QUANTIFYING HETEROGENEITY

IN META-ANALYSIS

Martina Mittlböck and Harald Heinzl

Core Unit for Medical Statistics and Informatics
Medical University of Vienna

Meta-Analyses:

aims at estimation of summary effects

$$\bar{\theta} = \left(\sum_{i=1}^{k} w_i \hat{\theta}_i\right) \Big/ \left(\sum_{i=1}^{k} w_i\right)$$

$\hat{\theta}_i$ **are the observed treatment effects**

(e.g. log odds ratios, mean differences,…)

w_i **denote the precisions of the studies**

Meta-Analyses:

aims at estimation of summary effects

$$\bar{\theta} = \left(\sum_{i=1}^{k} w_i \hat{\theta}_i\right) \Big/ \left(\sum_{i=1}^{k} w_i\right)$$

► **fixed effect approach**

$\hat{\theta}_i \sim N\left(\theta, \sigma_i^2\right)$ and $w_i = 1/\sigma_i^2$

► **random effects approach**

$\hat{\theta}_i \sim N\left(\theta_i, \sigma_i^2\right)$ and $\theta_i \sim N\left(\theta, \sigma_B^2\right)$

so that $w_i = 1/\left(\sigma_i^2 + \sigma_B^2\right)$

Points to consider in Meta-Analysis:

- between-study heterogeneity
- heterogeneity may complicate the interpretation of results
- usually determines whether fixed or random effects model should be used
- decision usually by a test for heterogeneity

Test for Heterogeneity

Cochrane's test for heterogeneity

$$Q = \sum_{i} w_i \left(\hat{\theta}_i - \bar{\theta} \right)^2$$

- ► under homogeneity $Q \sim \chi^2$ with (k-1) df's
- ► tests if the between-study variance $\sigma_B^2 = 0$
- ► Cochrane's test has low power

Meta-Analyses:

Simulations with normaly distributed outcome

θ is a mean difference

$$0 \le \sigma_B^2 \le 0.5 \qquad \sigma_W^2 = 0.1,\ 0.2$$

$$1)\ \ \theta_i \sim N\left(\theta, \sigma_B^2\right)$$

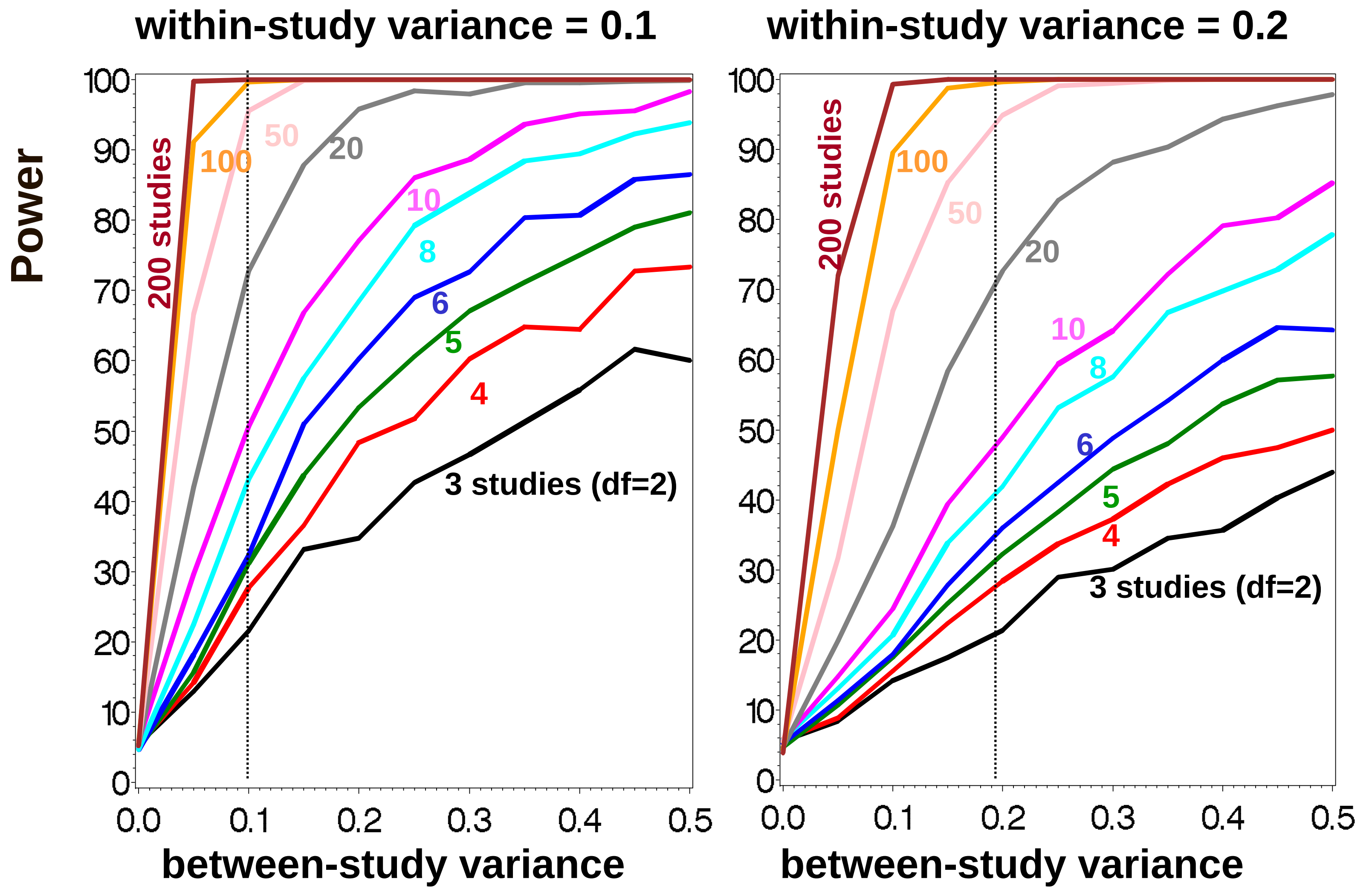
$$2)\ \ \sigma_i^2 \sim N\left(\sigma_W^2, 0.01^2\right)$$

$$3)\ \ \hat{\theta}_i \sim N\left(\theta_i, \sigma_i^2\right)$$

1000 simulation runs were performed

consequently $w_i = 1 / \left(\sigma_i^2 + \sigma_B^2\right)$

Test for Heterogeneity (normal outcome)



Meta-Analyses:

