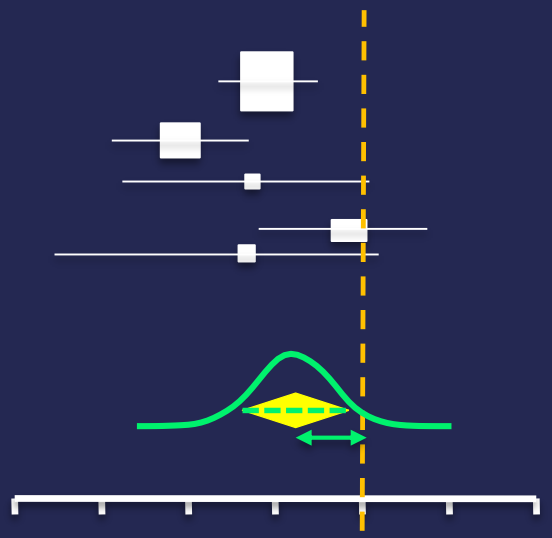
Methods for statistical inference in random effects meta-analysis for use in systematic reviews of medical interventions



Areti Angeliki Veroniki, MSc, PhD

Prepared for: 61 Biometric Colloquium, Germany International Biometrical Society

March 16, 2015

Knowledge Translation
Li Ka Shing Knowledge Institute
St. Michael's Hospital
Toronto, Canada

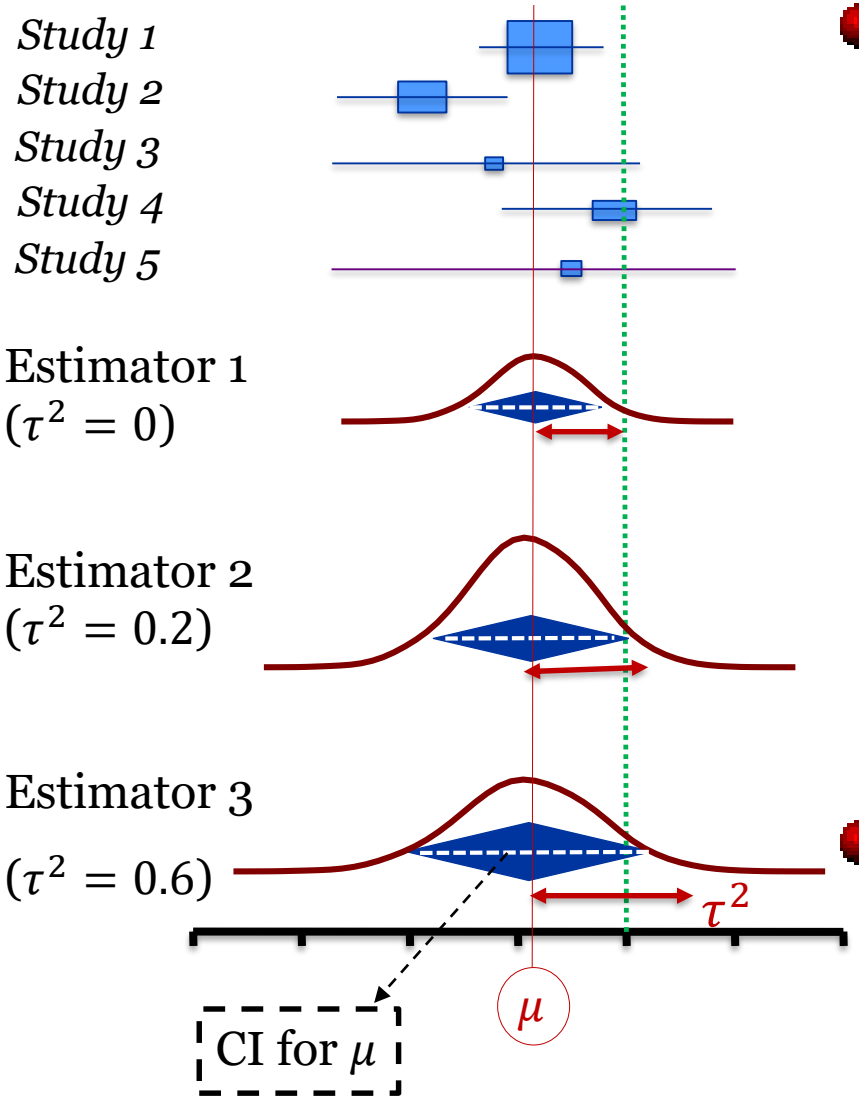
St. Michael's
Inspired Care.
Inspiring Science.

UNIVERSITY OF TORONTO

Aims of the presentation

- To summarize different estimation methods for **heterogeneity**, its **uncertainty**, and **variance** for the **summary treatment effect.**
  - Can different methods impact our decision-making?
- To discuss how the estimated **magnitude** for heterogeneity and the uncertainty of summary treatment effect affect meta-analysis' results
  - How can we infer on the significance of the heterogeneity magnitude?
- To present **empirical** and **simulation** findings for evaluating heterogeneity, its uncertainty, and variance for the summary effect
  - Which method is the most appropriate to apply?
- To identify **factors** that might control the estimated magnitude of heterogeneity and uncertainty of summary effect according to published studies.
  - Under which circumstances the methods perform best?

