One-stage dose-response meta-analysis

Nicola Orsini, Alessio Crippa

Biostatistics Team
Department of Public Health Sciences
Karolinska Institutet
http://ki.se/en/phs/biostatistics-team

2017 Nordic and Baltic Stata Users Group meeting

September 1, 2017

Outline

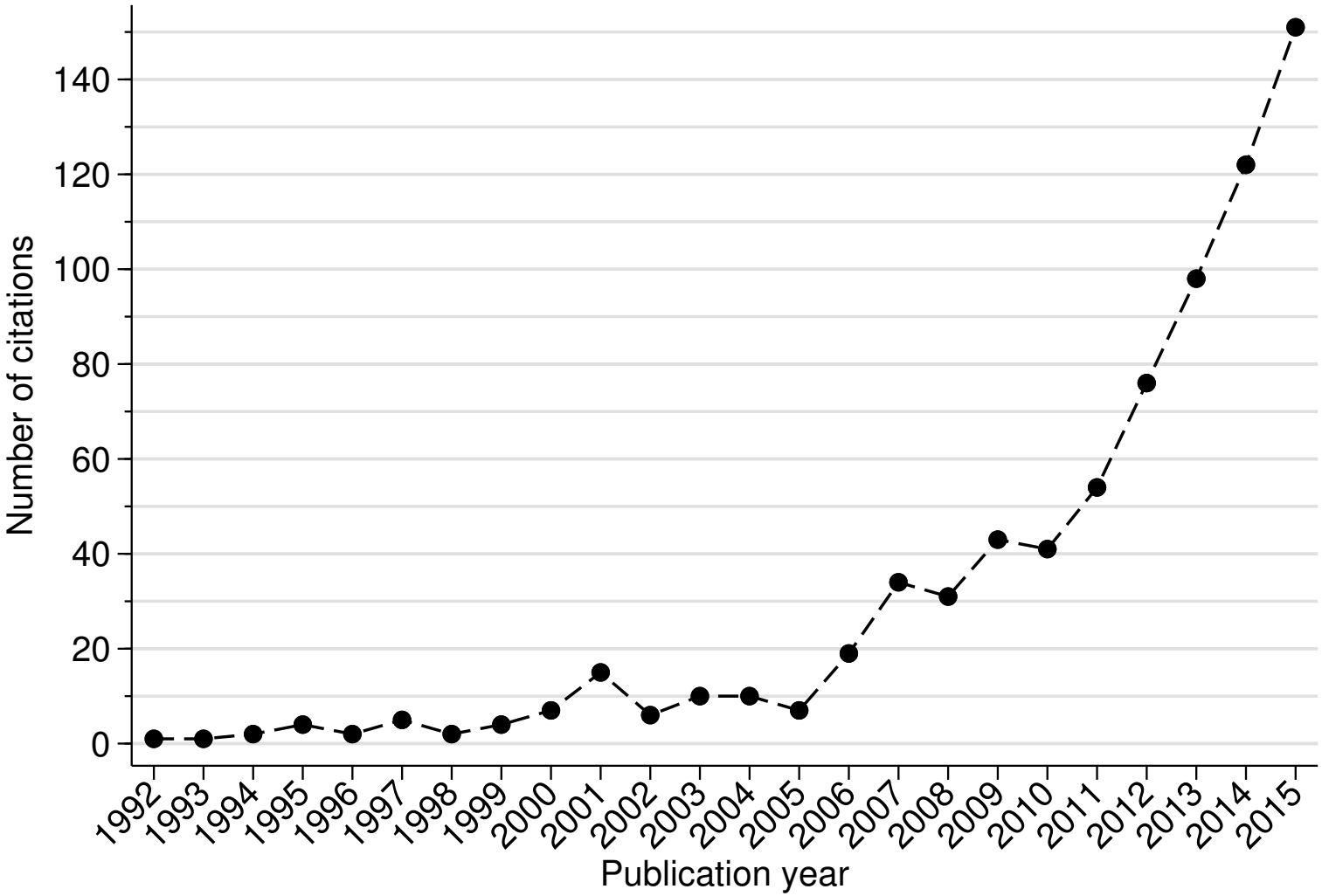
- Goal
- Data
- Model
- Estimation
- Examples
- Summary

Goal

- A dose-response analysis describes the changes of a response across levels of a quantitative factor. The quantitative factor could be an administered drug or an exposure.

- A meta-analysis of dose-response (exposure-disease) relations aims at identifying the trend underlying multiple studies trying to answer the same research question.

