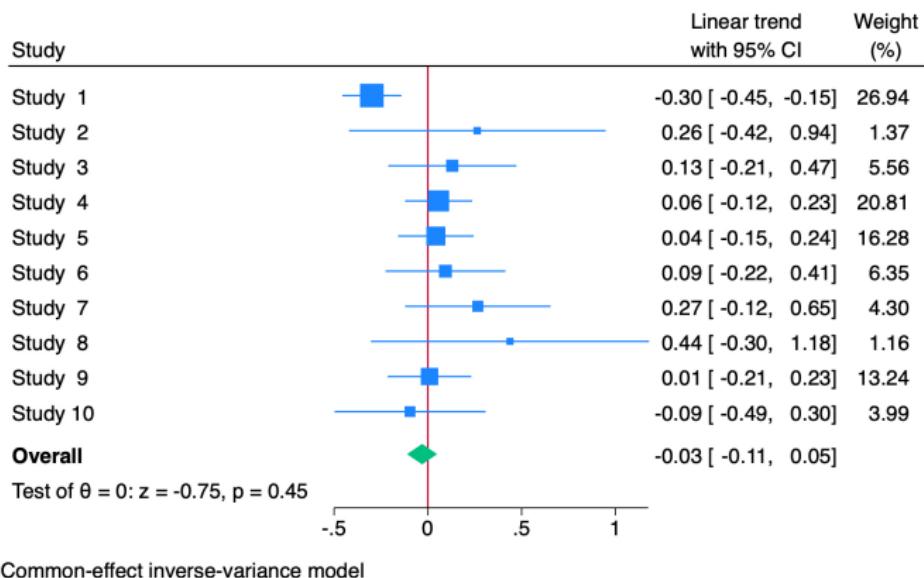
- ▶ Test of hypothesis/Confidence about the **summary** effect

$$H : \beta = 0 \text{ vs } \bar{H} : \beta \neq 0$$

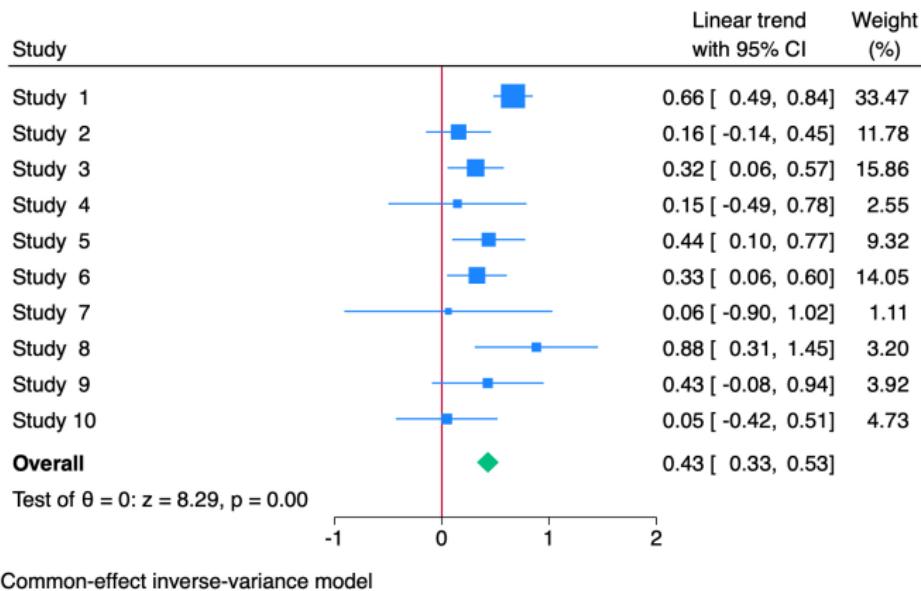
$$\text{Marginal Quantile } C(\beta < \hat{Q}_p^M) = p$$

$$\text{Conditional Quantile } C(\beta < \hat{Q}_p^C) = p$$

What can we learn from it? Case #1



What can we learn from it? Case #2



Common-effect inverse-variance model

IPD common effect meta-analysis of linear trends

Consider $I = 10$ studies of size $n = 1000$, investigating the conditional effect of right-skewed dose distribution $X \sim \chi^2(5)$ on the mean outcome.

The std deviation of the individual outcome, for any value of the dose is $\sigma_{Y_i} = 10$.

Every 1 unit increase in the dose is associated with a $\beta = 0.5$ unit increase in the average outcome.

All studies are arising from a **common** effect scenario where $\beta_{1i} = \beta = 0.5$

Study-specific standard error of the linear trend

An estimate of the study-specific standard error of the linear trend should be around

$$\widehat{SE}(\hat{\beta}_i) = \sigma_{Y_i}/(\sigma_{X_i}\sqrt{n - 1})$$

Of note, the std error for the i -th study is inversely related to both dose variability and sample size.