Simulations with binary outcome

θ **is the log-odds ratio and exp(θ)=1**

$$0 \le \sigma_B^2 \le 0.5 \qquad n_1 = n_2 \qquad p_1 = 0.5$$

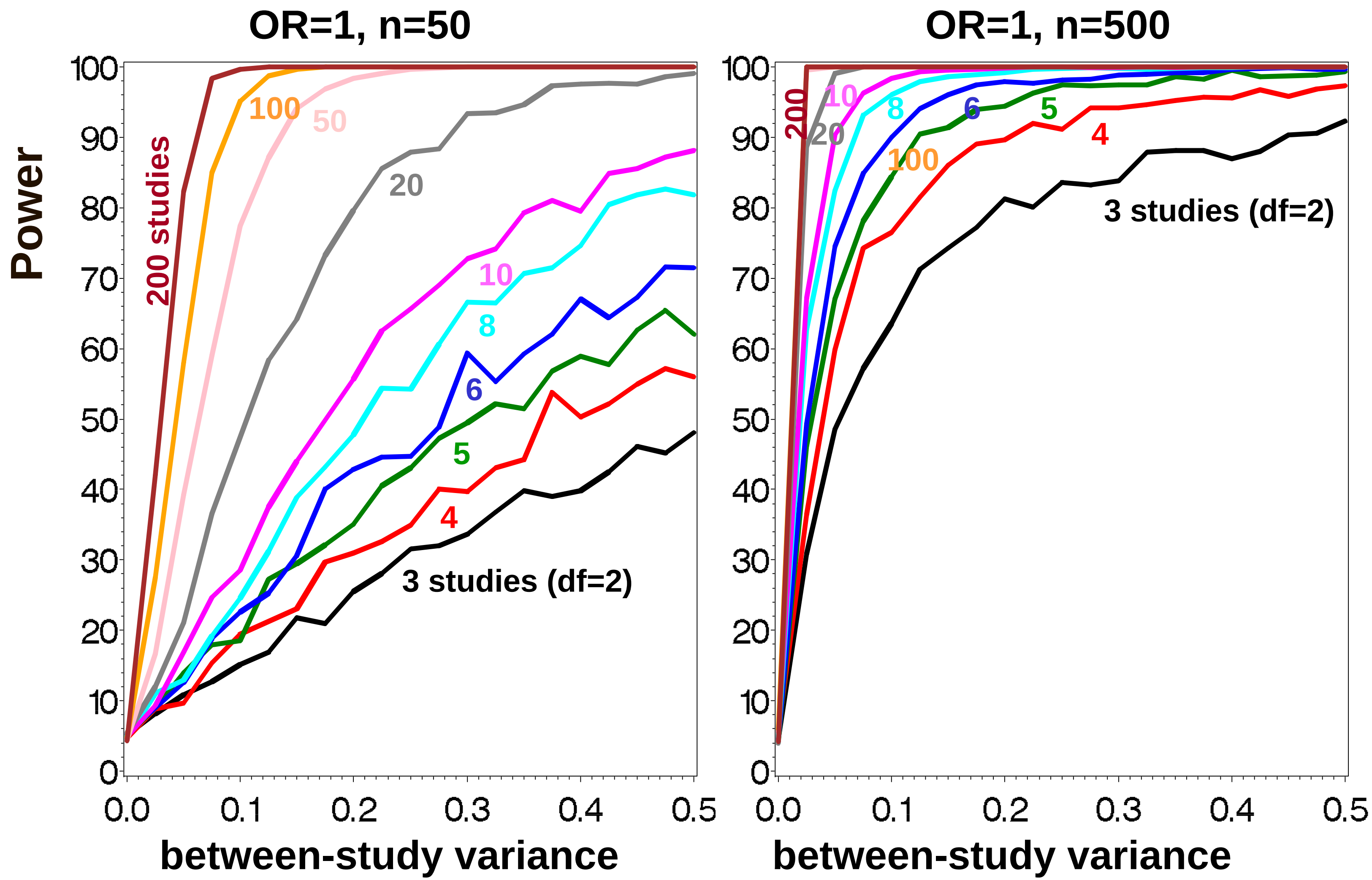
$$\theta_i \sim N\left(\theta, \sigma_B^2\right)$$

$$a_{1,i} \sim B\left(p_1, n_1\right) \quad \text{and} \quad a_{2,i} \sim B\left(p_2, n_2\right)$$

$$\exp\left(\hat{\theta}_i\right) = \left[a_{1,i}\left(n_2 - a_{2,i}\right)\right] / \left[a_{2,i}\left(n_1 - a_{1,i}\right)\right]$$

$$\operatorname{var}\left(\hat{\theta}_i\right) = 1/a_{1,i} + 1/a_{2,i} + 1/\left(n_1 - a_{1,i}\right) + 1/\left(n_2 - a_{2,i}\right)$$

Test for Heterogeneity (binary outcome)



Test for Heterogeneity

As Cochrane‘s test has low power

- ► **Petitti** (StatMed, 2001) **preferres α=0.1**

- ► **thus a sig. value must be considered as strong evidence of heterogeneity (Baujat et al.,** StatMed 2002**)**

- ► **measures of heterogeneity ?**

Estimation of variance

Estimation of between-study variance:

$$E(Q) = \sigma_B^2 \left(\sum_i w_i - \frac{\sum_i w_i^2}{\sum_i w_i} \right) + (k-1)$$

$$\Rightarrow \hat{\sigma}_B^2 = \frac{Q-(k-1)}{\sum_i w_i^2 - \left(\sum_i w_i^2 / \sum_i w_i^2 \right)}$$

If $Q<(k-1)$ then $\hat{\sigma}_B^2 < 0$ and

$\max\left(0, \hat{\sigma}_B^2\right)$ is used in practice

Estimation of variance

Estimation of between-study variance:

$$E(Q) = \sigma_B^2 \left(\sum_i w_i - \frac{\sum_i w_i^2}{\sum_i w_i} \right) + (k-1)$$

$$\Rightarrow \hat{\sigma}_B^2 = \frac{Q-(k-1)}{\sum_i w_i^2 - \left(\sum_i w_i^2 / \sum_i w_i^2 \right)}$$

Is e.g. $\hat{\sigma}_B^2 = 0.6$ of importance?

Estimation of variance

Estimation of within-study variance:

1) Inverse arithmetic mean weight

(Takkouche et al. (1999)):

$$\hat{\sigma}^2_{W,1} = k \Big/ \sum_i w_i$$

2) Another possibility is

$$\hat{\sigma}^2_{W,2} = \frac{(k-1)\sum_i w_i}{\left(\sum_i w_i\right)^2 - \sum_i w_i^2}$$

Estimation of variance

Estimation of within-study variance:

As $E(Q) = (k-1)\left[\left(\sigma_B^2/\sigma_{W,2}^2\right)+1\right]$

$\sigma_{W,2}^2$ **is directly related to both E(Q) and the power of the heterogeneity test**

Thus $\sigma_{W,2}^2$ **is preferable to** $\sigma_{W,1}^2$ **,**
to assess the impact of within-study variance on the power of heterogeneity tests

Overall Measure of Heterogeneity

I^2 by Higgins and Thompson **(StatMed, 2002)**

implemented in Review Manager Software (RevMan) of the Cochrane Collaboration

$$Q > df \qquad I^2 = \frac{Q - df}{Q}$$

$$Q \leq df \qquad I^2 = 0$$

► **has values between 0 and 1**

Overall Measure of Heterogeneity

I^2 by Higgins and Thompson (StatMed, 2002)

$$I^2 = \frac{Q - df}{Q} \qquad E\left(I^2\right) = \frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2}$$

- ► the proportion of the between-study variance on the total variance
- ► does not include the number of studies (k)

Overall Measure of Heterogeneity

I^2 was claimed to be independent on the number of studies

$$I^2 = \frac{Q - df}{Q} = 1 - \frac{df}{Q}$$

$$H_0 : Q \sim \chi^2_{k-1} \Rightarrow 1/Q \sim \text{inverse} - \chi^2_{k-1}$$

$$\Rightarrow E\left(\frac{1}{Q}\right) = \frac{1}{df - 2} \quad \textbf{for } df > 2 \textbf{ or } k > 3$$

