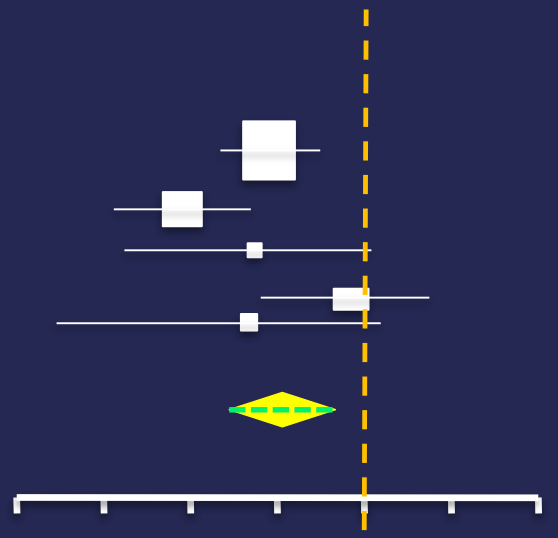
Methods to estimate the between-study variance and to calculate uncertainty in the estimated overall effect size



Areti Angeliki Veroniki, PhD

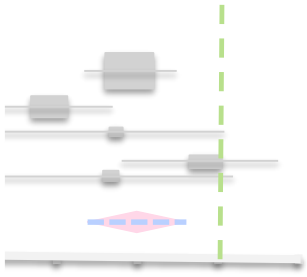
October 09, 2018

School of Education,
University of Ioannina,
Ioannina,
Greece

EVIDENCE SYNTHESIS METHODS
STATISTICS TEAM

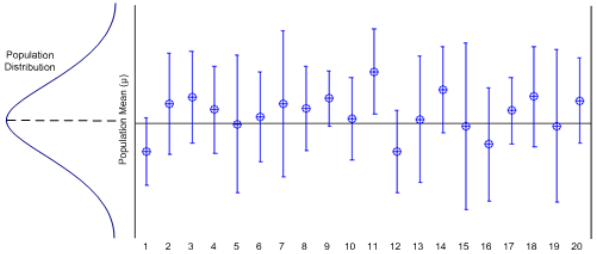
Competing Interests

I have no actual or potential conflict of interest in relation to this presentation



Webinar objectives

- To give an overview of the available methods for estimation of the between-study variance and its corresponding uncertainty
  - Can different methods impact our decision-making?
- To give an overview of the available methods to calculate confidence intervals for the overall effect size
  - What are the properties of the different methods?
- To present real-life and simulation findings that compare the methods
  - Which method is the most appropriate to apply? Are any methods preferable than others?
- To discuss potential issues surrounding the computation of prediction intervals



Work conducted on behalf of the Cochrane Statistical Methods Group

Invited Review

Research Synthesis Methods

Received 26 June 2014, Revised 20 May 2015, Accepted 24 June 2015 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/jrsm.1164

Methods to estimate the between-study variance and its uncertainty in meta-analysis

Areti Angeliki Veroniki,[a*] Dan Jackson,[b] Wolfgang Viechtbauer,[c] Ralf Bender,[d] Jack Bowden,[e] Guido Knapp,[f] Oliver Kuss,[g] Julian PT Higgins,[h,i] Dean Langan[i] and Georgia Salanti[j]

Meta-analyses are typically used to estimate the overall/m
inference about between-study variability, which is typically
parameter, is usually an additional aim. The DerSimonian ar

Cochrane Methods

Articles

Recommendations for quantifying the uncertainty in the summary intervention effect and estimating the between-study heterogeneity variance in random-effects meta-analysis

Areti Angeliki Veroniki, Dan Jackson, Wolfgang Viechtbauer, Ralf Bender, Guido Knapp, Oliver Kuss, Dean Langan

has also been suggested that the quantile-approximation[12], t, and Knapp and Hartung[17,19] (HKSJ for heterogeneity > 0) methods have coverage closer to the nominal level than the Wt method.[12] An advantage of the HKSJ method is that it is insensitive to the magnitude and estimator of heterogeneity, as well the number of studies included in a meta-analysis.[8]

A prediction interval of the possible intervention effect in an individual setting can also be calculated, to facilitate the interpretation of the meta-analysis result.[20–22]

Inference for the between-study heterogeneity variance

The heterogeneity variance can be estimated using various approaches, including the method proposed by DerSimonian and Laird (DL)[9] that is the most commonly implemented approach

Received: 9 November 2017 | Revised: 23 May 2018 | Accepted: 13 August 2018

DOI: 10.1002/jrsm.1319

ronto,

RESEARCH ARTICLE

WILEY Research Synthesis Methods

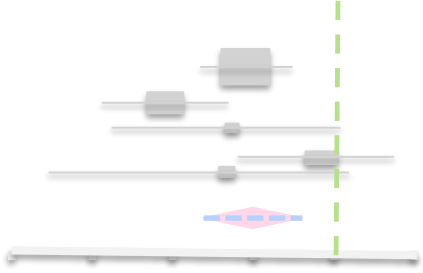
Methods to calculate uncertainty in the estimated overall effect size from a random-effects meta-analysis

Areti Angeliki Veroniki[1,2] | Dan Jackson[3] | Ralf Bender[4] | Oliver Kuss[5,6] | Dean Langan[7] | Julian P.T. Higgins[8] | Guido Knapp[9] | Georgia Salanti[10]

[1]Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, Canada

[2]Department of Primary Education,

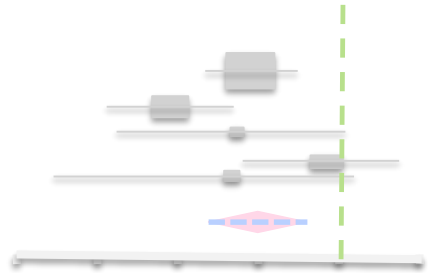
Meta-analyses are an important tool within systematic reviews to estimate the overall effect size and its confidence interval for an outcome of interest. If het-



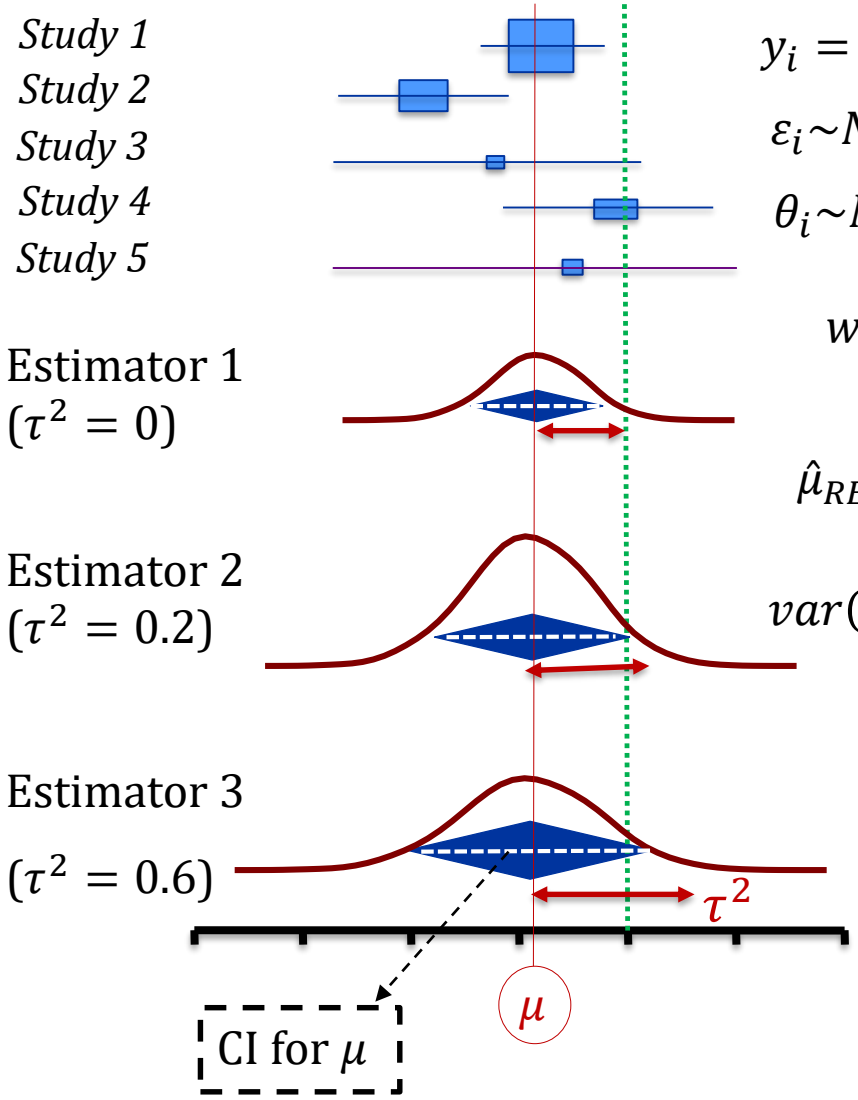
Work conducted on behalf of the Cochrane Statistical Methods Group

Acknowledgments:

- Dr. Dan Jackson
- Prof. Ralf Bender
- Dr. Oliver Kuss
- Dr. Dean Langan
- Prof. Julian PT Higgins
- Dr. Guido Knapp
- Dr. Jack Bowden
- Dr. Wolfgang Viechtbauer
- Dr. Georgia Salanti



Introduction



$$y_i = \theta_i + \varepsilon_i$$

$$\varepsilon_i \sim N(0, v_i)$$

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

$$w_{i,RE} = \frac{1}{v_i + \hat{\tau}^2}$$

$$\hat{\mu}_{RE} = \frac{\sum y_i w_{i,RE}}{\sum w_{i,RE}}$$

$$var(\hat{\mu}_{RE}) = \frac{1}{\sum w_{i,RE}}$$

- The choice of the method for estimating
  - between-study variance (heterogeneity) and its uncertainty
  - uncertainty for the overall effect size

is important when conducting a meta-analysis.

- When no appropriate methods are used, this can seriously jeopardize results, and interpretation difficulties may occur.

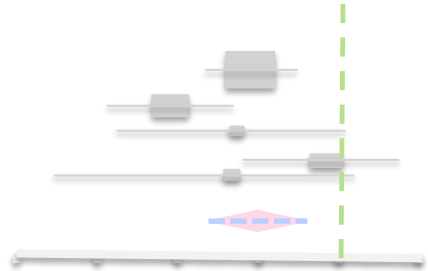
Have you ever used a different, other than the default option, between-study variance estimator?

a) Yes, I have used different methods to assess the robustness of the meta-analysis results

b) Yes, I have used different methods, as deemed appropriate, in different meta-analyses

c) No, I always use the default option in the relevant meta-analytic software

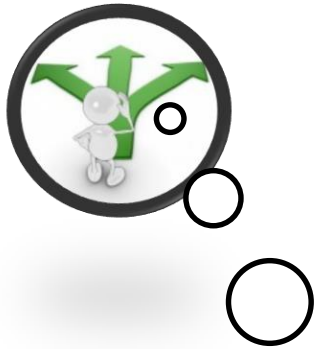
d) No, I was not aware that different methods exist

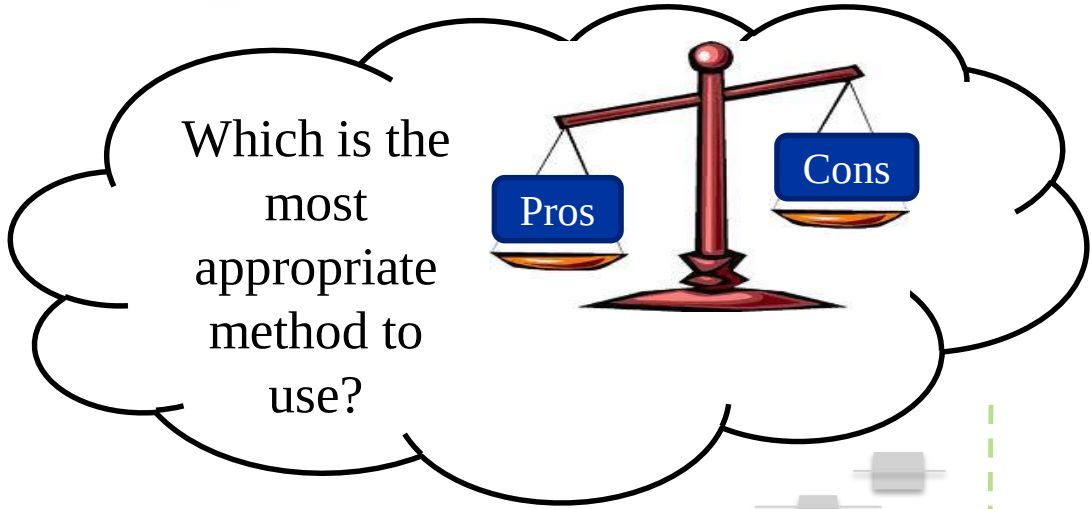


Illustrative example

Thorlund et al. RSM 2011

RE meta-analysis of corticosteroids for preventing death caused by tuberculosis meningitis.





