Goodness of fit in dose-response meta-analysis

Alessio Crippa[1] Nicola Orsini[1]

[1]Institute of Environmental Medicine, Karolinska Institutet

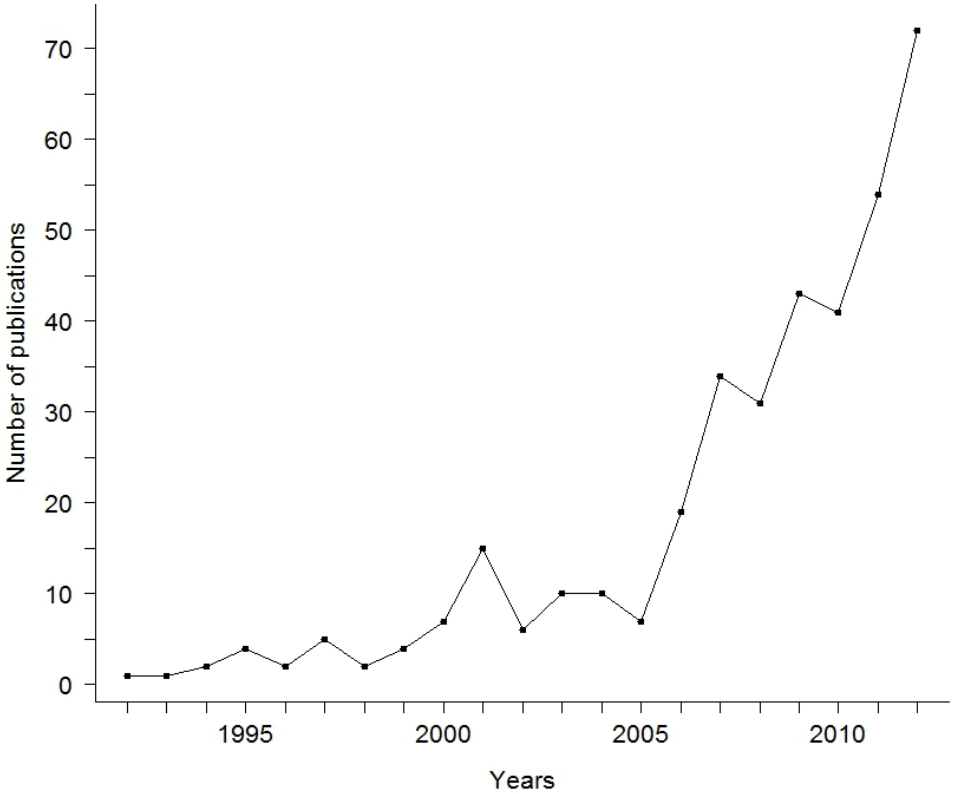
4th Nordic-Baltic Biometric Conference, Stockholm

Contents

Introduction

Dose-response meta-analysis

- Specific type of meta-analysis
- Define an overall trend from summarized data, where the exposure is categorized and the results presented in a tabular way
- Method first formalized by Greenland and Longnecker (1992)



- Number of published dose-response meta-analyses increases exponentially
- 42 in the first 4 months of 2013 (2 every week)
- 26 (60%) estimated linear trend
- Only 17 (40%) investigated non-linearity and provided a graphical presentation
- 39 (93%) assessed heterogeneity

- None overlaid the observed data points and the summary exposure-disease trend
- Define the degree of consistency of prior knowledge around a summary or pooled trend

Aims

- Describe how to estimate dose-response relation
- Clarify how observed and fitted relative risks can be compared
- Propose a measure of goodness of fit

Estimate dose-response relation

Two stage procedure:

First stage

Define and estimate the dose-response association for each study j (linear, fractional polynomial, splines):
p-th vector of estimates β_j and accompanying $p \times p$ estimated covariance matrix $\mathbf{V}_j$

Second stage

Combine these estimates to obtain an overall measure of association

Model definition

Log linear model for a single study (linear trend):

$$y_i = \beta_1 X_i + \epsilon_i \tag{1}$$

where y_i are the log of non referent relative risks, X_i the corresponding levels of exposure ($x = 0$ correspond to the reference category).