Increasing number of dose-response meta-analyses



Data source: ISI Web of Knowledge

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

Common practice in statistical analysis

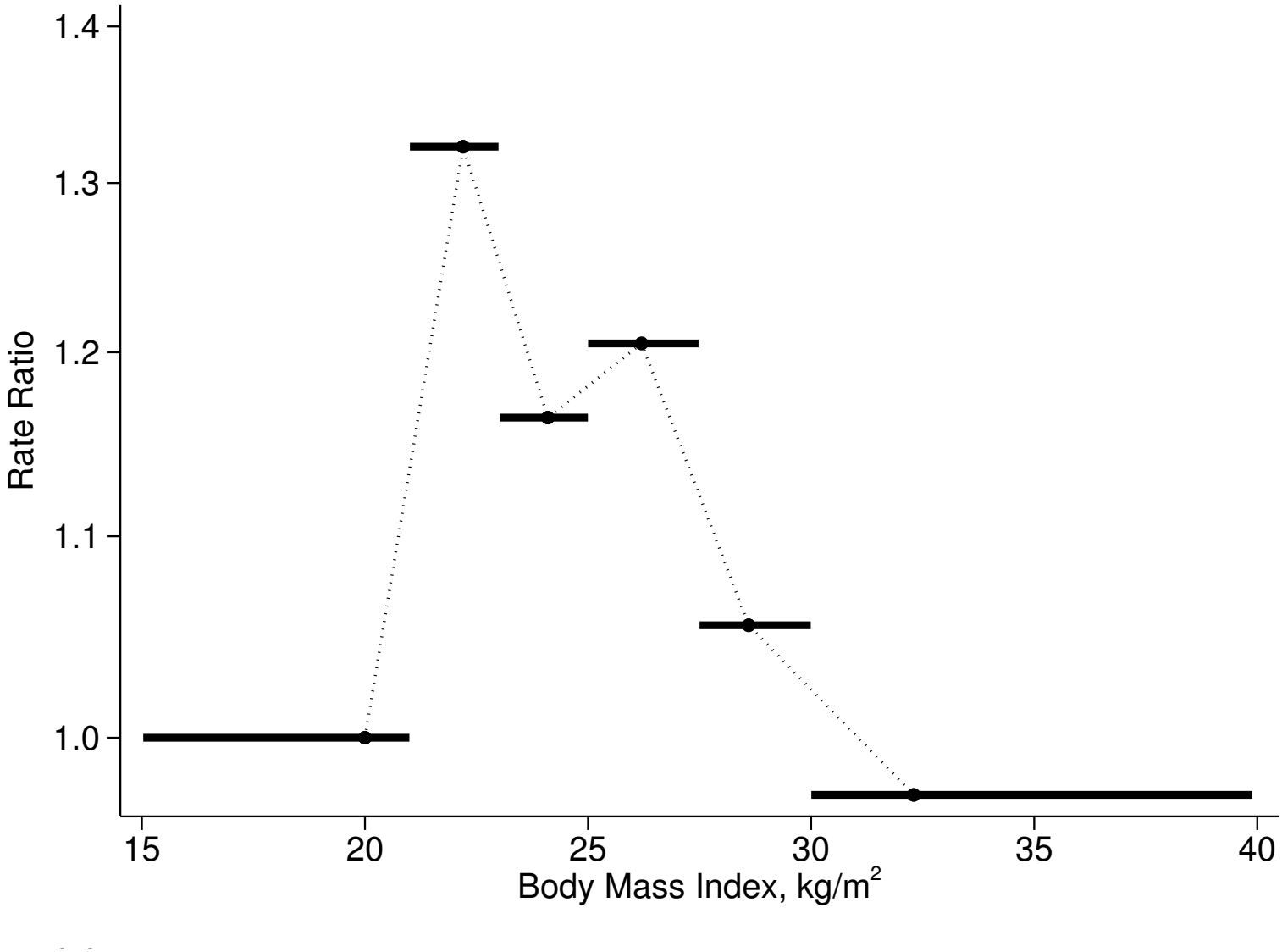
- Plot summarized data and connect the dots with a line
- First estimate a curve within each study and then average regression coefficients across studies (two-stage approach)
- Exclude studies with less than 3 exposure groups
- Linear vs non-linear relationships
- Find the "best" fitting dose-response model

Data for a single study

Table: Rate ratios of prostate cancer according to categories of body mass index (kg/m^2). Data from a cohort of 36,143 middle-age and elderly men followed for 446,699 person-years during which 2,037 were diagnosed with prostate cancer.

Alcohol Intake	Median, grams/day	No. of cases	Person-years	Rate Ratio (95% CI)
< 21.00	20.0	84	21,289	1.00 Ref.
$[21.00; 23.00)$	22.2	323	61,895	1.32 (1.04, 1.68)
$[23.00; 25.00)$	24.1	532	115,885	1.16 (0.92, 1.46)
$[25.00; 27.50)$	26.2	651	136,917	1.21 (0.96, 1.51)
$[27.50; 30.00)$	28.6	283	68,008	1.05 (0.83, 1.35)
≥ 30	32.3	164	42,704	0.97 (0.75, 1.27)

