Comparison of meta-analysis methods for synthesizing heterogeneous studies

Anna Chaimani

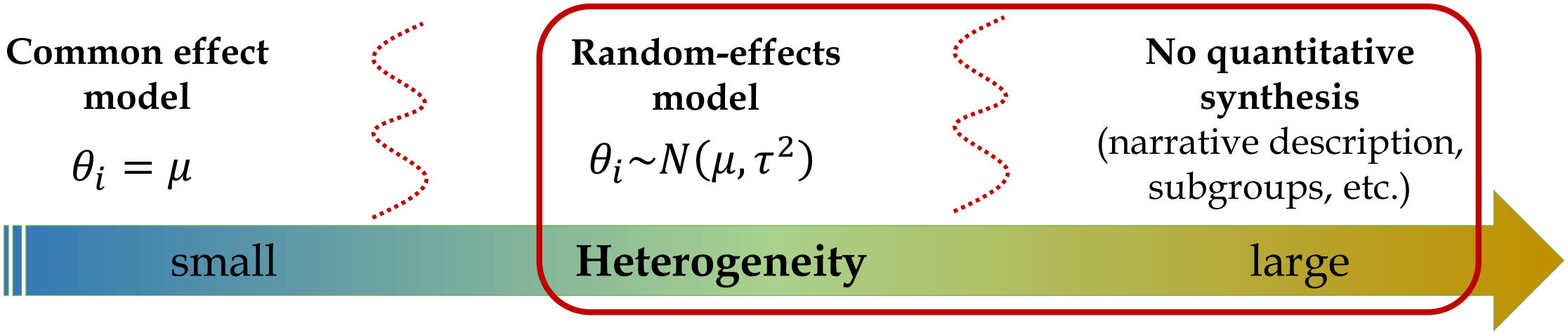
Centre of Research in Epidemiology and Statistics (CRESS-U1153), Inserm, Paris Cité University, France

Göttingen, 29 August 2024

Introduction

- Meta-analysis generally requires that studies are sufficiently homogeneous to be synthesized
- In the presence of heterogeneity, a random effects model is usually considered more appropriate
- "meta-analyses of very diverse studies can be misleading"
- "the presence of heterogeneity affects the extent to which generalizable conclusions can be formed"

Cochrane Handbook



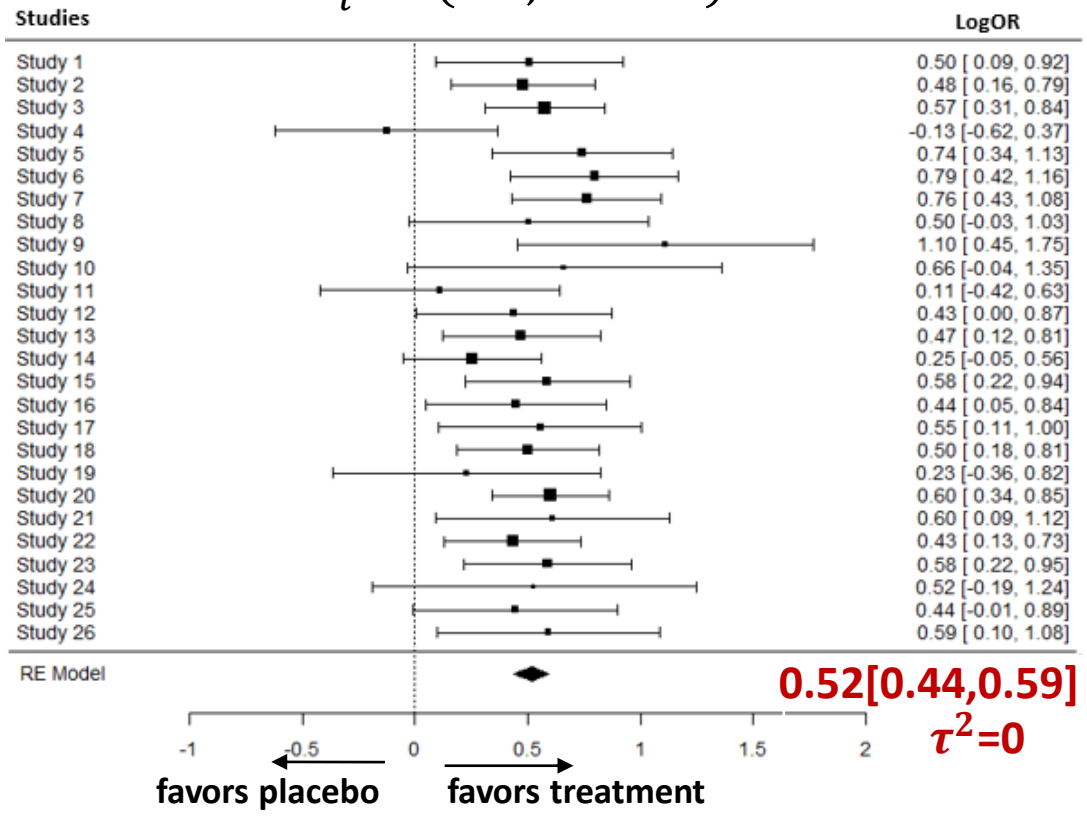
- collection of meta-analyses combining RCTs with observational studies: τ^2=369, 419, 1152!