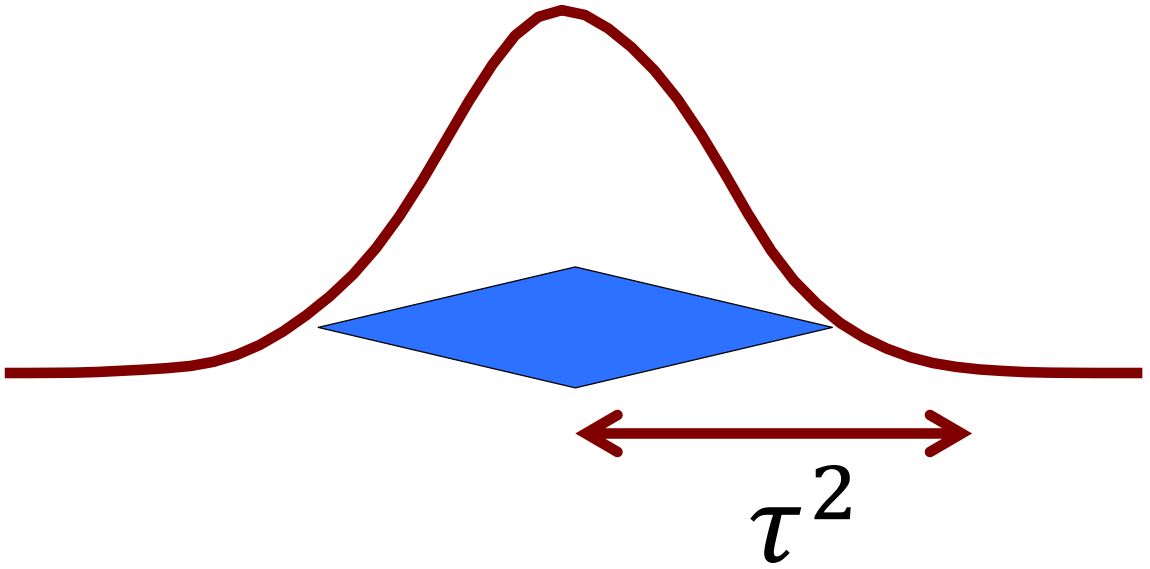
Study (year)	Corticosteroids n/N	Control n/N	Risk ratio (random) (95% CI)	Weight (%)	Risk ratio (random) (95% CI)
O'Toole (1969)	6/11	9/12		5.4	0.73 (0.39 to 1.37)
Girgis (1991)	72/145	79/135		45.5	0.85(0.68 to 1.05)
Kumarvelu (1994)	5/20	7/21		2.3	0.75 (0.28 to 1.98)
Chotmongkol (1996)	5/29	2/30		0.8	2.59 (0.54 to 12.29)
Schoeman (1997)	4/67	13/67		1.9	0.31 (0.11 to 0.90)
Lardizabal (1998)	4/29	6/29		1.6	0.67 (0.21 to 2.12)
Thwaites (2004)	87/274	112/271		42.5	0.77 (0.61 to 0.96)
Total DL (95% CI)					**0.79 (0.69 to 0.92)**

0.2 0.5 1 2 5

Favors corticosteroids Favors control

DL (normal) - Test for overall effect: P=0.002; Heterogeneity: D_{DL}^2=0%	
DL (t-dist) - Test for overall effect: P=0.02	0.79 (0.66 to 0.96)
HM (normal) – Test for overall effect: P=0.021; Heterogeneity: D_{HM}^2=52.2%	0.78 (0.63 to 0.96)
HM (normal) – Test for overall effect: P=0.051	0.78 (0.61 to 1.00)
REML (normal) - Test for overall effect: P=0.002; Heterogeneity: D_{REML}^2=0%	0.79 (0.69 to 0.92)
REML (t-dist) - Test for overall effect: P=0.02	0.79 (0.66 to 0.95)
HE (normal)- Test for overall effect: P=0.168; Heterogeneity: D_{HE}^2=86.7%	0.75 (0.50 to 1.13)
HE (t-dist)- Test for overall effect: P=0.135	0.75 (0.51 to 1.13)
SJ (normal) - Test for overall effect: P=0.117; Heterogeneity: D_{SJ}^2=82.1%	0.76 (0.54 to 1.07)
SJ (t-dist) - Test for overall effect: P=0.113	0.76 (0.53 to 1.09)

1. Inference on the heterogeneity



Literature Review of the between-study variance methods

Our search identified:

- 16 methods to estimate the between-study variance (grouped in 5 broad categories).
- 9 methods to calculate the confidence interval for the between-study variance (grouped in 6 broad categories)

The properties of the methods were evaluated in multiple simulation studies and/or real-life data evaluations of ≥2methods

Categories

A. Method of moments estimators
   i. DerSimonian and Laird (DL)
   ii. Positive DerSimonian and Laird (DLp)
   iii. Hedges and Olkin (HO)
   iv. Hartung and Makambi (HM)
   v. Hunter and Schmidt (HS)
   vi. Two-step Dersimonian and Laird (DL2)
   vii. Two-step Hedges and Olkin (HO2)
   viii. Paule and Mandel (PM)
B. Maximum likelihood estimators
   i. Maximum Likelihood (ML)
   ii. Restricted Maximum Likelihood (REML)
   iii. Approximate Restricted Maximum Likelihood (AREML)
C. Model error variance estimators
   i. Sidik and Jonkman (SJ)
D. Bayes estimators
   i. Bayes Modal (BM)
   ii. Rukhin Bayes (RB)
   iii. Full Bayesian (FB)
E. Bootstrap estimators
   i. Non-parametric bootstrap DL (DLb)

Veroniki et al. Res Synth Methods. 2015.
https://doi.org/10.1002/jrsm.1164

Select the most appropriate estimator

1. Is a **zero value** possible?

- Estimators can be <u>positive</u> (with solutions excluding the zero value) or <u>non-negative</u> (with solutions including the zero value)

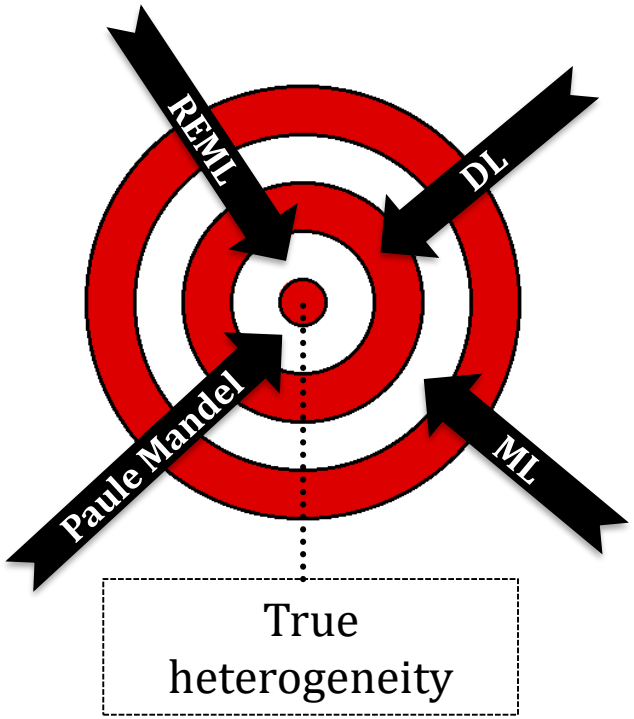
2. Is the estimator **unbiased**?

$$Bias(\hat{\tau}^2) = E(\hat{\tau}^2) - \tau^2 = 0$$

3. Is the estimator **efficient**?

- Low Mean Squared Error (MSE):

$$MSE(\hat{\tau}^2) = E[(\hat{\tau}^2 - \tau^2)^2] = Var(\hat{\tau}^2) + \left(Bias(\hat{\tau}^2)\right)^2$$

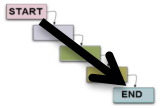


Select the most appropriate estimator

Be aware of the different properties of each estimator!

4. Ease of **computation**

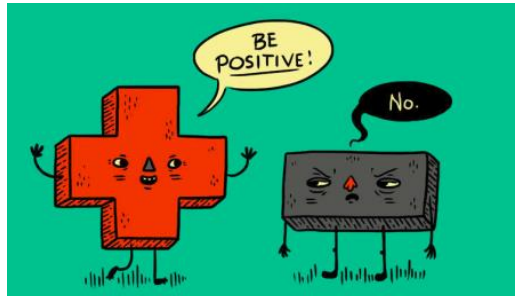
- Does the method include many and complex steps to estimate heterogeneity?
- Is the method **direct** or **iterative**?



Direct methods: provide an estimator in predetermined number of steps

Iterative methods: converge to a solution when a specific criterion is met

- *Iterative* methods do not always produce a result because of **failure to converge** during iterations – e.g., ML depends on the choice of maximization method



Method of Moments Estimators

- The method of moments estimators can be categorized to:

  a) Cochran's Q-based methods

$$Q = \sum_{i=1}^{k} w_{i,FE}(y_i - \hat{\mu}_{FE})^2 \sim \chi^2_{k-1}$$

  b) Generalized Q-based methods

$$Q_{gen} = \sum_{i=1}^{k} w_{i,RE}(y_i - \hat{\mu}_{RE})^2 \sim \chi^2_{k-1}$$

- The Cochran's Q-statistic and generalized Q-statistic, belong to the 'Generalized Cochran between-study variance statistics':

$$Q_a = \sum_{i=1}^{k} a_i(y_i - \hat{\mu}_a)^2 \sim \chi^2_{k-1}$$

with a_i the study weights.