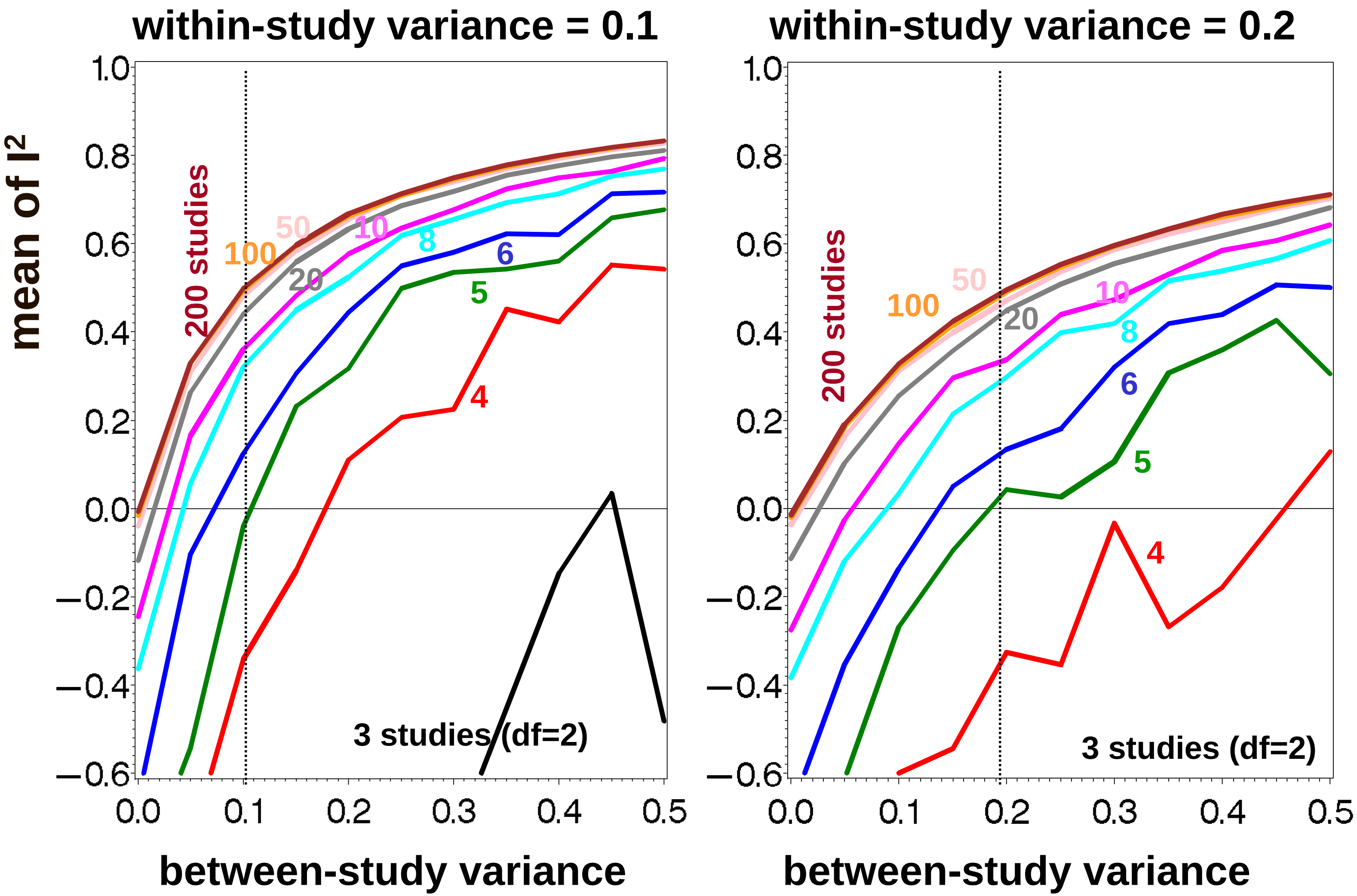
$$\Rightarrow E\left(I^2\right) = \frac{-2}{df - 2} \quad \textbf{for } \sigma^2_B = 0 \textbf{ and } k > 3$$

Overall Measure of Heterogeneity

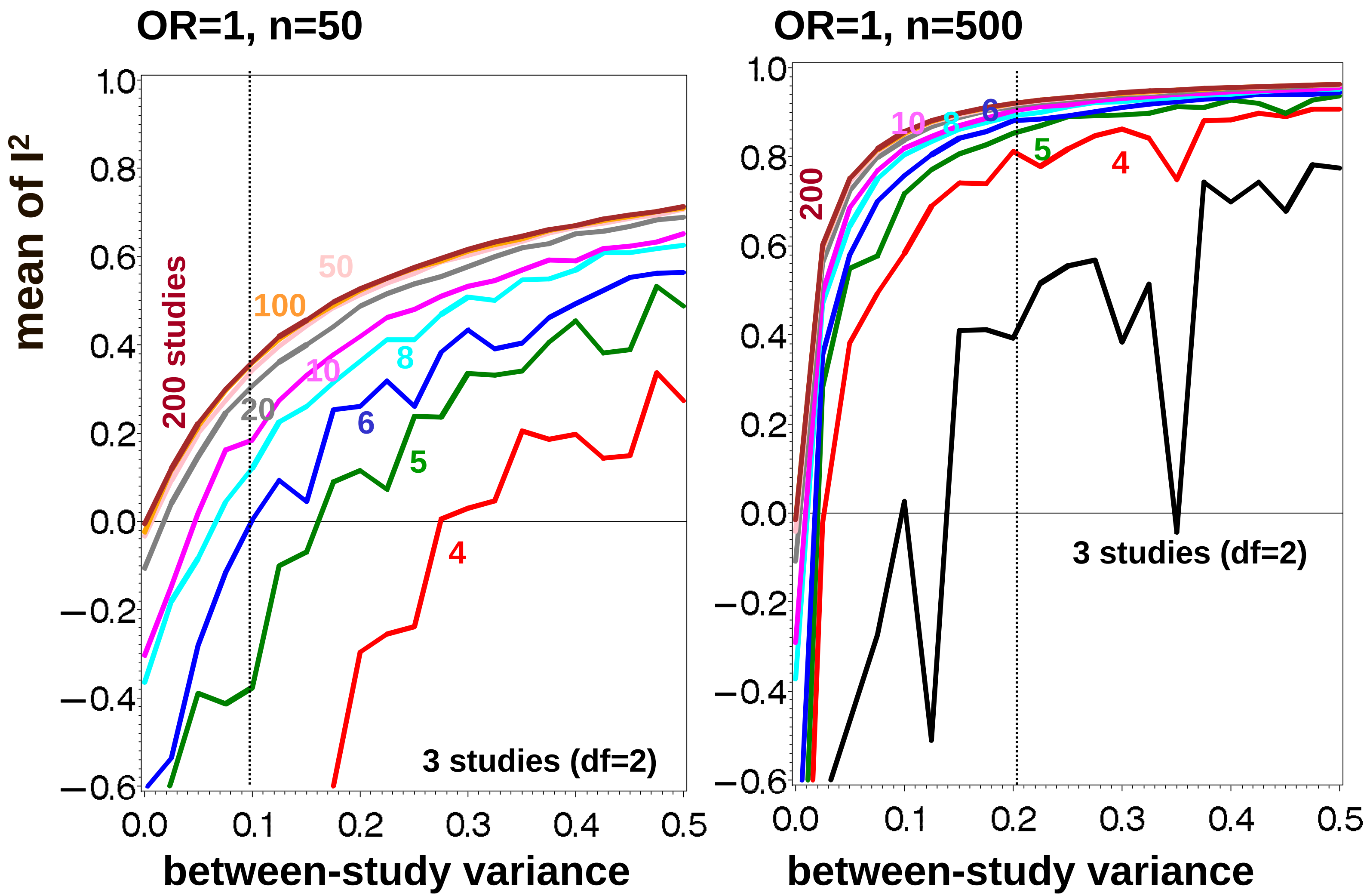
I^2 by Higgins and Thompson (StatMed, 2002)