In our specific example the typical standard error would be

$$\widehat{SE}(\hat{\beta}_i) = 10/(\sqrt{5(2)}\sqrt{1000 - 1}) = 0.1$$

Study-specific sampling distribution of the linear trend

Justified by the central limit theorem, the distribution of study-specific estimated linear trends can be approximated by a bell-shaped and symmetric normal distribution.

$$\hat{\beta}_i \sim \mathcal{N}(\beta, \widehat{SE}(\hat{\beta}_i))$$

In our specific example, we have

$$\hat{\beta}_i \sim \mathcal{N}(0.5, 0.1)$$

Estimate the common linear trend

An estimate $\hat{\beta}$ of the common linear trend β underlying the I studies can be obtained with the inverse-variance method

$$\hat{\beta} = \frac{\sum_{i=1}^I \hat{\beta}_i w_i}{\sum_{i=1}^I w_i}$$

where the weights are the inverse of the estimated variance of the study-specific linear trends

$$w_i = \frac{1}{\widehat{SE}(\hat{\beta}_i)^2}$$

Standard error of the common linear trend

An estimate of the standard error associated with the estimated $\hat{\beta}$ common linear trend β can be obtained as one over the square root of the sum of study-specific weights

$$\widehat{SE}(\hat{\beta}) = \frac{1}{\sqrt{\sum_{i=1}^I w_i}}$$

The sampling distribution of the estimated common linear trend, target of statistical inference, can be approximated by a bell-shaped and symmetric normal distribution.

$$\hat{\beta} \sim \mathcal{N}(\beta, \widehat{SE}(\hat{\beta}))$$

Sampling distribution in our example

In our specific example, we have

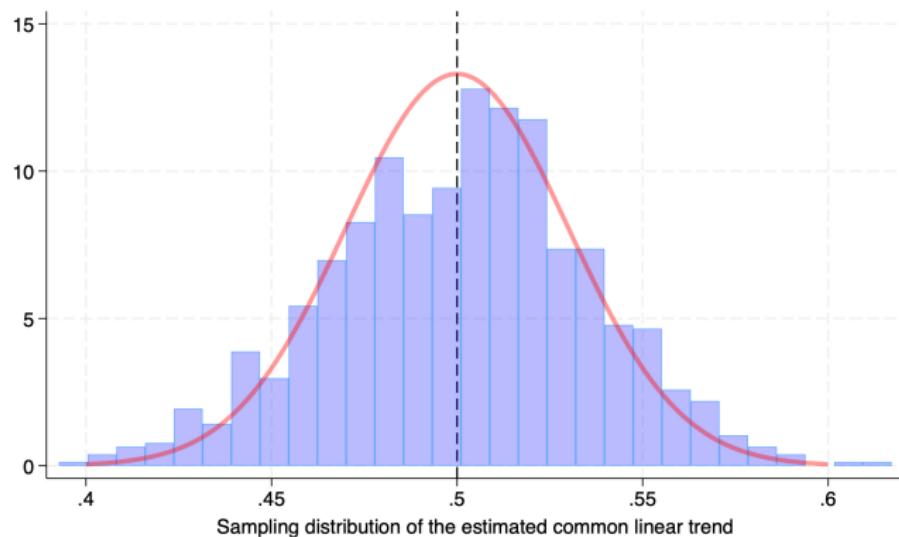
$$\widehat{SE}(\hat{\beta}_i) = 10 / (\sqrt{5(2)} \sqrt{1000 - 1}) = 0.1$$

$$\widehat{SE}(\hat{\beta}) = 1 / \sqrt{(1/0.1^2)10} = 0.03$$

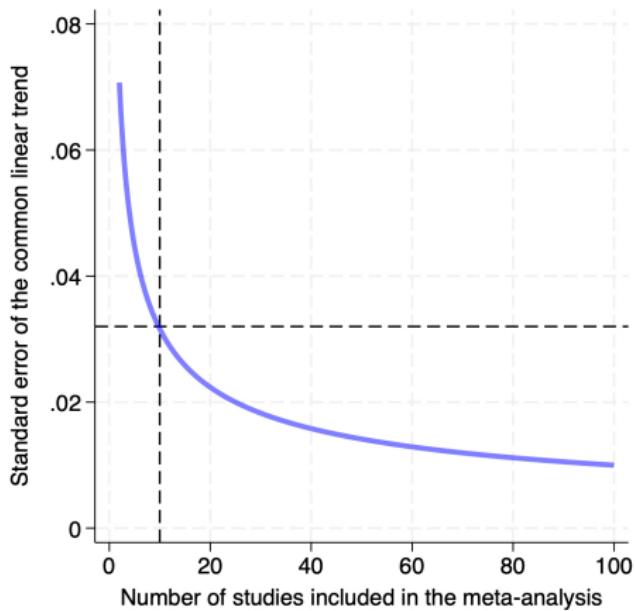
$$\hat{\beta} \sim \mathcal{N}(0.5, 0.03)$$

A simulation study can be used to verify the above approximations by replicating the meta-analysis a large number of times.

Simulated sampling distribution



Standard error decreases slowly with the number of studies



The logic of the statistical test

An empirical estimate is located in the sampling distribution centered about the parameter value used to divide the parameter space (i.e. null and alternative hypothesis).

$$H : \beta = 0$$

$$\bar{H} : \beta \neq 0$$

The statistical test is conducted assuming the validity of the hypothesis of no common effect of the dose. A test result (incompatible, compatible) is obtained by specifying the frequency of the type I error and the direction of the test (one-sided, two-sided).

Two-sided Wald-type test

A Wald-type statistic w is computed as follows $w = \frac{\hat{\beta} - 0}{\widehat{SE}(\hat{\beta})}$

If the computed statistic w is exceeding the extreme quantiles (saying 0.025 and 0.975) of the standard normal distribution, the test would provide a strong indication of discrepancy between the sample of data and the null hypothesis.