*Cheurfa et al. **Syst Rev** 2024*

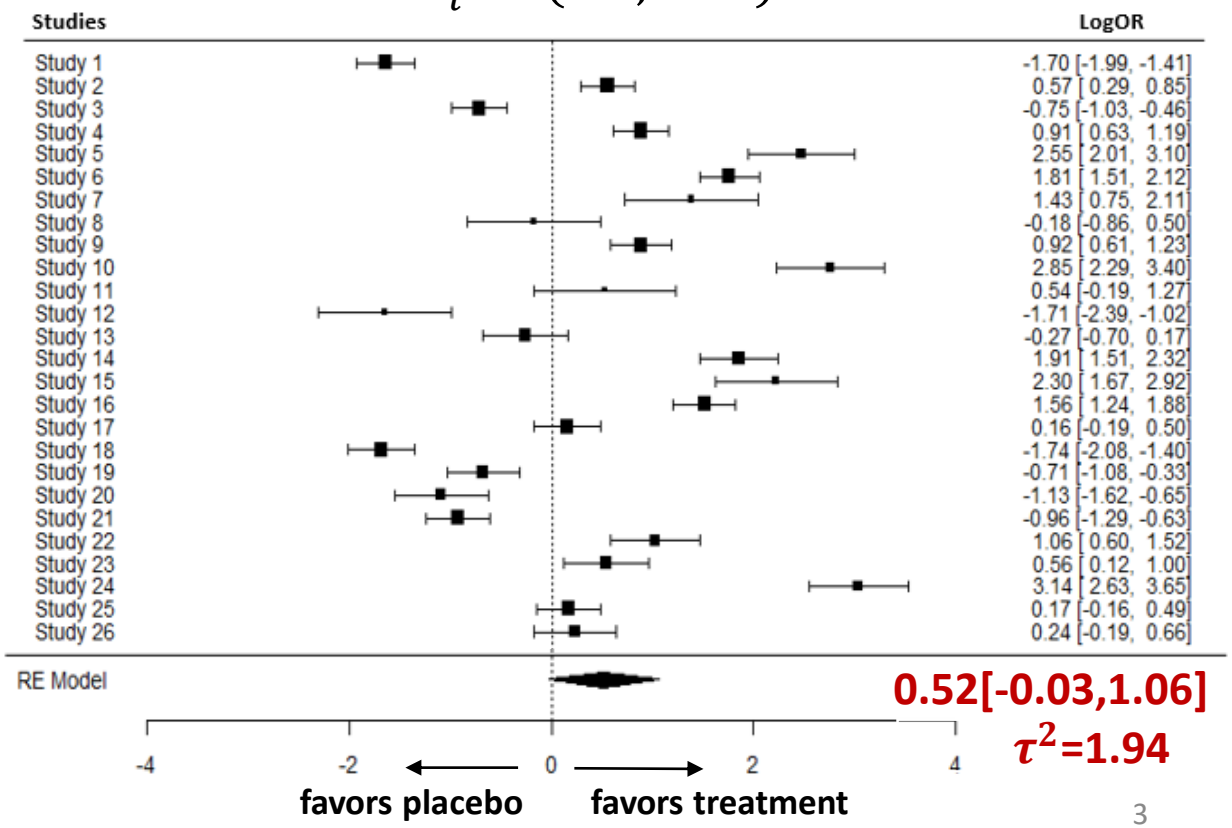
Common practice in published meta-analyses

- A tendency in the literature to ignore the extent of variation of study results and to focus on the estimated summary effect
- e.g. several meta-epidemiological studies comparing results from observational studies and RCTs