- ►**How severe is the dependence on the number of studies of I^2?**
- ►**How is the scaling of I^2?**

Measure of for Heterogeneity I^2



Measure of Heterogeneity I^2 (bin. Outc.)



Overall Measure of Heterogeneity

Another measure:

$$Q > df \qquad H_M^2 = \frac{Q - df}{df}$$

$$Q \leq df \qquad H_M^2 = 0$$

- ► **has values between 0 and ∞**
- ► $E\left(H_M^2\right) = \sigma_B^2 / \sigma_{W,2}^2$

  gives the between-study variance proportional to within-study variance

Overall Measure of Heterogeneity

Expectation of H_M^2 under H_0 :

$$H_M^2 = \frac{Q - df}{df} = \frac{Q}{df} - 1$$

$$Q \sim \chi_{k-1}^2 \quad \Rightarrow \quad E(Q) = df$$

$$\Rightarrow \quad E\left(H_M^2\right) = \frac{df}{df} - 1 = 0 \quad \textbf{for} \quad \sigma_B^2 = 0$$

Overall Measure of Heterogeneity

Measure of variance inflation:

$$Q > df \qquad H^2 = \frac{Q}{df}$$

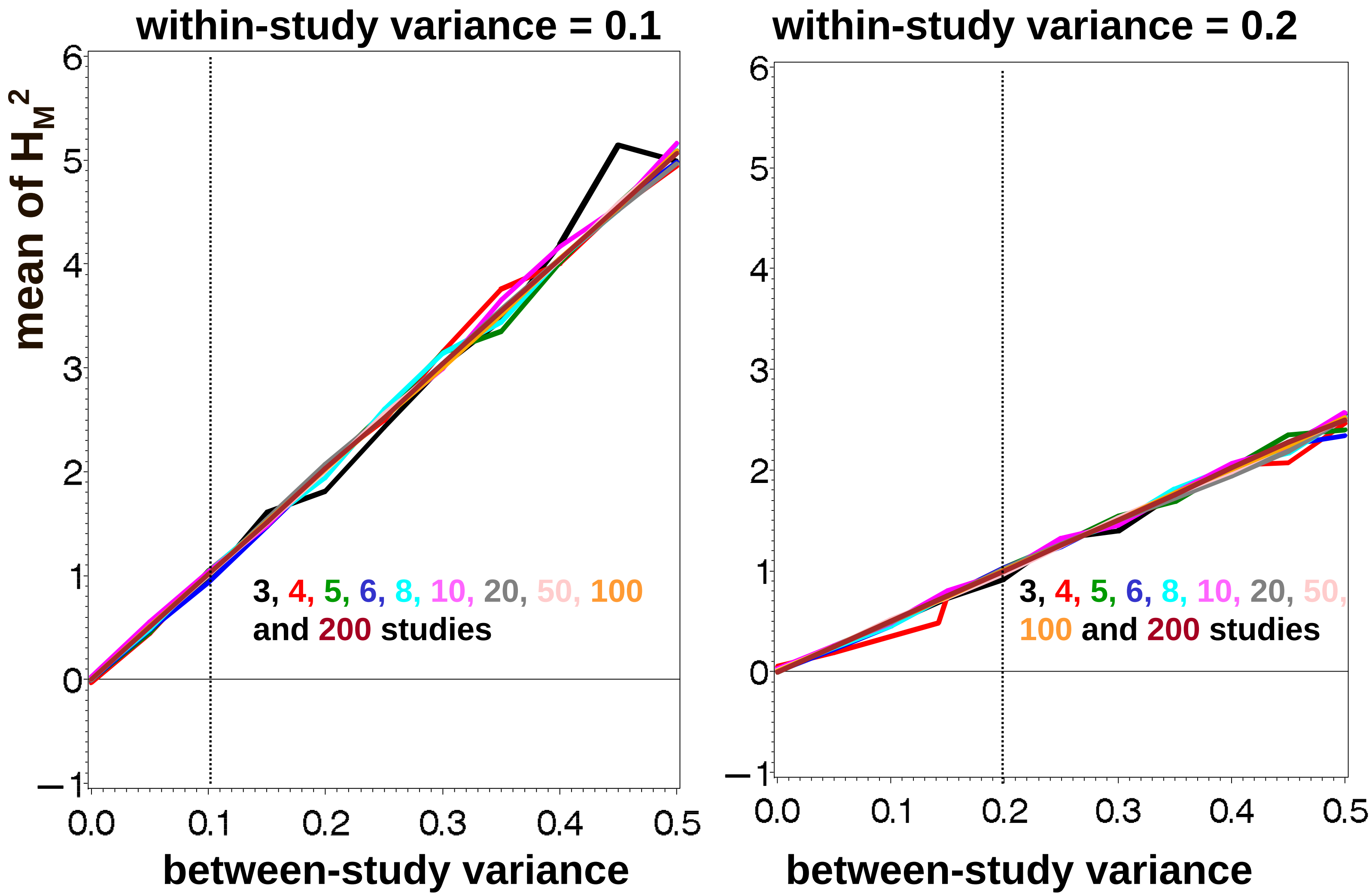
$$Q \leq df \qquad H^2 = 1$$

► **has values between 1 and ∞**

► $E\left(H^2\right) = \left(\sigma_B^2 + \sigma_{W,2}^2\right) / \sigma_{W,2}^2$