A two-sided p -value is derived from a standard normal distribution
 $p = 2\phi(-|w|)$

The logic of confidence

The degree of confidence (denoted here as C) or rational belief an investigator can place in the claim that the unknown parameter β , common effect, is less than or equal to the empirical p -quantile is actually p .

$$C(\beta \leq \hat{Q}_p^{\hat{\beta}}) = p$$
$$p \in (0, 1)$$

The p -quantile of confidence, $\hat{Q}_p^{\hat{\beta}}$, is obtained by shifting and rescaling the p -quantile of a standard normal distribution.

Schweder, Tore, and Nils Lid Hjort. *Confidence, likelihood, probability*. Cambridge University Press, 2016.

Confidence interval

A symmetric 95% confidence interval for the common effect is derived by computing the 0.025 and 0.975 quantiles of the sampling distribution centered about the empirical estimate of the common effect.

$$\hat{Q}_{0.025}^{\hat{\beta}} = \hat{\beta} + \phi^{-1}(0.025) \widehat{SE}(\hat{\beta})$$

$$\hat{Q}_{0.975}^{\hat{\beta}} = \hat{\beta} + \phi^{-1}(0.975) \widehat{SE}(\hat{\beta})$$

It follows

$$C\left(Q_{0.025}^{\hat{\beta}} \leq \beta \leq Q_{0.975}^{\hat{\beta}}\right) = 0.95$$

Visualizations of meta-analytical findings

- ▶ Forest plot. Univariate meta-analysis.

- ▶ Dose-response plot. Univariate/Multivariate meta-analysis.

Forest plot

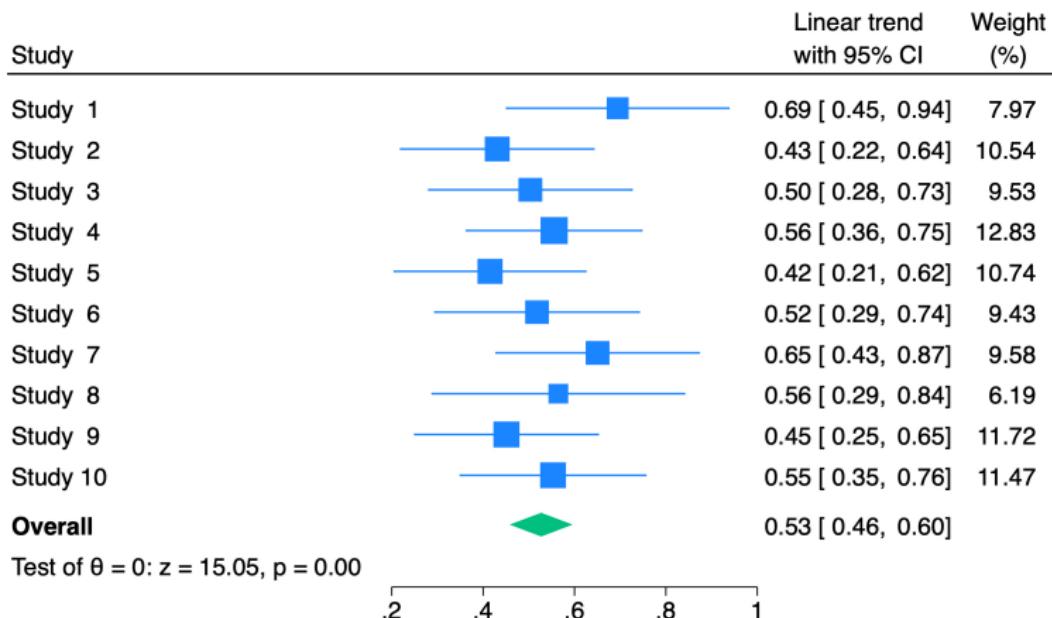
The forest plot shows study-specific inference (estimate and CI) together with statistical inference (test, CI) for the common dose effect.

The size s_i of the marker for a particular study is proportional to its study weight w_i divided by the sum of the weights.

$$s_i = w_i / \sum_{i=1}^I w_i$$

This relative weight (as %) is typically shown in the software output and graph.

Forest plot for 10 studies



Common-effect inverse-variance model

Dose-response plot

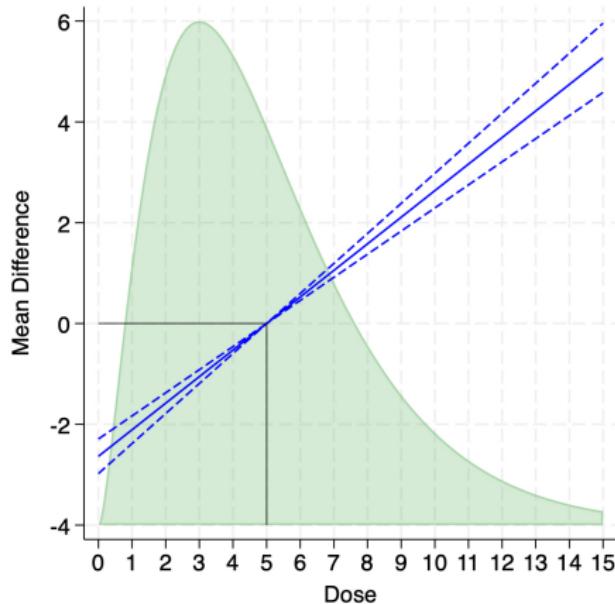
The dose-response plot shows statistical inference (typically 0.025, 0.500, and 0.975 estimated quantiles) for the common dose-response effect for a fine grid of plausible dose values x using an appropriate referent x_0 .