NB: The model in equation (1) has no intercept: the log relative risk for the referent exposure is set equal to 0 (RR=1).

GLS estimation

ϵ_i are not independent, $Cov(\epsilon_i) = \Sigma$
Σ can be estimated from the published data.

β can be efficiently estimated by gls:

$$\hat{\beta} = (\mathbf{X}'\Sigma\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{y} \tag{2}$$

$$\mathbf{V} = Cov(\hat{\beta}) = (\mathbf{X}'\Sigma\mathbf{X})^{-1} \tag{3}$$

Second stage

Pool the estimates from the first stage: $\hat{\beta} = \left[\hat{\beta}_1, \ldots \hat{\beta}_S\right]$

Multivariate random-effect meta-analysis:

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

Different methods for estimation: (full) maximum likelihood, restricted maximum likelihood or methods of moments

Fit statistics

To assess and quantify presence of heterogeneity (second stage analysis):

- $$Q = \sum_{j=1}^{S} \left[(\beta_j - \hat{\beta})' V_j^{-1} (\beta_j - \hat{\beta}) \right] \quad (5)$$

- $$I^2 = \frac{Q - df}{Q} \quad (6)$$

- Information Criteria, such as

$$AIC = -2l(\hat{\beta}, \hat{\psi}) + 2p \quad (7)$$

Motivating example

Table : Case-control data on alcohol and breast cancer risk (Rohan and Michael 1988)

gday	dose	case	control	n	crudeor	adjrr	lb	ub
Ref.	0	165	172	337	1.00	1.00	1.00	1.00
$<$2.5	2	74	93	167	0.83	0.80	0.51	1.27
2.5-9.3	6	90	96	186	0.98	1.16	0.73	1.85
$>$9.3	11	122	90	212	1.41	1.57	0.99	2.51

Linear trend:

$$log(adjrr) = \beta_1 X_i + \epsilon_i$$

```
library(dosresmeta)
data(cc_ex)

 mod <- dosresmeta(formula = logrr~0 + dose, study="cc",
            cov =c(case, n), se=c(loglb, logub), data=cc_ex)

 mod$Param

  id      Estimate Std. Error  z value   Pr(>z)
1  1 dose 0.046    0.02051     2.24      0.025

 mod$fit.stat

  id Q    Pr(>chi2) log ll
1  1 1.93 0.382     0.790
```

