European Research Council (IMMA, Grant nr 260559)

Methods to estimate the heterogeneity variance, its uncertainty and to draw inference on the meta-analysis summary effect

21st Cochrane Colloquium, Quebec
23/09/2013
Statistical Methods Group

School of Medicine
ΙΑΤΡΙΚΗ ΣΧΟΛΗ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ

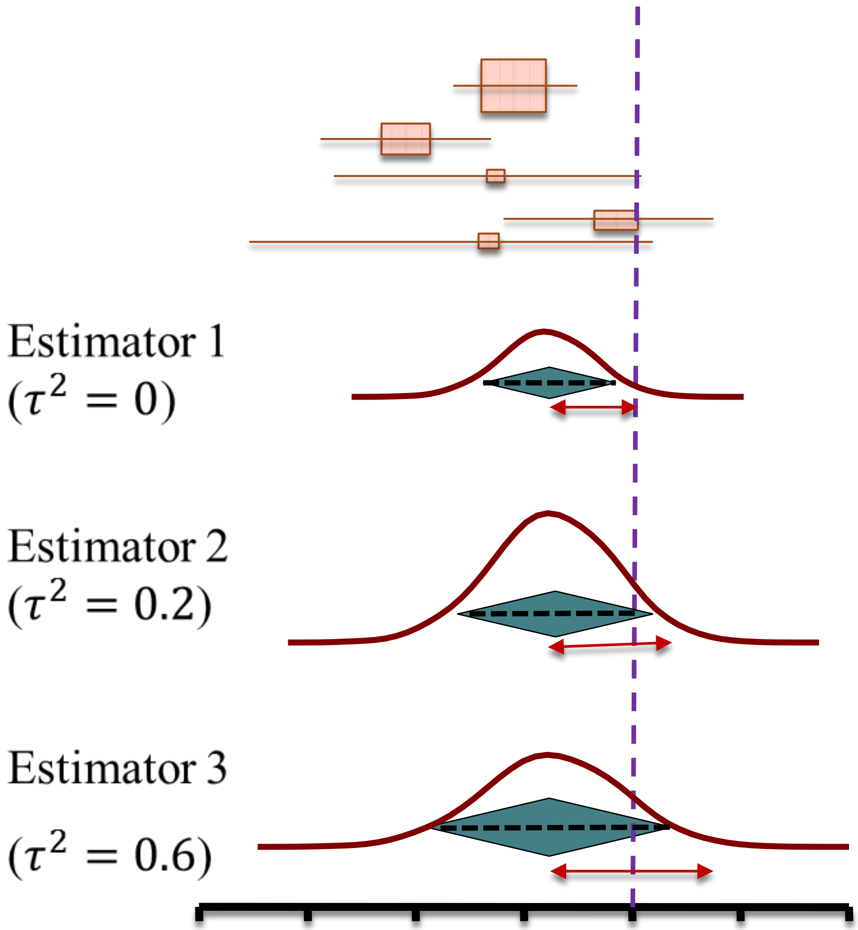
Prepared by Areti Angeliki Veroniki and Georgia Salanti

Discussed and Revised by the RevMan Heterogeneity Working Group; R. Bender, J. Bowden, JP. Higgins, D. Jackson, O. Kuss, G. Knapp, K. Thorlund, W. Viechtbauer

THE COCHRANE COLLABORATION®

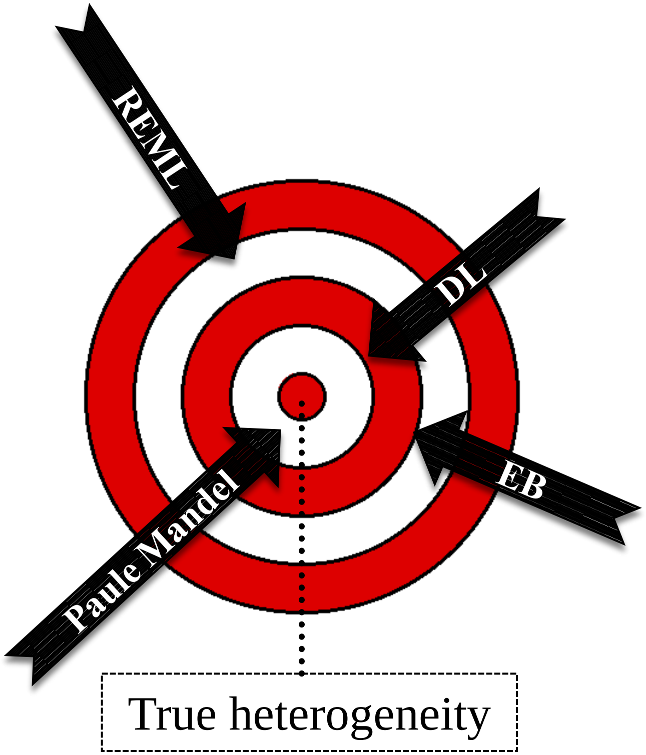
I have no actual or potential conflict of interest in relation to this presentation