$$Q_{0.025} = \hat{\beta}(x - x_0) + \phi^{-1}(0.025)\sqrt{\widehat{SE}(\hat{\beta})^2(x - x_0)^2}$$

$$Q_{0.500} = \hat{\beta}(x - x_0)$$

$$Q_{0.975} = \hat{\beta}(x - x_0) + \phi^{-1}(0.975)\sqrt{\widehat{SE}(\hat{\beta})^2(x - x_0)^2}$$

Dose-response plot for 10 studies



Study-specific dose-response plot

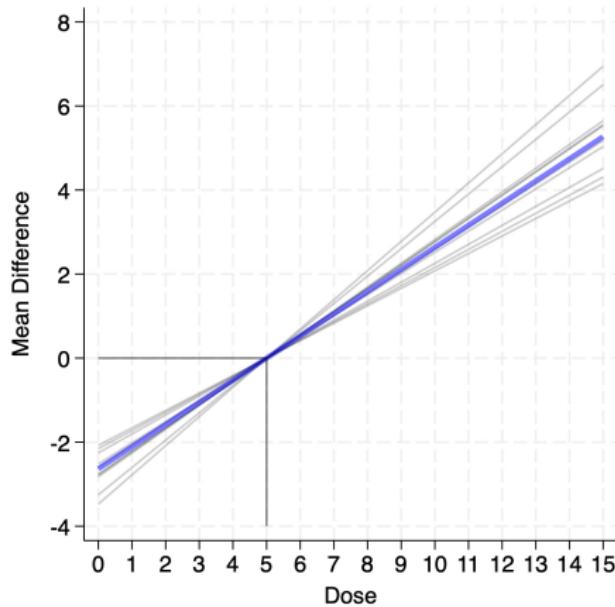
Study-specific dose-response plot ($Q_{0.500}$ is sufficient) can be used to show the variation across studies, in similar spirit to the forest plot, for a fine grid of plausible dose values x using an appropriate and common referent x_0 .

$$Q_{0.500} = \hat{\beta}_i(x - x_0)$$

$$i \in 1, 2, \dots, I$$

The common dose-response effect can be overlaid with the study-specific linear trends.

Study-specific dose-response plots for 10 studies



Random-effects linear dose-response model

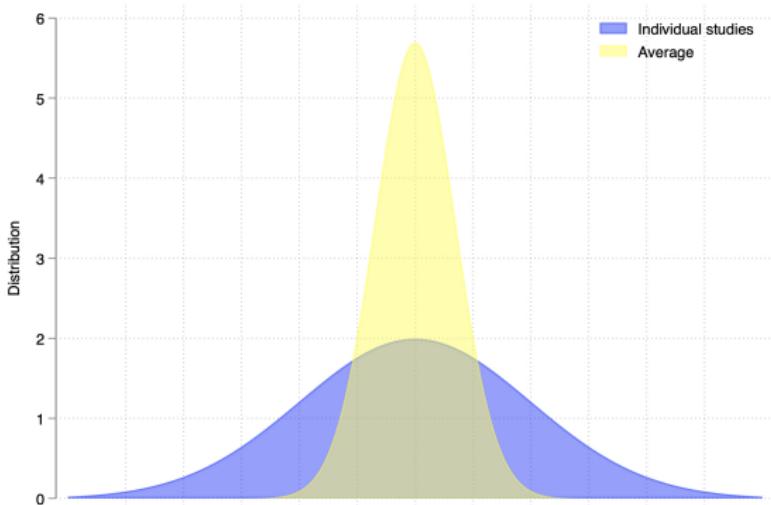
This model assumes a bell-shaped and symmetric distribution of linear dose-response trends across studies.

$$\beta_i \sim \mathcal{N}(\beta, \tau)$$

Target of statistical inference can be the summary of this distribution β and/or its spread τ .

Uncertainty of individual studies and its average

The graph below shows the distribution of the study-specific effects β_i and the sampling distribution of the estimated average $\hat{\beta}$.



Estimate the summary linear trend

An estimate $\hat{\beta}$ of the random-effects linear trend β can be obtained with the Der Simonian and Laird method (1986).

$$\hat{\beta} = \frac{\sum_{i=1}^I \hat{\beta}_i w_i}{\sum_{i=1}^I w_i}$$

where the weights are the inverse of the estimated variance of the study-specific linear trends

$$w_i = \frac{1}{\widehat{SE}(\hat{\beta}_i)^2 + \tau^2}$$

Standard error of the summary linear trend

An estimate of the standard error associated with the estimated $\hat{\beta}$ summary linear trend β can be obtained as one over the square root of the sum of study-specific weights

$$\widehat{SE}(\hat{\beta}) = \frac{1}{\sqrt{\sum_{i=1}^I w_i}}$$

The sampling distribution of the estimated summary linear trend, target of statistical inference, can be approximated by a bell-shaped and symmetric normal distribution.

$$\hat{\beta} \sim \mathcal{N} \left(\beta, \sqrt{\widehat{SE}(\hat{\beta})^2 + \tau^2} \right)$$

Statistical inference

The logic of the test of hypothesis and confidence is similar in common and random-effects meta-analysis.