DerSimonian and Kacker 2007, Jackson 2013

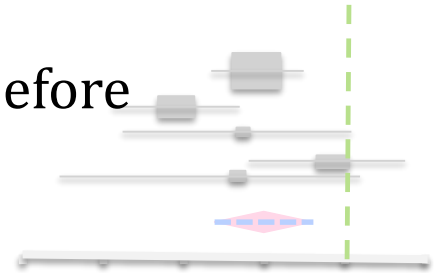
Notation
w_i: weight in study i
y_i: effect size in study i
μ: pooled estimate
k: number of studies in meta-analysis
τ^2 : heterogeneity
FE: fixed-effect model
RE: random-effects model

Method of Moments Estimators

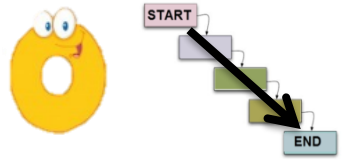
- A method of moments estimator can be derived by equating the expected value of Q_a and its observed value
- Equating Q_a to its expected value and solving for τ^2 we can obtain the generalised method of moments (GMM) estimator:

$$\hat{\tau}^2_{GMM} = \max\left\{0, \frac{Q_a - \left(\sum a_i v_i - \frac{\sum a_i^2 v_i}{\sum a_i}\right)}{\sum a_i - \frac{\sum a_i^2}{\sum a_i}}\right\}$$

- Each method of moments estimator is a special case of the general class of method of moments estimators with different weights a_i
- Under the assumptions of the RE model, known within-study variances, and before truncation of negative values the generalized method moments estimator is unbiased



Method of Moments Estimators –Cochran's Q-based methods



i. DerSimonian and Laird (DL)

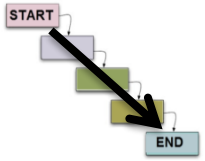
- ☐ The weights used are the inverse of the within-study variances
- ☒ The truncation to zero may lead to biased estimators [1]
- ☑ Performs well with low MSE when τ^2 is small [1, 2, 3]
- ☒ Underestimates true heterogeneity when τ^2 is large and particularly when the number of studies is small [1, 2, 6]

ii. Hedges and Olkin (HO)

- ☐ The weights used are the inverse of the number of studies
- ☑ Performs well in the presence of substantial τ^2 especially when the number of studies is large [1, 2, 3]
- ☒ **But** produces large MSE [4, 5]
- ☒ Not widely used and produces large estimates

1:Viechtbauer JEBS 2005, 2: Sidik and Jonkman Stat Med 2007, 3: Chung et *al* Stat Med 2013, 4: Thorlund et *al* RSM 2012, 5: DerSimonian and Laird Control Clin Trials 1986, 6: Novianti et al Contemp Clin Trials 2014

Method of Moments Estimators –Cochran's Q-based methods



iii. Hartung and Makambi (HM)

- ☐ A modification of DL with weights the inverse of the within-study variances - produces **positive** estimates [1]
- ☑ Is more efficient than DL and performs well for meta-analyses with small and large studies [4]
- ☐ Estimates higher τ^2 values compared to DL estimator [2]
- ☒ For small to medium study sizes and small τ^2 it produces substantial positive bias [4]

iv. Hunter and Schmidt (HS)

- ☐ A modification of DL with weights the inverse of the within-study variances
- ☑ Simple to compute
- ☑ Is more efficient than DL and HO methods [3]
- ☒ The method is associated with substantial negative bias [3]

DL: DerSimonian and Laird

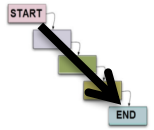
HO: Hedges and Olkin

1:Hartung & Makambi Commun in Stati-Simul and Comp 2003, 2: Thorlund et *al* RSM 2012, 3:Viechtbauer JEBS 2005, 4: Langan et *al* RSM 2018

Method of Moments Estimators – Generalized Q-based methods

i. Two-step Dersimonian and Laird (DL2)

☑ Uses the RE weights, and decreases bias compared to DL [2]



ii. Two-step Hedges and Olkin (HO2)

☑ Uses the RE weights, decreases bias compared to DL and HO [2]

iii. Paule and Mandel (PM)

☐ Uses the RE weights and is equivalent to empirical Bayes method.

☑ Performs best in terms of bias for both dichotomous and continuous data compared to DL, DL2, HO, REML, and SJ [5]

☑ For $\tau^2 = 0$ both DL and PM perform well, but as heterogeneity increases PM approximates τ^2 better compared to DL [1,2, 3, 4, 5]

☒ For mix of small & large studies it may produce higher positive bias than DL, HM, & REML [7]

DL: DerSimonian and Laird
HO: Hedges and Olkin
DL2: Two-step DerSimonian and Laird
REML: Restricted maximum likelihood
SJ: Sidik Jonkman
HM: Hartung and Makambi

1: Bowden et al BMC Med Res Methodol 2011, 2: DerSimonian and Kacker Contemp Clin Trials 2007, 3: Rukhin et al J Stat Plan Inference 2000, 4: Rukhin Journal of the Royal Statistical Society 2012, 5: Novianti et al Contemp Clin Trials 2014, 6: Knapp and Hartung Stat Med 2003, 7 : Langan et *al* RSM 2018

Maximum Likelihood Estimators

i. Maximum Likelihood (ML)

☒ Although it has a small MSE, it is associated with substantial negative bias as τ^2 increases, the number and size of the included studies is small [1, 2, 3, 4]

ii. Restricted Maximum Likelihood (REML)

☑ REML is less downwardly biased than **DL** [1, 2, 5]

☒ For dichotomous data, and small τ^2 and number of studies **REML** tends to have greater MSE than **DL**, but for continuous data **DL** and **REML** have comparable MSEs [1,2, 5, 6]

☒ REML is less efficient than **ML** and **HS** [1]

☑ REML is more efficient with smaller MSE than **HO** [1]

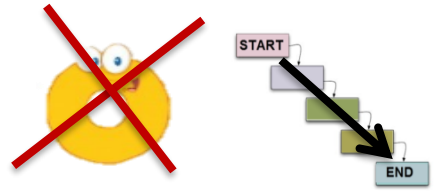
☑ It has relatively low bias and has comparable MSE with **HM** and **DL2** [7]

An ***approximate*** **REML** estimate is also available yielding almost the same results [2, 4]

1:Viechtbauer JEBS 2005, 2: Sidik and Jonkman Stat Med 2007, 3: Chung et al Stat Med 2013, 4: Thompson & Sharp Stat Med 1999, 5: Berkey et al Stat Med 1995, 6: Brockwell and Gordon Stat Med 2001, 7 : Langan et *al* RSM 2018

DL: DerSimonian and Laird
HS: Hunter and Schmidt
HO: Hedges and Olkin
HM: Hartung Makambi
DL2: Two-step DerSimonian and Laird

Model error variance estimators



i. Sidik and Jonkman (SJ)

- ☐ Yields always positive values
- ☐ Has methodological similarities with **PM**, but **SJ** is always positive and non-iterative [1]
- ☑ Has smaller MSE and substantially smaller bias than **DL** for large τ^2 and number of studies, and vice versa
- ☒ Produces larger estimates than the **DL** method [2]
- ☒ Large positive bias for small to moderate τ^2 and high MSE [3,4,5]

DL: DerSimonian and Laird

PM: Paule and Mandel

1: Sidik and Jonkman J Biopharm Stat 2005, 2: Thorlund et *al* RSM 2012, 3: Sidik and Jonkman Stat Med 2007, 4: Novianti et al Contemp Clin Trials 2014 , 7 : Langan et *al* RSM 2018

Bayes Estimators

i. Bayes Modal (BM)

- ☐ Yields always positive values
- ☑ When τ^2 is positive BM has very low MSE [1]
- ☒ Associated with large bias for small τ^2, especially for few and small studies
- ☒ For zero τ^2 it performs worse than **DL** and **REML** [1]

ii. Rukhin Bayes (RB)

- ☑ For small number of studies, RB with mean prior distribution of τ^2 equal to zero has lower bias than DL [2]

iii. Full Bayesian (FB)

- ☑ Allows incorporation of uncertainty in all parameters (including τ^2)
- ☒ The choice of prior for τ is crucial when the number of studies is small [3]

1: Chung et al Stat Med 2013, 2: Kontopantelis et al Plos One 2013, 3: Lambert et al Stat Med 2005

DL: DerSimonian and Laird

REML: Restricted maximum likelihood

BM: Bayes Modal

Bootstrap methods

i. Non-parametric bootstrap DL (DLb)

- ☑ DLb is associated with lower bias than **DL** and **RB positive** when the number of studies is greater than 5
- ☑ DLb performs better than DL in identifying the presence of heterogeneity even for few studies
- ☒ Non-parametric bootstrap methods perform well only for a large number of studies
- ☒ DLb has greater bias compared with DL and this is more profound in small meta-analyses

Kontopantelis et al 2013

DL: DerSimonian and Laird

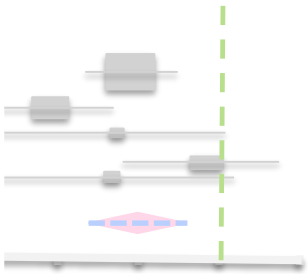
DLb: Non-parametric bootstrap DerSimonian and Laird

RB: Rukhin Bayes

Illustrative example

	I^2=0%	I^2=18%	I^2=45%	I^2=75%
Number of studies in the meta-analysis:	**14**	**18**	**17**	**11**
DerSimonian and Laird (DL)	**0.00**	0.01	**0.02**	0.13
Positive DerSimonian and Laird (DLp)	0.01	0.01	0.02	0.13
Two-step DerSimonian and Laird (DL2)	0.00	0.01	0.04	0.18
Hedges and Olkin (HO)	0.00	**0.00**	0.04	**0.22**
Two-step Hedges and Olkin (HO2)	0.00	0.01	0.04	0.19
Paule and Mandel (PM)	0.00	0.01	0.04	0.19
Hartung and Makambi (HM)	0.02	0.03	0.06	0.17
Hunter and Schmidt (HS)	0.00	0.01	0.02	**0.11**
Maximum likelihood (ML)	0.00	0.02	0.02	0.13
Restricted maximum likelihood (REML)	0.00	0.02	0.02	0.16
Sidik and Jonkman (SJ)	0.07	0.05	0.07	0.21
Positive Rukhin Bayes (RBp)	**0.15**	**0.11**	**0.12**	0.20
Full Bayes (FB) [Half normal prior for τ]	0.01	0.02	0.03	0.18
Bayes Modal (BM)	0.02	0.03	0.03	0.16
Non-parametric Bootstrap DerSimonian and Laird (DLb)	0.00	0.01	0.02	0.13

Veroniki et al. Res Synth Methods 2015



In summary…

	Direct	Zero value included	Simple to compute		Direct	Zero value included	Simple to compute
DL	☑	☑	☑	HS	☑	☑	☑
DLp	☑	☒	☑	ML	☒	☑	☒
DL2	☑	☑	☑	REML	☒	☑	☒
DLb	☒	☑	☒	AREML	☒	☑	☒
HO	☑	☑	☑	SJ	☑	☒	☑
HO2	☑	☑	☑	RB	☒	☑	☒
PM	☒	☑	☑	FB	☒	☑	☒
HM	☑	☑	☑	BM	☒	☒	☒

Simulation studies suggest in terms of **bias**:

- DL, DL2 , DLp, ML, HS, REML, RB with prior equal to zero, perform well for small τ^2
- HO, HO2, HM, SJ, PM, RBp, BM, perform well for large τ^2

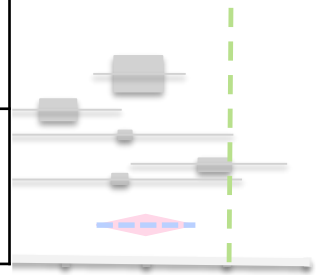
All methods decrease bias as k increases

Simulation studies suggest in terms of **efficiency**:

- DL, ML, HS, REML, perform well for small τ^2
- HO, BM, SJ, PM perform well for large τ^2

Software for the between-study variance estimator

Estimation Method	Software	Estimation Method	Software	Estimation Method	Software
DL	CMA, Excel (MetaEasy), Meta-Disc, Metawin, MIX, Open Meta Analyst, RevMan, R, SAS, STATA, SPSS	*ML*	CMA, Excel, HLM, Meta-Disc, Metawin, MLwin, Open Meta Analyst, R, SAS, STATA, SPSS	*REML*	HLM, Meta-Disc, MLwin, Open Meta Analyst, R, SAS, STATA
HO	R, Open Meta Analyst	*PM*	Open Meta Analyst, R, SAS, STATA	*FB*	Mlwin, R, SAS, BUGS, OpenBUGS, WinBUGS
HM	-	*SJ*	R, Open Meta Analyst	*RB*	-
HS	R	*AREML*	SPSS	*BM*	R, STATA
DL2	-	*HO2*	-		



Which software do you usually prefer to conduct your meta-analyses?