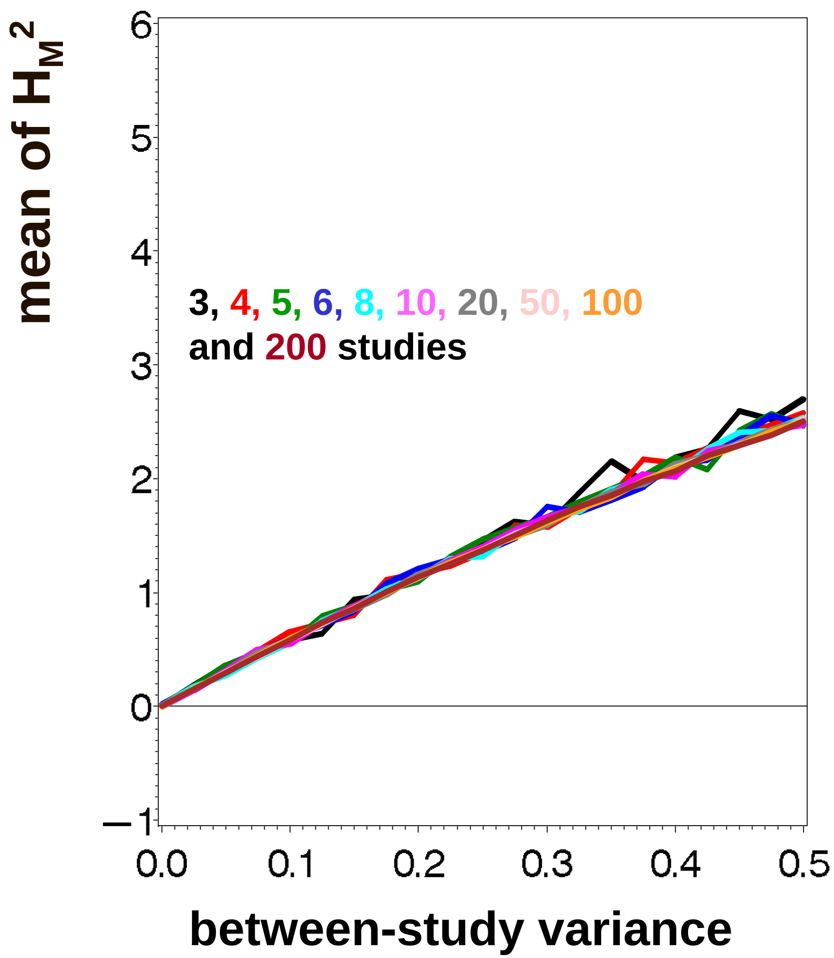
can be interpreted as ‚variance inflation factor‘ due to heterogeneity

Measure of for Heterogeneity H_M^2 (normal outc.)

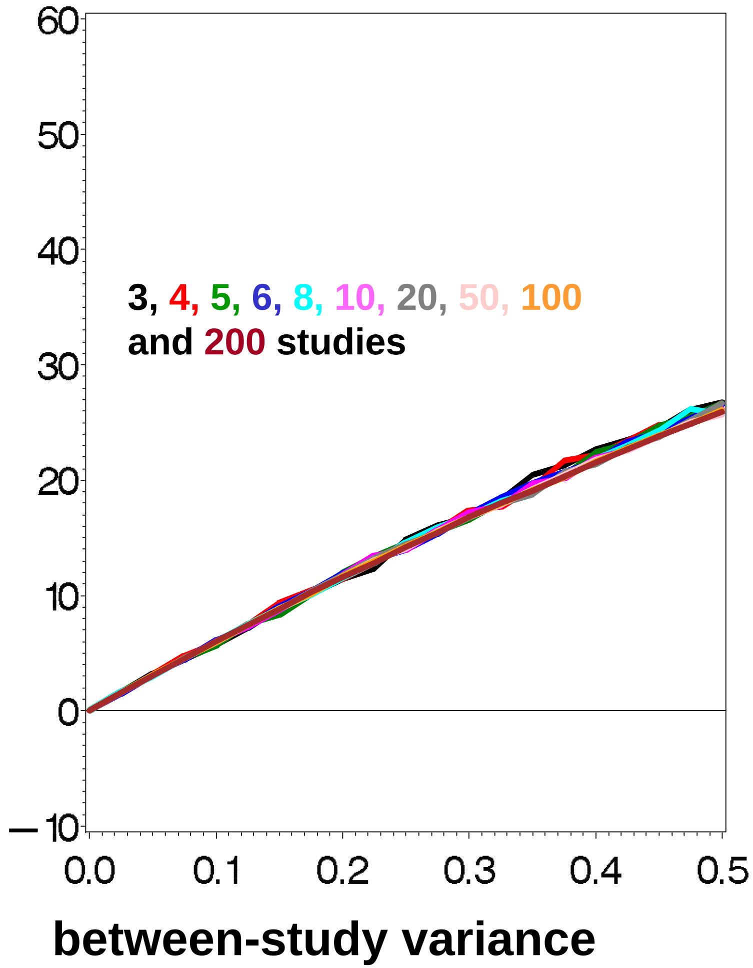


Measure of Heterogeneity H_M^2 (binary outcome)

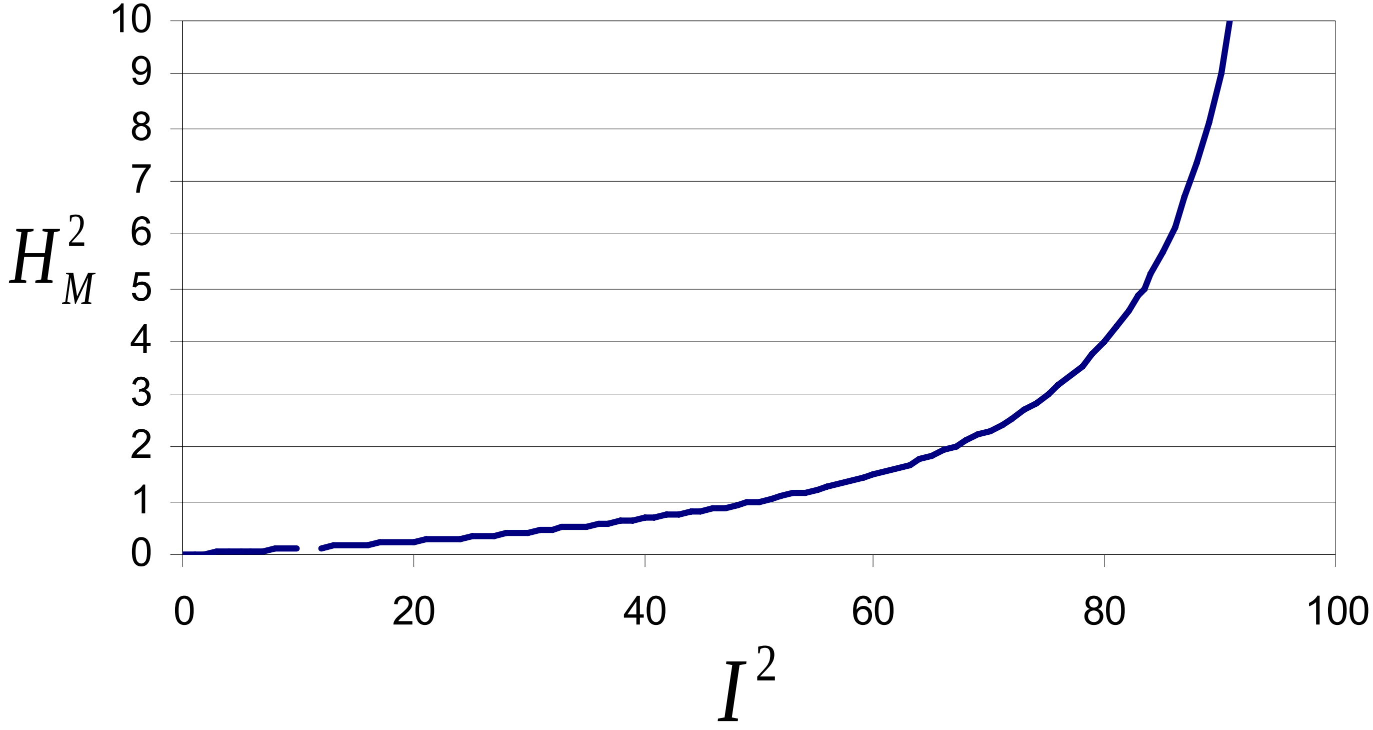
OR=1, n=50



OR=1, n=500



Relationship of the measures



Measures of Total Information

Hardy and Thompson (1998) have suggested a measue for total information, related to the power of heterogeneity test:

