Alternative distributions for the random effects meta-analysis model

Panagiotopoulou K[1], Evrenoglou T[1], Metelli S[1], Chaimani A[1,2]

[1] *Université Paris Cité, Center of Research in Epidemiology and Statistics, Inserm, Paris, France*

[2] Cochrane France, Paris

Introduction

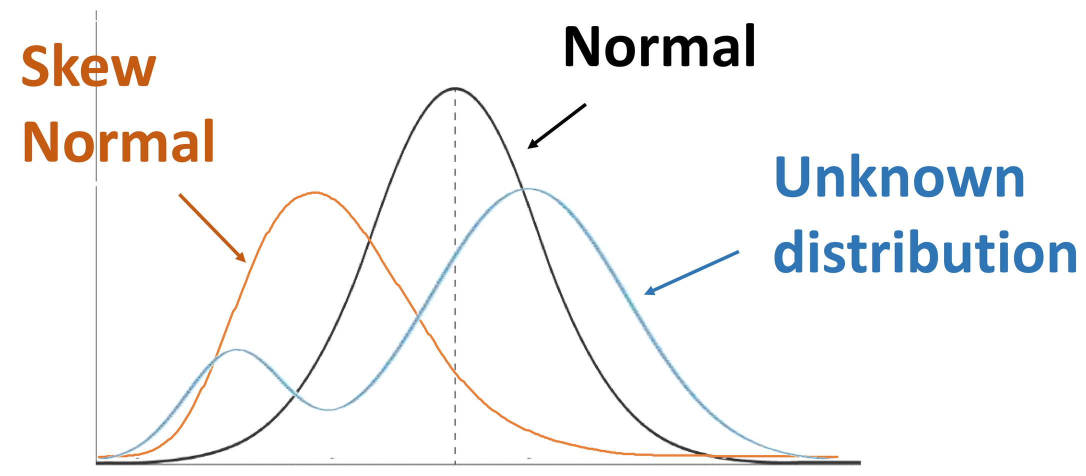
In the vast majority of random effects meta-analyses, the **between study variability is modelled through a normal distribution**.

However, in the presence **of skewed data or when substantially heterogeneous studies and populations** are synthesized, this normality assumption may threaten the validity of the results.

The main reasons for almost always assuming a normal between-study distribution are:

- Convenience,
- Model simplicity
- Tradition
- Software availability.

Potential true distributions of the underlying effects include



Aim

To identify through a systematic review, meta-analysis models that relax the between-study normality assumption and investigate their performance through a simulation study

Methods

Systematic review

Search for published articles presenting new methods or reviewing and assessing articles evaluating existing methods for meta-analysis that avoid the between-study normality assumption.

- Main search database was **PubMed**
- Hand-searching was conducted in **Google Scholar** and in **related journals** (such as Journal of the American Statistical Association, Annals of Statistics, etc.)

- **Interested in** articles introducing methodological reviews and simulation studies or commentaries on the properties and characteristics of the methods of interest. Articles only presenting applications of meta-analysis methods and overviews of reviews were excluded.

Simulation study

We explored in total 20 scenarios with varying the number of studies included (k=14, 26), true treatment effect (μ, 0 for no effect and 0.5 or 1 otherwise)and magnitude of heterogeneity (τ^2, 0.07 for low to moderate and 1.38 for high). The data generation was based on empirical evidence.