Introduction



- The choice of the method for estimating the heterogeneity is an important aspect when conducting a meta-analysis.
- Imprecise or biased estimation methods may lead to inappropriate results.
- We are going to review:
  1. Estimators and uncertainty of the **heterogeneity**
  2. Uncertainty of the **overall** treatment effect

Introduction

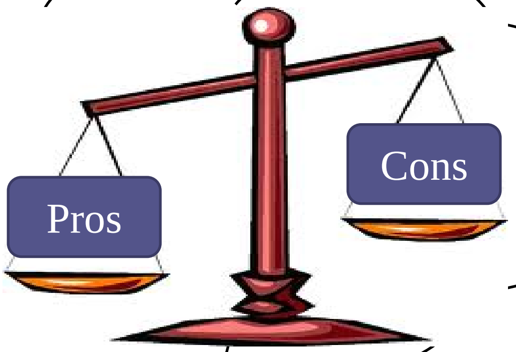


Aim

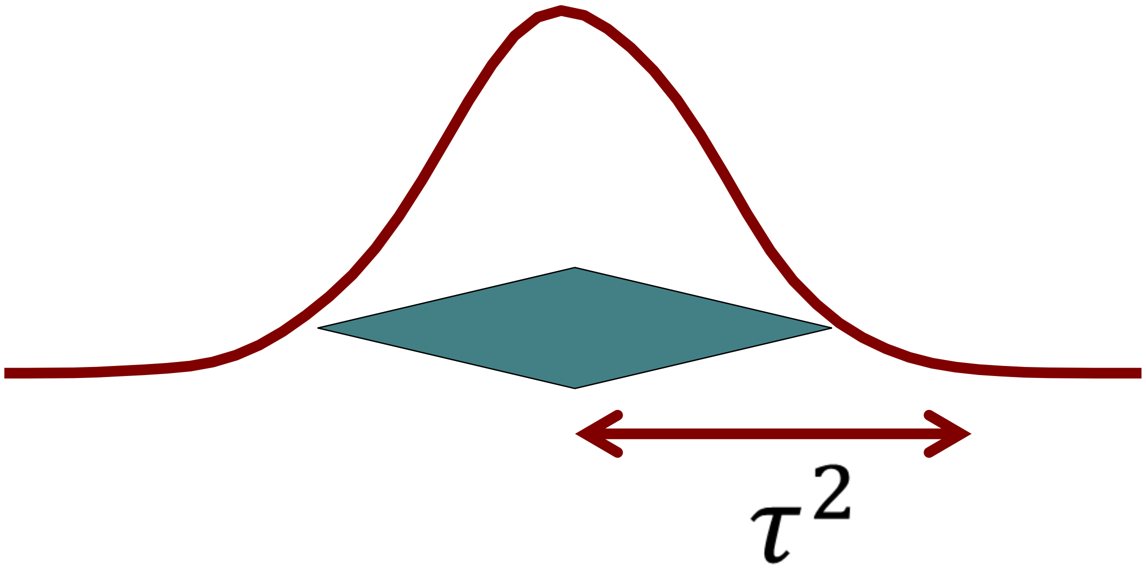
To review the available methods for estimating the heterogeneity and inferences on the summary effect in order to make recommendations for a possible inclusion in RevMan.

We aim to summarize the *differences* and *properties* of all the methods.