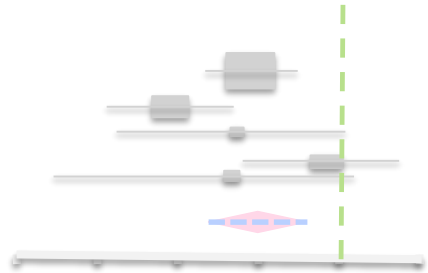
a) Review Manager

b) Stata and/or R

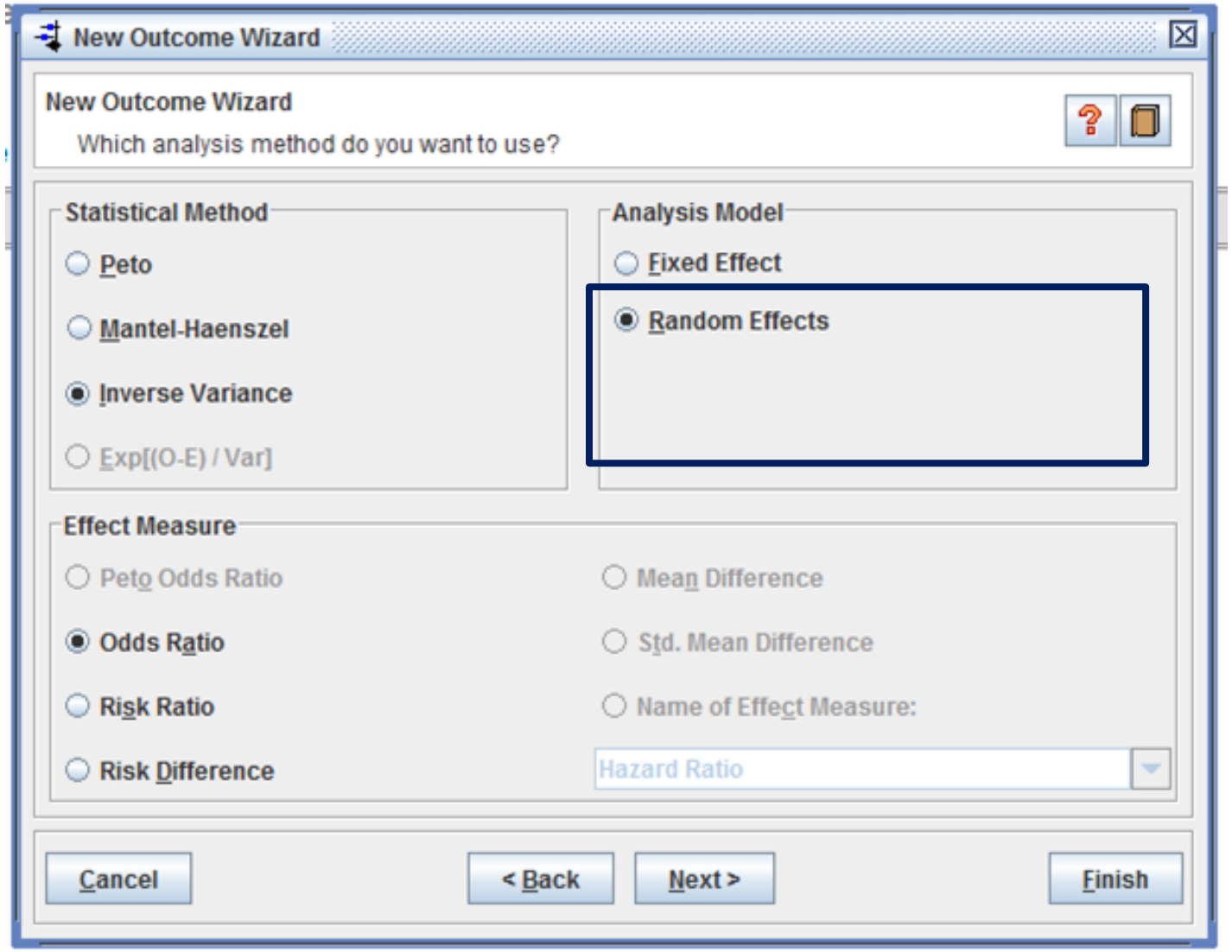
c) WinBUGS/OpenBUGS

d) All of the above

e) None of the above



Should we consider additional options in RevMan?



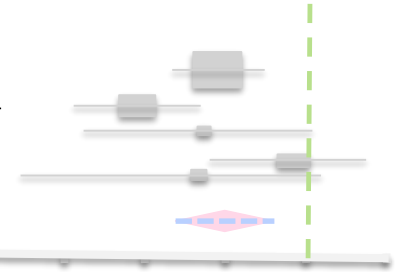
Which estimation method for the between-study variance should we consider adding in the Cochrane Review Manager?

RevMan

Recommendations based on published studies

According to simulation and empirical findings, the main factors that may affect the between-study variance estimation are:

- Number and size of studies included in the meta-analysis
- Magnitude of heterogeneity
- Distribution of true treatment effects
- Type of data (e.g., dichotomous, continuous)
- Choice of effect measure
- Frequency of events (for dichotomous outcomes)
- How well study-specific weights, variances and treatment effects are estimated – we often assume these are known.



Recommendations based on published studies

An empirical study using 57,397 Cochrane meta-analyses with $k \geq 2$ showed that:
→ The mean τ^2 is higher than generally assumed but fails to be detected, especially for small k!

Kontopantelis et al. 2013

A descriptive analysis of Cochrane systematic reviews found that 75% of meta-analyses contained 5 or fewer studies

Davey et al. 2011

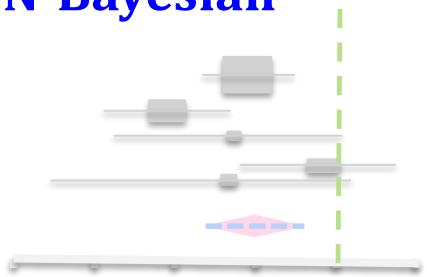
The majority of the pairwise meta-analyses have:

$$k \leq 10 \text{ and } \tau^2 \leq 0.4$$

Turner et al 2012
Pullenayegum et al 2011
Rhodes et al 2014

Summarizing study results in specific scenarios, we make recommendations mostly on **NON-Bayesian** estimators

- The fully Bayesian estimator has not been evaluated extensively in comparative studies



Recommendations based on published studies

Alternative methods are needed!

For the most common scenario for pairwise meta-analyses research studies have shown ($k \leq 10$, $\tau^2 \leq 0.4$):

- ☒ DL underestimates τ^2 when k is small and for rare events [1, 2, 3, 7]
- ☒ DLp, HM, RBp, BM and SJ overestimate τ^2 [2, 4, 5, 6]
- ☐ DLp has good coverage for the overall effect size [8]
- ☐ HM has a good coverage for the overall effect size when $\tau^2 \cong 0.07$ for dichotomous outcomes, and for $0.01 \leq \tau^2 \leq 0.05$ for continuous outcomes [8]
- ☒ DL has lower bias and MSE than HO and SJ [1, 2]
- ☒ BM performs worse than DL and REML when $\tau = 0$ [3]

Implement in RevMan?	
DL	Implemented
DLp	☒
HM	☒
RBp	☒
BM	☒
SJ	☒
HO	☒

1:Viechtbauer JEBS 2005, 2: Sidik &Jonkman Stat Med 2007, 3: Chung et al Stat Med 2013, 4: Thorlund et *al* RSM 2012, 5:Novianti et al Contemp Clin Trials 2014, 6: Kontopantelis et al Plos One 2013, 7: Langan et al RSM 2018, 8: Petropoulou & Mavridis Stat Med 2017

Recommendations based on published studies

"One should probably avoid the biased HS and ML estimators because they can potentially provide quite misleading results" [6]

- ☒ HS and ML are associated with substantial negative bias [6]
- ☒ DLb has higher bias than DL for small k
- ☐ DLb has good coverage for the overall effect size [10]
- ☐ DL2 approximates PM, inherits most of the best properties of DL and PM and is simple to compute. For rare events underestimates τ^2 [3, 4, 9]
- ☐ HO2 approximates PM [3]
- ☑ REML is less downwardly biased than DL and ML, but has greater MSE [1, 2]
  - REML is recommended for continuous data [5, 7]
  - REML has similar properties with the DL2 [9]
- ☑ AREML yields almost identical estimates with REML [1]

Implement in RevMan?	
HS	☒
ML	☒
DLb	☒
DL2	?
HO2	?
REML	☑
AREML	☑

1: Berkey et al Stat Med 1995, 2: Sidik &Jonkman Stat Med 2007, 3: DerSimonian & Kacker Contemp Clin Trials 2007, 4: Bhaumik et al J Amer Stat Assn 2012, 5:Viechtbauer JEBS 2005, 6 Viechtbauer JEBS 2005, 7:Novianti et al Contemp Clin Trials 2014, 8: Panityakul et al 2013, 9: Langan et al RSM 2018, 10: Petropoulou & Mavridis Stat Med 2017

Recommendations based on published studies

- ☒ PM is positively biased when study sizes differ importantly [9]
- ☐ it is often approximately unbiased when DL is negatively biased [9]
- ☑ PM outperforms DL and REML in terms of bias [3, 4, 6, 8]
  - PM performs better than DL, DL2, PM, HO,REML, SJ in terms of bias for both continuous and dichotomous data [7]
- ☑ Easy to obtain
- ☑ An improved PM is available for rare events [4]

Implement in RevMan?	
PM	☑

BUT

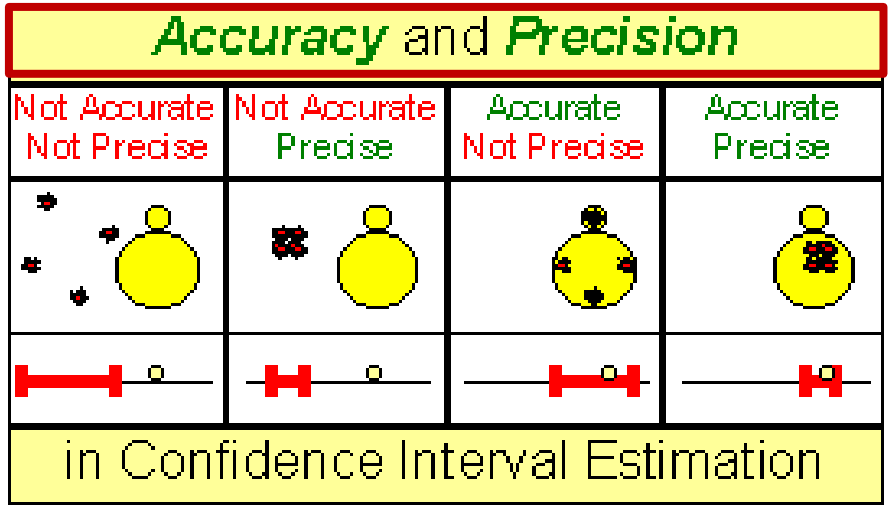
- Estimation of between-study variance in meta-analyses with <10 studies may be imprecise, especially when study sizes are small and events are rare
- Hence, it is rarely appropriate to rely on one between-study variance estimator!