Meta-analysis data sets are generated based on

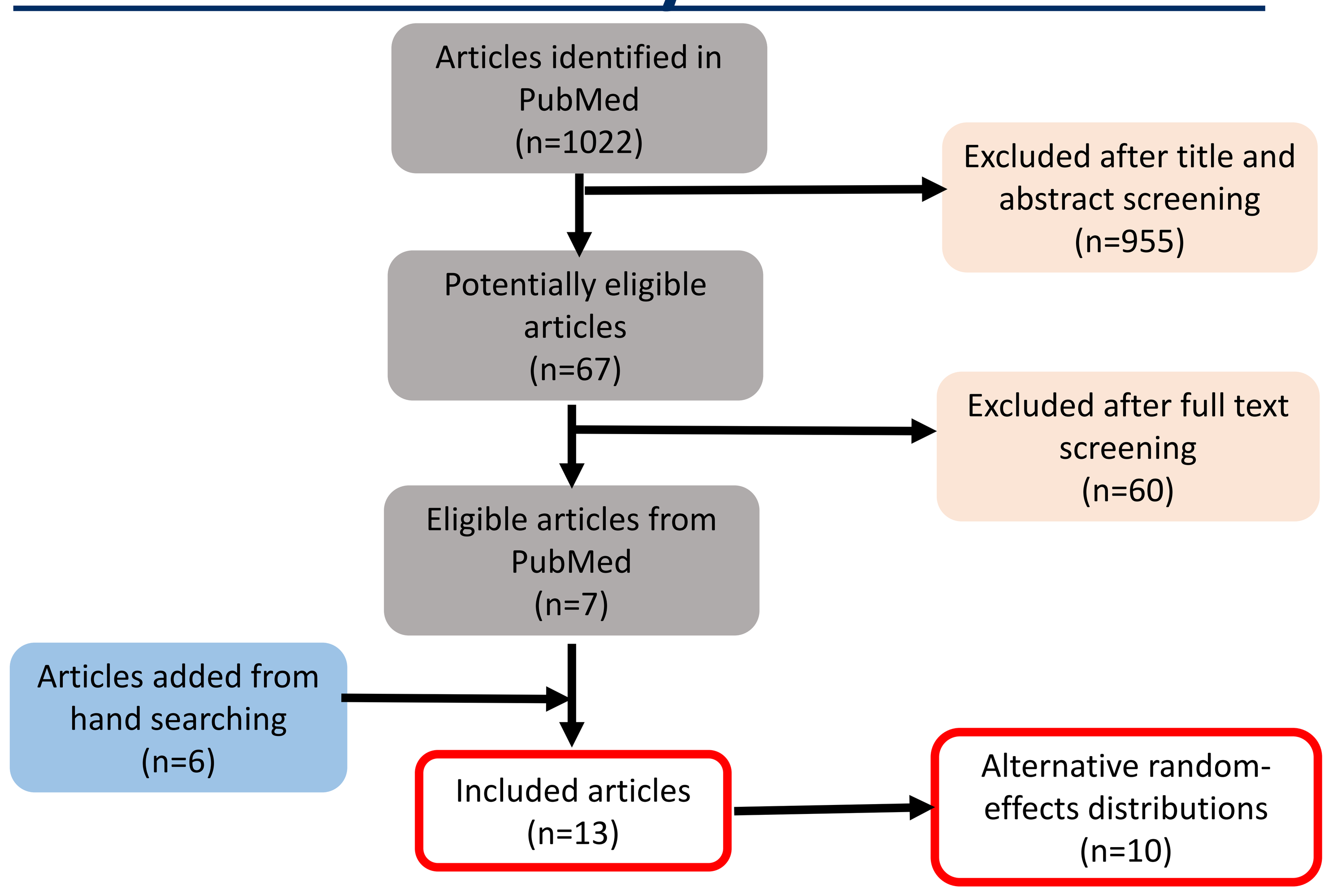
- Normal distribution, $N(\mu, \tau^2)$
- Skew normal distribution, $N(\mu, \tau^2, 0.5)$
- Mixture of two normal distributions, $0.3N(\mu_1, \tau_1^2) + 0.7N(\mu_2, \tau_2^2)$

Meta-analysis models under evaluation

- **Normal model**
- **Skew normal model**
- **Dirichlet Process model**

For each scenario we used 1000 data sets

Results from the Systematic Review



Summary of the identified alternative random effects distributions

Skewed extensions of normal or t-distribution

Positively or negatively skewed

$$\theta_i \sim SN(\mu, \tau^2, v)$$

- (v): parameter which regulates the skewness

Mixture of distributions

Combines two or more distributions

$$\text{e.g. } \theta_i \sim w_1 N(\mu_1, \tau_1^2) + w_2 N(\mu_2, \tau_2^2)$$

- (w_1, w_2): mixing proportions

Dirichlet process mixture models

Creates clusters of the underlying effects.

$$\theta_i \sim DP(a, G_0)$$

- (a):concentration parameter, regulates the concentration of data points within each cluster
- (G_0): base distribution, controls the mean of the process