Plot of the data for a single study



Summarized vs Individual Data

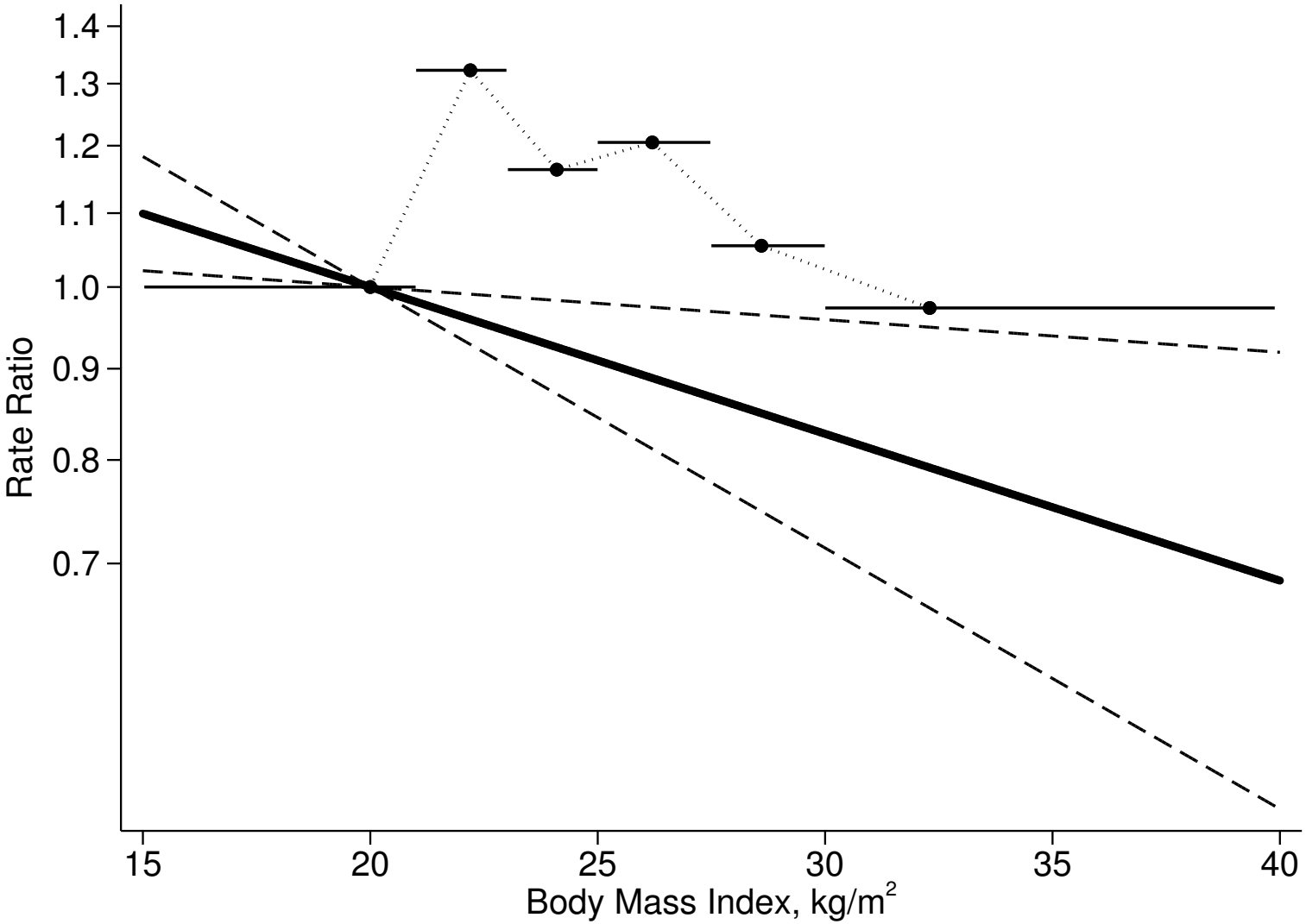
```
. glst logrr bmic , cov(py case) se(se) ir

Generalized least-squares regression                Number of obs    =         5
Goodness-of-fit chi2(4)     =     9.62              Model chi2(1)    =      6.35
Prob > chi2                 =   0.0473              Prob > chi2      =    0.0117
------------------------------------------------------------------------------
       logrr |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        bmir |  -.0189389   .0075147    -2.52   0.012    -.0336675   -.0042103
------------------------------------------------------------------------------
```

Every 5 kg/m^2 increase in body mass index is associated with 9% (95% CI=0.85-0.98) lower prostate cancer risk.

```
. streg bmi , dist(exp) nohr

Exponential PH regression

No. of subjects =       36,143                  Number of obs    =       36,143
No. of failures =        2,037
Time at risk    =   446698.5243
                                                LR chi2(1)       =         7.92
Log likelihood  =    -9249.8404                 Prob > chi2      =       0.0049

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         bmi |  -.0197478   .0070674    -2.79   0.005    -.0335997   -.0058959
       _cons |   -4.88436   .1817654   -26.87   0.000    -5.240614   -4.528106
------------------------------------------------------------------------------
```

Every 5 kg/m^2 increase in body mass index is associated with 9% (95% CI=0.85-0.97) lower prostate cancer rate.

Challenge of comparing alternative parametrizations expressed in relative terms



Features of the data

- The response variable is a vector of contrasts relative to a common referent
- Correlation among study-specific contrasts
- Graphical comparison of alternative models is not straightforward
- Number of contrasts is varying across studies
- Exclusion of studies with not enough contrasts to fit more complicated model