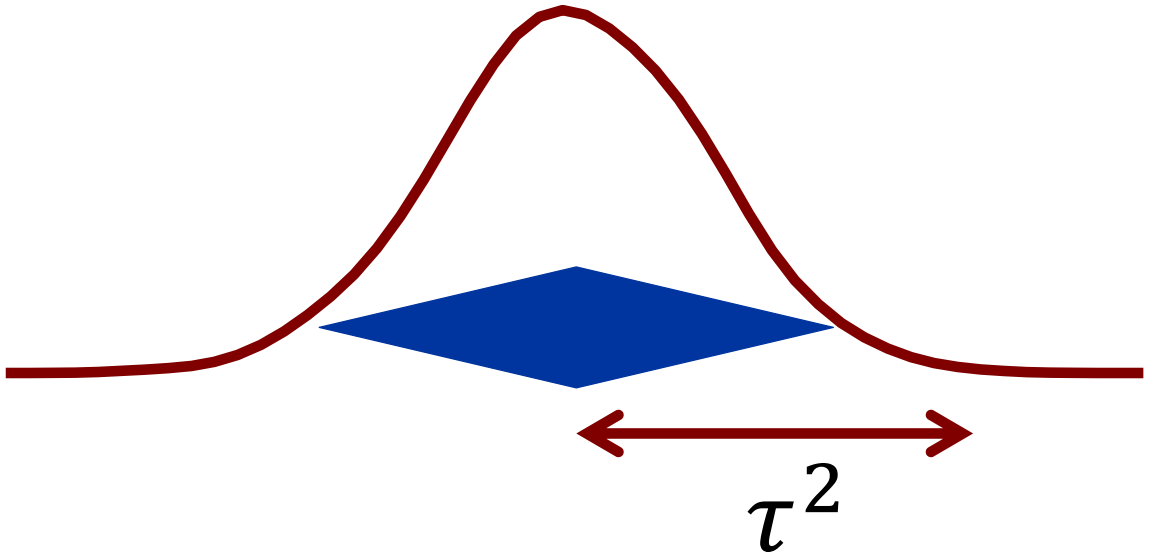
Meta-analysis



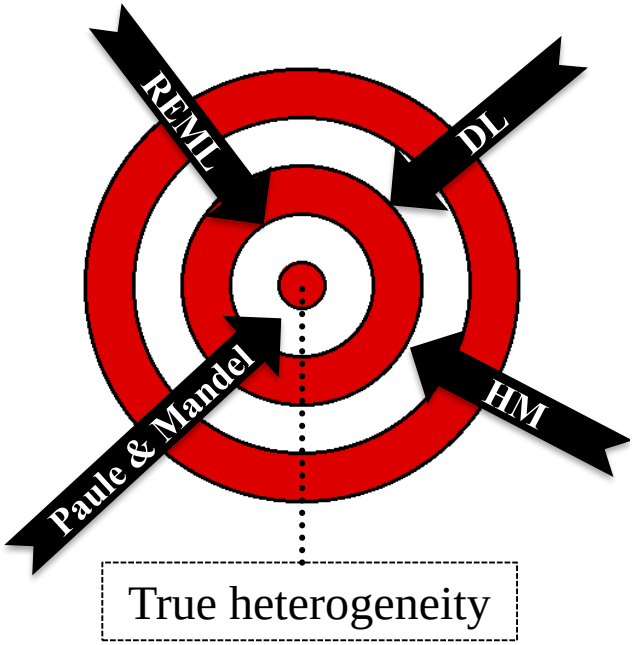
- The choice of the method for estimating
  1. between-study variance (heterogeneity) and its uncertainty
  2. uncertainty for the summary treatment effect

  is important when conducting a meta-analysis.
- Imprecise or biased estimation methods may lead to inappropriate results.

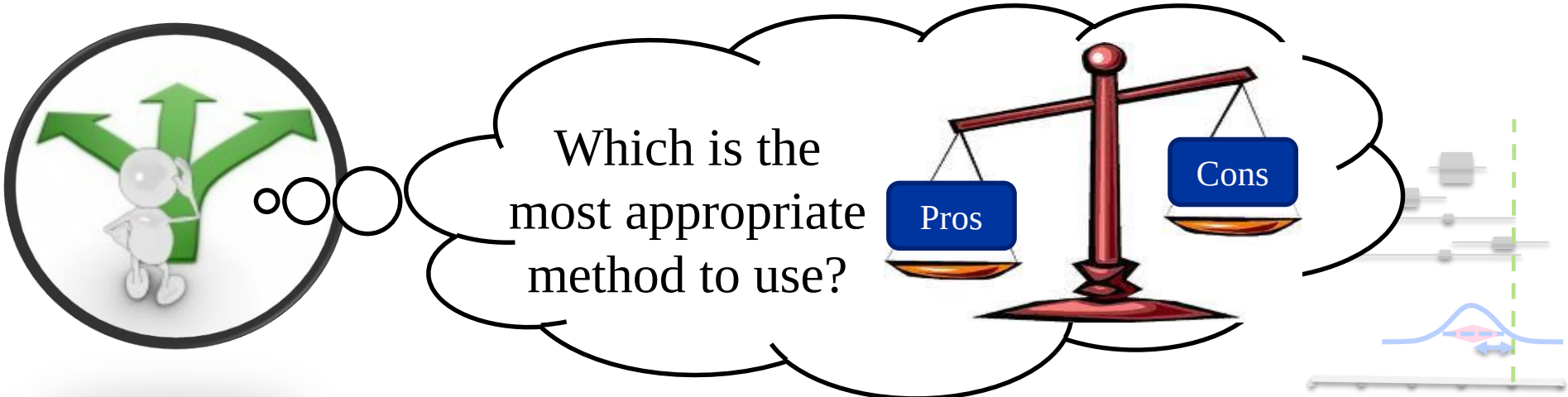
1. Inference on between-study variance



Between-study variance

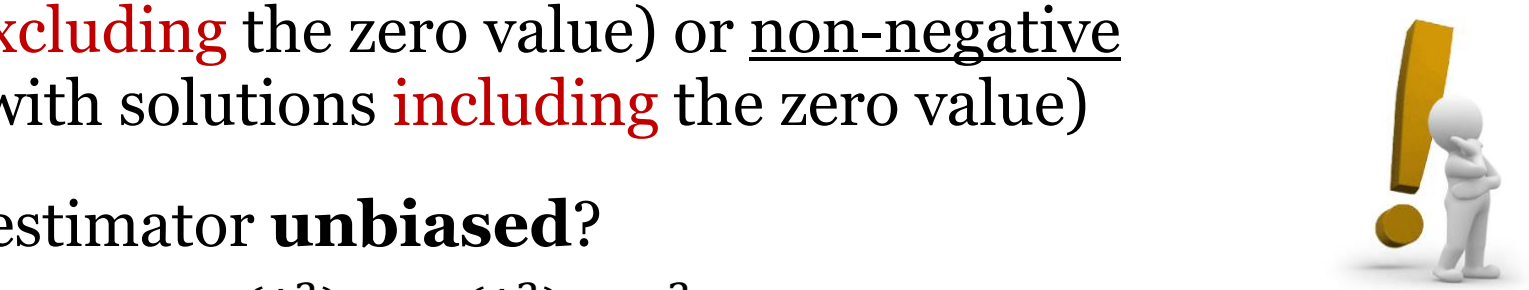


- Plethora of methods, associated with different properties, exist to estimate the between-study variance.
- The performance of between-study variance estimators may vary in various meta-analysis settings.
- An erroneous choice of the method could lead to misleading results.