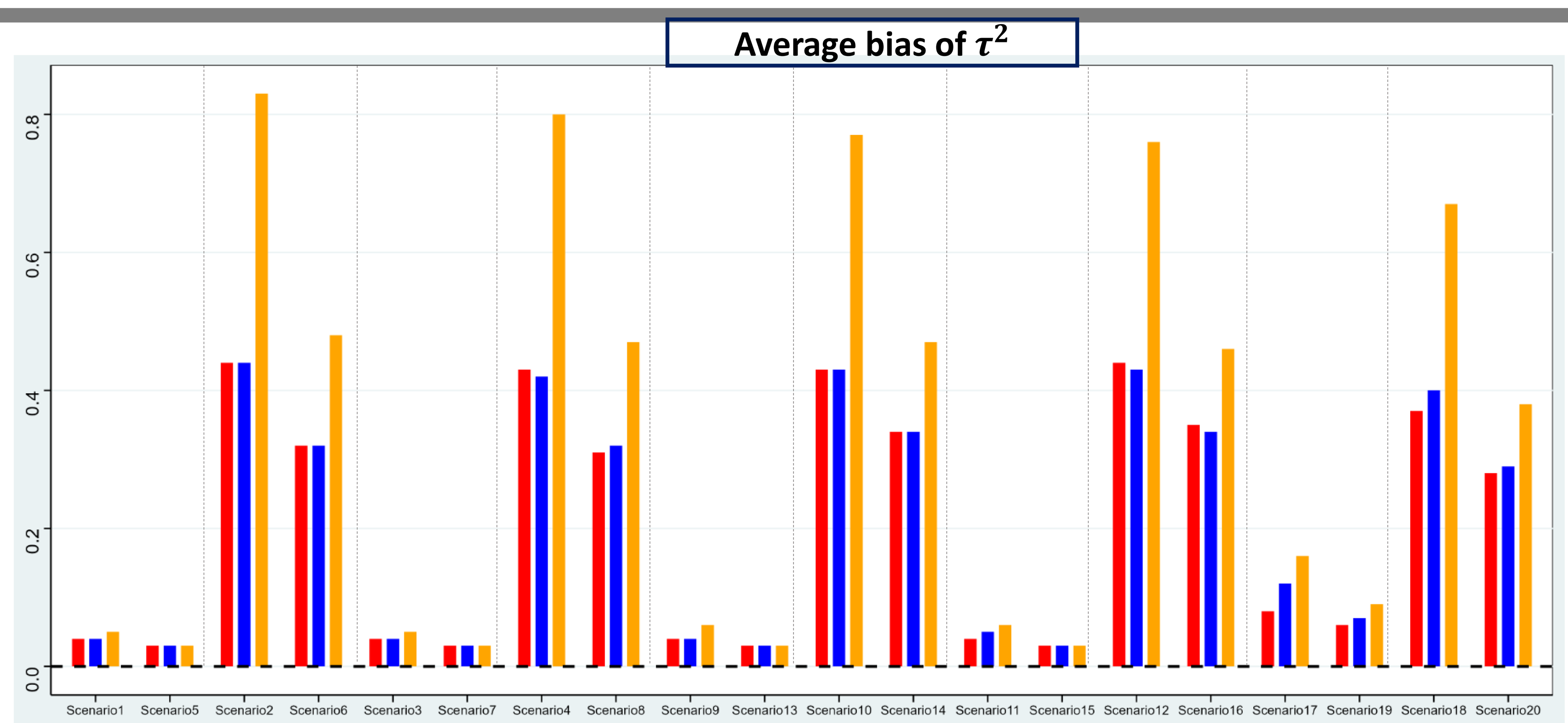
e.g. $G_0 = N(\mu, \tau^2)$

Simulation Results

Data \ (μ, τ^2)	(0,0.07)	(0,1.38)	(0.5,0.07)	(0.5,1.38)
Normal	Scenario1,5	Scenario2,6	Scenario3,7	Scenario4,8
Skew normal	Scenario 9,13	Scenario10,14	Scenario11,15	Scenario12,16
	N(0,0.07)+N(1,0.07)		N(0,0.07)+N(1,1.38)	
Mixture	Scenario17,19		Scenario18,20	

Number of studies:
k1=14: Scenario 1-4, 9-12 & 17, 18
k2=26: Scenario 5-8, 13-16 & 19, 20

Models
- NORMAL
- SKEW NORMAL
- DP



Discussion

- The Dirichlet process model performs well in case of normal and/or non-normal data.
- The Skew normal model appears to be the worst in case of high heterogeneity.
- Alternative random effects distributions provide more information about the structure of the data set.

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