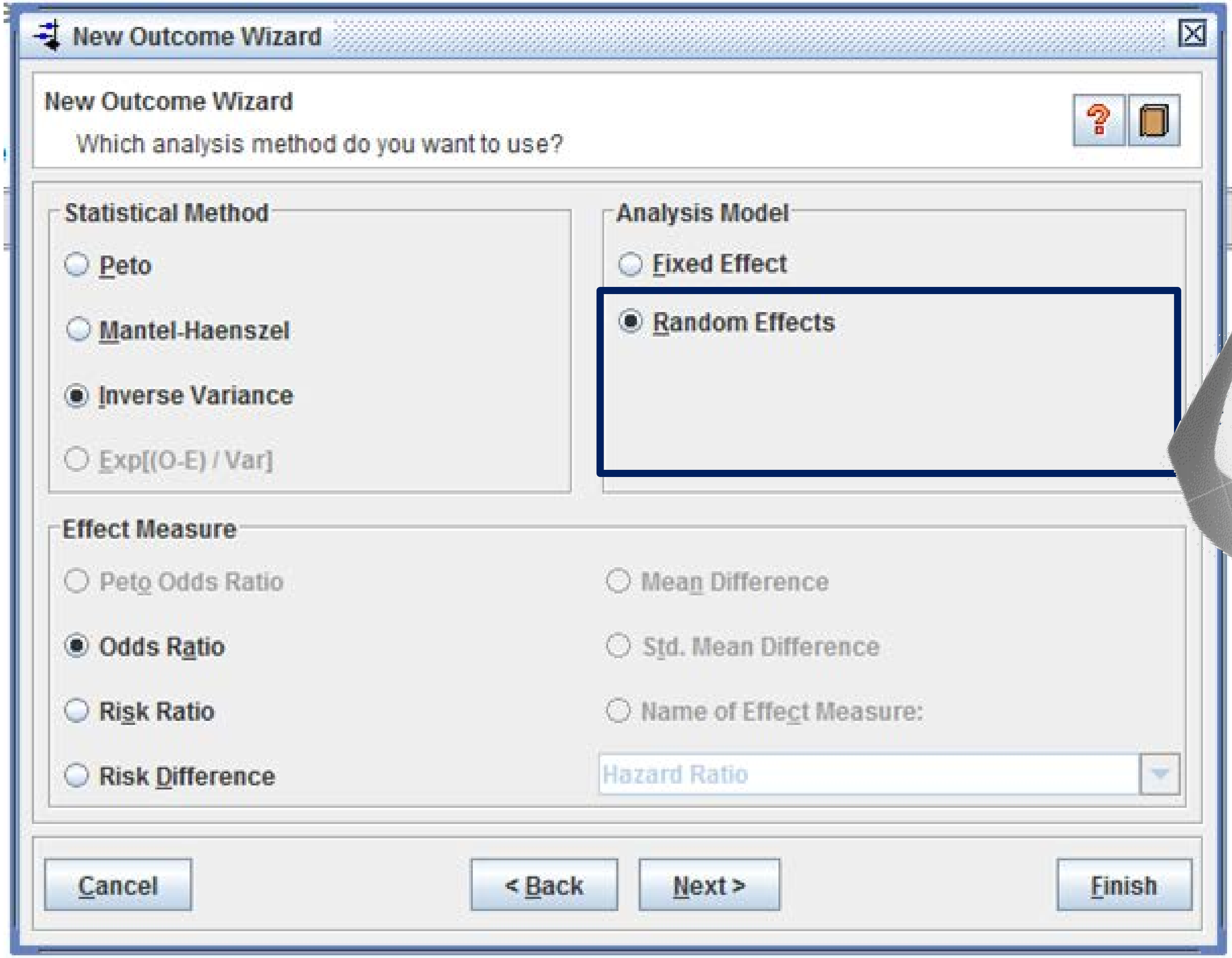
Which is the best method to use?



Inference on the heterogeneity



Introduction



New Outcome Wizard

Which analysis method do you want to use?

Statistical Method
- ○ Peto
- ○ Mantel-Haenszel
- ● Inverse Variance
- ○ Exp[(O-E) / Var]

Analysis Model
- ○ Fixed Effect
- ● Random Effects

Effect Measure
- ○ Peto Odds Ratio
- ● Odds Ratio
- ○ Risk Ratio
- ○ Risk Difference
- ○ Mean Difference
- ○ Std. Mean Difference
- ○ Name of Effect Measure: Hazard Ratio

Extra options: **estimator & CI** for τ^2

RevMan

Select the best estimator

Be aware of the different properties of each estimator!

A good estimator should be:

- **Unbiased**

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

- **Accurate** with low Mean Squared Error (MSE)

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

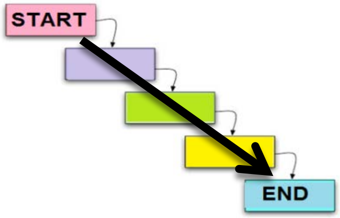
- **Efficient**: Not affected by the sampling fluctuation

- If $MSE(\hat{\tau}_1^2) < MSE(\hat{\tau}_2^2)$ then $\hat{\tau}_1^2$ is said to be more efficient than $\hat{\tau}_2^2$

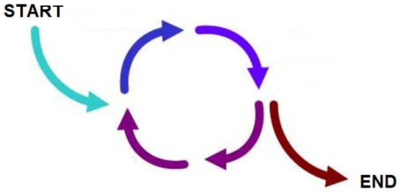
Introduction

Estimators are

- **Direct methods**: provide a parameter 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.