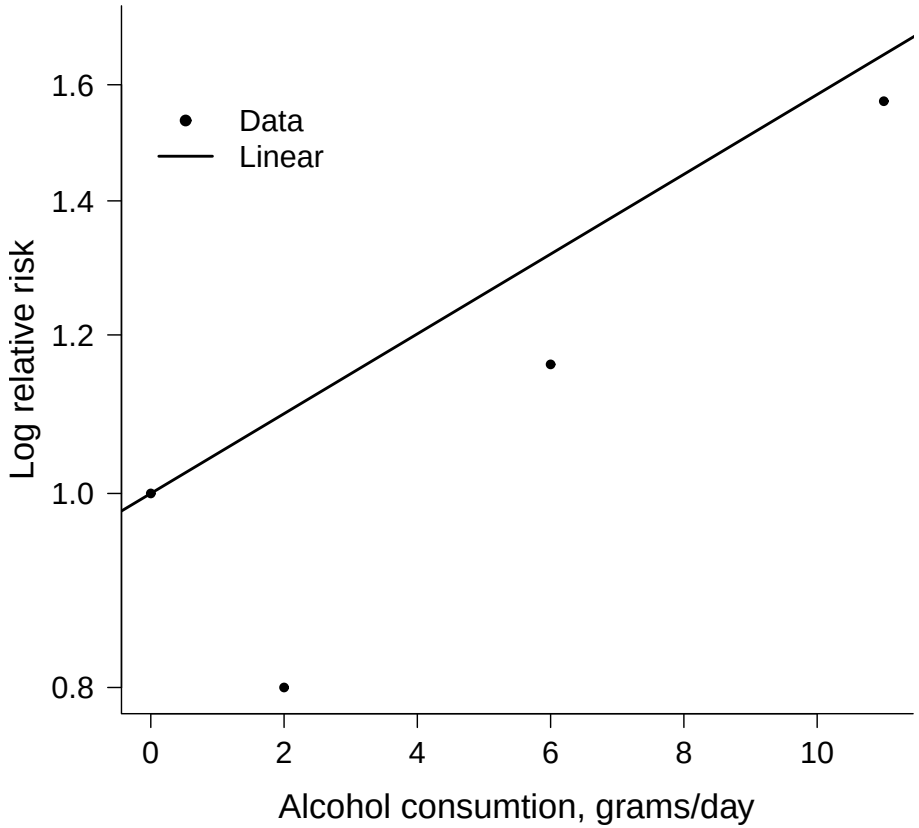


Figure : Comparison between corrected and uncorrected prediction

De-correlate data points: a single study

Consider y, $\mathbf{X}$ and Σ:
$\mathbf{L}$ is the Cholesky decomposition of Σ, $\Sigma = \mathbf{L}\mathbf{L}'$

$$\begin{aligned} y^* &= \mathbf{L}^{-1} y \\ \mathbf{X}^* &= \mathbf{L}^{-1} \mathbf{X} \end{aligned} \tag{8}$$

Model in equation (1) can be re-formulated as:

$$y^* = \mathbf{X}^* \beta^* + \epsilon^* \tag{9}$$

NB: Parameter estimates do not change: $\hat{\beta}^* = \hat{\beta}$

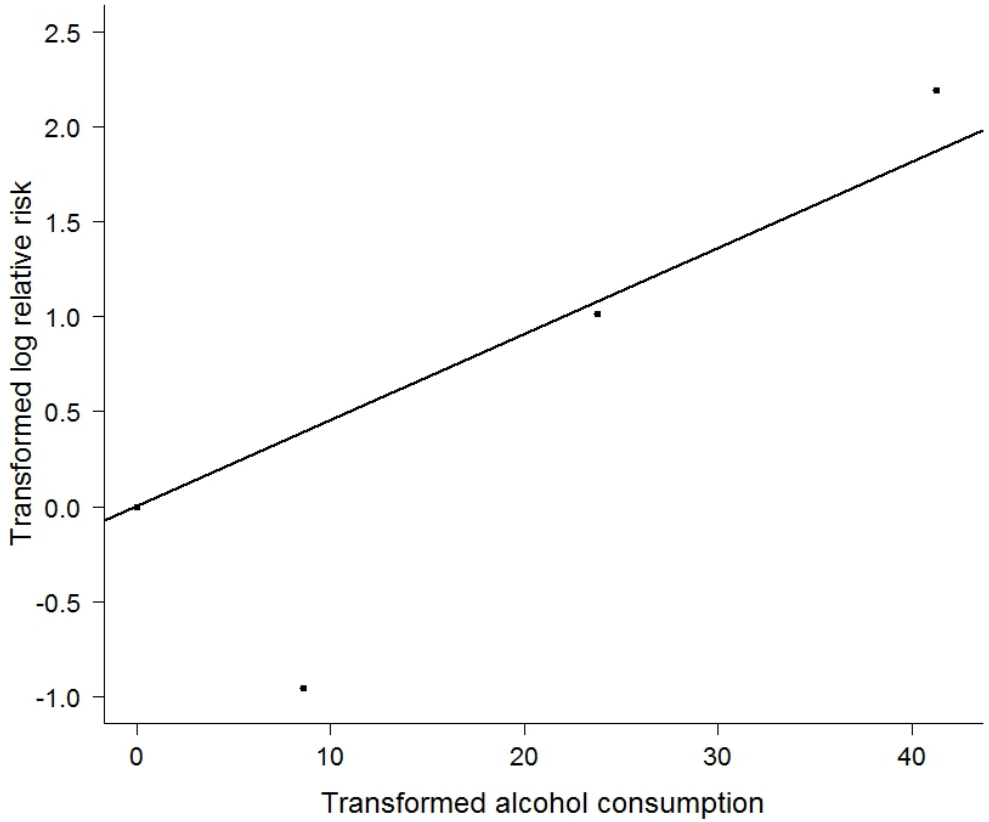


Figure : Data points and fitted trend corrected for covariance of log relative risks, based on decorralate data