1: Berkey et al Stat Med 1995, 2: Sidik &Jonkman Stat Med 2007, 3: DerSimonian & Kacker Contemp Clin Trials 2007, 4: Bhaumik et al J Amer Stat Assn 2012, 5:Viechtbauer JEBS 2005, 6: Bowden et al BMC Med Res Methodol 2011, 7:Novianti et al Contemp Clin Trials 2014, 8: Panityakul et al 2013, 9: Langan et al RSM 2018, 10: Petropoulou & Mavridis Stat Med 2017

Confidence Interval (CI) for the between-study variance

Estimation of between-study variance

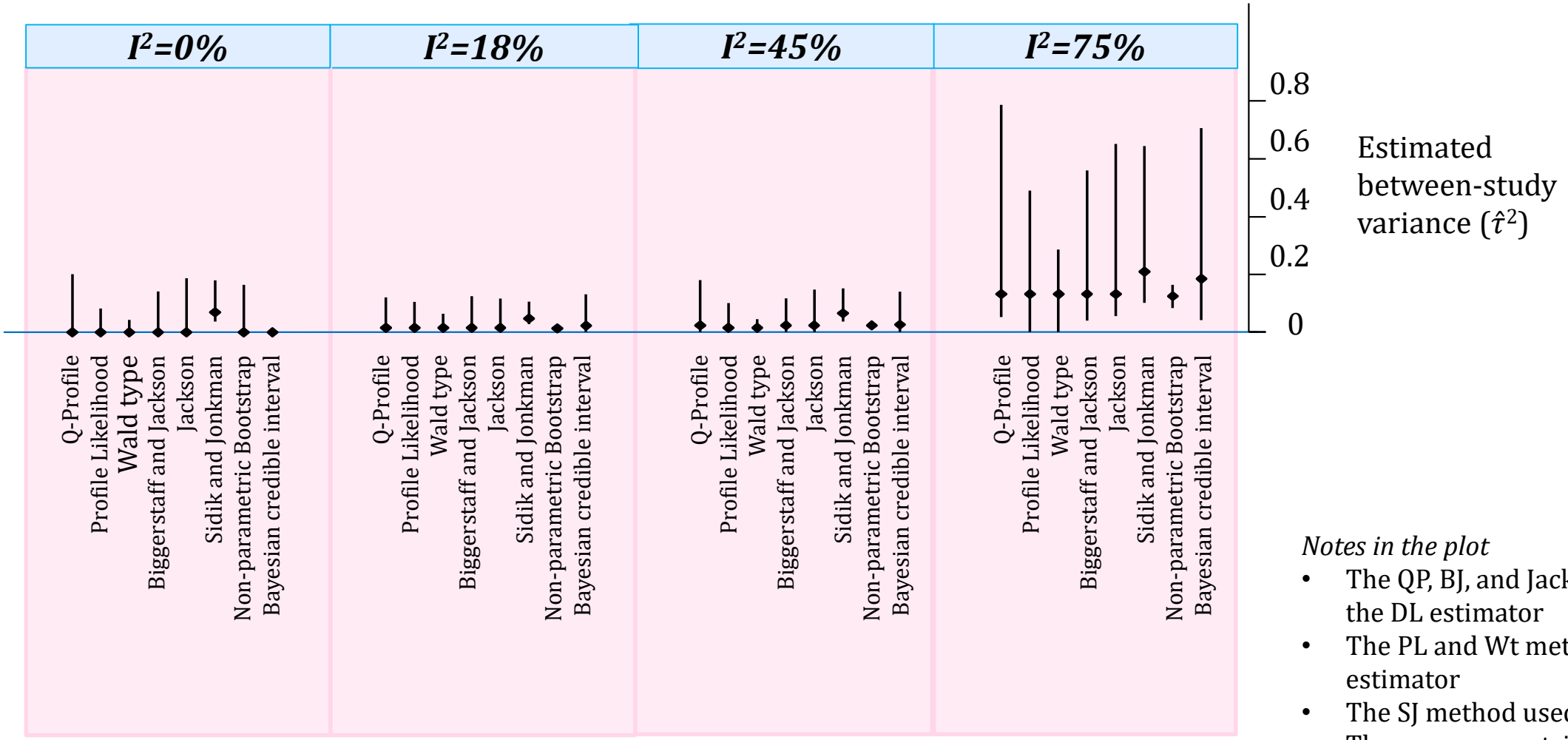
- Point Estimation
- Interval Estimation



Desirable properties

- ✓ Accuracy = High Coverage Probability – P(τ∈ CI)
  - The closer the coverage is to the nominal level (usually 0.95) the better the CI.
- ✓ Precision = Narrow CI
  - Narrower CIs retaining the correct coverage are preferable because they increase precision.

Different CI methods may suggest different results…



Veroniki et al. Res Synth Methods 2015

Notes in the plot

- The QP, BJ, and Jackson methods used the DL estimator
- The PL and Wt methods used the ML estimator
- The SJ method used the SJ estimator
- The non-parametric bootstrap method used the DLb estimator
- The Bayesian CrI used the FB estimator

Confidence Intervals (CIs) for the between-study variance

- ☒ **Bootstrap** CIs have less than adequate coverage probabilities [1]
- ☒ The **PL** and **Wt** CIs require a large number of studies to perform well [1]
- ☒ **SJ** has very poor coverage probability when τ^2 is small [1]
- ☑ **QP** is preferable to **PL**, **Wt**, **BT** and **SJ** methods regarding coverage even for a small number of studies [1, 2, 4, 6]
- ☑ Both **QP** and **BJ** provide narrow CIs [7]

1:Viechtbauer Stat Med 2007, 2: Knapp et al Biom J 2006, 4: Viechtbauer Journal of Statistical Software 2010, 5:Bowden et al BMC Med Res Methodol 2011, 6: Tian Biom J 2008, 7: Jackson RSM 2013

Categories

A. Likelihood-based CIs
   a) Profile likelihood (PL)
B. Asymptotically normal based CIs
   a) Wald type (Wt)
C. Generalized Cochran Q - based CIs
   a) Biggerstaff and Tweedie (BT)
   b) Jackson (J) (including Biggerstaff and Jackson (BJ))
   c) Q-profile (QP)
D. Sidik and Jonkman CIs (SJ)
E. Bootstrap CIs
F. Bayesian Credible Intervals

Confidence Intervals (CIs) for the between-study variance

☒ **QP**, **BJ**, and **Jackson** methods can result in null sets for the CI of τ^2 when heterogeneity and the number of studies are small

- **QP** provides is more accurate CIs than **BJ** for large τ^2, and vice versa for small τ^2**.** For moderate τ^2 **Jackson's** method is recommended using weights equal to the reciprocal of the within-study standard errors [1, 7]

☑ **QP** is simple to compute

Categories

A. Likelihood-based CIs
   a) Profile likelihood (PL)

B. Asymptotically normal based CIs
   a) Wald type (Wt)

C. Generalized Cochran Q - based CIs
   a) Biggerstaff and Tweedie (BT)
   b) Jackson (J) (including Biggerstaff and Jackson (BJ))
   c) Q-profile (QP)

D. Sidik and Jonkman CIs (SJ)

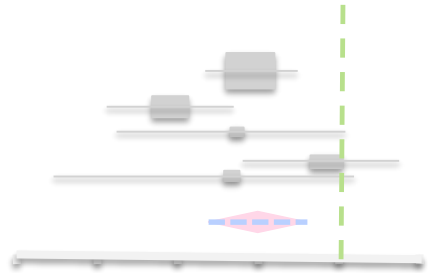
E. Bootstrap CIs

F. Bayesian Credible Intervals

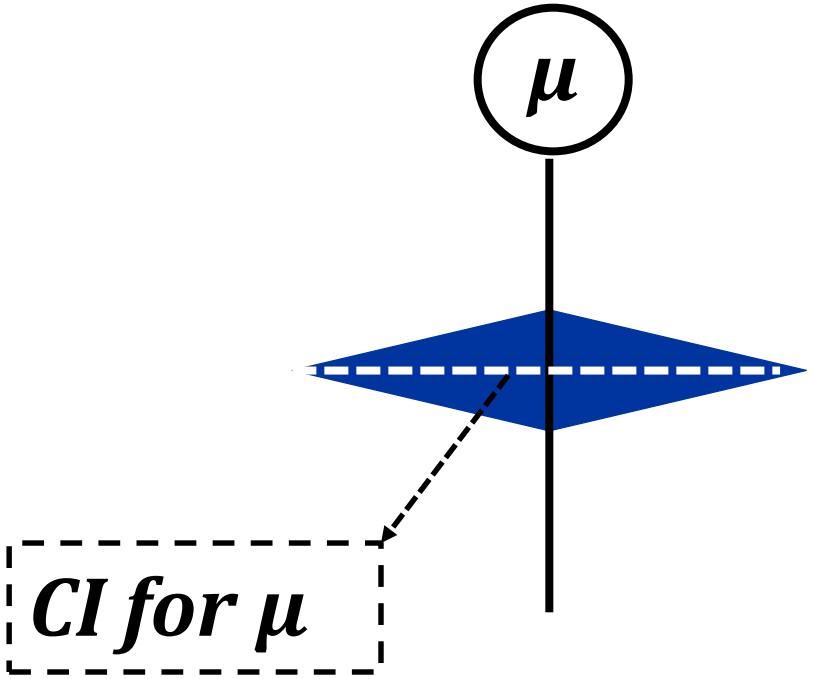
1:Viechtbauer Stat Med 2007, 2: Knapp et al Biom J 2006, 4: Viechtbauer Journal of Statistical Software 2010, 5:Bowden et al BMC Med Res Methodol 2011, 6: Tian Biom J 2008, 7: Jackson RSM 2013

	PL	Wt	BT, BJ, Jackson	QP	SJ	Bootstrap	Bayesian CrI
DL	--	✓	✓	(✓)	--	✓	--
DLp	--	✓	✓	(✓)	--	✓	--
DL2	--	✓	✓	(✓)	--	✓	--
HO	--	✓	✓	(✓)	--	✓	--
HO2	--	✓	✓	(✓)	--		
PM	--	✓	(✓)	✓			
HM	--	✓	✓				--
HS	--	✓	(✓)			✓	--
ML	✓	✓			--	✓	--
REML	✓				--	✓	--
AREML				(✓)	--	✓	--
SJ			(✓)	(✓)	✓	✓	--
RB		✓	(✓)	(✓)	--	✓	✓
RBp	--	✓	(✓)	(✓)	--	✓	--
FB	--	--	--	--	--	--	✓
BM	--	✓	(✓)	(✓)	--	✓	--
DLb	--	✓	(✓)	(✓)	--	✓	--

Not all confidence intervals are appropriate for all of the available between-study variance estimators



2. Inference on the overall effect size



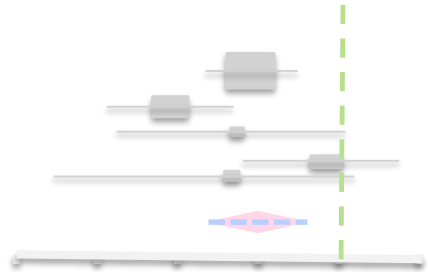
Have you ever used different methods to calculate the uncertainty in the overall effect size?

a) Yes, I have used different methods to assess the robustness of the meta-analysis results

b) Yes, I have used different methods, as deemed appropriate, in different meta-analyses

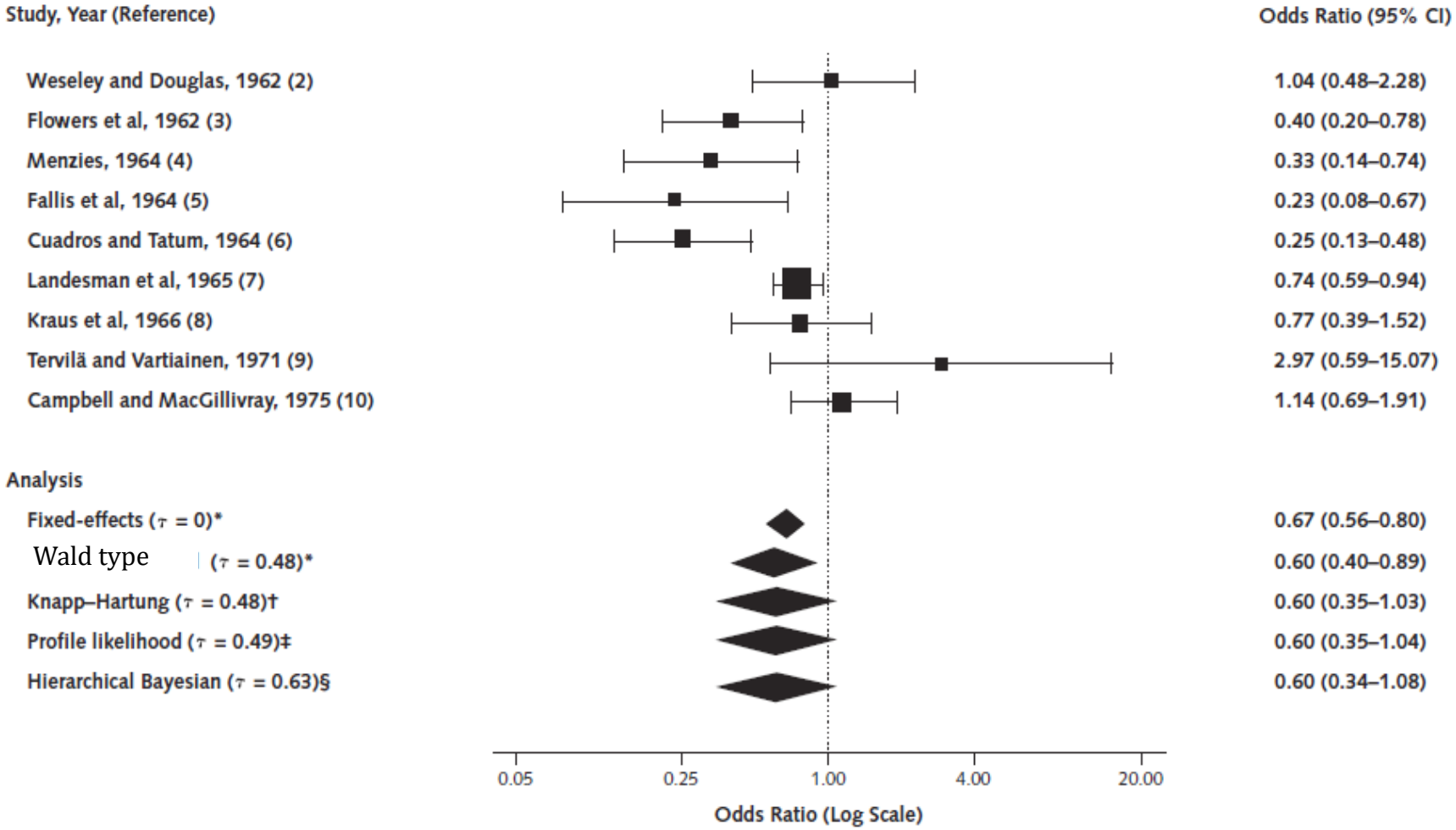
c) No, I always use the default option in the relevant meta-analytic software

d) No, I was not aware that different methods exist



Various CIs can lead to different conclusions

Figure. Heterogeneous evidence from Collins and colleagues' meta-analysis of the effects of diuretics on preeclampsia (11).