- **Positive methods**: provide solutions in **(0, +∞]**
- **Non-negative methods**: provide solutions in **[0, +∞]**

Introduction

Categories of the estimators for τ^2

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

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

B. Maximum Likelihood Estimators

C. Weighted Least Squares Estimators

D. Bayes estimators

Method of Moments Estimators

Cochran's Q-based methods

DerSimonian and Laird 1986

i. DerSimonian and Laird (DL)

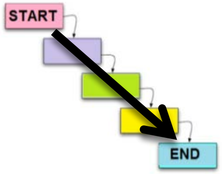
- ✖ The truncation to zero may lead to biased estimators [1]
- ☑ Performs well with low MSE when τ^2 is small [1, 2, 3]
- ✖ Underestimates the true heterogeneity when τ^2 is large and particularly when the number of studies is small[1, 2]

Already implemented in RevMan

Cochran 1954 and Hedges 1983

ii. General form of Hedges-Olkin (GHO)

- ☑ 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 for small τ^2



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

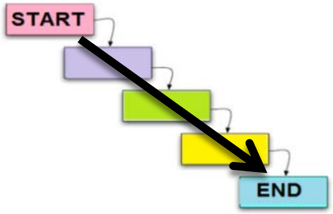
Method of Moments Estimators

Cochran's Q-based methods

iii. Hartung and Makambi (HM)

Hartung and Makambi 2003

- A modification of DerSimonian and Laird
- Produces **positive** estimates [1]
- ✖ Overestimates τ^2 for small to moderate heterogeneity [2]



iv. Hunter and Schmidt (HS)

Hunter and Schmidt 2004

- ☑ Simple to compute
- ☑ Is more efficient than DerSimonian and Laird and General Hedges-Olkin[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

Generalised Q-test

i. Two-step Dersimonian and Laird (DL2)

De *007*

☑ Downwards bias compared to DL

ii. Two-step General form of Hedges-Olkin (GHO

☑ Downwards bias compared to DL and GHO

DL: DerSimonian and Laird
REML: Restricted Maximum Likelihood
EB: Empirical Bayes

iii. Paule and Mandel (PM)

Paule and Mandel 1982

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

☑ Under the normality assumption PM approximates REML and EB [2, 3, 4]

An *improved* PM is also available for *rare* events that reduces bias compared to DL, DL2 and PM estimators [5]

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: Bhaumik et al J Amer Stat Assn 2012

Maximum Likelihood Estimators

i. Maximum Likelihood (ML)

Hardy and Thompson 1996

DL: DerSimonian and Laird

HS: Hunter- Schmidt

GHO: General

Hedges-Olkin

- ✖ Although it has a small MSE, it is associated with substanti increases, the number and size of the included studies is sma

ii. Restricted Maximum Likelihood (REML)

- ☑ REML is less downwardly biased than **DL**[1, 2, 5]
- ✖ For small τ^2 and number of studies **REML** tends to have greater MSE than **DL**[2, 5, 6]
- ✖ **REML** less efficient than **ML** and **HS**[1]
- ☑ **REML** is more efficient with smaller MSE than **GHO**[1]

An ***approximate*** **REML** estimate is also available but it yields 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

Weighted Least Squares Estimators

i. Sidik and Jonkman (SJ)

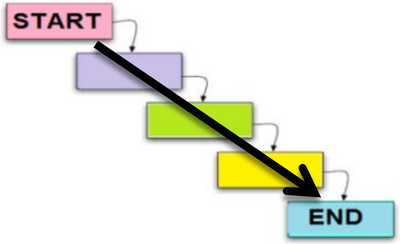
Sidik and Jonkman 2005

- Yields always **positive** values

☑ 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]



1: Sidik and Jonkman J Biopharm Stat 2005, 2: Thorlund et *al* RSM 2012, 3: Sidik and Jonkman Stat Med 2007

Bayes Estimators

i. Empirical Bayes (EB)

Morris 1983

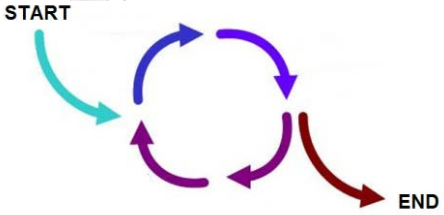
ii. Bayes Modal (BM)