An estimate of the variability across studies in the standard error has an impact of the estimated summary dose-response effect as well as its standard error.

The interpretation given to the β , however, should accord with the chosen meta-analytical model (i.e. common vs summary/average) dose-response effect).

Quantiles of confidence

- ▶ **Marginal quantiles.** What is the degree of confidence that can be assigned to an *inequality* regarding the *unknown* effect of the dose on a *typical study* in light of the data and specified statistical model?

- ▶ **Conditional quantiles.** What is the degree of confidence that can be assigned to an *inequality* regarding the *unknown study-specific effects* of the dose in light of the data and specified statistical model?

Extending the common effect example

The random-effect linear dose-response mechanism is $\beta_i \sim N(0.5, 0.2)$. Consider $I = 10$ studies of the same size $n = 1000$, equal dose distribution $X \sim \chi^2(5)$, and equal conditional outcome std deviation $\sigma_{Y_i} = 10$. Using a dose of 5 units as referent we have that

$$\beta_i(x - 5) \sim N(0.5(x - 5), 0.2(x - 5))$$

The typical standard error of the slope for the summary linear trend would be

$$\widehat{SE}(\hat{\beta}) = 1/\sqrt{1/(0.1^2 + 0.2^2)10} = 0.07$$

Marginal vs Conditional Quantiles

The degree of confidence (C) in the inequality below is p

$$C(\beta(x - 5)) \leq Q_p^M(\hat{\beta}(x - 5)) = p$$

where the **marginal** quantile would be

$$Q_p^M(\hat{\beta}(x - 5)) = 0.5(x - 5) + \phi^{-1}(p)\sqrt{(0.07^2)(x - 5)^2}$$

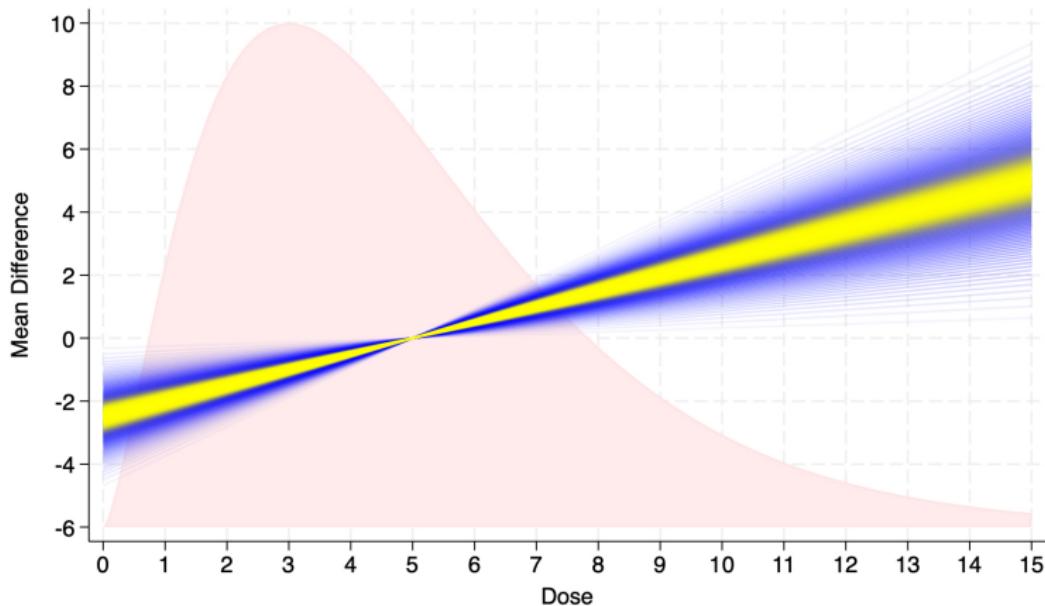
And similarly, the degree of confidence in the proposition below is p

$$C(\beta_i(x - 5)) \leq Q_p^C(\hat{\beta}_i(x - 5)) = p$$

where the **conditional** quantile would be

$$Q_p^C(\hat{\beta}_i(x - 5)) = 0.5(x - 5) + \phi^{-1}(p)\sqrt{(0.07^2 + 0.2^2)(x - 5)^2}$$

Marginal (yellow) and conditional (blue) quantiles



Extension to multiplicative models (logistic, survival)

- ▶ The logic of the meta-analysis and inference presented for the difference in mean outcomes is applicable to the difference in transforms of the mean outcomes (logit, log).
- ▶ To interpret the result one has to exponentiate the estimated regression coefficient $\hat{\beta}$ and estimated quantiles of confidence

What comes next

- ▶ Dose—response meta—analysis of aggregated data.
- ▶ The structure of aggregated dose—response data.
- ▶ How to estimate a linear trend of aggregated dose—response data.
- ▶ How to synthesize study-specific linear trend.

Dose-response meta-analysis on aggregated data

- ▶ Often accessible and efficient.
- ▶ Broader scope, with a larger number of studies and diverse populations.
- ▶ Validation of findings from individual studies.
- ▶ Higher consistency and transparency in research synthesis.
- ▶ Greater applicability to real-world settings and decision-making processes.