De-correlate data points: several studies

Consider m studies:
First decorrelate observations in each study (eq. 8)
Pool data by concatenating y_j^* and $\mathbf{X}_j^*$:

$$y^* = \begin{bmatrix} y_1^* \\ \vdots \\ y_j^* \\ \vdots \\ y_m^* \end{bmatrix} \quad \mathbf{X}^* = \begin{bmatrix} \mathbf{X}_1^* \\ \vdots \\ \mathbf{X}_j^* \\ \vdots \\ \mathbf{X}_m^* \end{bmatrix}$$

A measure of goodness of fit

(fixed effect) model in eq. 4 can be re-formulated as:

$$y^* = \mathbf{X}^* \beta^* + \epsilon^* \tag{10}$$

R^2 can be adopted to assess the fit of the analysis:

$$R^2 = 1 - \frac{\sum_{j=1}^{S} \sum_{i=1}^{n_j} (y_{ij}^* - \mathbf{X}_{ij}^* \beta)^2}{\sum_{j=1}^{S} \sum_{i=1}^{n_j} y_{ij}^{*2}} \tag{11}$$

where β is estimated from the fixed effect model in eq. 4.

Properties

- Well known measure of goodness of fit in traditional context
- Simple computation
- Based on all data points
- Simple and intuitive interpretation
- Unit-less measure, range: [0,1]
- Evaluate the agreement low, moderate, considerable and high to R^2 in the range of [0,25], (25,50],(50,75] and (75,100]