Model

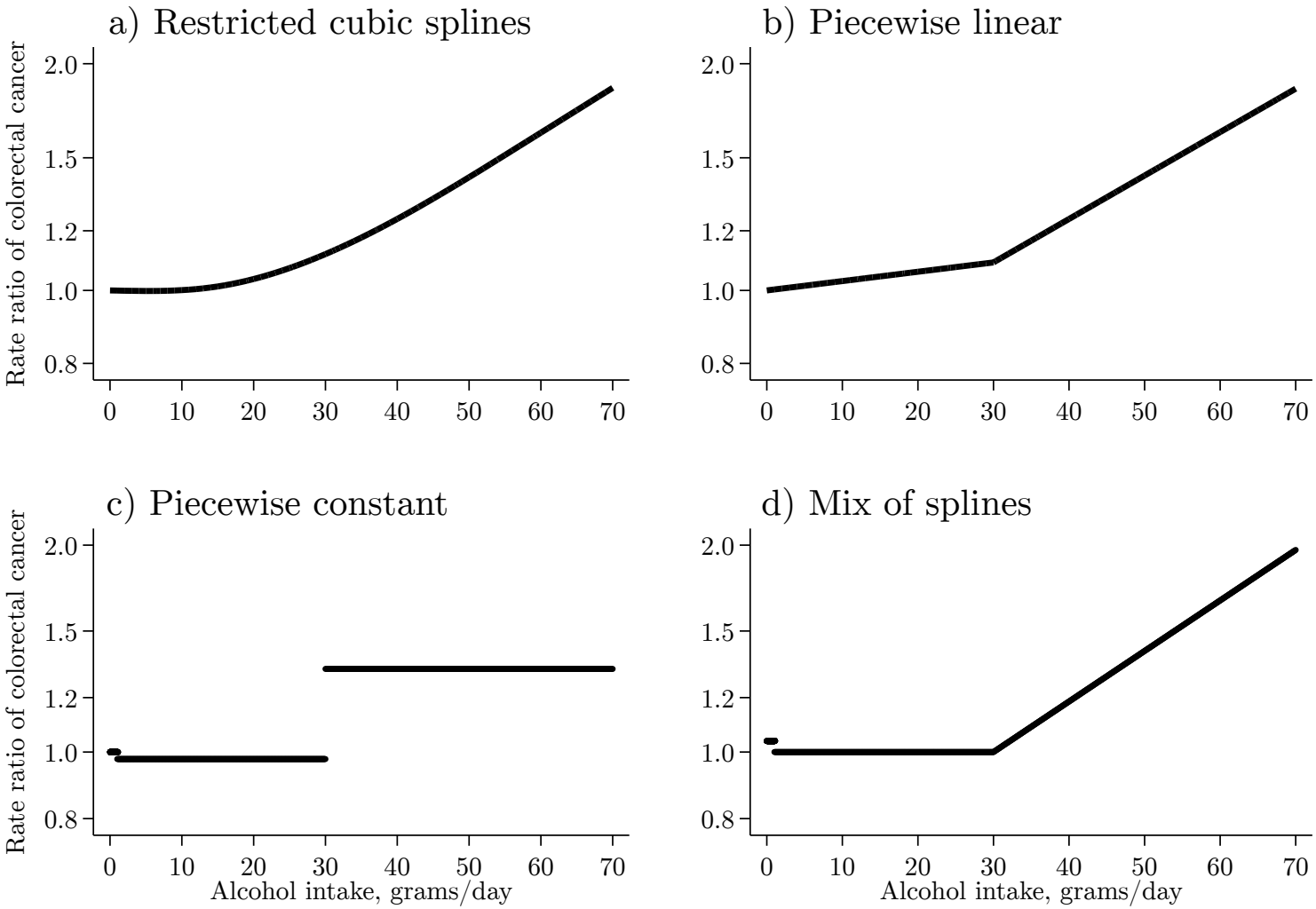
A one-stage model for meta-analysis of aggregated dose-response data can be written in the general form of a linear mixed model

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i \tag{1}$$

$\mathbf{y}_i$ is the $n_i \times 1$ outcome vector in the i-th study $\mathbf{X_i}$ is the corresponding

$n_i \times p$ design matrix for the fixed-effects $\boldsymbol{\beta}$, consisting of the p transformations able to answer a variety of research questions.

Splines according to the research question



Design matrix

Since the $\mathbf{y}_i$ is a set of response contrasts relative to the baseline dose x_{0i}, $\mathbf{X}_i$ needs to be constructed in a similar way by centering the p transformations of the dose levels to the corresponding values in x_{0i}.

Let consider, for example, a transformation f; the generic j-th row of $\mathbf{X}_i$ would be defined as $f(x_{ji}) - f(x_{0i})$.

As a consequence $\mathbf{X}$ does not contain the intercept term ($y = 0$ for $x = x_{0i}$).

Random effects and residual error term

$\mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi})$

The random-effects $\mathbf{b}_i$ represent study-specific deviations from the population average dose-response coefficients $\boldsymbol{\beta}$.

$\mathbf{Z}_i$ is the analogous $n_i \times q$ design matrix for the random-effects.

The residual error term $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{S}_i)$, whose variance matrix $\mathbf{S}_i$ is assumed known.

$\mathbf{S}_i$ can be either given or approximated using available summarized data.

Marginal and conditional model

The marginal model of Equation 1 can be written as

$$\mathbf{y}_i \sim \mathcal{N}\left(\mathbf{X}_i\beta, \mathbf{Z}_i\boldsymbol{\Psi}\mathbf{Z}_i^\top + \mathbf{S}_i\right) \tag{2}$$

with $\mathbf{Z}_i\boldsymbol{\Psi}\mathbf{Z}_i^\top + \mathbf{S}_i = \boldsymbol{\Sigma}_i$. The marginal variance $\boldsymbol{\Sigma}_i$ can be separated in two parts: the within-study component $\mathbf{S}_i$, that can be reconstructed from the available data and the between-study variability as a quadratic form of $\boldsymbol{\Psi}$. Alternatively the conditional model can be written as

$$\mathbf{y}_i \mid \boldsymbol{b}_i \sim \mathcal{N}\left(\mathbf{X}_i\beta + \mathbf{Z}_i\boldsymbol{b}_i, \mathbf{S}_i\right) \tag{3}$$

The dose-response model in Equation 1 can be extended to the case of meta-regression by including an interaction terms between the p dose transformations and the study-levels variables in the fixed-effect design matrix $\mathbf{X}_i$.

