Cochrane Training

Webinar, 26.01.2023

IQWiG

Performing Meta-Analyses in the Case of Very Few Studies

Ralf Bender

IQWiG, Cologne, Germany

Outline

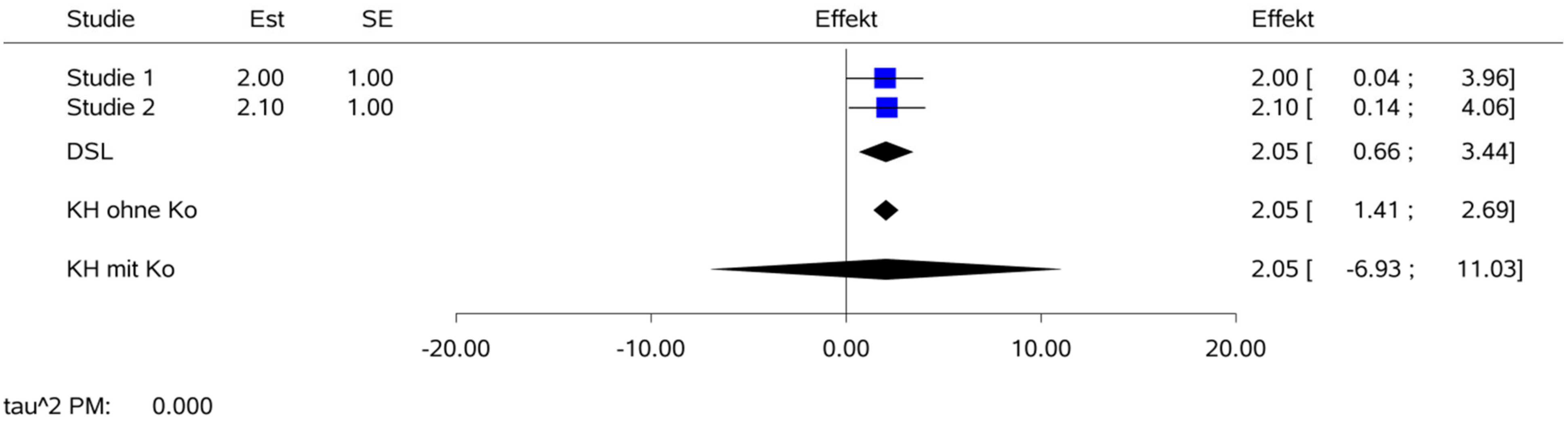
- Introduction
  - Models
  - Estimation methods
  - Qualitative summary of study results
- Meta-analysis with very few studies
  - Problems, examples
  - Qualitative summary of study results
  - Procedure, examples
- Discussion
- Outlook
  - Beta-binomial Model
  - Bayesian meta-analysis
- Summary
- Conclusion
- References

Poll 1: Continent

Poll 2: Affiliation

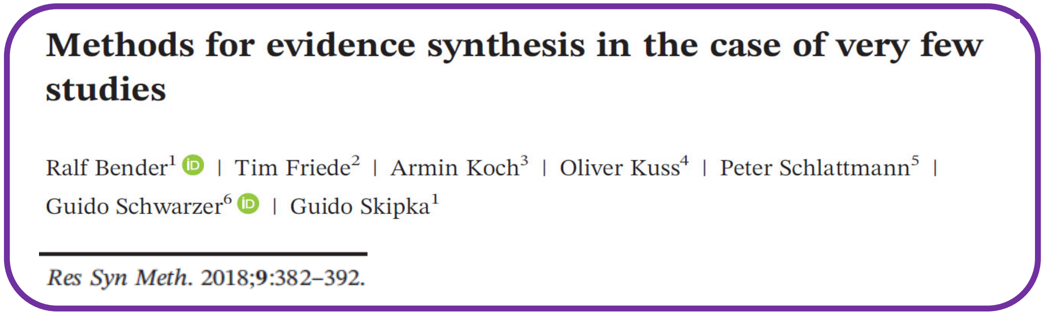
Topic for today:

Meta-analyses with very few studies



Studie	Est	SE	Effekt
Studie 1	2.00	1.00	2.00 [0.04 ; 3.96]
Studie 2	2.10	1.00	2.10 [0.14 ; 4.06]
DSL			2.05 [0.66 ; 3.44]
KH ohne Ko			2.05 [1.41 ; 2.69]
KH mit Ko			2.05 [-6.93 ; 11.03]