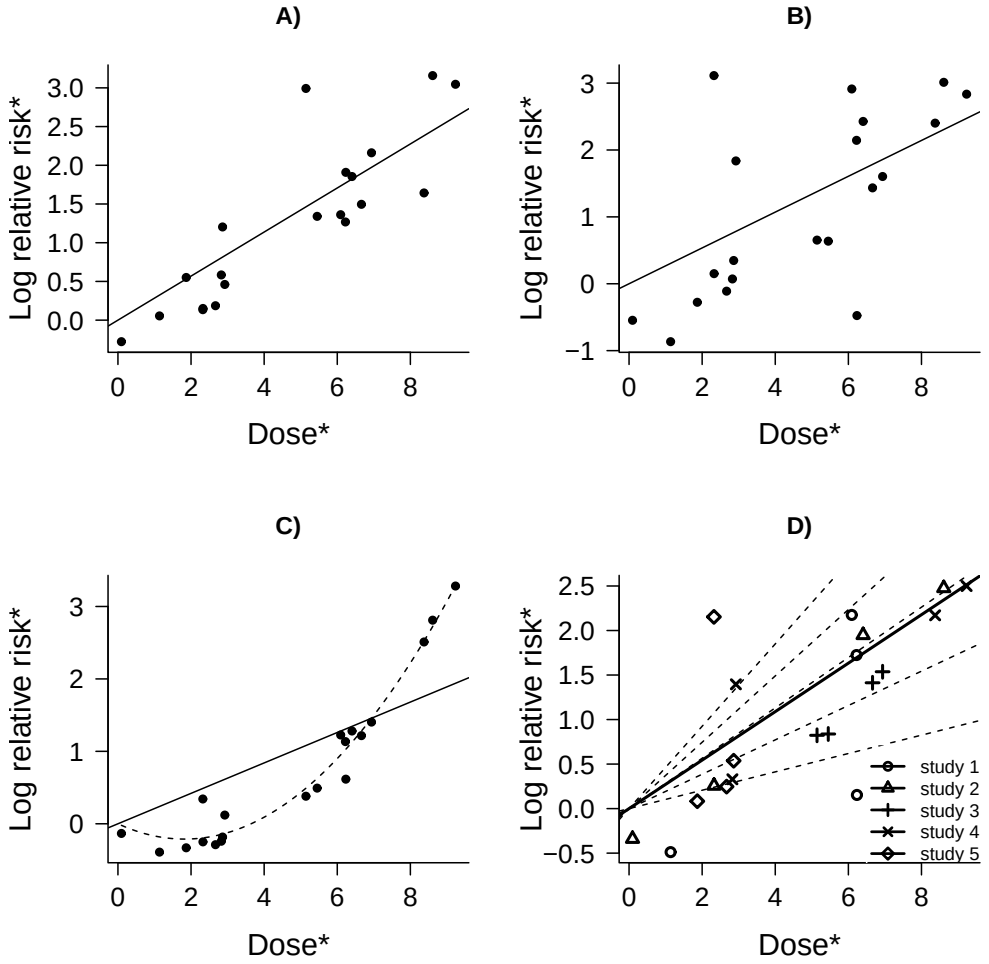


Figure : Different causes for disagreement

Body Mass Index and renal cell cancer risk

R^2 can help to compare the fit of different analyses:

- Linear Trend

$$log(RR_{ij}) = \beta_{1j} X_{ij} + \epsilon_{ij} \quad (12)$$

```
dosresmeta(formula = logor~dose, study=c(id, studyt),
                cov=c(case ,n), se=selogor, data=bmi_rc)
```

- Non-linear relation (restricted cubic spline)

$$log(RR_{ij}) = \beta_{1j} X_{1ij} + \beta_{2j} X_{2ij} + \epsilon_{ij} \quad (13)$$

```
dosresmeta(formula = logor~dose+doses, study=c(id, studyt),
                cov=c(case,n),se=selogor,data=bmi_rc)
```

Table : Estimated coefficients for linear and non-linear dose-response meta-analysis of BMI and renal cancer risk

Estimate	Parameter	Estimate	Std. Error	z	Pr(>|z|)
Linear	β_1	0.076	0.013	5.6	<0.001
Non-linear	β_1	0.038	<0.001	1.6	0.100
	β_2	0.056	0.033	1.7	0.084

Table : Fit statistics for linear and non-linear dose-response for dose-response meta-analysis of BMI and renal cancer risk