Chung et al 2013

ML: Maximum Likelihood

GHO: General Hedges-Olkin

- o Yields always **positive** values
- ☑ When τ^2 is *positive* it has very low MSE[1]
- ✖ Associated with large bias for small τ^2, especially for few stu
- ✖ For zero τ^2 it performs worse than DL, ML and REML[1]
- ✖ For zero τ^2 it performs better than GHO [1]



iii. Full Bayesian (FB)

Smith et al 1995

- ○ Needs MCMC methodology
- ✖ The choice of the prior for τ is crucial when the number of studies is small [2]
- ✖ A strictly positive prior for τ^2 may produce inflated estimates when τ^2 is close to zero [3]

1: Chung et al Stat Med 2013, 2: Lambert et al Stat Med 2005, 3: Thompson and Sharp Stat Med 1999

Software

Estimator	Software	Estimator	Software
DL	RevMan, R, STATA SAS, SPSS, MIX, Excel, CMA, Metawin, Meta-Disc	*PM*	R
GHO	R	*SJ*	R
HM	-	*ML*	R, STATA SAS, SPSS, HLM, MLwin, Excel, CMA, Metawin, Meta-Disc
HS	R	*REML*	R, STATA, SAS, SPSS, HLM, MLwin, Metawin, Meta-Disc
DL2	-	*EB*	R, STATA, SAS, Meta-Disc
GHO2	-	*BM*	R, STATA
FB	R, SAS, MLwin, BUGS, OpenBUGS, WinBUGS		

tor should be included in RevMan?

HM: Hartung-Makambi
SJ: Sidik-Jonkman
GHO: General Hedges-Olkin
BM: Bayes Modal
FB: Full Bayes

e pairwise meta-analyses have:

$k \leq 10$ and $\tau^2 \leq 0.4$

Turner et al 2012
Pullenayegum et al 2011

In such cases research studies have shown:

- ✖ HM and SJ overestimate τ [4]
- ✖ DL has lower bias and MSE than GHO and SJ [1, 2]
- ✖ BM performs worse than DL and REML when $\tau = 0$ [3]
- ✖ FB needs MCMC methodology

DL	implemented
HM	☒
SJ	☒
GHO	☒
BM	☒
FB	☒

1:Viechtbauer JEBS 2005, 2: Sidik &Jonkman Stat Med 2007, 3: Chung et al Stat Med 2013, 4:Thorlund et *al* RSM 2012

HS: Hunter-Schmidt
ML: Maximum Likelihood
REML: Restricted Maximum Likelihood
EB: Empirical Bayes
PM: Paule and Mandel

should be included in RevMan?

✖ ciated with ias.

the biased HS and M an potentially pro quite misleading results”[5]

- REML is less downwardly biased than DL and ML, but has greater MSE [1, 2]
  - REML is recommended as the best approach[5, 6]

☑ PM is less downwardly biased than DL.

  - The estimator is a better method than DL[3, 4, 7]

DL	implemented
HS	☒
ML	☒
REML	?
EB	?
PM	☑

“DL is very easy to calculate but it may be a misleading estimate of τ^2. Likelihood-based methods (e.g. REML) or Bayesian methods may be preferred, but are more computationally demanding to calculate” [7]

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:Thompson and Sharp Stat Med 1999, 7: Bowden et al BMC Med Res Methodol 2011

Which estimator should be included in RevMan?

Advantages of PM estimator

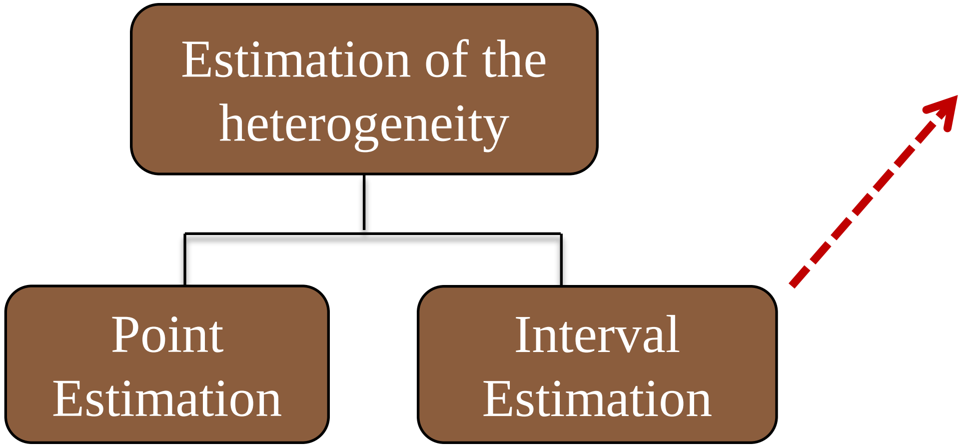
✓ It does not require distributional assumptions and it is more robust for the estimation of τ^2 compared to DL estimator which is dependent on large sample sizes [1]

✓ Mirrors both the REML and EB estimates [1, 2, 3, 4]

✓ Very easy to obtain.