Select the most appropriate estimator

Be aware of the different properties of each estimator!



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

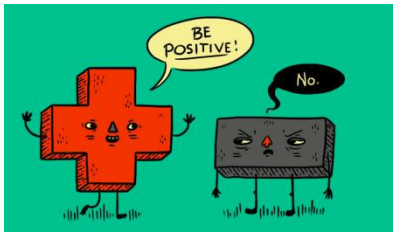
- Estimators can be positive (with solutions excluding the zero value) or non-negative (with solutions including the zero value)

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)

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

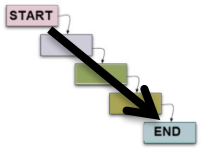


Select the most appropriate estimator

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

– ML depends on the choice of maximization method.

Families of heterogeneity estimators

A. Method of moments estimators

   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

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

B. Maximum likelihood estimators

C. Model error variance estimators

D. Bayes estimators

E. Bootstrap estimators

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

- 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

- A method of moments estimator can be derived by equating the expected value of Q_a and its observed value.
- 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)

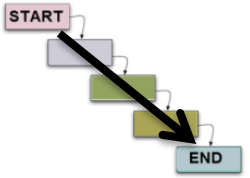
DerSimonian and Laird 1986

- ☐ 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)

Cochran 1954, and Hedges 1983

- ☐ 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



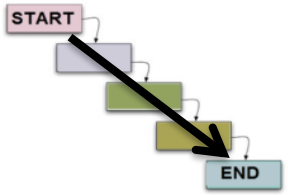
1:Viechtbauer JEBS 2005, 2: Sidik and Jonkman Stat Med 2007, 3: Chung et *al* Stat Med 2013, 4: Thorlund et *al* RSM 2012, 5: DerSimonian and Laird Control Clin Trials 1986, 6: Novianti et al Contemp Clin Trials 2014

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

iii. Hartung and Makambi (HM)

Hartung and Makambi 2003

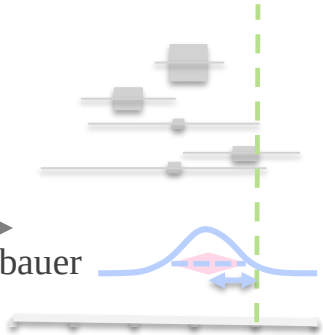
- ☐ A modification of DL with weights the inverse of the within-study variances [1]
- ☑ Simple to compute
- ☐ Produces **positive** estimates
- ☐ Estimates higher τ^2 values compared to DL estimator [2]



iv. Hunter and Schmidt (HS)

Hunter and Schmidt 2004

- ☐ 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]



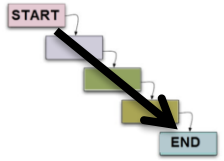
1:Hartung & Makambi Commun in Stati-Simul and Comp 2003, 2: Thorlund et *al* RSM 2012, 3:Viechtbauer JEBS 2005

Method of Moments Estimators – Generalized Q-based methods

DerSimonian and Kacker 2007

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)

Paule and Mandel 1982

- ☐ 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]
- ☑ Robust even when the RE model assumptions do not hold [3]

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

Maximum Likelihood Estimators

i. Maximum Likelihood (ML)

Hardy and Thompson 1996

☒ 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]

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

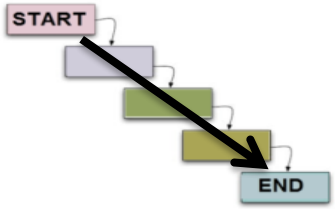
1:Viechtbauer JEBS 2005, 2: Sidik and Jonkman Stat Med 2007, 3: Chung et al Stat Med 2013, 4: Thompson & Sharp Stat Med 1999, 5: Berkey et al Stat Med 1995, 6: Brockwell and Gordon Stat Med 2001

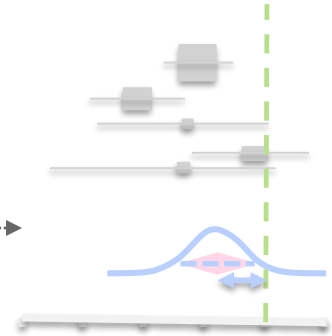
Model error variance estimators

i. Sidik and Jonkman (SJ)

Sidik and Jonkman 2005

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





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

Bayes Estimators

i. Bayes Modal (BM)

Chung et al 2013

- ☐ 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)

Rukhin 2013

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

iii. Full Bayesian (FB)

Smith et al 1995

- ☑ 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]

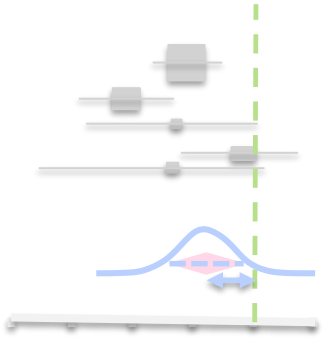
1: Chung et al Stat Med 2013, 2: Kontopantelis et al Plos One 2013, 3: Lambert et al Stat Med 2005

Bootstrap methods

Kontopantelis et al 2013

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.