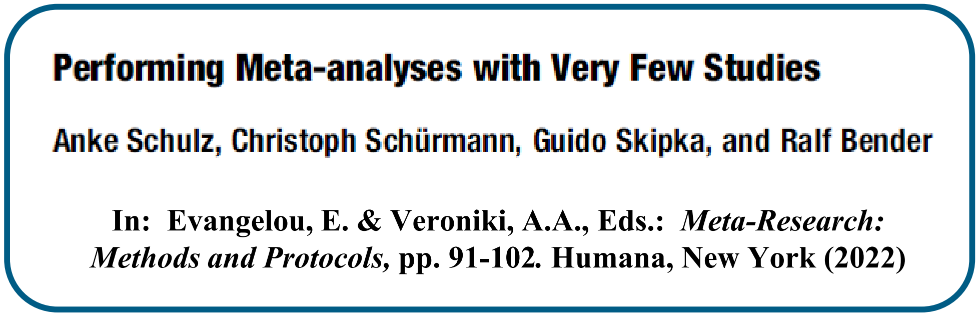
tau^2 PM: 0.000



Methods for evidence synthesis in the case of very few studies

Ralf Bender[1] | Tim Friede[2] | Armin Koch[3] | Oliver Kuss[4] | Peter Schlattmann[5] | Guido Schwarzer[6] | Guido Skipka[1]

Res Syn Meth. 2018;**9**:382–392.



Performing Meta-analyses with Very Few Studies

Anke Schulz, Christoph Schürmann, Guido Skipka, and Ralf Bender

In: Evangelou, E. & Veroniki, A.A., Eds.: *Meta-Research: Methods and Protocols,* pp. 91-102. Humana, New York (2022)

Introduction

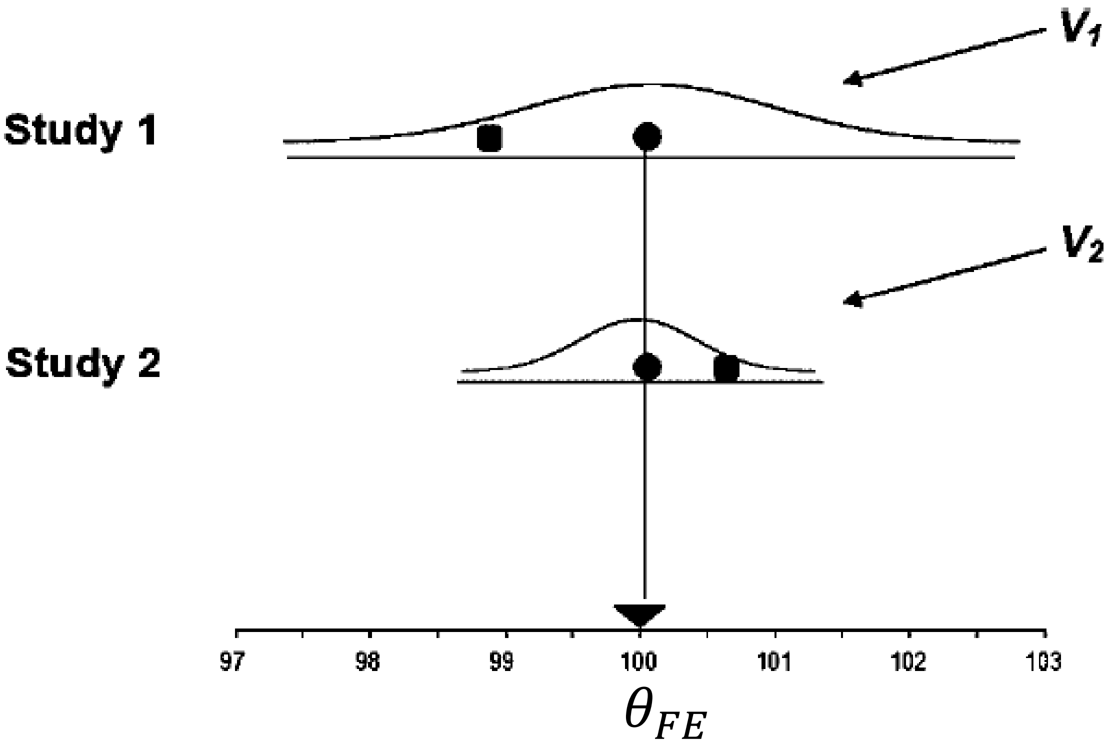
2 main meta-analytic models:

- Model with fixed effect (FEM)
  - Assumption:
    All studies estimate the same effect
  - Better term: *"Common-effect model"*
- Model with random effects (REM)
  - Assumption:
    The studies estimate different effects
  - For illustrating heterogeneity:
    Prediction intervals (PIs) are useful

Note: There are more models and approaches for meta-analysis. However, in practice, these do not play a major role (see Bender et al., *RSM* 2018).