Which is the most appropriate method to use?

Cornel et al. Annals of Internal Medicine 2014

Literature Review of CI methods

Our search identified:

- 69 relevant publications
- 15 methods to compute a CI for the overall effect size (grouped in 7 broad categories)

The properties of the methods were evaluated in 31 papers:

- including 30 simulation studies and 32 real-life data evaluations of ≥2methods

The most popular technique is WTz

Veroniki et al. Res Synth Methods. 2018.
doi: 10.1002/jrsm.1319.

Categories

A. Wald-type (WT) CIs
   a) Wald-type normal distribution (WTz)
   b) Wald-type t-distribution (WTt)
   c) Quantile approximation (WTqa)
B. Hartung-Knapp/Sidik-Jonkman (HKSJ) CIs
C. Likelihood-based CIs
   a) Profile likelihood (PL)
   b) Higher-order likelihood inference methods
D. Henmi and Copas (HC) CIs
E. Biggerstaff and Tweedie (BT) CIs
F. Resampling CIs
   a) Zeng and Lin (ZL)
   b) Bootstrap
   c) Follmann and Proschan (FP)
G. Bayesian Credible Intervals

Confidence Interval methods

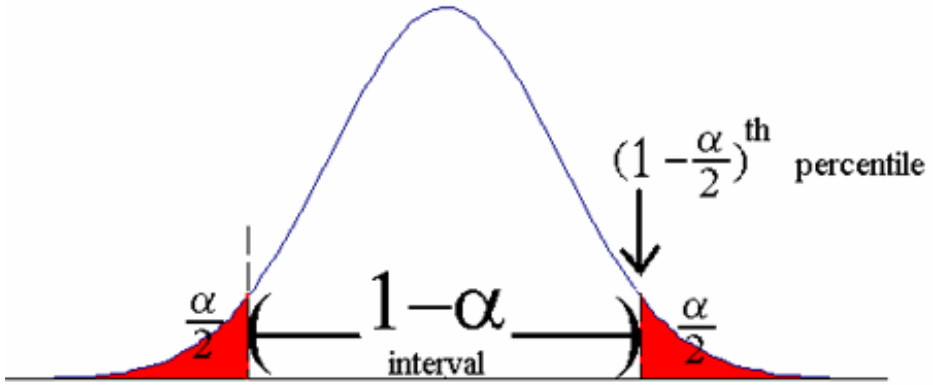
No	Method	Confidence Interval
1	Wald-type normal distribution (WTz)	$\hat{\mu}_{RE} \pm z_{0.975}\sqrt{var(\hat{\mu}_{RE})}$
2	Wald-type t-distribution (WTt)	$\hat{\mu}_{RE} \pm t_{k-1,0.975}\sqrt{var(\hat{\mu}_{RE})}$
3	Quantile approximation (WTqa)	$\hat{\mu}_{RE} \pm b_k\sqrt{var(\hat{\mu}_{RE})}$, with b_k the quantile approximation function of the distribution of the statistic $M = \frac{\hat{\mu}_{RE}-\mu}{\sqrt{var(\hat{\mu}_{RE})}}$
4	Hartung-Knapp/Sidik-Jonkman (HKSJ)	$\hat{\mu}_{RE} \pm t_{k-1,0.975}\sqrt{\sigma^2_{w,\hat{\mu}_{RE}}}$, with $\sigma^2_{w,\hat{\mu}_{RE}} = q \cdot var(\hat{\mu}_{RE})$, $q = \frac{Q_{gen}}{k-1}$, and $Q_{gen} = \sum w_{i,RE}(y_i - \hat{\mu}_{RE})^2$
5	Modified HKSJ	HKSJ, but use q^* instead of q: $q^* = \max\{1, q\}$
6	Profile likelihood (PL)	Profile log-likelihood for μ: $lnL_p(\mu) = lnL(\mu, \hat{\tau}^2_{ML}(\mu))$, $lnL_p(\mu) > lnL_p(\hat{\mu}_{RE}) - \frac{\chi^2_{1,0.05}}{2}$

Confidence Interval methods

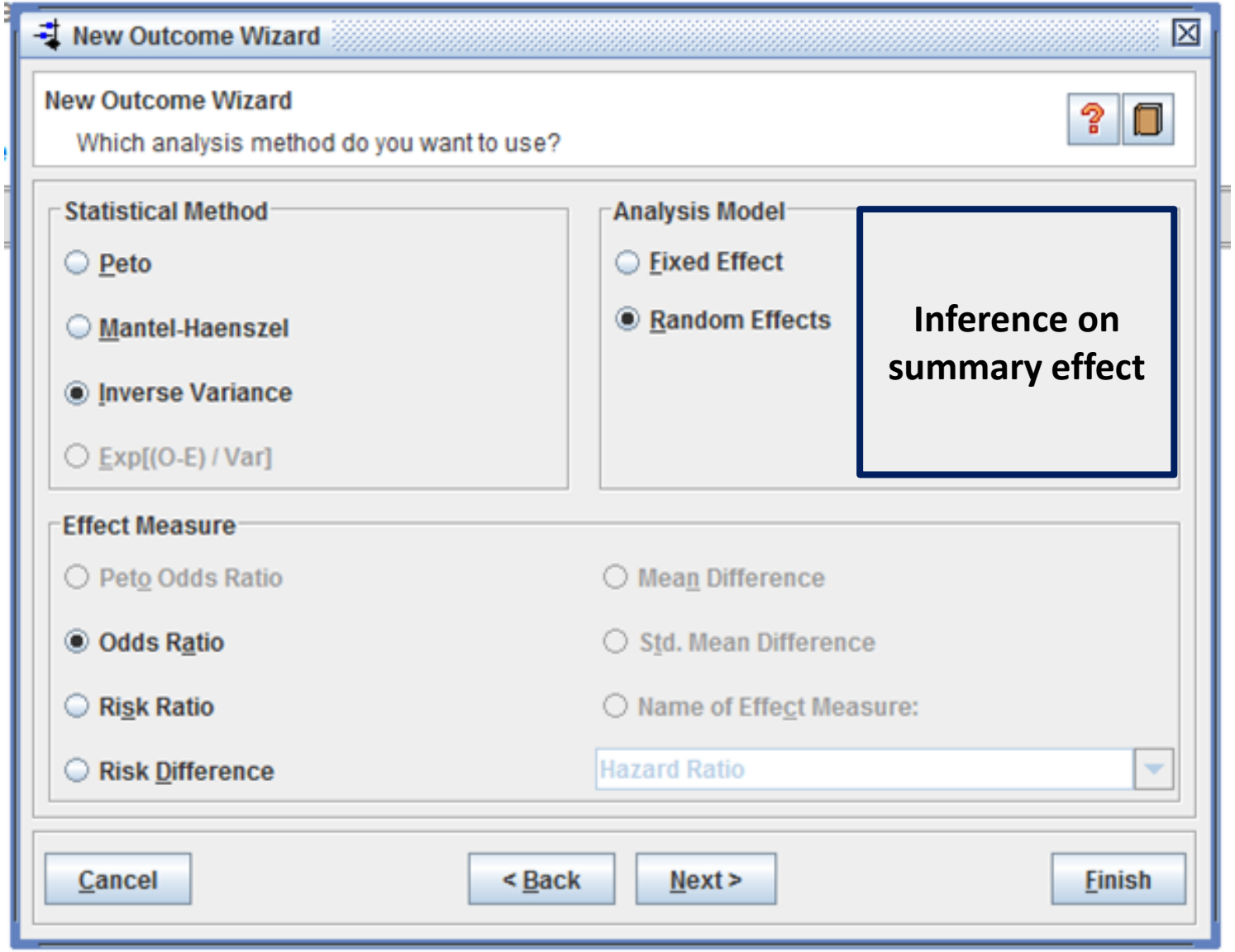
No	Method	Confidence Interval
7, 8	Higher-order likelihood inference methods	The Bartlett-type adjusted efficient score statistic (BES) (No 7) and Skovgaard's statistic (SS) (No 8) use a higher-order approximation than the PL
9	Henmi and Copas (HC)	Hybrid approach: the FE estimate is accompanied by a CI that allows for τ^2 under the assumptions of a RE model
10	Biggerstaff and Tweedie (BT)	$\hat{\mu}_{RE}^{BT} \pm z_{0.975}\sqrt{var(\hat{\mu}_{RE}^{BT})}$, with $var(\hat{\mu}_{RE}^{BT}) = \frac{1}{(\sum w_{i,RE}^{BT})^2}\sum(w_{i,RE}^{BT})^2(v_i + \hat{\tau}^2)$ and $w_{i,RE}^{BT} = E(w_{i,RE})$
11	Resampling methods: Zeng and Lin (ZL)	Simulate values of τ^2 using DL, then simulate estimated average effect sizes using the sampled τ^2 to calculate the weights in $\hat{\mu}_{RE} = \frac{\sum y_i w_{i,RE}}{\sum w_{i,RE}}$. Repeat both aspects B times, get empirical distribution of $\hat{\mu}_{RE}$ and compute CI
12, 13	Resampling methods: Bootstrap confidence intervals	Non-parametric bootstrap CI (No 12) with resampling from the sample itself with replacement, and Parametric bootstrap CI (No 13) with resampling from a fitted model

Confidence Interval methods

No	Method	Confidence Interval
14	Resampling methods: Follmann and Proschan (FP)	Permutation tests can be extended to calculate CIs for the effect size. CIs are constructed by inverting hypothesis test to give the CI bounds - parameter values that are not rejected by the hypothesis test lie within the corresponding CI
15	Bayesian credible intervals	Bayesian credible intervals for the overall effect size can be obtained within a Bayesian framework



Should we consider additional options in RevMan?



Should we consider adding an extra method to calculate the uncertainty in the overall effect size in the Cochrane Review Manager?