Two-stage approach

Similar strategy to meta-analysis of individual patient data.

- ▶ Stage 1: Estimate a dose–response model on *aggregated data*.
- ▶ Stage 2: Meta–analysis of study-specific dose-response coefficients.

It is possible to combine the two stages using a one-stage dose–response meta–analysis.

Aggregated dose-response data

- ▶ Also referred to as *summarized* or *tabular* or *published* dose-response data.
- ▶ They consist of contrasts (Odds Ratio, Mean Differences, Hazard Ratio, Relative Risks, ...) of different exposure categories to a common referent.
- ▶ They are easily accessible from publications or can be provided by principle investigators (there is no sensible information).

Case-control data on alcohol and breast cancer risk, AJE 1992

Category (g/day)	Assigned dose	cases	n	Adjusted OR	Lower bound	Upper bound
Ref.	0	165	337	1.00	1.00	1.00
≤ 2.5	2	74	167	0.80	0.51	1.27
2.5-9.3	6	90	186	1.16	0.73	1.85
≥ 9.3	11	122	212	1.57	0.99	2.51

$$y = \log(\text{OR})$$

$$\text{Var}(y) = \left(\frac{\log(\text{ub}) - \log(\text{lb})}{2 \cdot z_{1-\alpha/2}} \right)^2$$

Experimental data on aripiprazole dosage and severity of schizophrenia, *BMC Med Res Methodol* 2016

Assigned dose	Mean PANSS score	SD PANSS score	n	Mean difference	SE
Cutler 2006					
0	5.30	18.31	85	0.00	0.00
2	8.23	18.32	92	2.93	2.76
5	10.60	18.31	89	5.30	2.78
10	11.30	18.32	94	6.00	2.74
McEvoy 2007					
0	2.33	26.10	107	0.00	0.00
10	15.04	27.60	103	12.71	3.65
15	11.73	26.20	103	9.40	3.65
20	14.44	25.90	97	12.11	3.71

Observational data on alcohol intake (gr/day) and colorectal cancer rates, AJE 2012

Assigned dose	cases	peryears	logRR	SE
atm				
0.000	28	22185.73	0.000	NA
1.829	38	43030.54	-0.417	0.251
9.199	43	53088.96	-0.396	0.246
22.857	32	45348.09	-0.488	0.263
35.667	16	19790.79	-0.279	0.321
58.426	27	19919.85	0.202	0.286
hpm				
0.000	100	103002.30	0.000	NA
2.100	65	106825.51	-0.416	0.160
9.500	104	119845.65	-0.099	0.143
18.800	63	58034.23	0.094	0.165
36.700	46	33081.40	0.205	0.185
59.400	30	18454.93	0.343	0.220

Features

- ▶ Each row of data represents large samples of individual data.
- ▶ Outcome are estimated comparisons (OR, MD RR, HR).
- ▶ Difficult to model as comparisons depend on the chosen referent and it may vary across studies
- ▶ Comparisons are likely to be positively correlated within each study.

Study-specific linear dose-response model

For each study indexed by $i = 1, \dots, I$

$$y_{ij} = \beta_{1i} (x_{ij} - x_{i0}) + \varepsilon_{ij}$$

y_{ij} is the contrast for the j -th exposure category in the i -th study on the modeling scale (MD, log(OR), log(HR)).

x_{ij} is the assigned dose in the j -th exposure category, while x_{i0} is the referent dose in the i -th study.

ε_{ij} is the residual error.

- ▶ No intercept because ($y_{ij} = 0$ for $x_{ij} = x_{i0}$)
- ▶ Variance/covariance is known $\text{Cov}(\varepsilon_i) = \Sigma_i$

Covariance reconstruction

Two popular methods have been proposed for reconstructing the covariance structure Σ_i : Greenland & Longnecker (1992), and Hamling (2008).

Both utilize additional descriptive information such as the sample size (or total person-time) and number of cases within each exposure interval.

A similar strategy have be developed for approximating Σ_i for a continuous outcome based on the sample size and the SD (Crippa & Orsini, 2016).

Estimation of study-specific linear trend

Generalized Least Square (GLS) estimation

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

$$\hat{V}_i = \text{Var}(\beta_i) = (X_i^\top \Sigma_i^{-1} X_i)^{-1}$$

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

Ignoring the covariance:

- ▶ Less efficient estimates of regression slope.
- ▶ Less accurate variance estimates.
- ▶ Invalid heterogeneity tests.

Statistical inference

As for the inference on study-specific linear trend based on IPD, the estimators based on aggregated data are: consistent, efficient, and asymptotically normally distributed.

Wald-type test of hypothesis for $H_0 : \beta_i = b_0$ (such as no linear association $b_0 = 0$).

$1 - \alpha$ confidence intervals based on the normal approximation:

$$\hat{\beta}_i \pm z_{1-\alpha/2} \cdot \sqrt{\hat{v}_i}$$

Comparison between IPD and AD

There is a huge difference in the amount of data and effort for collecting and analysing the data between the type of analyses.

It has been shown a good agreement between the linear trend estimated on aggregated data and individual patient data.

Differences between the analyses on IPD and AD data:

- ▶ Number of exposure categories and used cutpoints.
- ▶ Assigned dose to each exposure category.
- ▶ Model definition for the analysis of IPD.
- ▶ Set of modelled confounders.