Meta-analysis: FEM

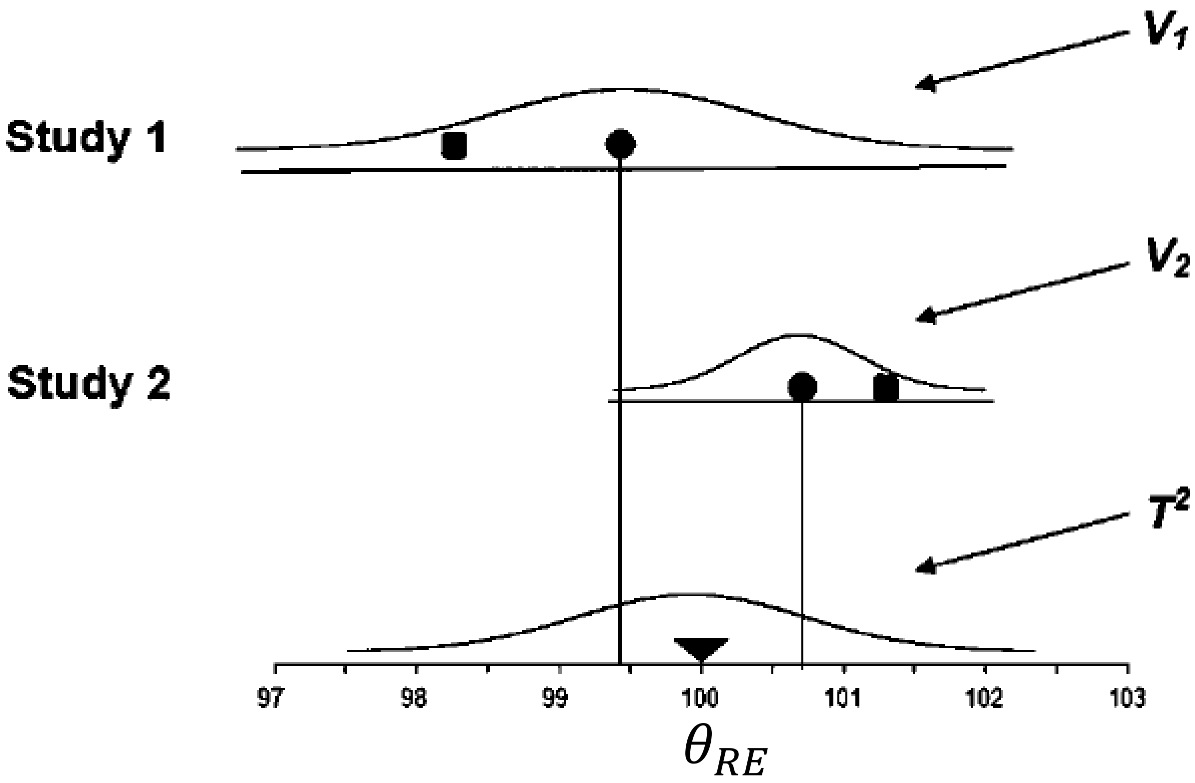
- $y_i = \theta_{FE} + \varepsilon_i \;,\;\; \varepsilon_i \sim N(0, v_i) \;,\;\; Var(y_i) = v_i$
- Assumption: All studies estimate the same effect.
- Parameter of interest: **Fixed effect $\boldsymbol{\theta_{FE}}$**



From: Borenstein et al. (2010): *RSM* **1**, 97-111.

Meta-analysis: REM

- $y_i = \theta_i + \varepsilon_i, \theta_i = \theta_{RE} + \delta_i, \ \varepsilon_i \sim N(0, v_i)\ , \delta_i \sim N(0, \tau^2), Var(y_i) = v_i + \tau^2$
- Assumption: Each study estimates a study-specific true effect.
- Parameter of interest: **Expected value $\boldsymbol{\theta_{RE}}$ of the effects**



From: Borenstein et al. (2010): *RSM* **1**, 97-111.