RevMan

Comparative evaluation of the methods

i. Wald-type methods (WTz, WTt, WTqa)

- ☑ For large number of studies WTz, WTt, and WTqa perform well [1, 2]
- ☒ WTz performs worse in terms of coverage for small number of studies (k<16) compared with the PL and the WTt methods [1]
- ☒ WTz and WTt depend on the number of studies, the τ^2 estimator, and the τ^2 magnitude [4]
- ☒ Coverage of WTz has been found to be as low as 65% (at 95% nominal level) when I^2=90% and k=2,3 [3]
- ☒ Coverage of WTt may be below the 95% nominal level, but it becomes conservative (close to 1) when k is small [1, 2, 3]
- ☒ WTqa and WTt have on average similar coverage, but WTqa outperforms WTz, PL, and ZL CIs – but it is very conservative [2, 6]
- ☒ WTqa has been criticized that it is very difficult to obtain suitable critical values b_k that apply to all meta-analyses [5]

Implement in RevMan?	
WTz	Implemented
WTt	☒
WTqa	☒

WTz: Wald type – normal distr

WTt: Wald type – t distr

WTqa: Wald type – quantile approximation

1: Jackson et al J Stat Plan Infer 2010, 2: Brockwell and Gordon Stat Med 2007, 3: Langan et al RSM 2018, 4: Sanchez-Meca and Marin-Martinez Psychol Methods 2008, 5: Jackson and Bowden Stat Med. 2009, 6: Zeng and Lin Biometrika. 2015

Comparative evaluation of the methods

ii. Hartung-Knapp/Sidik-Jonkman methods (HKSJ, modified HKSJ)

- ☑ HKSJ on average produces wider CIs with more coverage than the WTz and WTt methods [1, 2, 3]
- ☑ HKSJ has coverage close to the nominal level, is not influenced by the magnitude or estimator of τ^2, and is insensitive to the number of trials [1, 2, 3, 4, 5]
- ☑ Simulations suggest HKSJ has good coverage for the odds ratio, risk ratio, mean difference, and standardized mean difference effect measures [3, 7]
- ☐ Real-life data studies showed that the WTz method yielded more often statistically significant results compared with the HKSJ method [1, 6]
- ☒ HKSJ is suboptimal than the WTz and WTt CIs when binary outcomes with rare events are included in a meta-analysis [2]
- ☒ Caution is needed for the HSKJ CI when <5 studies of unequal sizes are included in a meta-analysis [4, 6]
- ☒ In the absence of heterogeneity it may be: HKSJ coverage < WTz coverage [6]

WTz: Wald type – normal distr

WTt: Wald type – t distr

1:IntHout et al BMC Med Res Methodol. 2014, 2: Langan et al RSM 2018, 3: Makambi J Biopharm Stat. 2004, 4: Hartung Biom J 1999, 5: Sanchez-Meca and Marin-Martinez Psychol Methods 2008, 6: Wiksten et al Stat Med. 2016, 7: Sidik and Jonkman Stat Med. 2002

Comparative evaluation of the methods

ii. Hartung-Knapp/Sidik-Jonkman methods (HKSJ, modified HKSJ)

- ☑ The modified HKSJ is preferable when few studies of varying size and precision are available [1]
- ☒ For small k (particularly for k=2) and small τ^2 the modified HKSJ tends to be over-conservative [1, 2, 3]

Implement in RevMan?	
HKSJ	☑
mHKSJ	☑

1: Röver et al BMC Med Res Methodol. 2015, 2: Jackson et al Stat Med. 2017, 3: Viechtbauer Psychol Methods. 2015, 4: Brockwell and Gordon Stat Med. 2007, 5: Kosmidis Biometrika. 2017, 6: Noma Stat Med 2011, 7: Guolo & Varin Stat Methods Med Res. 2015

Comparative evaluation of the methods

iii. Likelihood-based methods (PL, BES, SS)

- ☑ PL has higher coverage closer to the nominal level than WTz and WTt, even when k is relatively small ($k \leq 8$) [4, 5]
- ☑ BES improves coverage over WTz, WTt, and PL CIs as τ^2 increases and/or k decreases [6]
- ☑ SS yields similar results with BES, and has better coverage than WTz and PL CIs [6, 7]
- ☒ Caution is needed for k≤5 as BES tends to be over-conservative [6]

Implement in RevMan?	
PL	?
BES	?
SS	?

WTz: Wald type – normal distr

WTt: Wald type – t distr

PL: Profile Likelihood

BES: Bartlett-type adjusted efficient score statistic

SS: Skovgaard's statistic

1: Röver et al BMC Med Res Methodol. 2015, 2: Jackson et al Stat Med. 2017, 3: Viechtbauer Psychol Methods. 2015, 4: Brockwell and Gordon Stat Med. 2007, 5: Kosmidis Biometrika. 2017, 6: Noma Stat Med 2011, 7: Guolo & Varin Stat Methods Med Res. 2015

Comparative evaluation of the methods

iv. Henmi and Copas method (HC)

- ☑ For k>10 HC yields better coverage than WTz, HKSJ, PL, and BT methods, irrespective the absence/presence of publication bias [1]
- ☒ For k<10 the HKSJ and PL methods perform better than HC, WTz, and BT methods [1]

v. Biggerstaff and Tweedie method (BT)

- ☒ WTz and BT methods have comparable coverage (below the nominal level), but coverage increases for the exact weights [2, 3]

vi. Resampling methods (ZL, FP)

- ☑ ZL outperforms both WTz and PL for small k in terms of coverage [4]
- ☑ FP controls coverage better than WTz, WTt, PL, and is closely followed by BES [5]
- ☒ BES is slightly more powerful than FP especially for small k [5]

Implement in RevMan?	
HC	☒
BT	☒
ZL	?
FP	?

1: Henmi and Copas Stat Med. 2010, 2: Brockwell and Gordon Stat Med 2007, 3: Preuß and Ziegler Methods Inf Med. 2014, 4: Zeng and Lin Biometrika. 2015, 5: Huizenga et al Br J Math Stat Psychol. 2011

WTz: Wald type – normal distr
WTt: Wald type – t distr
HKSJ: Hartung-Knapp/Sidik-Jonkman
PL: Profile Likelihood
BES: Bartlett-type adj score statistic
ZL: Zeng and Lin
FP: Follmann and Proschan

Comparative evaluation of the methods

vii. Bayesian credible intervals

- ☑ Bayesian intervals produce intervals with coverage closer to the nominal level compared to the HKSJ, modified HKSJ, and PL CIs [1, 2]
- ☑ Bayesian intervals tend to be smaller than the HKSJ CI even in situations with similar or larger coverage [1]
- ☒ The performance of the Bayesian intervals may vary depending on the prior assigned to the between-study variance [3]

Implement in RevMan?	
Bayes	?

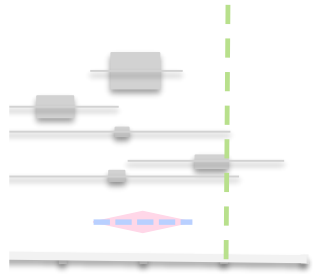
HKSJ: Hartung-Knapp/Sidik-Jonkman

PL: Profile Likelihood

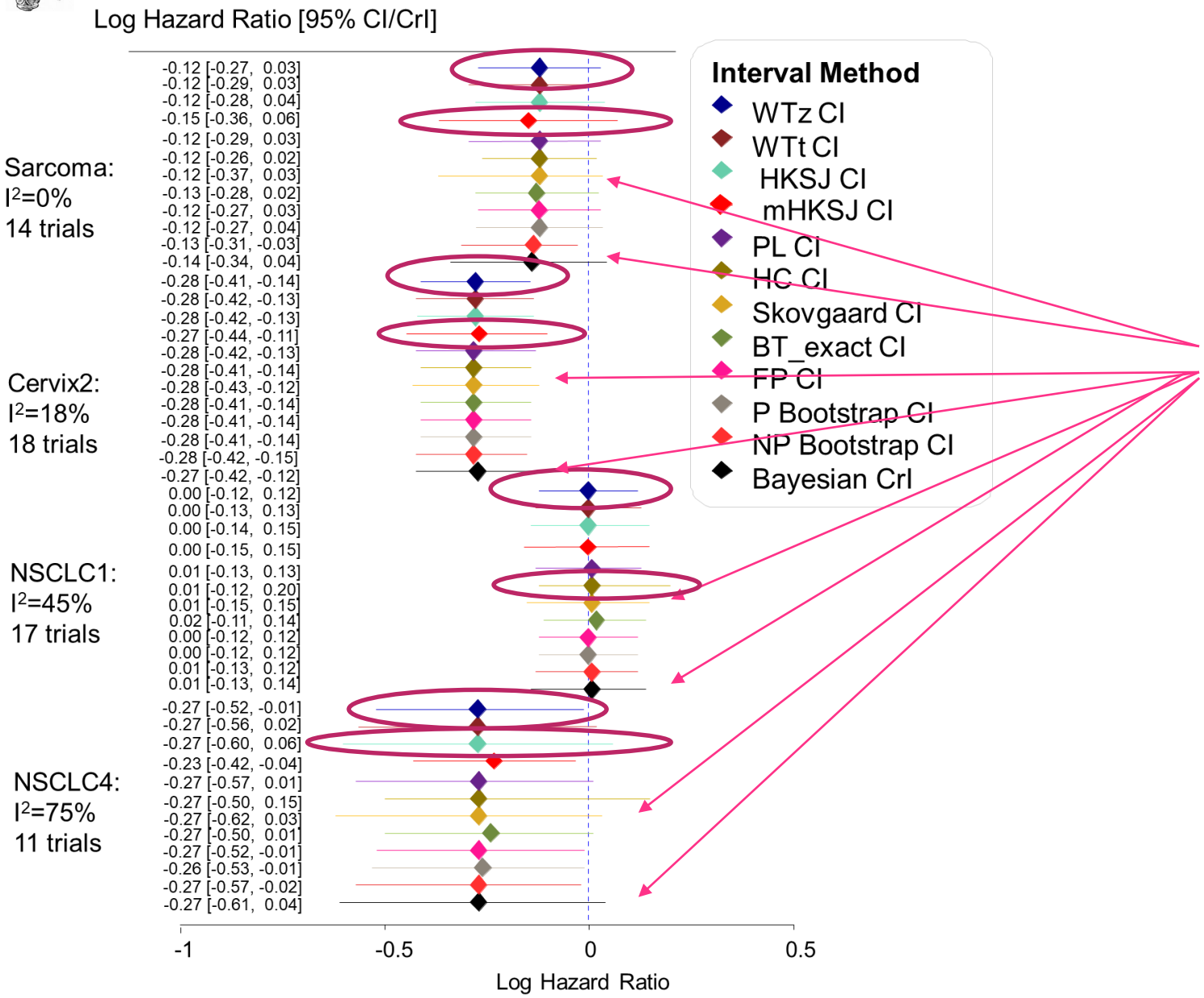
1: Friede et al RSM 2017, 2: Bodnar et al Stat Med. 2017, 3: Lambert et al Stat Med. 2005

Software for CIs for the overall effect size

CI Method	Software	CI Method	Software	CI Method	Software
WTz	CMA, Excel (MetaEasy, MetaXL), Meta-Disc, Metawin, MIX, MLwin, Open Meta Analyst, RevMan, R, SAS, Stata, SPSS	*PL*	Excel (MetaEasy), HLM, Meta-Disc, MLwin, R, SAS, Stata	*Bootstrap (parametric and non-parametric)*	Metawin, MLwin, R, Stata
WTt	Excel (MetaEasy), R, SAS	*BES*	-	*FP*	Excel (MetaEasy), R, Stata
WTqa	-	*SS*	R	*ZL*	-
HKSJ	CMA, R	*HC*	R	*Bayes*	MLwin, R, SAS, BUGS, OpenBUGS, WinBUGS
Modified HKSJ	Stata	*BT*	R		