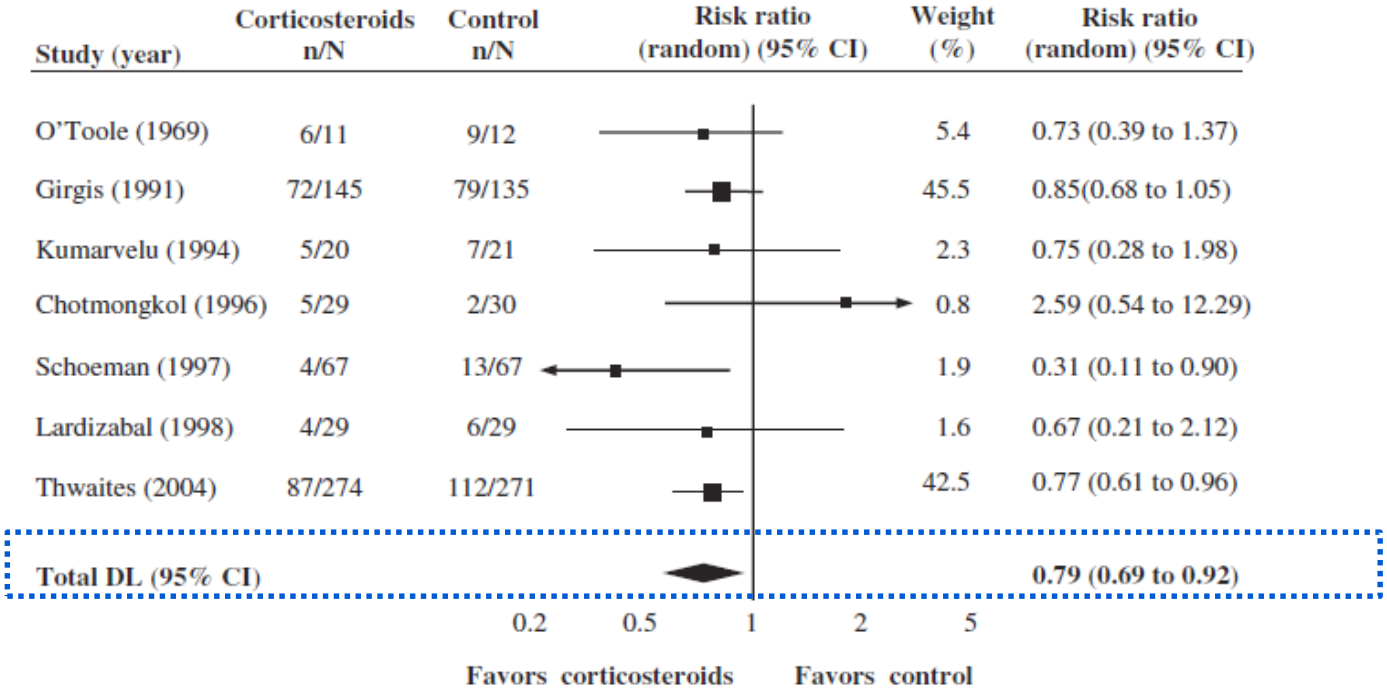
Illustrative example

Bowden et al., 2011,
Veroniki et al. (under review) 2015

	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

Illustrative example

Thorlund et al. 2011



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)

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

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, and HO, BM, SJ, PM perform well for large τ^2.

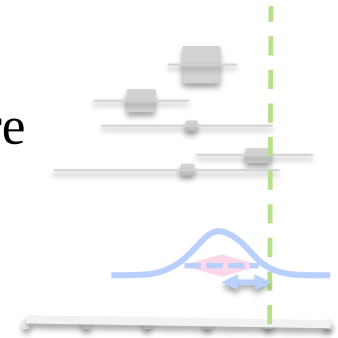
Software

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

Recommendations based on published studies

According to simulation and empirical findings, the main factors that may affect the heterogeneity 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 [1, 2, 3]
- ☒ DL positive, HM, RB positive, BM and SJ overestimate τ^2 [2, 4, 5, 6]
- ☒ DL has lower bias and MSE than HO and SJ [1, 2]
- ☒ BM performs worse than DL and REML when $\tau = 0$ [3]
- ☒ HS and ML are associated with substantial negative bias.[1]

> "One should probably avoid the biased HS and ML estimators because they can potentially provide quite misleading results" [1]

DL	
DLp	
HM	
RBp	
BM	
SJ	
HO	
HS	
ML	

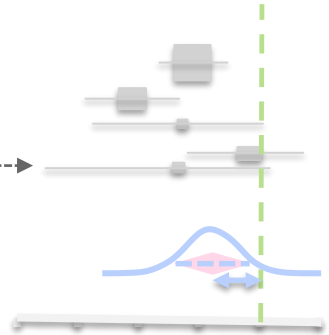
1:Viechtbauer JEBS 2005, 2: Sidik &Jonkman Stat Med 2007, 3: Chung et al Stat Med 2013, 4: Thorlund et *al* RSM 2012, 5:Novianti et al Contemp Clin Trials 2014, 6: Kontopantelis et al Plos One 2013

Recommendations based on published studies

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

- ☒ DLb has higher bias than DL for small k
- ☐ DL2 approximates PM. For rare events underestimates τ^2 [3, 4]
- ☐ 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]
- ☑ AREML yields almost identical estimates with REML [1]
- ☑ 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]

DLb	
DL2	
HO2	
REML	
AREML	
PM	



1: Berkey et al Stat Med 1995, 2: Sidik &Jonkman Stat Med 2007, 3: DerSimonian and 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

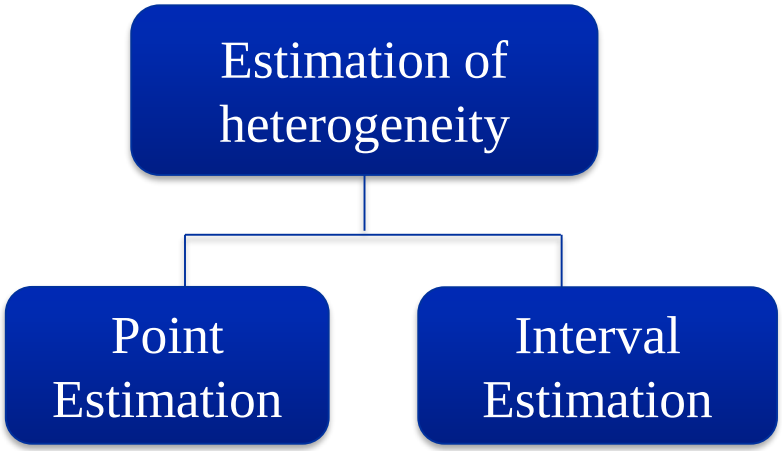
Advantages of Paule and Mandel estimator