$\theta_i \sim N(0.5, 0.0001)$



$\theta_i \sim N(0.5, 2.63)$

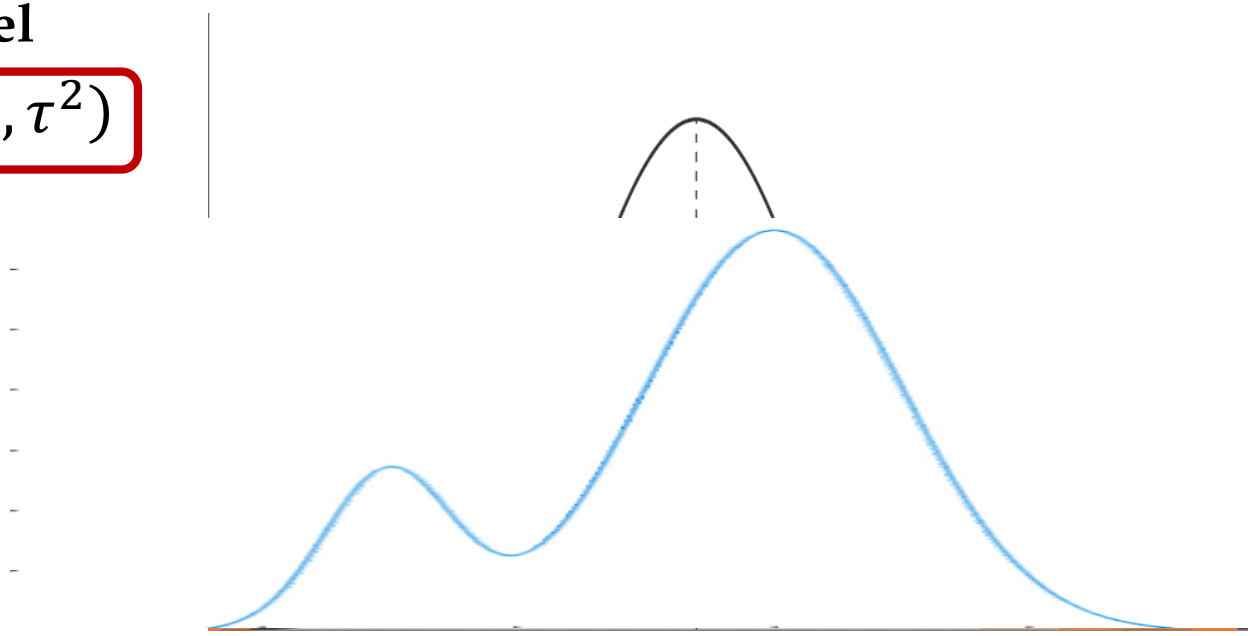


Conventional random effects meta-analysis

Random-effects model

$$\theta_i \sim N(\mu, \tau^2)$$

- outlying studies
- effect modifiers
  - known and observed
  - unknown or unobserved



- A systematic review to identify alternative flexible meta-analysis models that relax the between study normality assumption
- A simulation study to investigate and compare their performance under the presence of substantial heterogeneity for normal and non-normal data

Results of systematic review

16 eligible articles suggesting 14 alternative between-study distributions:

1. long-tail and skewed extensions
2. mixture of distributions
3. models using based on Dirichlet Process priors

- most in Bayesian framework
- most provided code and few accompanied by an R package

Identified alternative random effects models
long-tail and skewed distributions (1)

- **t-distribution model** $\theta_i \sim t(\mu, \omega, \nu)$
  - with $\tau^2 = \frac{\nu}{(\nu-2)}\omega^2$
  - more weight in the tails – outliers less influential
  - *metaplus* package in R

ω: scale parameter
ν: degrees of freedom
(determining the weight of the tails)
ξ: location parameter
γ: shape parameter
(determining the level of skewness)

- **skew-normal model** $\theta_i \sim SN(\xi, \omega, \gamma)$
  - with $\mu = \xi + \omega b\delta$, $\tau^2 = \omega^2[1 - (b\delta)^2$, $b = \sqrt{\frac{2}{\pi}}$, $\delta = \frac{\gamma}{\sqrt{1+\gamma^2}}$
  - *flexmeta* package in R

Identified alternative random effects models
long-tail and skewed distributions (2)

- **skew-t model** $\theta_i \sim ST(\xi, \omega, \nu, \gamma)$
  - with $\mu = \xi + \omega b_\nu \delta$, $\tau^2 = \omega^2 \left[\frac{\nu}{\nu-2} - (b_\nu\, \delta)^2\right]$,

    $b_\nu = \frac{\sqrt{\nu}\Gamma\left(\frac{\nu-1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}$, $\Gamma(\nu) = (\nu-1)!$

  - *flexmeta* package in R
  - bivariate extensions (assuming correlation of treatment effect and baseline risk or DTA meta-analysis) *Lee & Thompson* ***Stat Med*** *2008, Negeri et al.* ***Biom J*** *2020*

ω: scale parameter
ν: degrees of freedom
(determining the weight of the tails)
ξ: location parameter
γ: shape parameter
(determining the level of skewness)

- **other skewed distributions**
  - asymmetric Subbotin II $\theta_i \sim AS2(\xi, \omega, \nu, \gamma)$
    sharper skewness and excess kurtosis for very small ν
  - Jones–Faddy $\theta_i \sim JF(\xi, \omega, \gamma, d)$ – equivalent to a t-distribution with $\gamma + d$ dof
  - sinh-arcsinh $\theta_i \sim SAS(\xi, \omega, \gamma, d)$
    allows for both symmetric and skewed shapes as well as heavy or light tail-weight
  - *flexmeta* package in R

Identified alternative random effects models
mixture of distributions (1)