Estimation

We consider estimation methods based on maximum likelihood (ML) and restricted maximum likelihood (REML). The marginal likelihood for the model in Equation 2 is defined as

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

where $n = \sum_{i=1}^{k} n_i$ and ξ is the vector of the variance components in $\mathbf{\Psi}$ to be estimated. Assuming ξ is known, ML estimates of β and $V(\beta)$ are obtained by generalized least square estimators.

Estimation

An alternative is provided by REML estimation that maximizes the following likelihood

$$\ell_R(\boldsymbol{\xi}) = -\frac{1}{2}(n-p)\log(2\pi) - \frac{1}{2}\sum_{i=1}^{k}\log|\boldsymbol{\Sigma}_i(\boldsymbol{\xi})| + \\ -\frac{1}{2}\log\left|\sum_{i=1}^{k}\mathbf{X}_i^T\boldsymbol{\Sigma}_i(\boldsymbol{\xi})^{-1}\mathbf{X}_i\right| - \frac{1}{2}\sum_{i=1}^{k}\left[\left(\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}\right)^{\top}\boldsymbol{\Sigma}_i(\boldsymbol{\xi})^{-1}\left(\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}\right)\right] \tag{4}$$

where $\hat{\boldsymbol{\beta}}$ indicates the estimates obtained by generalized least squares. Both ML and REML estimation methods have been implemented in the new `drmeta` Stata package. The additional fixed-effects analysis constrains the variance components $\boldsymbol{\xi}$ in $\boldsymbol{\Psi}$ to be all equal to zero.

Hypothesis testing, goodness-of-fit, model comparison

Hypothesis testing and confidence intervals for single coefficients can be constructed using standard inference from linear mixed models, based on the approximate multivariate distribution of $\boldsymbol{\beta}$.

Multivariate extensions of Wald-type or likelihood ratio tests can be adopted to test the hypothesis $H_0 : \beta_1 = \cdots = \beta_p = 0$.

An absolute measure of the fit of the model (Discacciati et al *Res Synt Meth*, 2015) is the deviance $D = \sum_{i=1}^{k} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^\top \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})$

The coefficients of determination R^2 and a visual assessment of the decorrelated residuals may complement the previous measure.

The fit of the separate analyses can be also compared using fit statistics such as the Akaike information criterion.

Prediction

The average dose-response curve can be presented pointwisely as predicted (log) relative responses for selected dose values $\mathbf{x}^*$ using one value x_0 as referent

$$\hat{\mathbf{y}}^* = (\mathbf{X}^* - \mathbf{X}_0)\,\hat{\boldsymbol{\beta}} \tag{5}$$

where $\mathbf{X}^*$ and $\mathbf{X}_0$ are the design matrices evaluated at $\mathbf{x}^*$ and x_0 respectively. An approximate 95% confidence interval for the predicted (log) relative measures can be constructed as

$$(\mathbf{X}^* - \mathbf{X}_0)\,\hat{\boldsymbol{\beta}} \pm z_{1-\frac{\alpha}{2}} \sqrt{\operatorname{diag}\left((\mathbf{X}^* - \mathbf{X}_0)\, V\left(\hat{\boldsymbol{\beta}}\right)(\mathbf{X}^* - \mathbf{X}_0)^\top\right)} \tag{6}$$

Best Linear Unbiased Prediction (BLUP)

The multivariate normal assumption for unobserved random-effects can be used for making inference on the study-specific curves.

Henderson (*Biometrics*, 1950) showed that the (asymptotic) best linear unbiased prediction (BLUP) of $\mathbf{b}$ can be computed as

$$\hat{\mathbf{b}}_i = \hat{\boldsymbol{\Psi}}\mathbf{Z}_i^\top \hat{\boldsymbol{\Sigma}}_i^{-1}\left(\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}\right) \tag{7}$$

The conditional study-specific curves are given by $\mathbf{X}_i\hat{\boldsymbol{\beta}} + \hat{\mathbf{b}}_i$, that is a combination of the study-specific and population-average associations.

Interestingly, the study-specific curve can be predicted also for studies with p transformations of the dose with $p > n_i$.