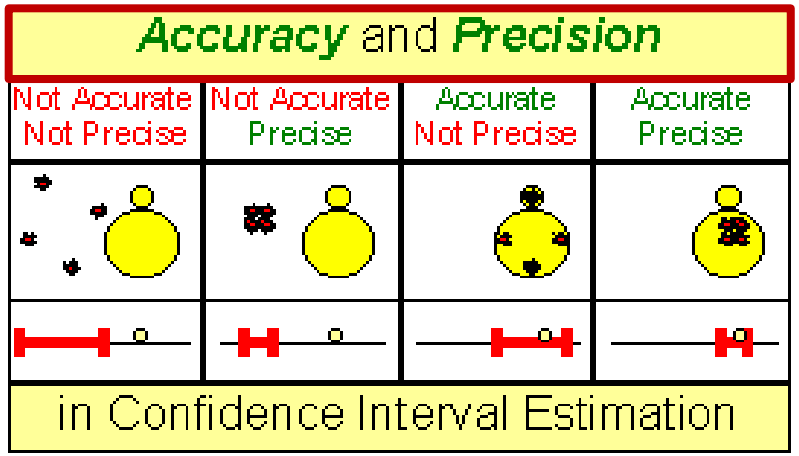
- ☑ When the assumptions underlying the method do not hold, PM is robust for the estimation of τ^2 compared to DL estimator, which is dependent on large sample sizes [1, 2, 3]
- ☑ Outperforms other competitive estimators in terms of bias for both dichotomous and continuous data[4, 5]
- ☑ Easy to obtain.[6]
- ☑ An improved PM is available for rare events.[7]

We suggest using a new estimator!

1:Rukhin et al J Stat Plan Inference 2000, 2: Rukhin Journal of the Royal Statistical Society 2012, 3: DerSimonian and Kacker Contemp Clin Trials 2007, 4: Panityakul et al 2013, 5:Novianti et al Contemp Clin Trials 2014, 6: Bowden et al BMC Med Res Methodol 2011, 7: Bhaumik et al J Amer Stat Assn 2012

Confidence Intervals (CIs) for heterogeneity

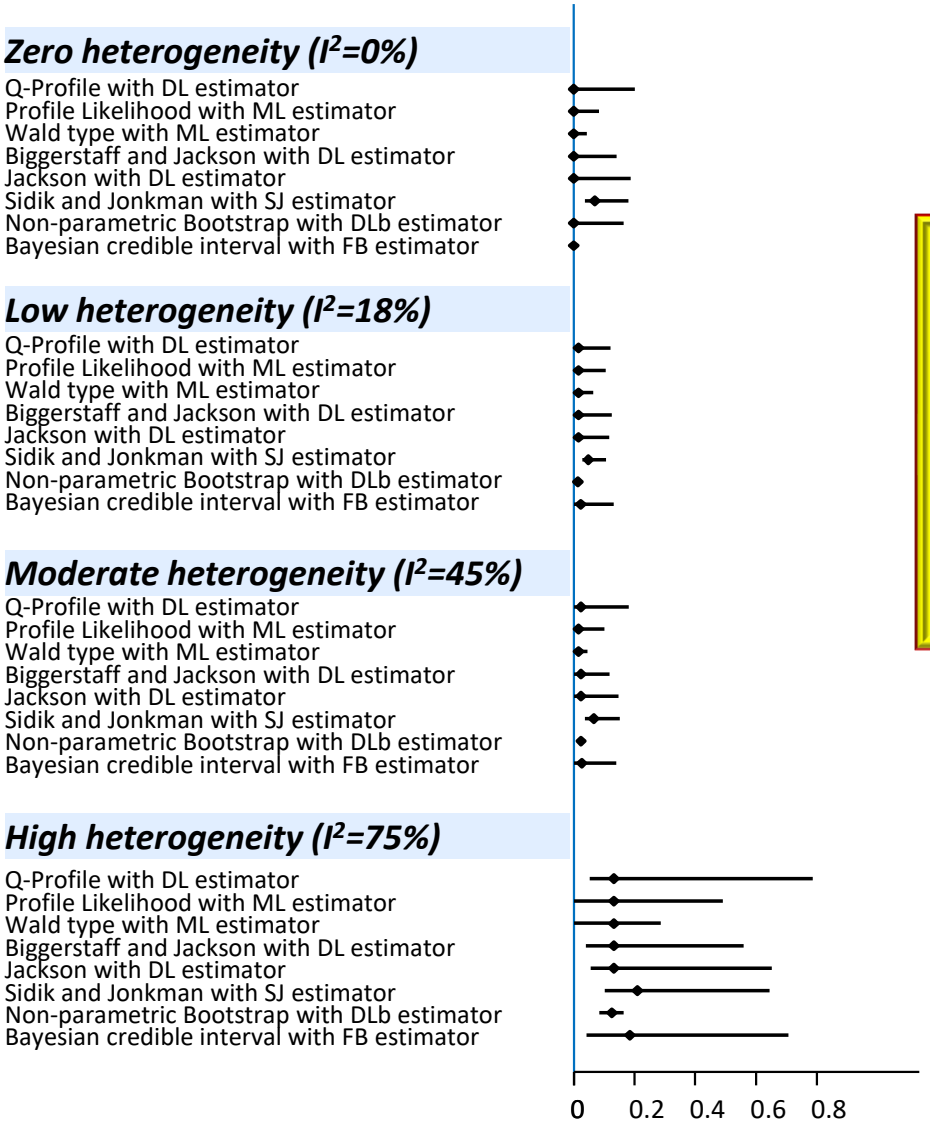




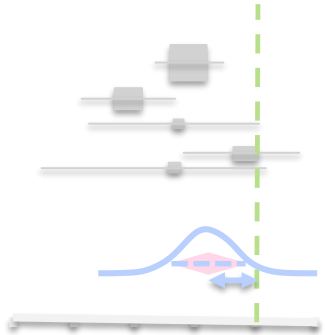
Desirable properties

- ✓ Accuracy = High Coverage Probability – P($\tau^2 \in$ CI)
- ✓ Precision = Narrow CI.

Illustrative example



Various estimation methods may suggest different results!