$$TI = \sum_i w_i = k / \hat{\sigma}^2_{W,1}$$

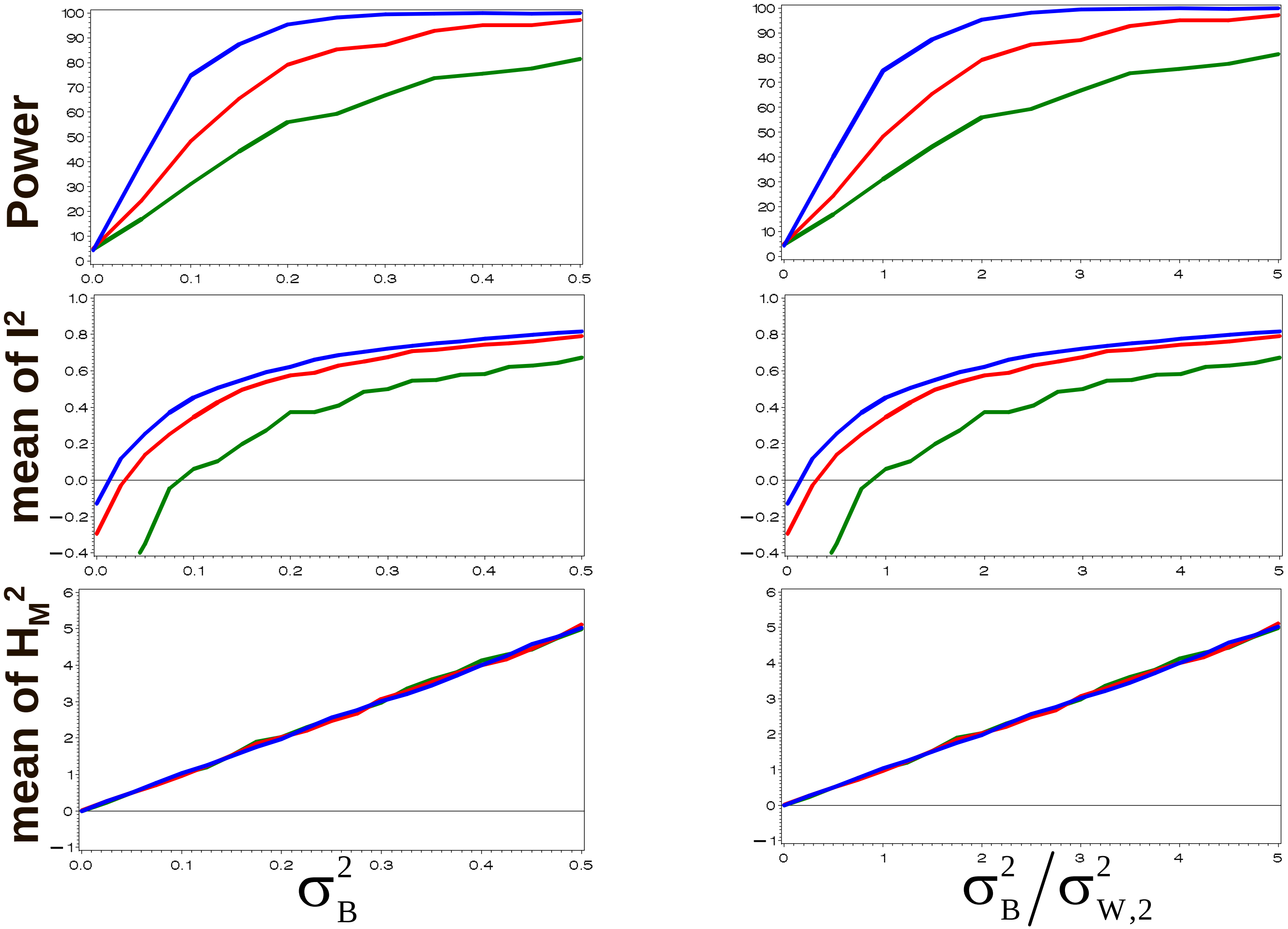
Another possibility is

$$TI_M = k / \hat{\sigma}^2_{W,2}$$

Test for Heterogeneity (Scenario 1)

		k	w_1	$w_2,\ldots,w_k$	$\sigma^2_{W,1}$	$\sigma^2_{W,2}$	TI	TI_M
Scenario 1	**(a)**	**5**	**10**	**10**	**0.1**	**0.1**	**50**	**50**
	(b)	**10**	**10**	**10**	**0.1**	**0.1**	**100**	**100**
	(c)	**20**	**10**	**10**	**0.1**	**0.1**	**200**	**200**

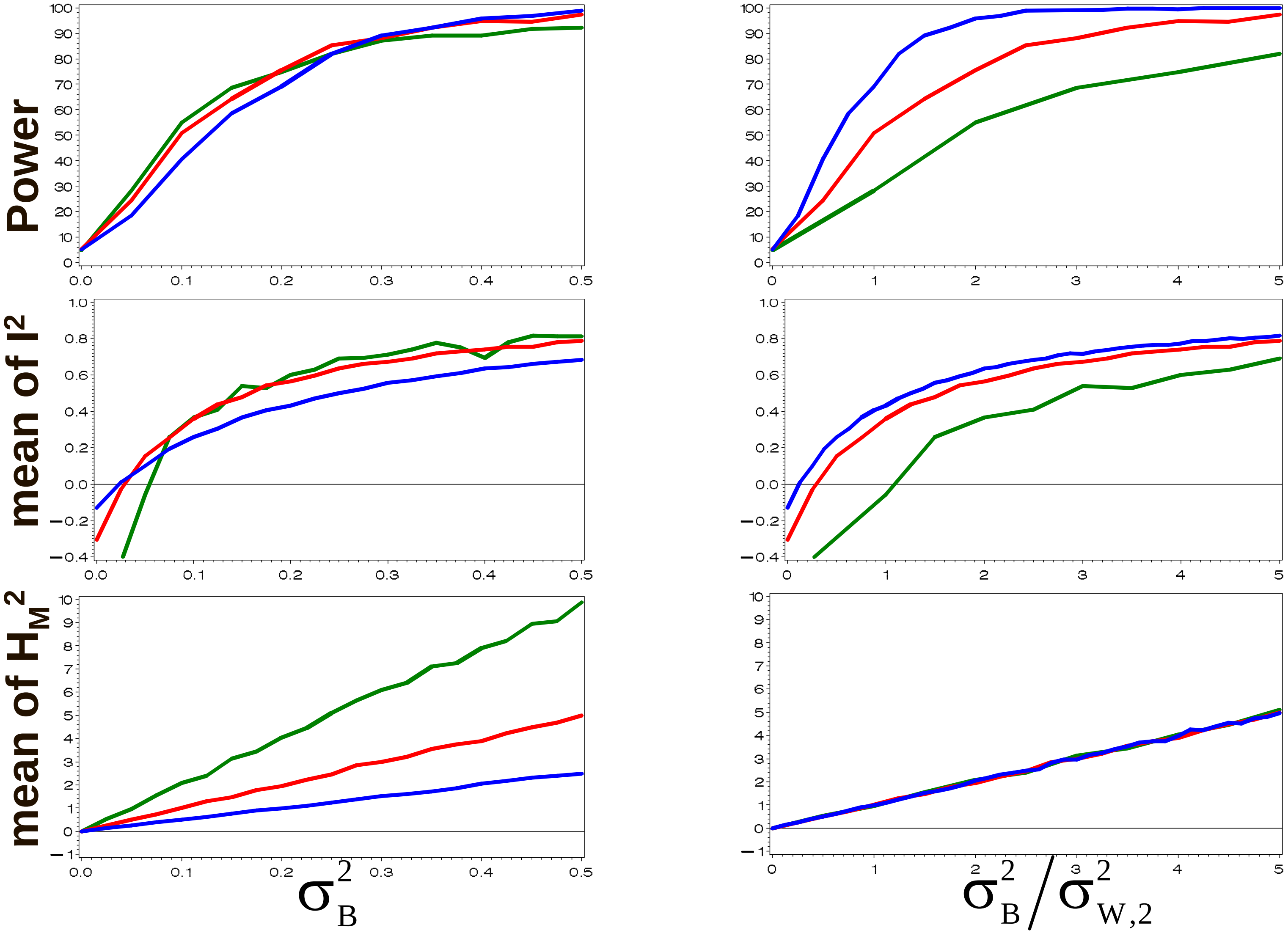
Test for Heterogeneity (Scenario 1)