Summarized data for 9 simulated case-control studies

id	rr	lrr	urr	dose	n	cases
1	1.00	1.00	1.00	2.4	2260	42
1	0.89	0.62	1.28	5.2	6136	102
2	1.00	1.00	1.00	1.7	651	39
2	0.68	0.47	0.97	5.1	3962	164
2	1.13	0.68	1.89	8.8	387	26
3	1.00	1.00	1.00	0.8	224	11
3	0.75	0.40	1.43	3.9	2639	99
3	0.79	0.42	1.51	6.7	2031	80
3	2.02	0.83	4.91	9.8	106	10
4	1.00	1.00	1.00	3.7	4306	89
4	0.69	0.50	0.96	6.2	4316	62
5	1.00	1.00	1.00	2.0	849	22
5	1.03	0.64	1.65	5.2	3638	97
5	1.76	0.97	3.20	8.6	513	23
6	1.00	1.00	1.00	0.1	112	7
6	0.42	0.19	0.95	3.5	2229	61
6	0.51	0.23	1.13	6.4	2515	83
6	1.12	0.41	3.04	9.6	144	10
7	1.00	1.00	1.00	3.2	3807	117
7	0.64	0.49	0.83	6.0	5295	105
8	1.00	1.00	1.00	1.5	442	14
8	1.20	0.69	2.11	4.7	3667	139
8	2.05	1.13	3.73	7.9	891	56
9	1.00	1.00	1.00	0.0	87	5
9	0.38	0.15	0.98	3.2	1943	44
9	0.46	0.18	1.17	6.0	2697	74
9	0.62	0.21	1.88	9.0	273	10

One-stage vs Two-stage model

The one-stage model can be written as

$$\mathbf{y}_i = (\beta_1 + b_{1i})(\mathbf{x}_i - x_{0i}) + (\beta_2 + b_{2i})(\mathbf{x}_i^2 - x_{0i}^2) + \boldsymbol{\epsilon}_i$$

where $\mathbf{y}_i$ is the vector of log odds ratios for the non-referent exposure levels in the i-th studies.
The alternative two-stage analysis estimates the same model separately for each study

$$\mathbf{y}_i = \beta_{1i}(\mathbf{x}_i - x_{0i}) + \beta_{2i}(\mathbf{x}_i^2 - x_{0i}^2) + \boldsymbol{\epsilon}_i$$

and obtaines the population-average dose-response coefficients by using multivariate meta-analysis on the study-specific $\hat{\boldsymbol{\beta}}_i$ estimated in the previous step.
Note that 3 studies (ID 1, 4, and 7) cannot be included in the two-stage analysis, since the quadratic models are not identifiable
($p = 2 > n_{i'} = 1, \text{for } i' = 1, 4, 7$).

drmeta - One-stage model

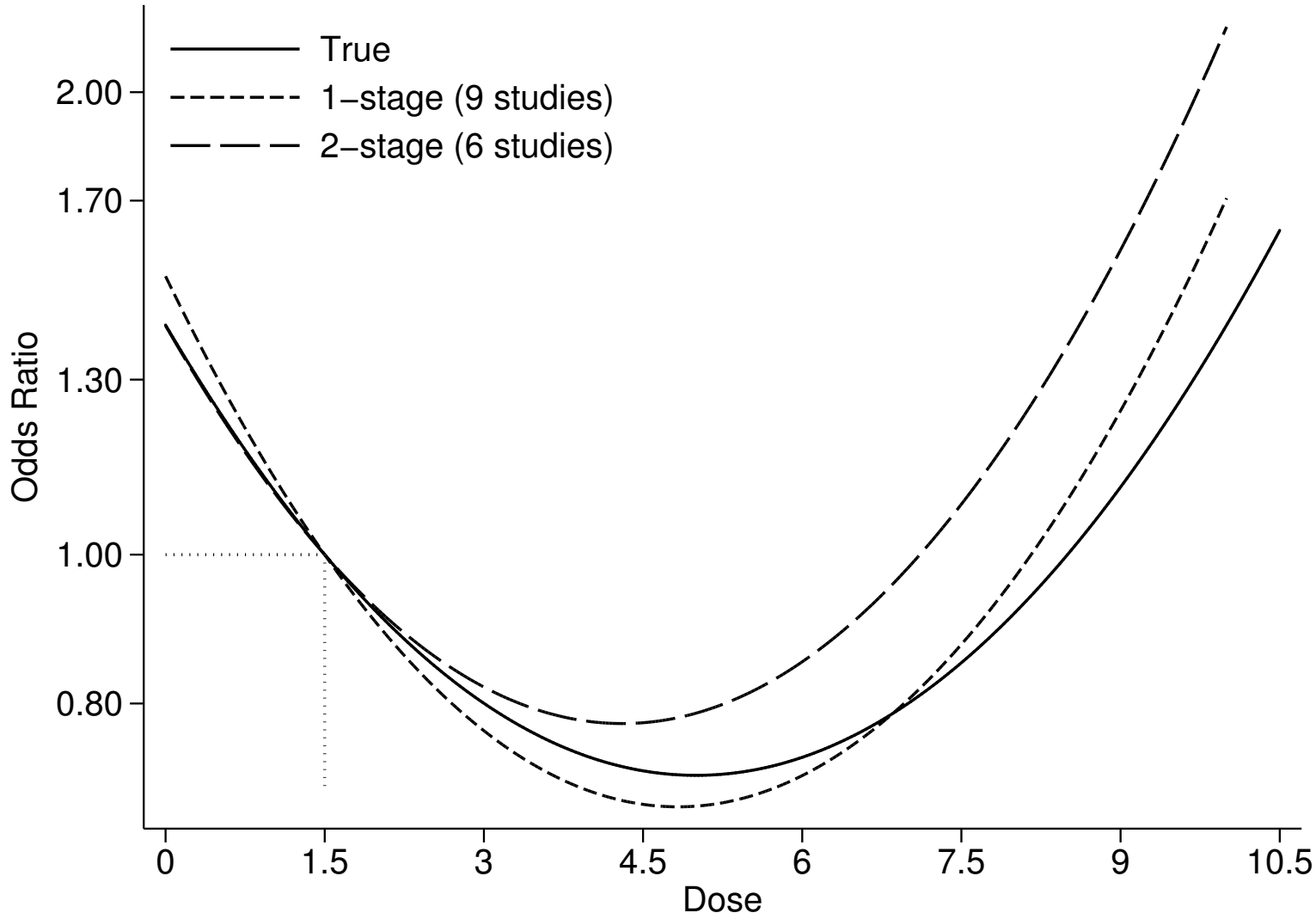
```
. gen dosesq = dose^2
. bysort id: gen dosec = dose-dose[1]
. bysort id: gen dosesqc = dosesq-dosesq[1]

. drmeta logrr dosec dosesqc ,  se(se)  data(n cases)  set(id typen) reml

One-stage random-effect dose-response model      Number of studies =         9
                                                     Number of obs =        18
Optimization   = reml                                Model chi2(2) =     36.67
Log likelihood = -8.6740999                          Prob > chi2 =     0.0000
------------------------------------------------------------------------------
       logrr |      Coef.   Std. Err.       z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        dose |  -.3294237   .0733105    -4.49   0.000    -.4731097   -.1857376
      dosesq |   .0341118   .0060608     5.63   0.000     .0222329    .0459907
------------------------------------------------------------------------------
```