Random-effects meta-analysis

$$\hat{\beta}_i = \beta + b_i + \epsilon_i \text{ or } \hat{\beta}_i \sim \mathcal{N}(\beta, \hat{\nu}_i + \tau^2)$$

$\hat{\beta}_i$ is the linear trend estimated from the aggregated data in the i -th study.

β is the mean of the (true) study-specific linear trend.

b_i is distance between the mean and the study-specific (true) linear trend: $\text{Var}(b_i) = \tau^2$.

ϵ_i is residual deviation from the true linear trend in the i -th study:
 $\text{Var}(\epsilon_i) = \hat{\nu}_i$.

Marginal quantiles

Estimation and inference is exactly the same as presented for meta-analysis of linear trend based on individual patient data.

Marginal predicted dose-response based on a reference value x_0

$$\hat{y} = \hat{\beta} \cdot (x - x_0)$$

With a $1 - \alpha$ confidence intervals:

$$\hat{y} \pm z_{1-\alpha/2} (x - x_0) \sqrt{\hat{v}_i}$$

In case of modelling log relative measures, such marginal predictions need to be exponentiated.

How to present the results

The study-specific and average linear trend can be presented:

- ▶ Forrest plot
 - ▶ Display the study-specific \hat{b}_i and average $\hat{\beta}$.
 - ▶ Focus is on the interpretation of the linear trend interpretation.
- ▶ Prediction of \hat{y}
 - ▶ Display the study-specific predictions \hat{y}_i and average predictions \hat{y} for a set of x values.
 - ▶ Focus is on the interpretation on the scale of the outcome (y).

Software

- ▶ **dosresmeta** R package
- ▶ **drmeta** Stata command
- ▶ **%metadose** SAS macro



Systematic Reviews and Meta- and Pooled Analyses

Coffee Consumption and Mortality From All Causes, Cardiovascular Disease, and Cancer: A Dose-Response Meta-Analysis

Alessio Crippa*, Andrea Discacciati, Susanna C. Larsson, Alicja Wolk, and Nicola Orsini

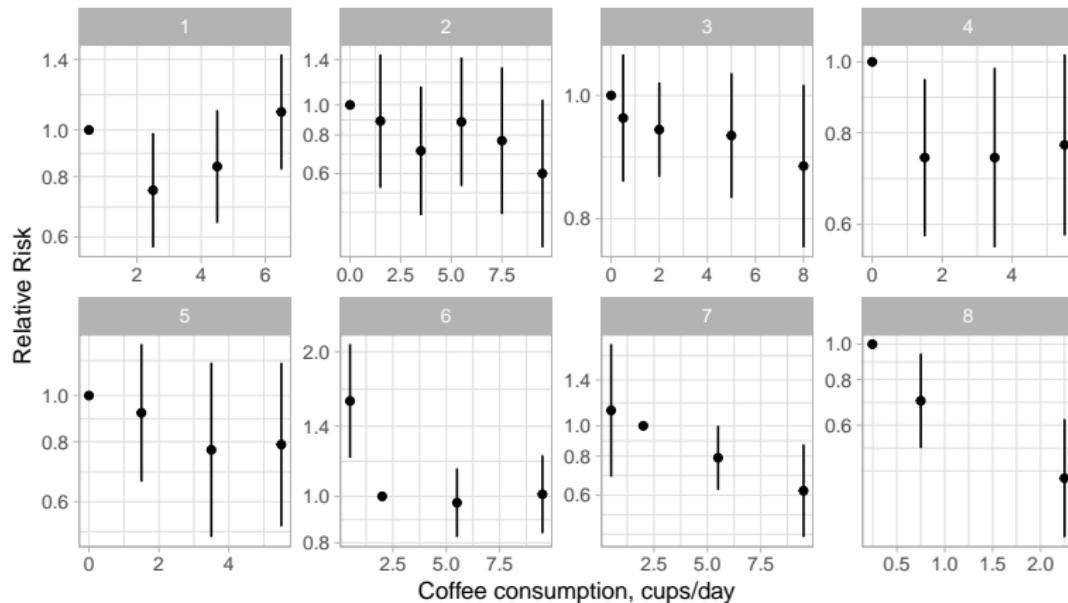
* Correspondence to Alessio Crippa, Institute of Environmental Medicine, Karolinska Institutet, Box 210, SE-171 77 Stockholm, Sweden (e-mail: alessio.crippa@ki.se).

Initially submitted March 11, 2014; accepted for publication July 1, 2014.

Several studies have analyzed the relationship between coffee consumption and mortality, but the shape of the association remains unclear. We conducted a dose-response meta-analysis of prospective studies to examine the dose-response associations between coffee consumption and mortality from all causes, cardiovascular disease (CVD), and all cancers. Pertinent studies, published between 1966 and 2013, were identified by searching PubMed and by reviewing the reference lists of the selected articles. Prospective studies in which investigators reported relative risks of mortality from all causes, CVD, and all cancers for 3 or more categories of coffee consumption were eligible. Results from individual studies were pooled using a random-effects model. Twenty-one prospective studies, with 121,915 deaths and 997,464 participants, met the inclusion criteria. There was strong evidence of nonlinear associations between coffee consumption and mortality for all causes and CVD (P for nonlinearity < 0.001). The largest risk reductions were observed for 4 cups/day for all-cause mortality (16%, 95% confidence interval: 13, 18) and 3 cups/day for CVD mortality (21%, 95% confidence interval: 16, 26). Coffee consumption was not associated with cancer mortality. Findings from this meta-analysis indicate that coffee consumption is inversely associated with all-cause and CVD mortality.