	Q	p-value	I^2	R^2
Linear	14.1	0.049	50	67
Non-linear	22.4	0.071	37	70

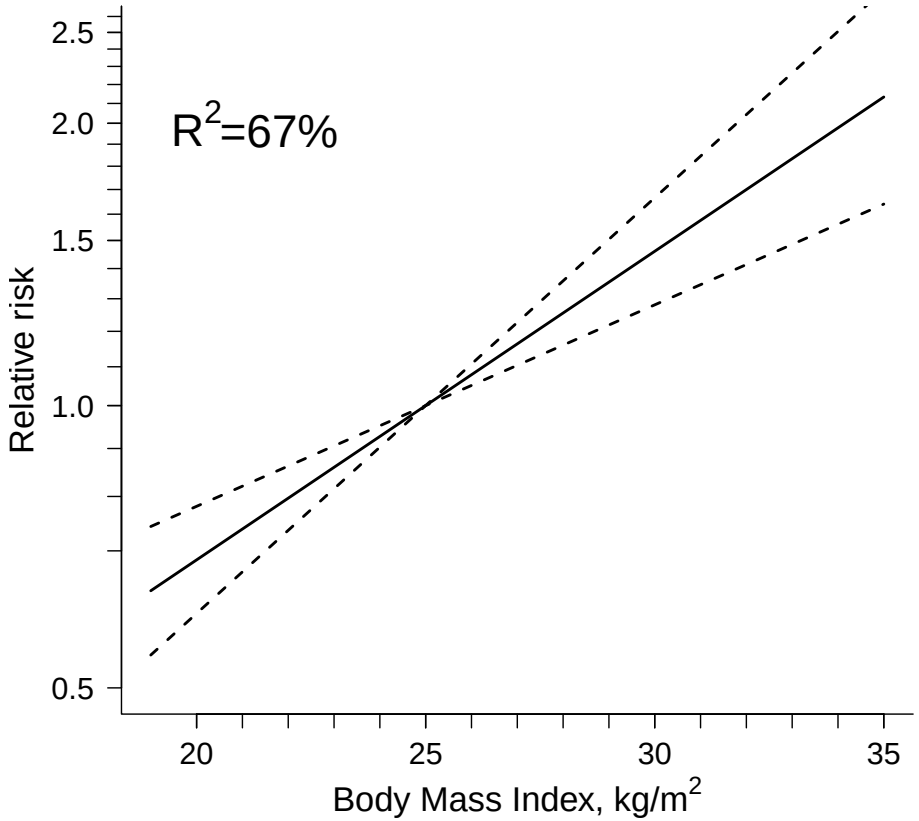


Figure : Predicted dose-response association between BMI and risk of renal cell cancer

Alcohol intake and colorectal cancer

R^2 may warn about lack of fit even if Q statistic and I^2 do not reveal any problems.

We compare two analyses:

- Linear trend

$$log(RR_{ij}) = \beta_j X_{ij} + \epsilon_{ij} \tag{14}$$

- Non-linear relation (restricted cubic spline):

$$log(RR_{ij}) = \beta_{1j} X_{1ij} + \beta_{2j} X_{2ij} + \epsilon_{ij} \tag{15}$$

Table : Estimated coefficients for linear and non-linear dose-response meta-analysis of alcohol intake and risk of colorectal cancer

Estimate	Parameter	Estimate	Std. Error	z	Pr(>|z|)
Linear	β_1	0.006	0.001	4.7	<0.001
Non-linear	β_1	-0.001	<0.001	-0.3	0.800
	β_2	0.021	0.010	2.0	0.045

Table : Fit statistics for linear and non-linear dose-response meta-analysis between alcohol intake and risk of colorectal cancer

	Q	p-value	I^2	R^2
Linear	4.7	0.702	0	32
Non-linear	14.2	0.432	2	38

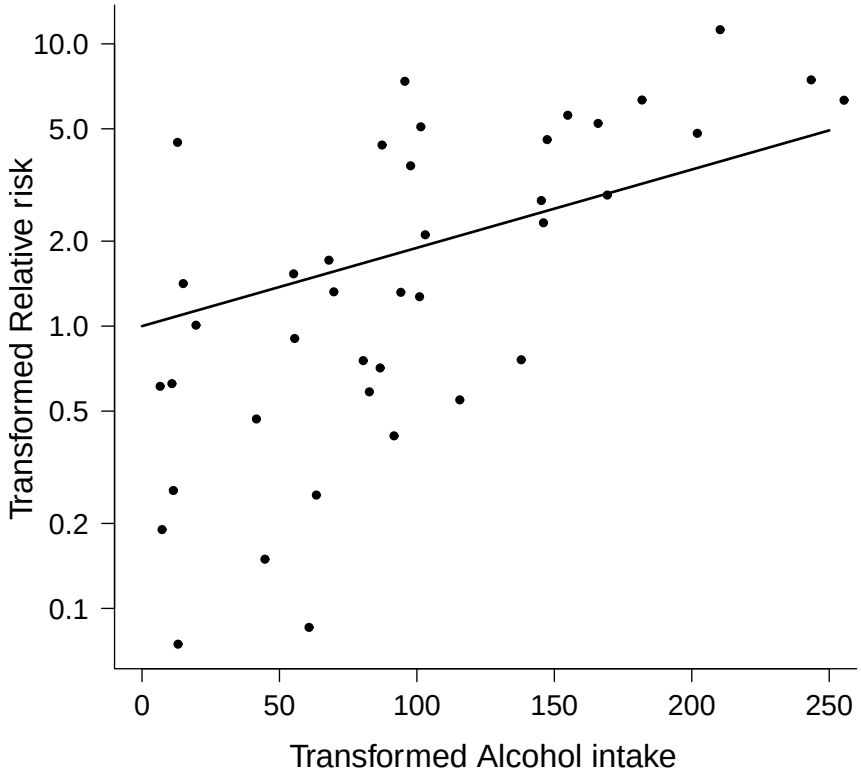


Figure : Predicted dose-response relation based on decorrelate data for dose-response meta-analysis between alcohol intake and colorectal cancer

Alcohol consumption and risk of esophageal cancer

R^2 provides a different information from the usual fit statistics.