Bowden et al., 2011,
Veroniki et al. (under review) 2015

Confidence Intervals (CIs) for heterogeneity

Categories

A. Likelihood-based CIs — *Hardy and Thompson 1996*
   a) Profile likelihood (PL)

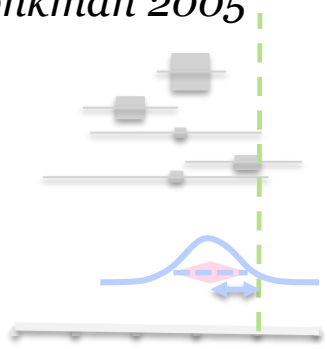
B. Asymptotically normal based CIs — *Biggerstaff and Tweedie 1997*
   a) Wald type (Wt)

C. Generalized Cochran Q - based CIs — *Biggerstaff and Tweedie 1997*
   a) Biggerstaff and Tweedie (BT) — *Biggerstaff and Jackson 2013, Jackson 2014*
   b) Jackson (J) (including Biggerstaff and Jackson (BJ))
   c) Q-profile (QP) — *Viechtbauer 2007*

D. Sidik and Jonkman CIs (SJ) — *Sidik and Jonkman 2005*

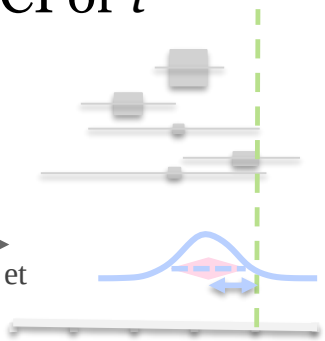
E. Bootstrap CIs

F. Bayesian Credible Intervals



Recommendations based on published studies

- ☒ **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]
  - **QP** provides is more accurate 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.
- ☒ **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.[1, 7]
- ☑ **QP** is simple to compute.

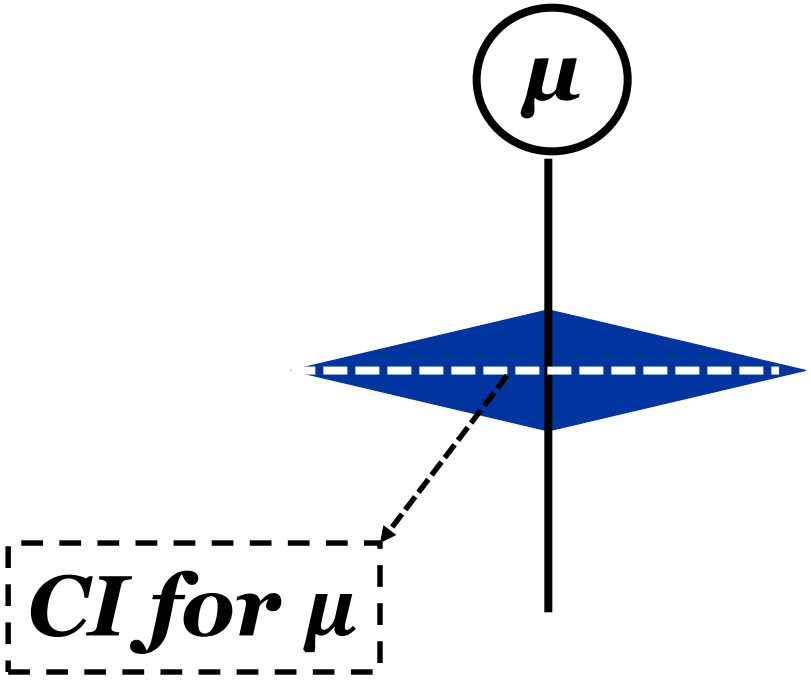


1:Viechtbauer Stat Med 2007, 2: Knapp et al Biom J 2006, 4: Viechtbauer Journal of Statistical Software 2010, 5:Bowden et al BMC Med Res Methodol 2011, 6: Tian Biom J 2008, 7: Jackson RSM 2013

	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 summary treatment effect



CIs for the overall treatment effect
Categories

A. Likelihood-based CIs
   a) Profile likelihood (PL) — *Hardy and Thompson 1996*

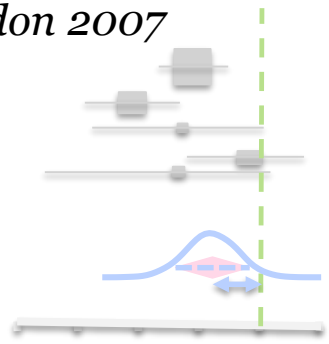
B. Asymptotically normal based CIs — *DerSimonian and Laird, 1986*
   a) Wald type (Wt)
   b) Biggerstaff and Tweedie (BT) — *Biggerstaff and Tweedie 1997*

C. CIs based on the t-distribution
   a) t-distribution with typical variance (t) — *Follmann and Proschan, 1999*
   b) Knapp and Hartung (KH) — *Hartung, 1999, and Knapp and Hartung 2003*

D. Quantile Approximation — *Brockwell and Gordon 2007*

E. Bootstrap CIs

F. Bayesian Credible Intervals



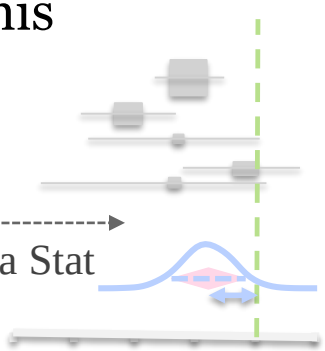
Likelihood-based CIs

i. Profile likelihood (PL)

Hardy and Thompson 1998

- ☑ The method has a good performance for large sample sizes – coverage close to 95%. [1]
- ☑ The method has higher coverage than the Wald type even for small number of studies. [2]
- ☐ **But**, for equal study sizes Wald type and PL have comparable coverage [1]
- ☒ Convergence is not always guaranteed! For few studies and small heterogeneity the process is improved. [3, 4]

A Bartlett-type correction to PL : improves coverage properties via multiplying a modifying factor to the likelihood ratio statistic. This achieves higher coverage than simple PL and Wald type. [5]

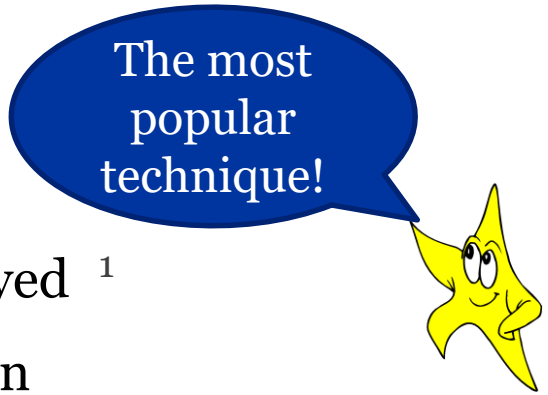


1: Jackson et al J Stat Plan Infer 2010, 2: Brockwell and Gordon Stat Med 2001, 3: Noma Stat Med 2011, 4: Bartlett Proceedings of the Royal Society1937, 5: Noma RSM 2011