type	Dose	cases	n	logRR	SE	rr	lb	ub
Klatsky et al.								
ci	0.5	57	249	0.00	0.00	1.00	1.00	1.00
ci	2.5	136	655	-0.29	0.14	0.75	0.57	0.99
ci	4.5	144	619	-0.17	0.14	0.84	0.64	1.10
ci	6.5	115	387	0.09	0.14	1.09	0.83	1.43
ci	0.0	17	192	0.00	0.00	1.00	1.00	1.00
LeGrady et al.								
ci	1.5	88	1121	-0.12	0.25	0.89	0.54	1.46
ci	3.5	155	2464	-0.34	0.24	0.71	0.44	1.15
ci	5.5	155	1986	-0.13	0.24	0.88	0.55	1.42
ci	7.5	39	575	-0.27	0.28	0.77	0.44	1.32
Rosengren et al.								
ci	9.5	24	427	-0.51	0.28	0.60	0.35	1.04
ci	0.0	832	34755	0.00	0.00	1.00	1.00	1.00
ci	0.5	564	18106	-0.04	0.06	0.96	0.86	1.08
ci	2.0	2081	53596	-0.06	0.04	0.94	0.86	1.02
ci	5.0	658	15541	-0.07	0.06	0.93	0.83	1.04
ci	8.0	274	5522	-0.13	0.08	0.88	0.76	1.02

Descriptive data visualization

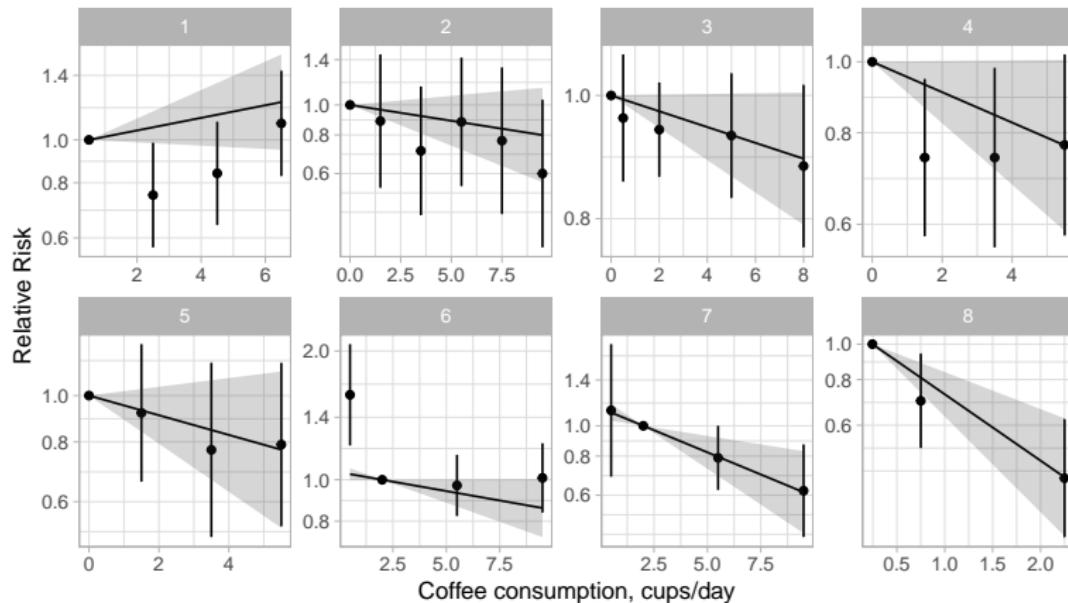


Study-specific linear trend (1st stage)

$$y_{ij} = \beta_{i1} (x_{ij} - x_{i0}) + \varepsilon_{ij}, \quad i = 1, \dots, 22$$

id	bi	sei	exp(bi)	lb	ub
1	0.033	0.021	1.033	0.991	1.077
2	-0.024	0.019	0.977	0.941	1.013
3	-0.014	0.008	0.986	0.971	1.001
4	-0.048	0.025	0.953	0.908	1.001
5	-0.047	0.035	0.954	0.891	1.021
6	-0.020	0.011	0.980	0.960	1.001
7	-0.065	0.021	0.937	0.900	0.975
:	:	:	:	:	:

Fitted data visualization



Meta-analytic model (2nd stage)

$$\hat{\beta}_i \sim N(\beta, \hat{v}_i + \tau^2), \quad i = 1, \dots, 22$$

Random-Effects Model (k = 22; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0003 (SE = 0.0002)
tau (square root of estimated tau^2 value): 0.0173
I^2 (total heterogeneity / total variability): 75.01%
H^2 (total variability / sampling variability): 4.00

Test for Heterogeneity:

$Q(df = 21) = 77.0088, p-val < .0001$

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub	
-0.0326	0.0051	-6.4424	<.0001	-0.0425	-0.0227	***

Every one cup of coffee per day increase was associated with a 3% decrease ($\exp(-0.0326) = 0.967$) in the mortality risk.

Forest plot