drmeta - Two-stage model

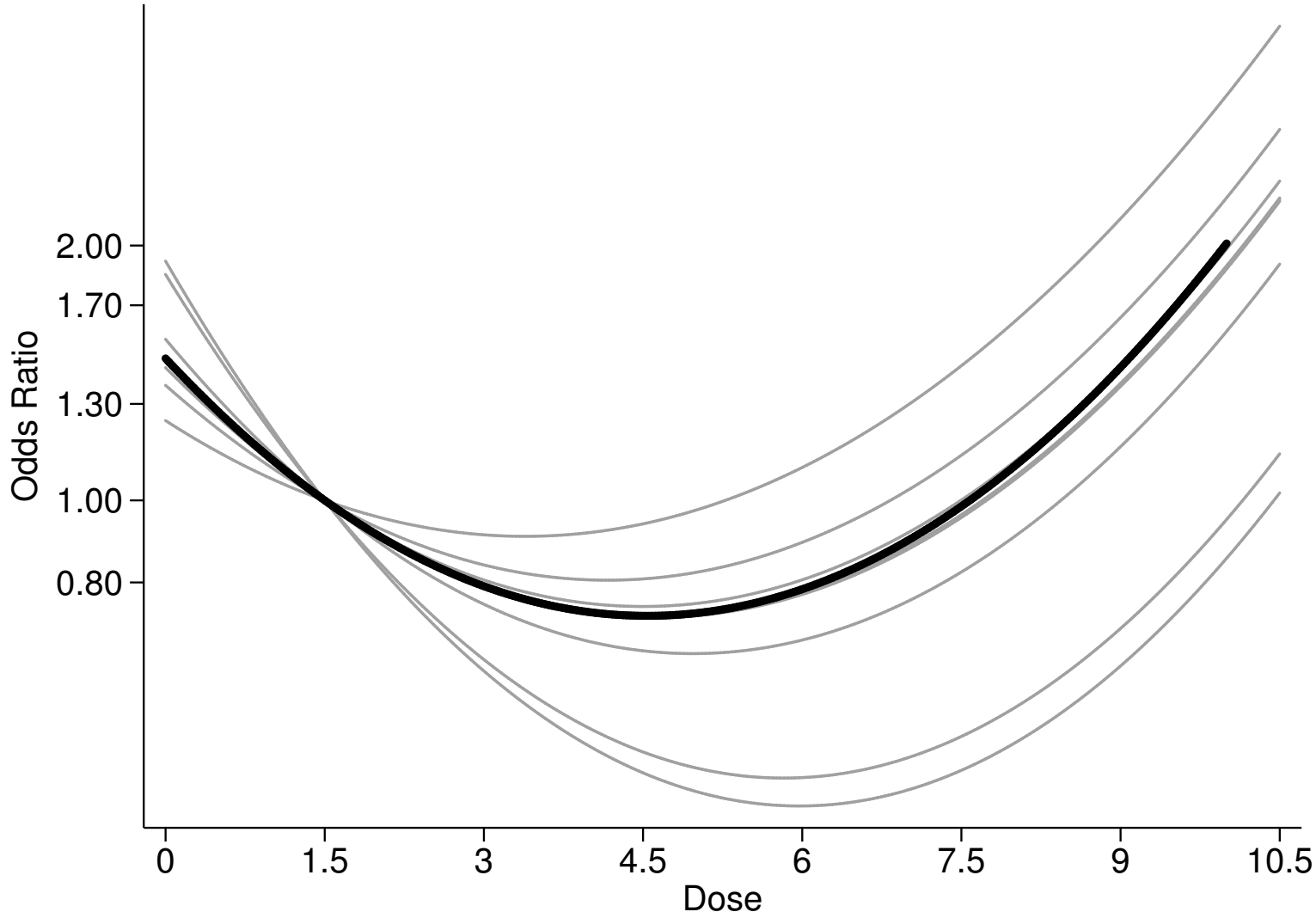
A traditional two-stage approach is feasible by excluding the three studies with only one non-reference catergory.

```
. drop if inlist(id, 1,4,7)
. drmeta logrr dosec dosesqc ,  se(se)  data(n cases)  set(id type) reml 2stage

Two-stage random-effect dose-response model      Number of studies =         6
                                                     Number of obs =        12
Optimization   = reml                                Model chi2(2) =     41.31
Log likelihood = 23.134639                             Prob > chi2 =    0.0000
------------------------------------------------------------------------------
       logrr |      Coef.   Std. Err.       z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       dosec |  -.2771024   .0642939    -4.31   0.000    -.4031161   -.1510887
     dosesqc |   .0321893   .0059277     5.43   0.000     .0205713    .0438073
------------------------------------------------------------------------------
```