Asymptotically normal-based CIs

i. Wald-type (Wt)

DerSimonian and Laird 1986

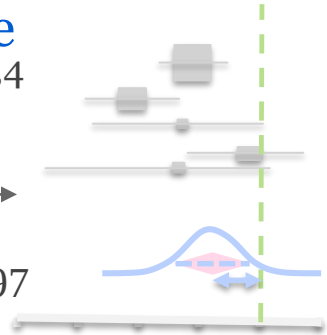


- ☒ The method has considerably low coverage probability, unless size and number of studies are large and τ^2 is low.
- ☒ Depends on the estimator for heterogeneity employed [1]
- ☑ The method using the **BM** estimator outperforms in coverage compared to the **Wt** with DL, ML, REML and HO [2]

ii. Biggerstaff and Tweedie (BT)

Biggerstaff and Tweedie 1997

- ☑ The method takes into account the variability of τ^2.
- ☒ The **Wt** (using DL estimator) and **BT** methods have the same coverage probability but the BT method provides wider CIs.[3,4]



1: Sanchez-Meca and Marin-Martinez Psychol Methods2008, 2: Chung et al Stat Med 2013, 3: Brockwell and Gordon Stat Med 2007, 4: Biggerstaff and Tweedie Stat Med 1997

CIs based on the t-distribution

i. *t*-distribution with typical variance (*t*)

Follmann and Proschan 1999

- ☒ Produces wider CIs than those obtained by Wald type method, especially when the heterogeneity and the number of studies are small [1]
- ☒ Depends on the estimator for τ^2 employed as well as on the number of studies [1]

ii. Knapp and Hartung (KH)

Knapp and Hartung 2003

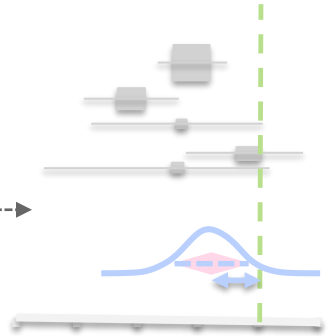
- ☐ Estimates the variance of the overall mean effect with a weighted extension of the usual formula.
- ☑ Not influenced by the magnitude and the estimator of the heterogeneity [1, 2, 3, 4, 5]
- ☑ Provides coverage close to the nominal level *irrespective* the magnitude of heterogeneity and the number of studies. [1, 4]
- ☑ Has a better coverage (except for the case that τ^2 equals zero) and control type I error than the Wald type method. [1, 6]

1: Sanchez-Meca and Marin-Martinez Psychol Methods2008, 2: Hartung Biometrical 1999, 3: Makambi J Biopharm Stat 2004, 4: Sidik and Jonkman Communications in Statistics 2003, 5: Knapp and Hartung Stat Med 2003, 6: Inhout et al. BMC Med. Res. Methodol. 2014

Quantile Approximation (QA)

Brockwell and Gordon 2007

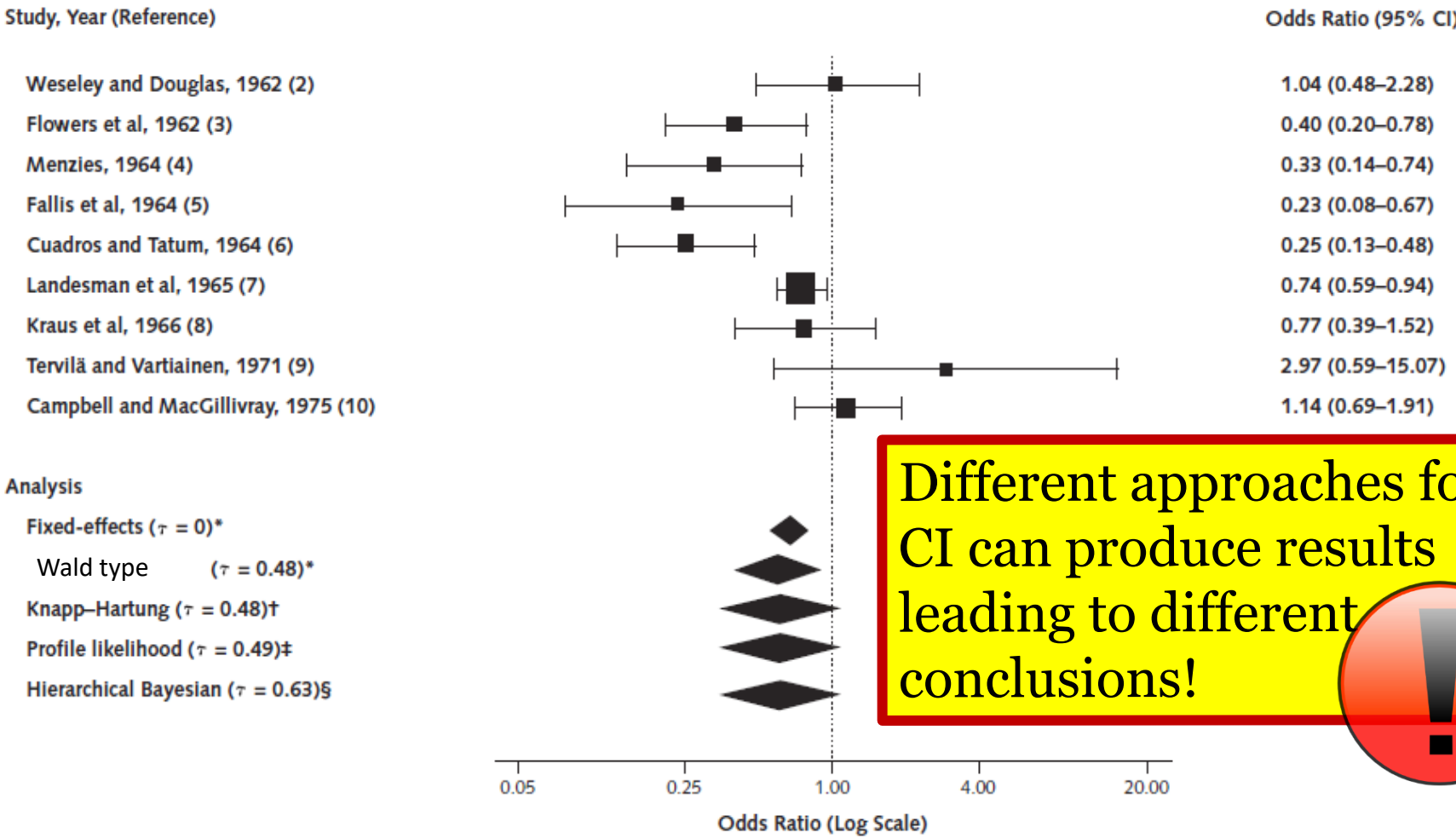
- ☑ Produces CIs with better coverage compared to Wald type .
- ☒ The number of studies, τ^2 and the sampling variances can impact on the quantiles of QA method[1, 2]
- ☒ Different estimators for the heterogeneity impact on the coverage probability of the method [3]