Test for Heterogeneity (Scenario 1)

		k	w_1	$w_2,\ldots,w_k$	$\sigma^2_{W,1}$	$\sigma^2_{W,2}$	TI	TI_M
Scenario 1	**(a)**	**5**	**10**	**10**	**0.1**	**0.1**	**50**	**50**
	(b)	**10**	**10**	**10**	**0.1**	**0.1**	**100**	**100**
	(c)	**20**	**10**	**10**	**0.1**	**0.1**	**200**	**200**
Scenario 2	**(a)**	**5**	**20**	**20**	**0.05**	**0.05**	**100**	**100**
	(b)	**10**	**10**	**10**	**0.1**	**0.1**	**100**	**100**
	(c)	**20**	**5**	**5**	**0.2**	**0.2**	**100**	**100**

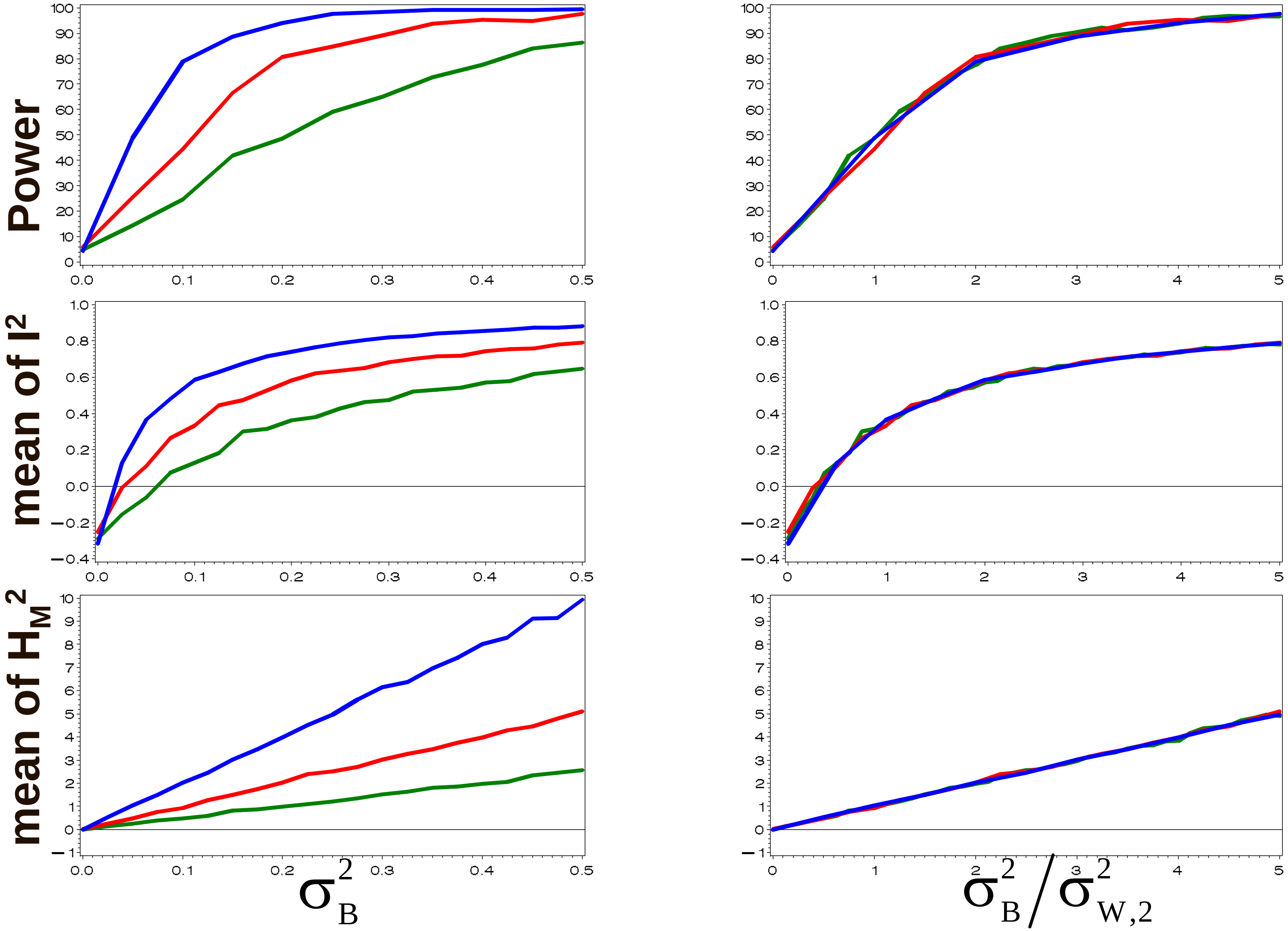
Test for Heterogeneity (Scenario 2)



Test for Heterogeneity (Scenario 1)

		k	w_1	$w_2, \ldots, w_k$	$\sigma^2_{W,1}$	$\sigma^2_{W,2}$	TI	TI_M
Scenario 1	**(a)**	**5**	**10**	**10**	**0.1**	**0.1**	**50**	**50**
	(b)	**10**	**10**	**10**	**0.1**	**0.1**	**100**	**100**
	(c)	**20**	**10**	**10**	**0.1**	**0.1**	**200**	**200**
Scenario 2	**(a)**	**5**	**20**	**20**	**0.05**	**0.05**	**100**	**100**
	(b)	**10**	**10**	**10**	**0.1**	**0.1**	**100**	**100**
	(c)	**20**	**5**	**5**	**0.2**	**0.2**	**100**	**100**
Scenario 3	**(a)**	**10**	**5**	**5**	**0.2**	**0.2**	**50**	**50**
	(b)	**10**	**10**	**10**	**0.1**	**0.1**	**100**	**100**
	(c)	**10**	**20**	**20**	**0.05**	**0.05**	**200**	**200**