1: DerSimonian and Kacker Contemp Clin Trials 2007, 2: Bowden et al BMC Med Res Methodol 2011, 3:Rukhin et al J Stat Plan Inference 2000, 4: Rukhin Journal of the Royal Statistical Society 2012

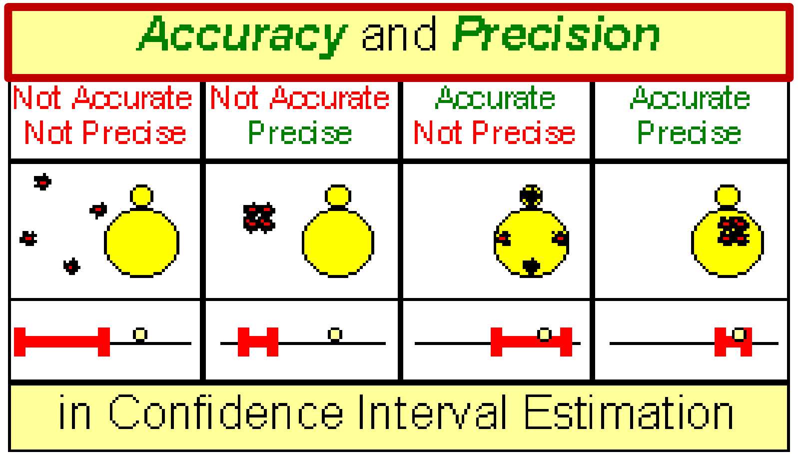
Confidence Intervals (CIs) for the heterogeneity



Estimation of the heterogeneity

Point Estimation

Interval Estimation



PROPERTIES

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

Confidence Intervals for the heterogeneity

Categories

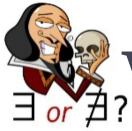
A. Likelihood-based CIs — *Hardy and Thompson 1996*
   a) Profile likelihood (PL)

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

C. Cochran’s Q-based CIs — *Biggerstaff and Tweedie 1997*
   a) Biggerstaff and Tweedie (BT)

D. Generalised Q-based CIs
   a) Biggerstaff and Jackson (BJ) — *Biggerstaff and Jackson 2013*
   b) Q-profile (QP) — *Viechtbauer 2007*

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



Which CI should be included in RevMan?

PL: Profile Likelihood
Wt: Wald type
QP: Q-profile
BJ: Biggerstaff & Jackson
BT: Biggerstaff & Tweedie

✖ The **PL** and **Wt** CIs rely on large number of studies

☑ **QP** is preferable to **PL**, **Wt**, **BT** and **SJ** methods re
even for a small number of studies [2, 4, 6]

☑ Both **QP** and **BJ** are accurate enough.[7]
- **BJ** is recommended for small τ^2 using weights equal to the re
within-study standard errors

✖ Both **QP** and **BJ** methods can result in null sets for the CI of τ^2 when the heterogeneity and the number of studies are small [1, 7]

☑ It is suggested to employ the QP method with the PM estimator[4, 5]

☑ **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

Confidence Intervals for I^2

i. Based on the Cochran's homogeneity statistic

- ✖ I^2 using DL depends on the size of the studies included[1]
- ✖ Empirical evidence suggests I^2 using DL estimates need to be interpreted with caution when the meta-analysis only includes a limited number of events or trials. CIs for I^2 using DL estimate provide good coverage as evidence accumulates [2]
- o Is already implemented in STATA (*heterogi*) and R (*metafor* package)

Higgins and Thompson 2002

ii. Based on the Generalised Q-statistic

- ☑ I^2 using PM maintains well the desired coverage compared to I^2 using DL[3]
- ☑ CIs for I^2 using PM are wider than those of I^2 using DL[3]