- when the data come from sub-populations or several outliers are present

- **common-mean mixture model** $\theta_i \sim w_1 N(\mu, \tau_1^2) + w_2 N(\mu, \tau_2^2)$
  - for outlier detection
  - weights propotional to the number of studies
  - summary effect is estimated including all studies but with outliers being down-weighted due to the larger variance assumed for their class
  - if $\tau_1^2 \approx \tau_2^2$, suggests the absence of outliers
  - extension for covariates
  - *metaplus* package in R

w: the weight in the mixture

Identified alternative random effects models
mixture of distributions (2)

- **mixture of bivariate normal distributions**

$$\begin{pmatrix}\theta_{i1}\\ \theta_{i2}\end{pmatrix} \sim w_1 N\left(\begin{pmatrix}\mu_{11}\\ \mu_{12}\end{pmatrix}\begin{matrix}\mu_{11}\\ \mu_{12}\end{matrix}, \Sigma_1\right)\begin{pmatrix}\mu_{11}\\ \mu_{12}\end{pmatrix}, \Sigma_1\Big) + \cdots + \; w_k N\left(\begin{pmatrix}\mu_{k1}\\ \mu_{k2}\end{pmatrix}\begin{matrix}\mu_{k1}\\ \mu_{k2}\end{matrix}, \Sigma_k\right)\begin{pmatrix}\mu_{k1}\\ \mu_{k2}\end{pmatrix}, \Sigma_k\Big)$$

  - DTA meta-analyses
  - to identify latent subgroups of studies and estimate sensitivity and specificity for each subgroup
  - *CAMAN* and *mada* R packages
  - extension for covariates predicting the latent subgroup classification

- **mixture of multivariate normal distributions**
  - longitudinal data

- **random-intercept mixture of regressions**
  - meta-analysis with multiple outcome reports nested within studies

k: number of mixture components (subgroups)
$w_k > 0$: the weight in the mixture

Identified alternative random effects models

Dirichlet process (DP) prior mixture models

- to identify the potential underlying clustering of the data (e.g. relevant subgroups of studies)

$$\theta_i \sim F$$
$$F \sim DP(\alpha, F_0)$$
$$F_0 \sim N(\mu_b, \tau_b^2)$$

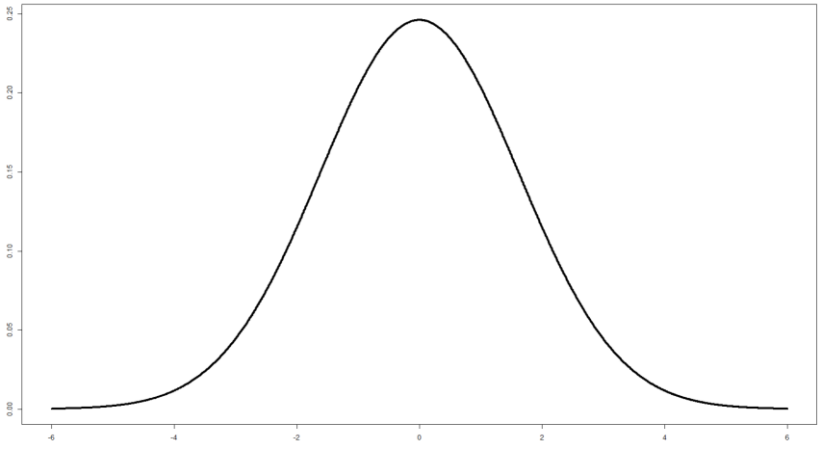
  - with α the concentration parameter that measures the variability of F around F_0 (high values suggest that F is 'close' to F_0)
  - F_0 is the base distribution that controls the mean of the process
  - a truncation that allows obtaining a plausible approximation to F is usually applied (e.g. the number of studies)
  - assumption of a discrete (mixture of points) or a continuous distribution (mixture of distributions)
  - conditional DP for small number of studies: conditional distribution for F given that the posterior median of F is μ_b
  - *bspmma, DPpackage* R packages