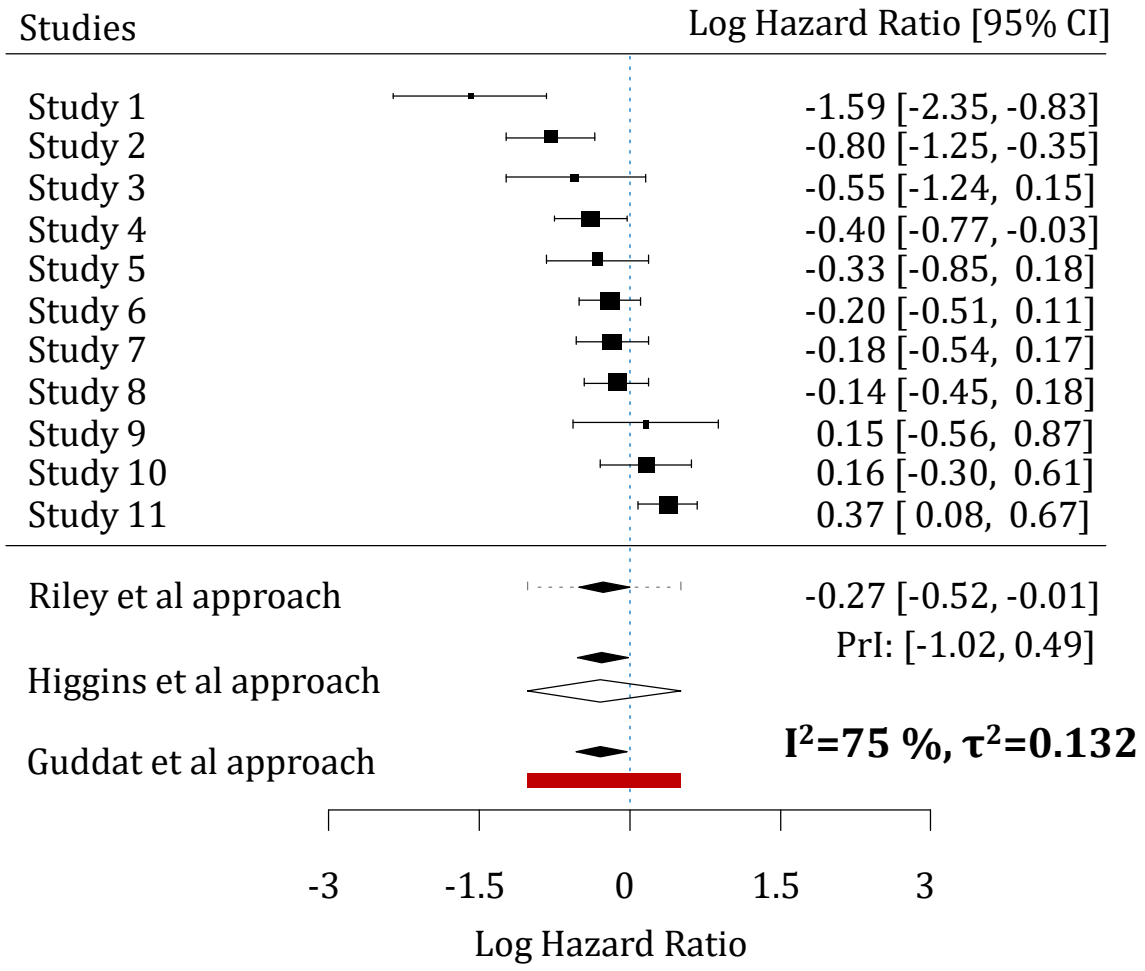
Illustrative example



- The WTz CI lies among the narrowest intervals
- The Skovgaard statistic CI and the Bayesian CrI lie among the largest intervals
- For very low (Sarcoma) and low (Cervix2) I^2 values, the modified HKSJ CI has the largest width across all intervals
- For moderate I^2 value (NSCLC1) the HC CI is associated with the highest uncertainty around the overall effect size
- For substantial I^2 value (NSCLC4)the HKSJ is the widest CI

Prediction Interval

- Although prediction intervals have not often been employed in practice they provide useful additional information to the confidence intervals



- A prediction interval provides a predicted range for the true effect size in a new study:

$$\hat{\mu}_{RE} \pm t_{k-1,0.975}\sqrt{\hat{\tau}^2 + var(\hat{\mu}_{RE})}$$

- Conclusions drawn from a prediction interval are based on the assumption the study-effects are normally distributed

Prediction Interval

- Prediction intervals are particularly helpful when excess heterogeneity exists, and the combination of individual studies into a meta-analysis would not be advisable

- The 95% prediction interval in >70% of the statistically significant meta-analyses in the Cochrane Database with $\hat{\tau}^2 > 0$, showed that the effect size in a new study could be null or even in the opposite direction from the overall result [1]

- The 95% prediction interval is only accurate when heterogeneity is large (I^2>30%) and the study sizes are similar [2]

- For small heterogeneity and different study sizes the coverage of prediction interval can be as low as 78% depending on the between-study variance estimator [2]

1: IntHout et al BMJ Open 2016, 2: Partlett and Riley Stat Med. 2017

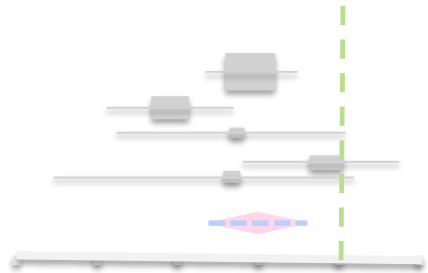
Should we consider more between-study variance estimators in Review Manager?

a) No because research has not concluded which one is the best

b) Yes because research has not concluded which one is the best

c) No because differences are negligible

d) Yes because results are sensitive



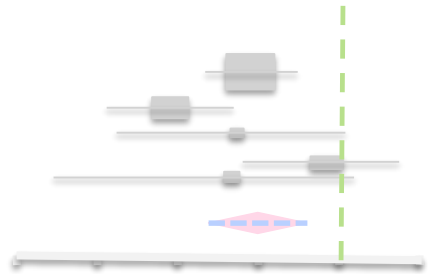
Should we consider more CI methods for the overall effect size in Review Manager?

a) No because research has not concluded which one is the best

b) Yes because research has not concluded which one is the best

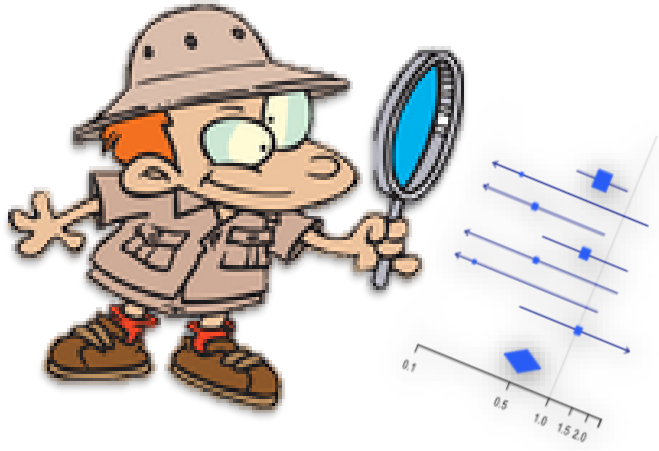
c) No because differences are negligible

d) Yes because results are sensitive



In Summary

- The WTz CI using the DL estimator for the between-study variance, are commonly used and are the default option in many meta-analysis software
- Simulations suggest that PM and REML estimators are better alternatives to estimate the between-study variance than DL
- The QP method and the alternative approach based on a 'generalized Cochrane between-study variance statistic' are among the best options to compute CI around the between-study variance
- Likelihood-based CIs yield coverage closer to the nominal level vs. WTz, but are computationally more demanding than WTz



In Summary

- Overall, studies suggest that the HKSJ method has one of the best performance profiles – performs well even for k<10 and is robust across different τ^2 estimators and values
- But, for $\hat{\tau}^2 = 0$ the HKSJ CI is too narrow. In such cases, the modified HKSJ can be used
- Caution is also needed in meta-analyses with rare events, with <5 studies, and different study precisions – the modified HKSJ can be used, but not for k=2
- Bayesian methods may be considered preferable when prior information is available
- A sensitivity analysis using a variety of methods may be needed, particularly when studies are few in number

<u>Time for CHANGE!</u>

It is rarely appropriate to rely on one estimation method when <10 studies are available!

References

1. Bender R, Friede T, Koch A, et al. Methods for evidence synthesis in the case of very few studies. Res Synth Methods. 2018;epub ahead of print:1-11.
2. Brockwell SE, Gordon IR. A comparison of statistical methods for meta-analysis. Stat Med. 2001;20(6):825-840.
3. Brockwell SE, Gordon IR. A simple method for inference on an overall effect in meta-analysis. Stat Med 2007; 26(25):4531-4543.
4. Cornell JE, et al. Random-effects meta-analysis of inconsistent effects: A time for change. Ann Intern Med. 2014.
5. DerSimonian R, Kacker R. Random-effects model for meta-analysis of clinical trials: an update. Contemp Clin Trials 2007; 28(2):105-114.
6. Guolo A. Higher-order likelihood inference in meta-analysis and meta-regression. Stat Med. 2012;31(4):313-327.
7. Follmann DA, Proschan MA. Valid inference in random effects meta-analysis. Biometrics. 1999;55(3):732-737.
8. Jackson D, Bowden J, Baker R. How does the DerSimonian and Laird procedure for random effects meta-analysis compare with its more efficient but harder to compute counterparts? J Stat Plan Infer 2010; 140(4):961-970.
9. Jackson D, White IR. When should meta-analysis avoid making hidden normality assumptions? Biometrical. 2018.
10. Hardy RJ, Thompson SG. A likelihood approach to meta-analysis with random effects. Stat Med. 1996;15(6):619-629.
11. Hartung J. An alternative method for meta-analysis. Biom J 1999;41(8):901-916.
12. Hartung J, Knapp G. On tests of the overall treatment effect in meta-analysis with normally distributed responses. Stat Med. 2001;20(12):1771-1782.
13. Henmi M, Copas JB. Confidence intervals for random effects meta-analysis and robustness to publication bias. Stat Med. 2010;29(29):2969-2983.
14. Higgins JP, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. J R Stat Soc Ser A Stat Soc. 2009;172(1):137-159.
15. IntHout J, Ioannidis JP, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. BMC Med Res Methodol. 2014.

References

16. Knapp G, Hartung J. Improved tests for a random effects meta-regression with a single covariate. Stat Med 2003; 22(17):2693-2710.
17. Langan D, Higgins JPT, Jackson D, et al. A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. Res Synth Methods. 2018; doi: 10.1002/jrsm.1316.
18. Makambi K.H. The Effect of the Heterogeneity Variance Estimator on Some Tests of Efficacy. J Biopharm Stat 2004; 2:439-449.
19. Noma H. Confidence intervals for a random-effects meta-analysis based on Bartlett-type corrections. Stat Med. 2011.
20. Petropoulou M., Mavridis D. A comparison of 20 heterogeneity variance estimators in statistical synthesis of results from studies: A simulation study. Statistics in Medicine 2017; 36(27): 4266-4280
20. Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. BMJ. 2011;342:d549.
21. Sanchez-Meca J, Marin-Martinez F. Confidence intervals for the overall effect size in random-effects meta-analysis. Psychol Methods 2008; 13(1):31-48.
22. Sidik K, Jonkman JN. A simple confidence interval for meta-analysis. Stat Med. 2002;21(21):3153-3159.
23. Sidik K, Jonkman JN. A comparison of heterogeneity variance estimators in combining results of studies. Stat Med 2007; 26(9):1964-1981.
24. Thorlund K., Wetterslev J., Thabane, Thabane L., Gluud C. Comparison of statistical inferences from the DerSimonian–Laird and alternative random-effects model meta-analyses – an empirical assessment of 920 Cochrane primary outcome meta-analyses. Research Synthesis Methods 2012; 2(4):238-253.
25. Veroniki, A. A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., Kuss, O., Higgins, J. PT., Langan, D., and Salanti, G. Methods to estimate the between-study variance and its uncertainty in meta-analysis. Res. Syn. Meth., 2016, 7: 55–79.
26. Veroniki, A. A., Jackson, D., Bender, R., Kuss, O., Langan, D., Higgins, J. PT., Knapp, G., and Salanti, G. Methods to calculate uncertainty in the estimated overall effect size from a random-effects meta-analysis. Res. Syn. Meth., 2019, doi: 10.1002/jrsm.1319.
27. Viechtbauer W. Bias and Efficiency of Meta-Analytic Variance Estimators in the Random-Effects Model. Journal of Educational and Behavioral Statistics 2005; 30(3):261-293.

Thank you for your attention!

Questions?

Areti Angeliki Veroniki MSc, PhD
Research Fellow, University of Ioannina, Ioannina, Greece
Post-doctoral Fellow, Faculty of Medicine, Imperial College, London, UK
Affiliate Scientist, St. Michael's Hospital, Toronto, Canada
E-mail: averonik@cc.uoi.gr