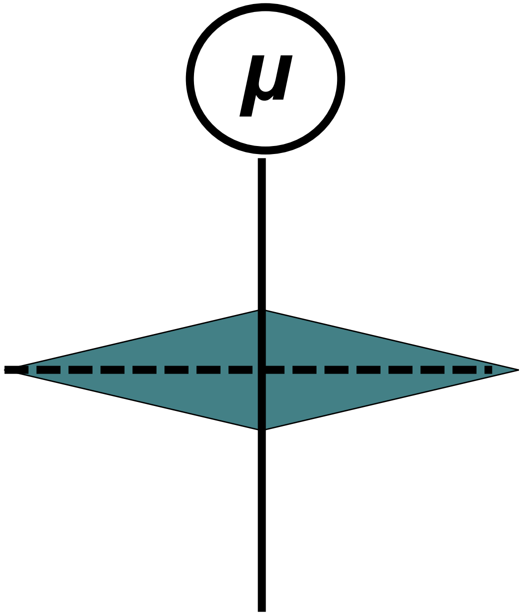
Bowden et al 2011

1: Rücker et al BMC Med Res Methodol 2008, 2: Thorlund et *al* RSM 2012, 3: Bowden et al BMC Med Res Meth 2011

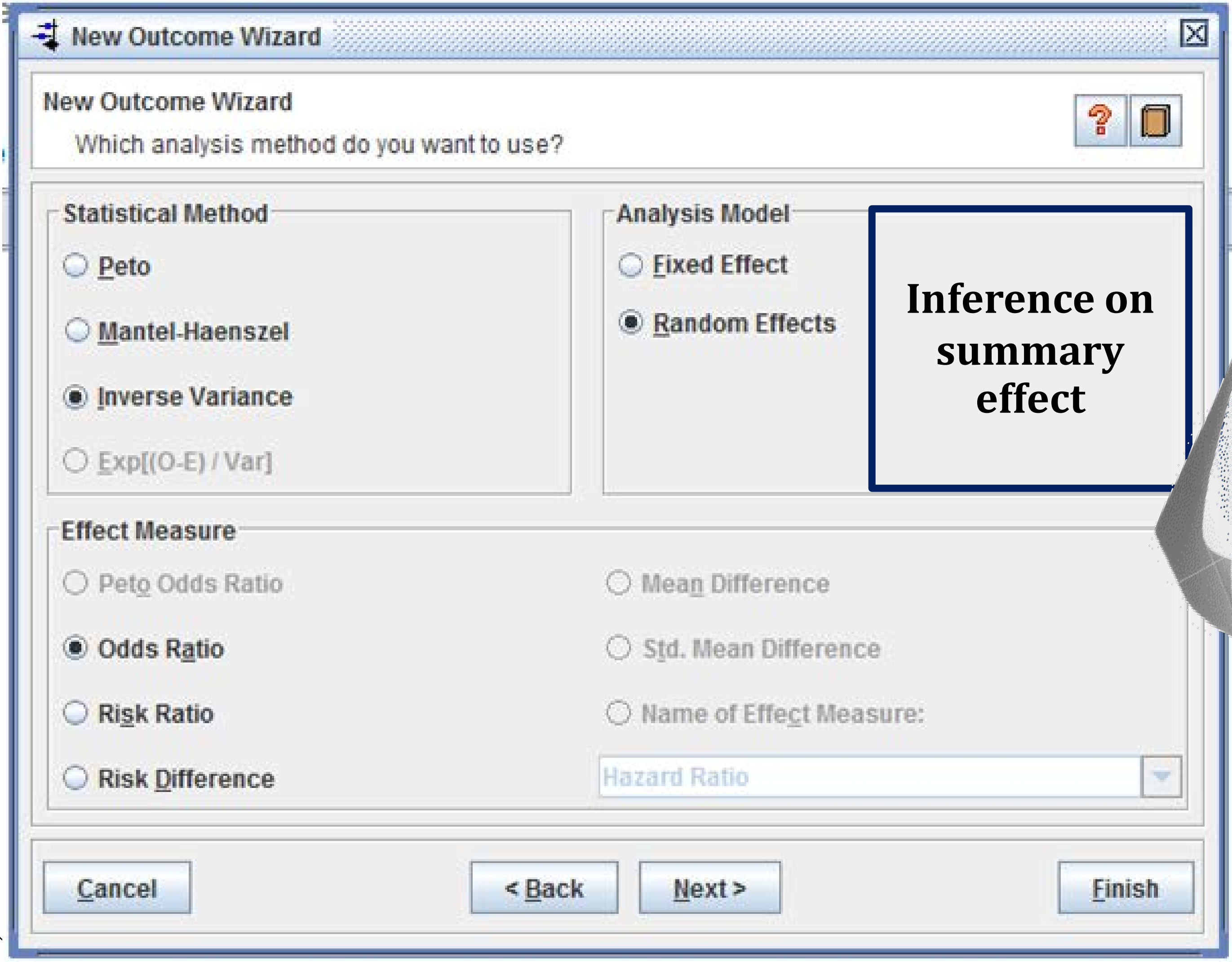
Summary of the estimators for the heterogeneity