- Fractional Polynomials:

$$log(RR_{ij}) = \beta_j X_{ij} + \beta_2 X_{ij} \log(X_{ij}) + \epsilon_{ij} \quad (16)$$

- Restricted cubic spline:

$$log(RR_{ij}) = \beta_{1j} X_{1ij} + \beta_{2j} X_{2ij} + \epsilon_{ij} \quad (17)$$

Table : Fit statistics for fractional polynomial and spline analysis in dose-response meta-analysis between alcohol and esophageal cancer

	R^2	AIC
Fractional Polynomial	70	-115.5
Spline	68	-44.9

- AIC tells us which one is better
- R^2 evaluates how much the fit differ

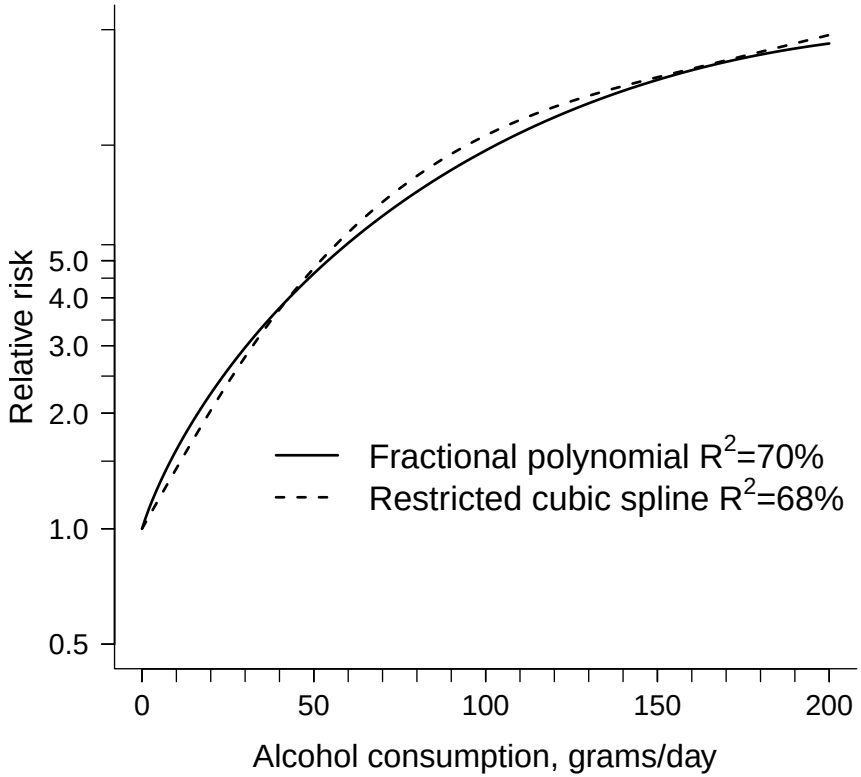


Figure : Predicted dose-response relations based on fractional polynomial and restricted cubic spline models for dose-response meta-analysis between alcohol and esophageal cancer

Conclusion

- Increasing number of published dose-response meta-analyses
- Fit statistics refer to statistical heterogeneity
- No measure of agreement between observed and modeled data

Strengths

- A possible graphical comparison
- R^2 as summary measure of agreement
- Improve the current practice
- **dosresmeta** R package available at `http://cran.r-project.org/`

Further investigations

- sensitivity analysis related to influential points;
- analysis of potential bias;
- development of robust methods;
- modeling risk instead of relative risk;
- including time dimension;
- improvements in the "**dosresmeta**" R package

Thank you!

Supervisors:
Nicola Orsini
Vincenzo Bagnardi
Rino Bellocco
Matteo Bottai
Alicja Wolk

Alessio Crippa
Unit of Nutritional Epidemiology
Institute of Environmental Medicine
Karolinska Institutet