Models compared in the simulation

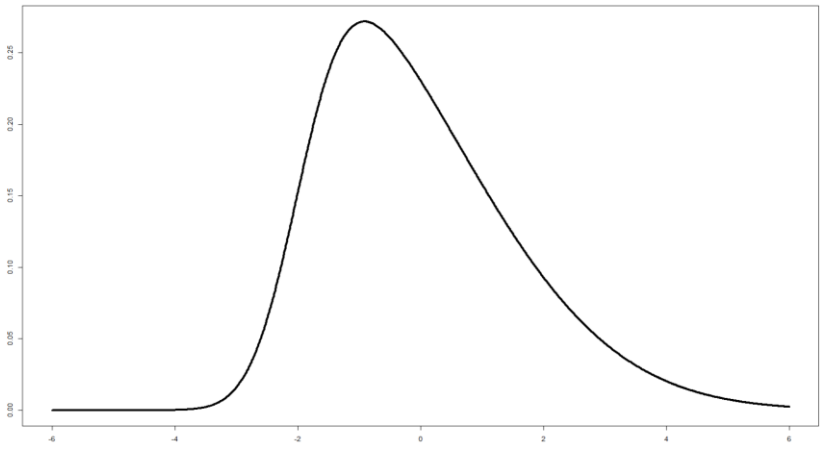
	Framework fitted	Within-study distribution	Between-study distribution	Prior distributions
1	Bayesian	Binomial	Normal	$\tau \sim HN(0,1)$
2	Bayesian	Binomial	Normal	$\tau \sim U(0,10)$
3	Frequentist	Normal	Normal	-
4	Bayesian	Binomial	t-distribution	$\omega \sim U(0,20)$
5	Frequentist	Normal	t-distribution	-
6	Bayesian	Binomial	Skew Normal	$\omega \sim U(0,20)$
7	Frequentist	Normal	Mixture of 2 normal distributions with common mean	-
8	Bayesian	Binomial	DP mixture of points	$\tau \sim HN(0,1)$, $\alpha \sim U(0.3,5)$
9	Bayesian	Binomial	DP mixture of points	$\tau \sim HN(0,1)$, $\alpha \sim U(0.3,10)$
10	Bayesian	Binomial	DP mixture of points	$\tau \sim U(0,10)$, $\alpha \sim U(0.3,5)$
11	Bayesian	Binomial	DP mixture of points	$\tau \sim U(0,10)$,$\alpha \sim \Gamma(1,1)$

Data generating process

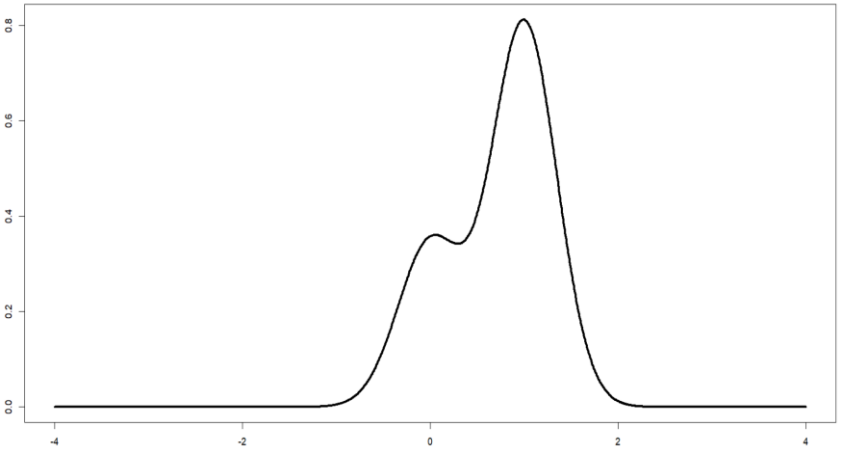
Shape of the true distributions



Normal



Skew normal



Mixture of two normal distributions

22 scenarios varying

- the number of studies $n = 14,26$
- the true mean treatment effect $\mu = 0,0.5$ or $\mu_1 = 0$ and $\mu_2 = 1$ (for the mixture scenarios)
- the heterogeneity, $\tau^2 = 0.12, 2.63$