Parameter	Estimation Method	Comments
Option 1		
Heterogeneity	DerSimonian and Laird based on Cochran's Q	Already implemented
CIs for heterogeneity	Q-Profile based on Generalized Q	
CIs for I^2	CIs based on Cochran's Q	As in *heterogi* in STATA and *metafor* in R
Option 2		
Heterogeneity	Paule and Mandel based on Generalized Q	
CIs for heterogeneity	Q-Profile based on Generalized Q	
CIs for I^2	CIs based on Generalized Q	As in Bowden et *al* 2011

Inference on the summary effect



Confidence Intervals for the overall mean effect



New Outcome Wizard

New Outcome Wizard

Which analysis method do you want to use?

Statistical Method

○ Peto

○ Mantel-Haenszel

● Inverse Variance

○ Exp[(O-E) / Var]

Analysis Model

○ Fixed Effect

● Random Effects

Effect Measure

○ Peto Odds Ratio

● Odds Ratio

○ Risk Ratio

○ Risk Difference

○ Mean Difference

○ Std. Mean Difference

○ Name of Effect Measure:

Hazard Ratio

Inference on summary effect

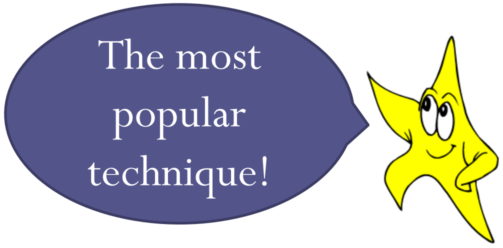
Extra options: **CI** for μ

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 the heterogeneity employed [1]
- ☑ The method using the BM estimator outperforms in coverage the Wt with DL, ML, REML and GHO[2]



Already implemented in RevMan

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

Likelihood-based CIs

i. Profile likelihood (PL)

Hardy and Thompson 1998

- ☑ The method has a good performance for large sample sizes -CP close to 95% [1]
- ☑ The method has higher coverage than Wald type even for small number of studies [2]
- o **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.

Bartlett-type correction to PL : improves the large sample approximation via multiplying a modifying factor to the likelihood ratio statistic. This achieves higher coverage than simple PL and Wald type [3, 4]

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

CIs based on *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

- o 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 than Wald type except for the case that τ^2 equals zero.

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

Quantile Approximation (QA)

- Approximates the 0.025 and 0.975 quantiles of the distribution of the statistic

$$M = \frac{\hat{\mu}_{RE} - \mu}{\sqrt{var(\hat{\mu}_{RE})}}$$

☑ 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]

Brockwell and Gordon 2007

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

Which CI should be included in Rev ?

PL: Profile Likelihood
Wt: Wald type
QA: Quantile-approximation
t: t-distribution
KH: Knapp & Hartung

- ✖ The **Wt** performs poorly for small samples in comparison
- ☑ The ***t*** method is associated with the highest coverage amo
- ✖ **PL** is computationally intensive involving iterative calcu
- ☑ The **QA** and ***t*** method have similar coverage and are associated coverage than **Wt** [2]
- ✖ The **QA** and ***t*** method depend on the estimator of the heterogeneity [3]
- ☑ Sanchez-Meca and Marin Martinez 2008 showed that **QA** and **KH** methods present good coverage in general. However, they suggest the use of **KH** method as it is insensitive to the 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 for the overall treatment effect

Parameter/Statistic	Estimation Method	Comments
Option 1		
CI for μ	Wald-type	already implemented
Test H_0:$\mu=0$	z-score	already implemented
Option 2		
CI for μ	Knapp-Hartung	
Test H_0:$\mu=0$	Knapp-Hartung t-test	

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. Knapp G, Hartung J. Improved tests for a random effects meta-regression with a single covariate. *Stat Med* 2003; 22(17):2693-2710.
5. Makambi K.H. The Effect of the Heterogeneity Variance Estimator on Some Tests of Efficacy. *J Biopharm Stat* 2004; 2:439-449.
6. 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.
7. Sidik K, Jonkman JN. A comparison of heterogeneity variance estimators in combining results of studies. *Stat Med* 2007; 26(9):1964-1981.
8. 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.
9. 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.