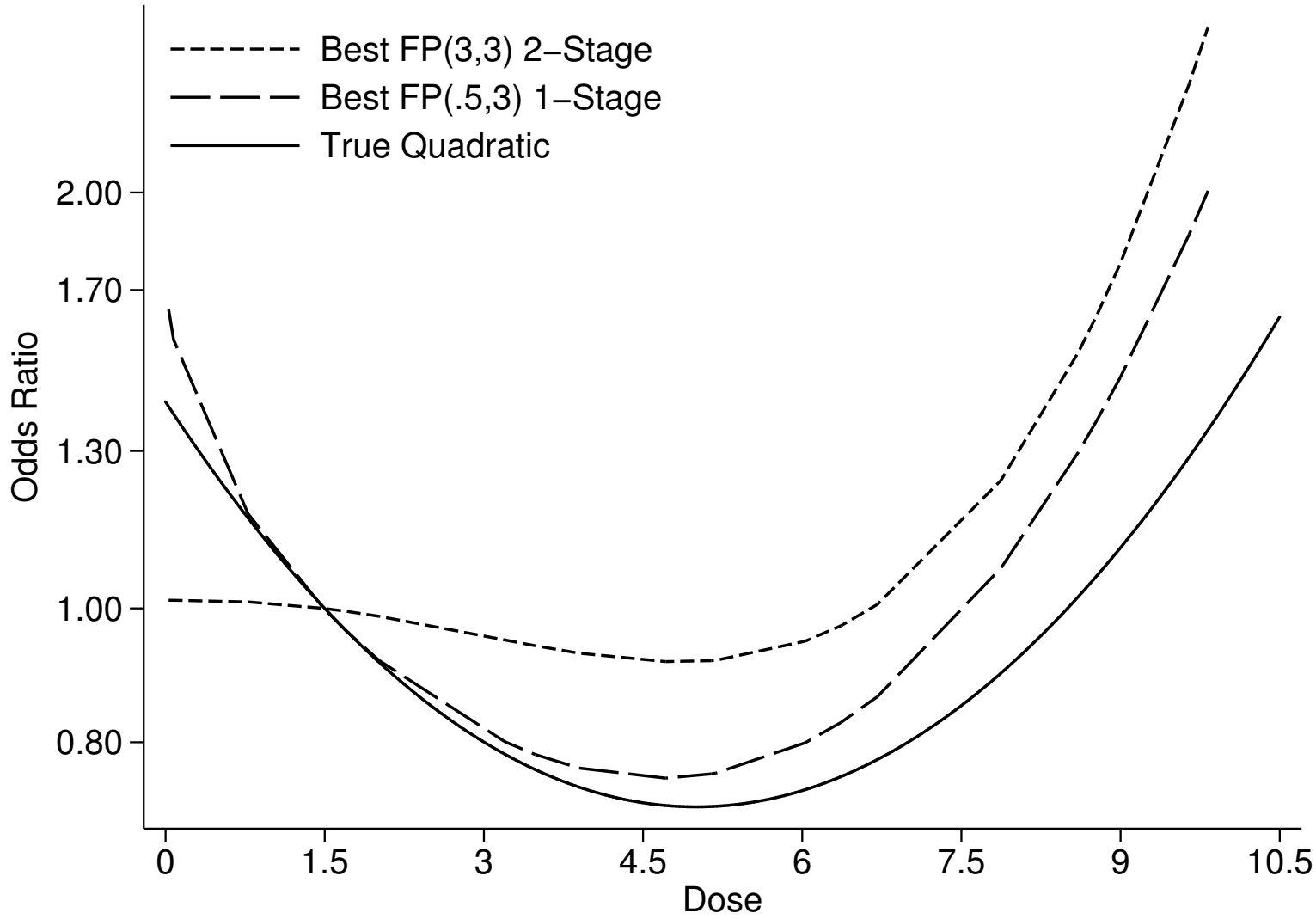
Population average dose-response curves



Study-specific dose-response trends based on predicted random effects



What is the "best" model?



Summary

- We introduced a one-stage approach for dose-response meta-analysis using linear mixed models for summarized data
- It includes all the available data in answering research questions
- It facilitates graphical comparison of study-specific and pooled dose-response relationship
- It seems to allow a better comparison of alternative models
- It is computationally more demanding than a classic two-stage approach