For each scenario we generated 1000 datasets

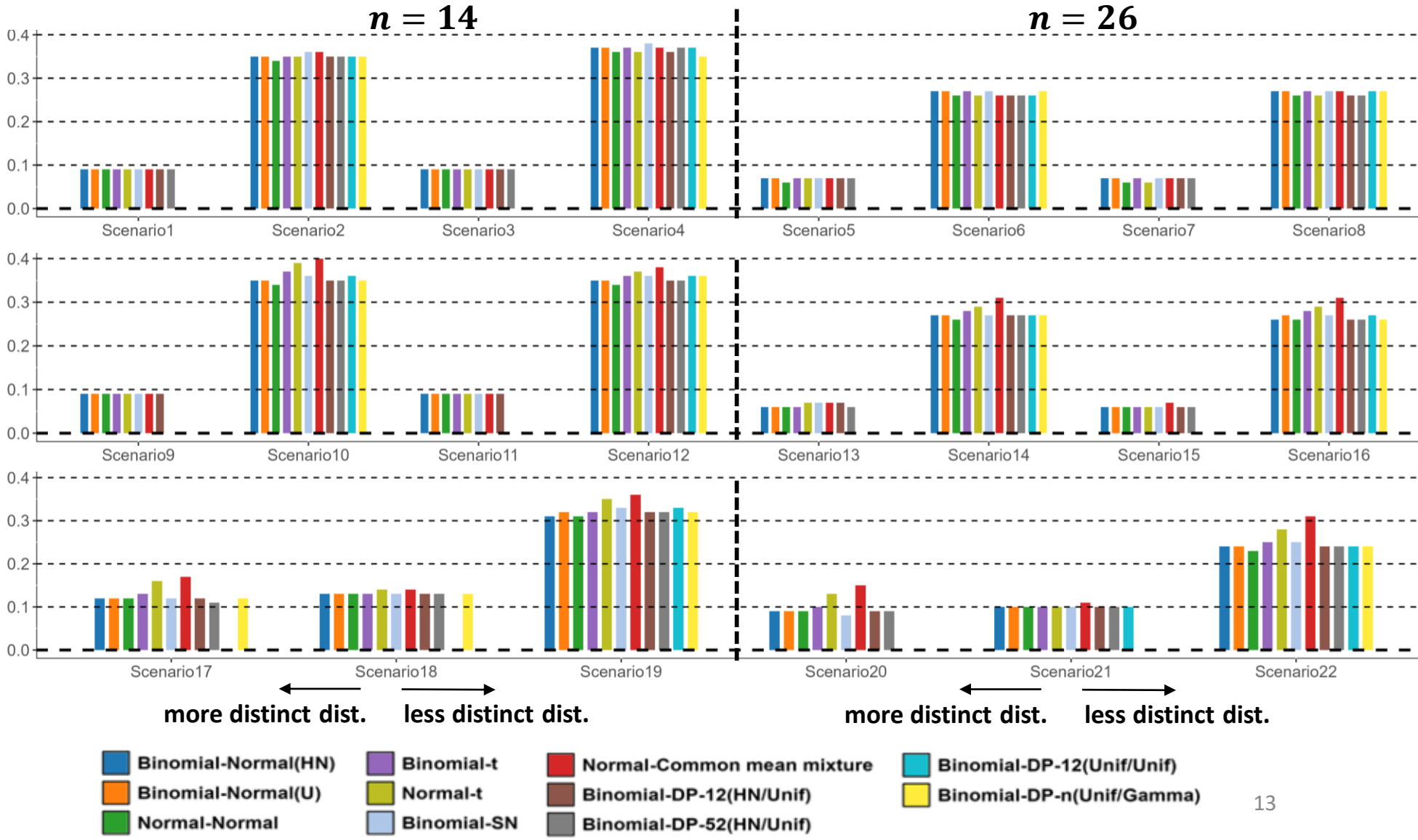
Simulation results

Average bias for the mean treatment effect (μ)

- **normal true dist.**
- sc. 1, 3, 5, 7: $\tau^2 = 0.12$
- sc. 2, 4, 6, 8: $\tau^2 = 2.63$
- sc. 1, 2, 5, 6: $\mu = 0$
- sc. 3, 4, 7, 8: $\mu = 0.5$

- **skew-normal true dist.**
- sc. 9, 11, 13, 15: $\tau^2 = 0.12$
- sc. 10, 12, 14, 16: $\tau^2 = 2.63$
- sc. 9, 10, 13, 14: $\mu = 0$
- sc. 11, 12, 15, 16: $\mu = 0.5$

- **normal mixture true dist.**
- $\mu_1 = 0, \mu_2 = 1$
- $\tau_1^2 = 0.12$
- sc. 17, 20: $\tau_2^2 = 0.005$
- sc. 18, 21:$\tau_2^2 = 0.12$
- sc. 19, 22:$\tau_2^2 = 2.63$