1: Brockwell and Gordon Stat Med 2007, 2: Jackson and Bowden Stat Med 2009, 3: Sanchez-Meca and Marin-Martinez Psychol Methods2008

Illustrative example

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



Different approaches for CI can produce results leading to different conclusions!

Cornel et al. Annals of Internal Medicine 2014

Software

CI Method	Software	CI Method	Software
CIs for the between-study variance			
PL	STATA [*metaan*]	*QP*	R [*metafor*]
Wt	R [*metaSEM*], STATA [*xtreg*]	*SJ*	-
Jackson	R [*metafor*]	*Bayesian CrI*	BUGS, OpenBUGS, WinBUGS
CIs for the overall treatment effect			
PL	STATA [*metaan*]	*t-distribution*	-
Wt	CMA, Excel (MetaEasy), Meta-Disc, Metawin, MIX, Open Meta Analyst, RevMan, R, SAS, STATA, SPSS	*KH*	R [*metafor*], STATA [*metareg*]
BT	-	*QA*	-

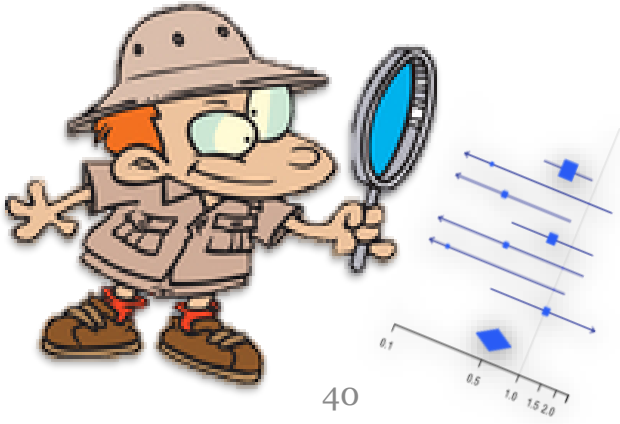
Recommendations based on published studies

- ☒ The **Wt** performs poorly for small samples in comparison to **PL** and ***t*** [1]
- ☑ The ***t*** method is associated with the highest coverage among **PL**, ***t*** and **Wt**.[1]
- ☒ **PL** is computationally intensive involving iterative calculations.
- ☑ The **QA** and ***t*** method have similar coverage and are associated with higher coverage than **Wt** [2]
- ☒ The ***t*** method depends on the estimator and the magnitude of the heterogeneity[3]
- ☑ **QA** and **KH** methods have good coverage in general, but only **KH** method is insensitive to heterogeneity and the number of studies [3]
- ☑ Knapp and Hartung 2003 suggested the use of **PM** estimator along with the **KH** method for obtaining CIs for μ so as to get a cohesive approach based on Q_{gen} [4]

1: Jackson et al J Stat Plan Infer 2010, 2: Brockwell and Gordon Stat Med 2007, 3: Sanchez-Meca and Marin-Martinez Psychol Methods2008, 4: Knapp and Hartung Stat Med 2003

Summary

- ☑ Simulations suggest that **PM** and **REML** estimators are better alternatives to estimate between-study variance than DL.
- ☑ Based on the scenarios and results presented in published studies, we recommend the **QP method** and alternative approach based on a '**generalized Cochrane between-study variance statistic**' to compute CI around heterogeneity.
- ☑ There is limited evidence to inform which method performs best, in particular when the **number of studies is low** (<5) and when the **normality assumption** does not hold.
- ☑ A **sensitivity analysis** using a variety of methods might be needed, particularly when studies are few in number.



References…

1. Brockwell SE, Gordon IR. A simple method for inference on an overall effect in meta-analysis. *Stat Med* 2007; 26(25):4531-4543.
2. DerSimonian R, Kacker R. Random-effects model for meta-analysis of clinical trials: an update. *Contemp Clin Trials* 2007; 28(2):105-114.
3. 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.
4. Jackson, D. Confidence intervals for the between-study variance in random effects meta-analysis using generalised Cochran heterogeneity statistics. *Res. Synth. Me*thods, 2013 4, 220–229
5. Knapp G, Hartung J. Improved tests for a random effects meta-regression with a single covariate. *Stat Med* 2003; 22(17):2693-2710.
6. Kontopantelis, E., Springate, D.A., Reeves, D. A Re-Analysis of the Cochrane Library Data: The Dangers of Unobserved Heterogeneity in Meta-Analyses. *PLoS ONE*, 2013, 8, e69930
7. Makambi K.H. The Effect of the Heterogeneity Variance Estimator on Some Tests of Efficacy. *J Biopharm Stat* 2004; 2:439-449.
8. 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.
9. Sidik K, Jonkman JN. A comparison of heterogeneity variance estimators in combining results of studies. *Stat Med* 2007; 26(9):1964-1981.
10. Veroniki AA, Bender R, Bowden J, Higgins J, Jackson D, Kuss O, Higgins JPT, Langan D, Salanti G. Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Res. Synth. Me*thods, (under review) 2015.
11. 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.

St. Michael's
Inspired Care.
Inspiring Science.

Special thanks to:

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

- My supervisor and mentors: Dr. Sharon Straus, Dr. Andrea Tricco

- Banting Postdoctoral Fellowship Program from the CIHR

E-mail: VeronikiA@smh.ca

UNIVERSITY OF TORONTO