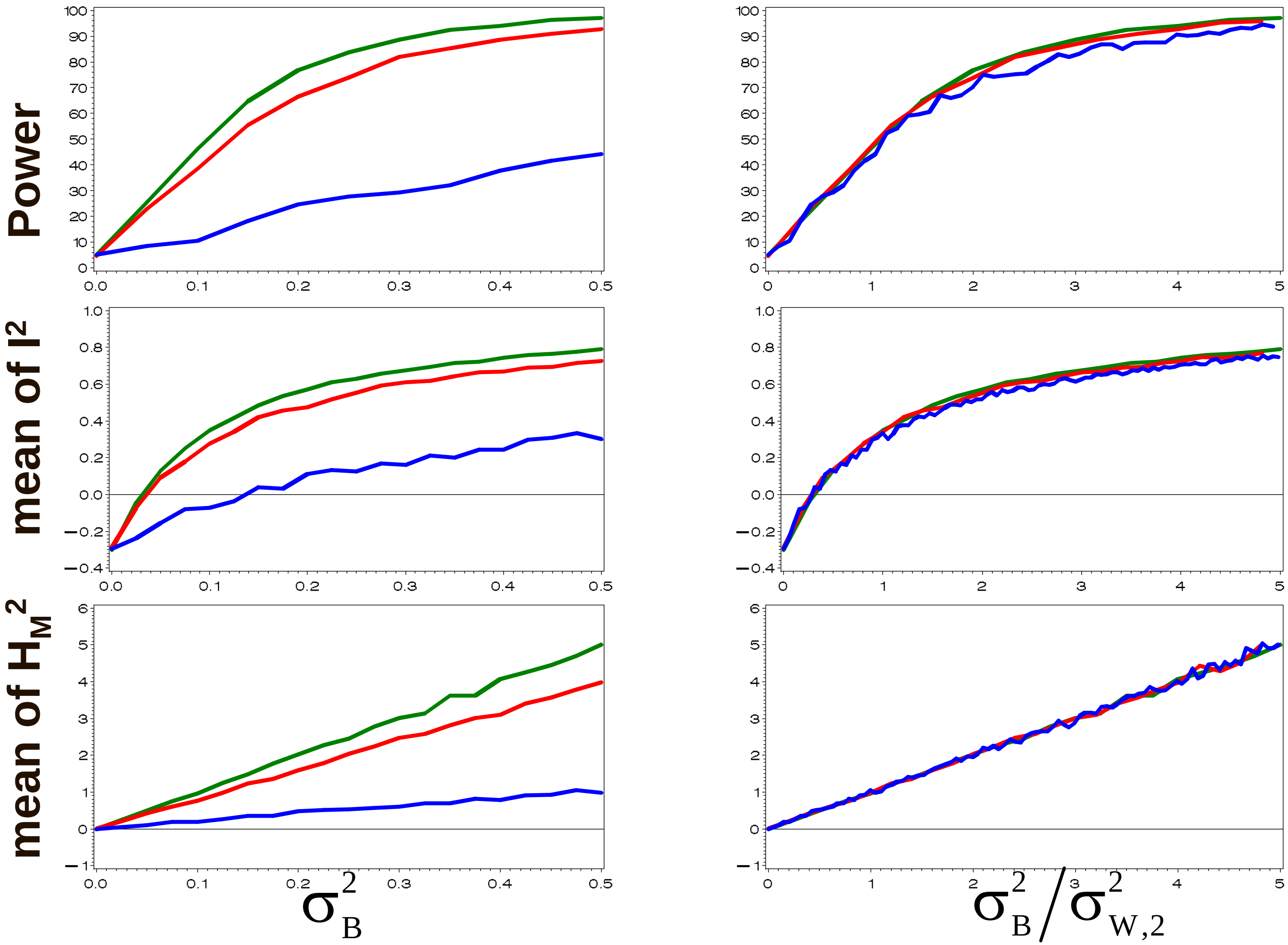
Test for Heterogeneity (Scenario 3)



Test for Heterogeneity (Scenario 1)

		k	w_1	$w_2,\ldots,w_k$	$\sigma^2_{W,1}$	$\sigma^2_{W,2}$	TI	TI_M
Scenario 1	(a)	5	10	10	0.1	0.1	50	50
	(b)	10	10	10	0.1	0.1	100	100
	(c)	20	10	10	0.1	0.1	200	200
Scenario 2	(a)	5	20	20	0.05	0.05	100	100
	(b)	10	10	10	0.1	0.1	100	100
	(c)	20	5	5	0.2	0.2	100	100
Scenario 3	(a)	10	5	5	0.2	0.2	50	50
	(b)	10	10	10	0.1	0.1	100	100
	(c)	10	20	20	0.05	0.05	200	200
Scenario 4	(a)	10	10	10	0.1	0.1	100	100
	(b)	10	50	5.56	0.1	0.12	100	80.2
	(c)	10	90	1.11	0.1	0.48	100	21.0

Test for Heterogeneity (Scenario 4)



Test for Heterogeneity (Scenario 1)

		k	w_1	$w_2,\ldots,w_k$	$\sigma^2_{W,1}$	$\sigma^2_{W,2}$	TI	TI_M
Scenario 1	(a)	5	10	10	0.1	0.1	50	50
	(b)	10	10	10	0.1	0.1	100	100
	(c)	20	10	10	0.1	0.1	200	200
Scenario 2	(a)	5	20	20	0.05	0.05	100	100
	(b)	10	10	10	0.1	0.1	100	100
	(c)	20	5	5	0.2	0.2	100	100
Scenario 3	(a)	10	5	5	0.2	0.2	50	50
	(b)	10	10	10	0.1	0.1	100	100
	(c)	10	20	20	0.05	0.05	200	200
Scenario 4	(a)	10	10	10	0.1	0.1	100	100
	(b)	10	50	5.56	0.1	0.12	100	80.2
	(c)	10	90	1.11	0.1	0.48	100	21.0

Overall Measure of Homogeneity:

$$Q > df \qquad I^2 = \frac{Q - df}{Q} \qquad H_M^2 = \frac{Q - df}{df}$$

$$Q = df \qquad I^2 = 0 \qquad H_M^2 = 0$$

$$Q < df \qquad H_M^2 = \frac{Q - df}{df}$$

extension for a measure for homogeneity, e.g. if studies are published more than once

Summary

- important to detect heterogeneity
- decision according to test or measure?
- I^2 and H_M^2 measure the amout of heterogeneity relative to total / within-study variance
- chance to investigate heterogeneity – better understanding of the data / therapy

References:

- Higgins J, Thompson S. Quantifying heterogeneity in a meta-analysis. Statistics in Medicine 2002; 21: 1539-1558.
- Higgins J, Thompson S, Deeks J, Altmann DG. Measuring inconsistency in meta-analyses. British Medical Journal 2003; 327: 557-560.
- Mittlböck M, Heinzl H. A simulation study comparing properties of heterogeneity measures in meta-analyses. Statistics in Medicine 2006; 25:4321-4333.