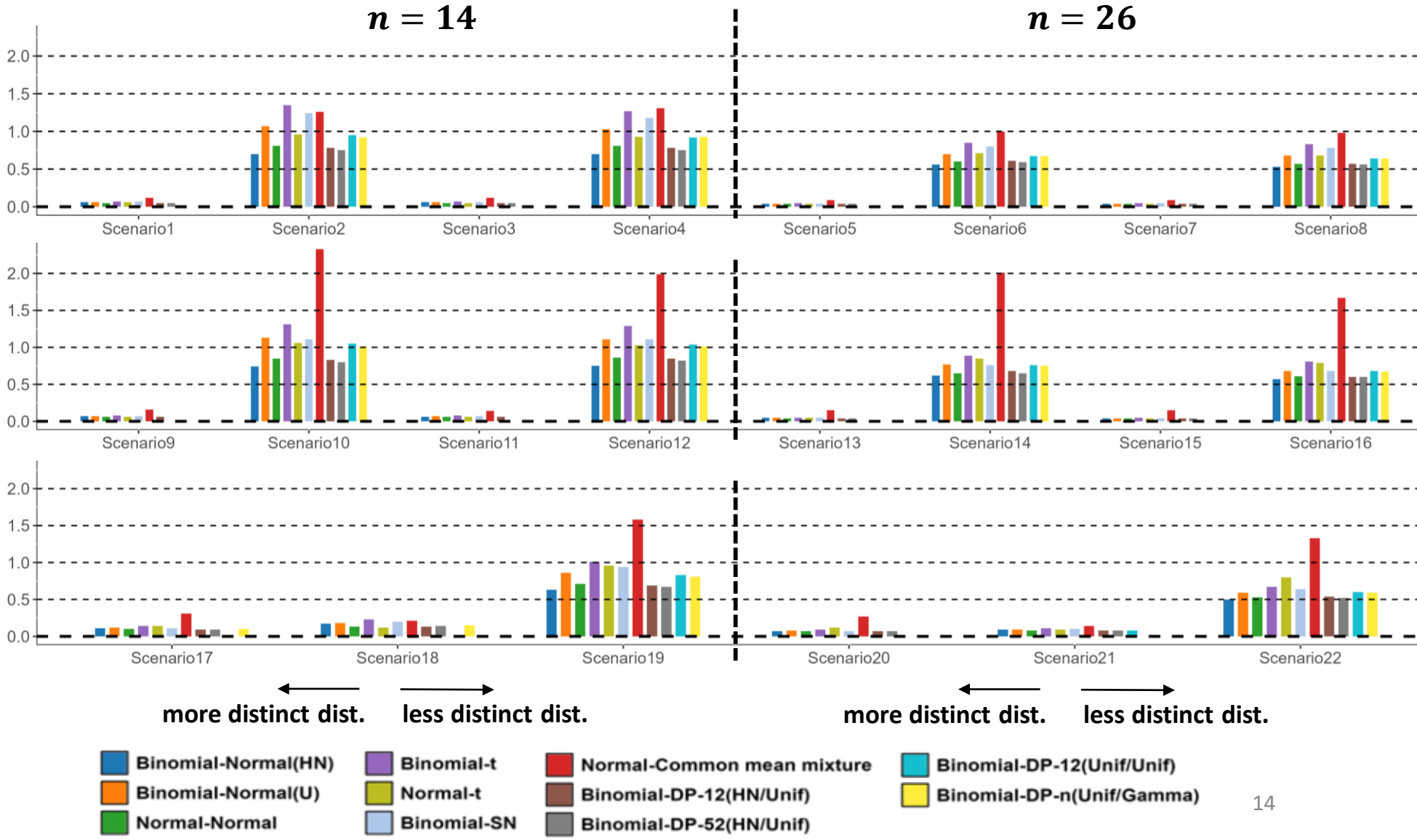
Simulation results

Average bias for the heterogeneity (τ^2)

- **normal true dist.**
- sc. 1, 3, 5, 7: $\tau^2 = 0.12$
- sc. 2, 4, 6, 8: $\tau^2 = 2.63$
- sc. 1, 2, 5, 6: $\mu = 0$
- sc. 3, 4, 7, 8: $\mu = 0.5$

- **skew-normal true dist.**
- sc. 9, 11, 13, 15: $\tau^2 = 0.12$
- sc. 10, 12, 14, 16: $\tau^2 = 2.63$
- sc. 9, 10, 13, 14: $\mu = 0$
- sc. 11, 12, 15, 16: $\mu = 0.5$

- **normal mixture true dist.**
- $\mu_1 = 0, \mu_2 = 1$
- $\tau_1^2 = 0.12$
- sc. 17, 20: $\tau_2^2 = 0.005$
- sc. 18, 21:$\tau_2^2 = 0.12$
- sc. 19, 22:$\tau_2^2 = 2.63$



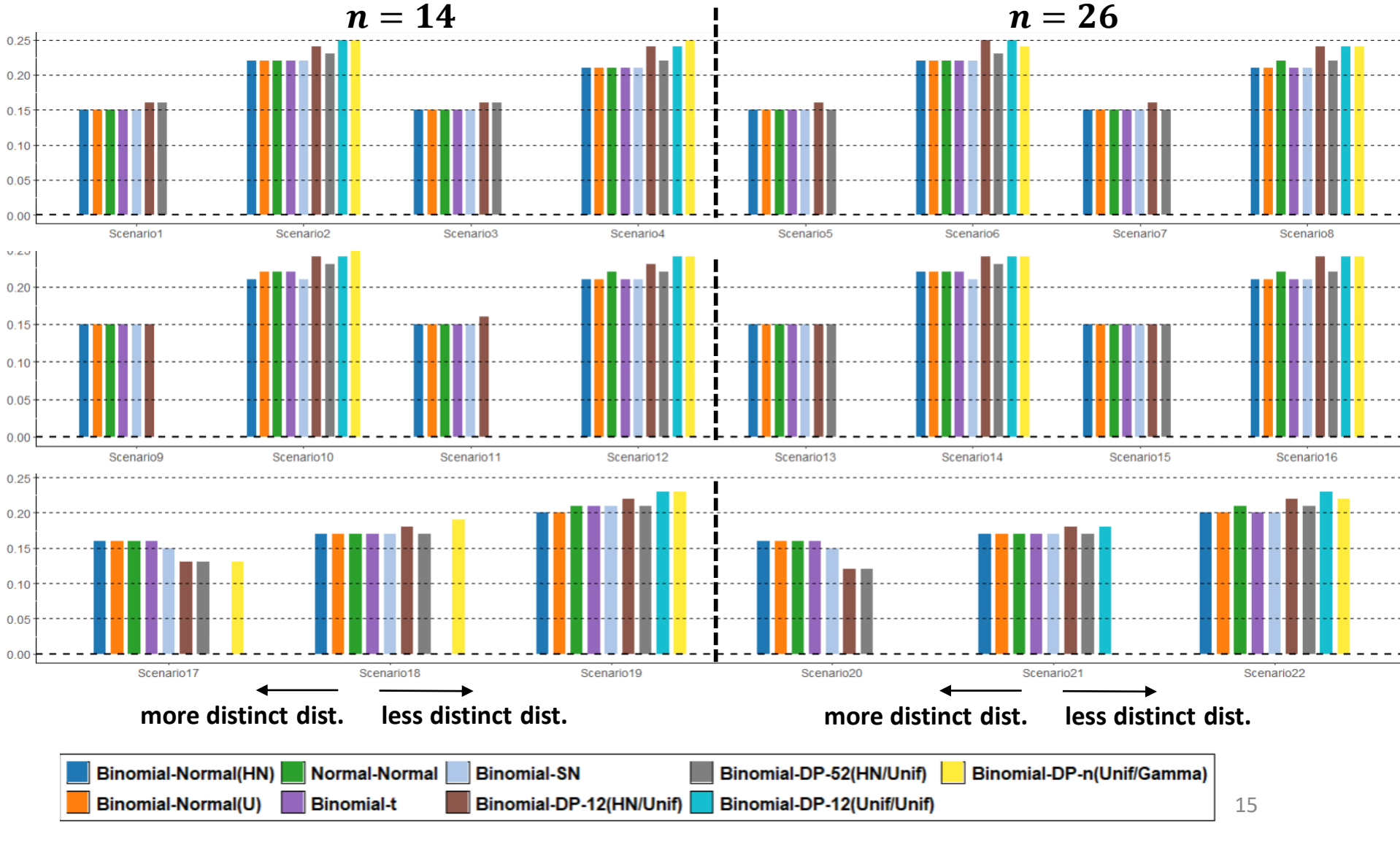
Simulation results

Average bias for the study-specific effects (θ_i)

- **normal true dist.**
- sc. 1, 3, 5, 7: $\tau^2 = 0.12$
- sc. 2, 4, 6, 8: $\tau^2 = 2.63$
- sc. 1, 2, 5, 6: $\mu = 0$
- sc. 3, 4, 7, 8: $\mu = 0.5$

- **skew-normal true dist.**
- sc. 9, 11, 13, 15: $\tau^2 = 0.12$
- sc. 10, 12, 14, 16: $\tau^2 = 2.63$
- sc. 9, 10, 13, 14: $\mu = 0$
- sc. 11, 12, 15, 16: $\mu = 0.5$

- **normal mixture true dist.**
- $\mu_1 = 0, \mu_2 = 1$
- $\tau_1^2 = 0.12$
- sc. 17, 20: $\tau_2^2 = 0.005$
- sc. 18, 21:$\tau_2^2 = 0.12$
- sc. 19, 22:$\tau_2^2 = 2.63$



Simulation results

Coverage probability and mean square error

Mean square error

- For scenarios with large heterogeneity
  - normal-normal and binomial-DP models – generally smaller MSE for the mean treatment effect estimate
  - binomial-normal(HN) – the smallest MSE for the heterogeneity estimate followed by the binomial-DP(HN) and the normal-normal models

Coverage probability

- For scenarios with large heterogeneity
  - normal and skew-normal scenarios with 26 studies: normal models – best coverage of the mean treatment effect
  - mixture scenarios: binomial-DP models best coverage overall of the mean treatment effect followed by the binomial-normal models
  - choice of prior more important than the choice of the model for heterogeneity

Discussion

- When substantial heterogeneity among studies is suspected or outlying studies are present, focusing on the mean treatment effect may lead to spurious conclusions

- Average bias of both the mean treatment effect and the heterogeneity is substantial in presence of high heterogeneity regardless of the model used

- These results are in agreement with smaller simulations studies when they found the alternative models mostly beneficial in terms of precision and model fit

- In meta-analyses where the distribution of the data seems multimodal, mixture models may result in more accurate estimates of the study-specific effects and should be considered

- Semi-parametric models (e.g. DP models) may assist identifying homogeneous subgroups of studies when potential effect modifiers are unobserved

Limitations

- Using scenarios with more studies might have improved the performance of some models
- We used a fixed skewness parameter that might have not resulted in many highly skewed datasets to properly assess the performance of the skew-normal model
- We did not compare all the identified models in our simulation
- We only used the normal distribution as base distribution in the DP models

Acknowledgements

Kanella Panagiotopoulou
Theodoros Evrenoglou
Christopher Schmid
Silvia Metelli

This project is supported by the French National Research Agency under the project ANR-22-CE36-0013-01

RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

anr agence nationale de la recherche

https://www.cer-methods.com/