REM: Prediction interval

- Confidence interval (CI):
  - 95%-CI: $\hat{\theta}_{RE} \pm\ t_{k-1,1-\frac{\alpha}{2}} \times SE(\hat{\theta}_{RE})$
  - Range, which includes with high certainty (95%) the true effect of the meta-analysis

- Prediction interval (PI):
  - 95%-PI: $\hat{\theta}_{RE} \pm\ t_{k-1,1-\frac{\alpha}{2}} \times \sqrt{\tau^2 + Var(\hat{\theta}_{RE})}$
  - Range, which includes with high certainty (95%) the true effect of a single study
  - Graphical illustration of heterogeneity in the REM

Methods for estimation

FEM: Inverse variance (IV)

- Continuous data: Method of inverse variance (IV)
- Point estimate: $\hat{\theta}_{FE} = \frac{\sum_{i=1}^{k} y_i w_{i,FE}}{\sum_{i=1}^{k} w_{i,FE}}$, with $w_{i,FE} = 1/\hat{v}_i$
- 95% CI: $\hat{\theta}_{FE} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{1}{\sum_{i=1}^{k} w_{i,FE}}}$, z_q: *q*-quantile of the normal distribution

FEM: Mantel-Haenszel (MH)

- Binary data: Mantel-Haenszel (MH) method
- Estimation performed by means of the fourfold tables (dependent on effect measure)

Methods for estimation

REM: DerSimonian & Laird (DSL)

- Historically, the standard approach for RE meta-analysis: DSL method (DerSimonian & Laird, *CCT* 1986)
- Point estimation: $\hat{\theta}_{RE} = \frac{\sum_{i=1}^{k} y_i w_{i,RE}}{\sum_{i=1}^{k} w_{i,RE}}$ with $w_{i,RE} = 1/(\hat{v}_i + \hat{\tau}^2)$
- Point estimation of τ by means of the method of moments
- 95% CI: $\hat{\theta}_{RE} \pm \boxed{z_{1-\frac{\alpha}{2}}} \sqrt{\frac{1}{\sum_{i=1}^{k} w_{i,RE}}}$, z_q: $\underline{q\text{-quantile of normal distribution}}$
- DSL has been criticized for some time (Cornell et al., *AIM* 2014)
- DSL ignores the uncertainty of variance estimations
- CIs are frequently too narrow (in the case of few studies)

Methods for estimation

REM: Hartung-Knapp-Sidik-Jonkman (HKSJ)

- Recommended by the Cochrane Collaboration:
  HKSJ method (Veroniki et al., *RSM* 2019)
- Estimation: $\hat{\theta}_{RE} = \frac{\sum_{i=1}^{k} y_i w_{i,RE}}{\sum_{i=1}^{k} w_{i,RE}}$ with $w_{i,RE} = 1/(\hat{v}_i + \hat{\tau}^2)$
- Estimation of τ by means of Paule-Mandel method
- 95% CI: $\hat{\theta}_{RE} \pm \boxed{t_{k-1,1-\frac{\alpha}{2}}} \sqrt{\frac{\sum_{i=1}^{k} w_{i,RE}(y_i - \hat{\theta}_{RE})^2}{(k-1)\sum_{i=1}^{k} w_{i,RE}}}$, $t_{m,q}$: *q*-quantile of *t*-distribution
- HKSJ holds type 1 error
- CIs frequently very wide (especially in the case of few studies)
- $z_{0.975}$ =**1.96**, $t_{1;0.975}$ =**12.7**, $t_{2;0.975}$ =**4.3**, $t_{3;0.975}$ =**3.2**, $t_{4;0.975}$ =**2.8**

Methods for estimation

REM: Hartung-Knapp-Sidik-Jonkman (HKSJ)

- Problems in homogeneous data situations
- 95% CI: $\hat{\theta}_{RE} \pm\ t_{k-1,1-\frac{\alpha}{2}} \sqrt{\frac{\sum_{i=1}^{k} w_{i,RE} \enclose{circle}{(y_i - \hat{\theta}_{RE})}^2}{(k-1)\sum_{i=1}^{k} w_{i,RE}}}$
- SE may be arbitrarily too small and CI too narrow
- Ad-hoc variance correction (Knapp & Hartung, *Stat. Med.* 2003)
- $Var(\hat{\theta}_{RE}) = max\left[\frac{1}{\sum_{i=1}^{k} w_{i,RE}}, \frac{\sum_{i=1}^{k} w_{i,RE}(y_i - \hat{\theta}_{RE})^2}{(k-1)\sum_{i=1}^{k} w_{i,RE}}\right]$
- Procedure required for the decision whether the ad-hoc variance correction (VC) should be used or not

Qualitative summary of results

Concept of conclusive effects (IQWiG, 2022):

- Data situation, in which an effect can be derived although a meaningful pooled effect estimation is not possible
- No pooled effect estimation when:
  - Heterogeneity is too large
  - Data are insufficient to apply the desired model (REM)

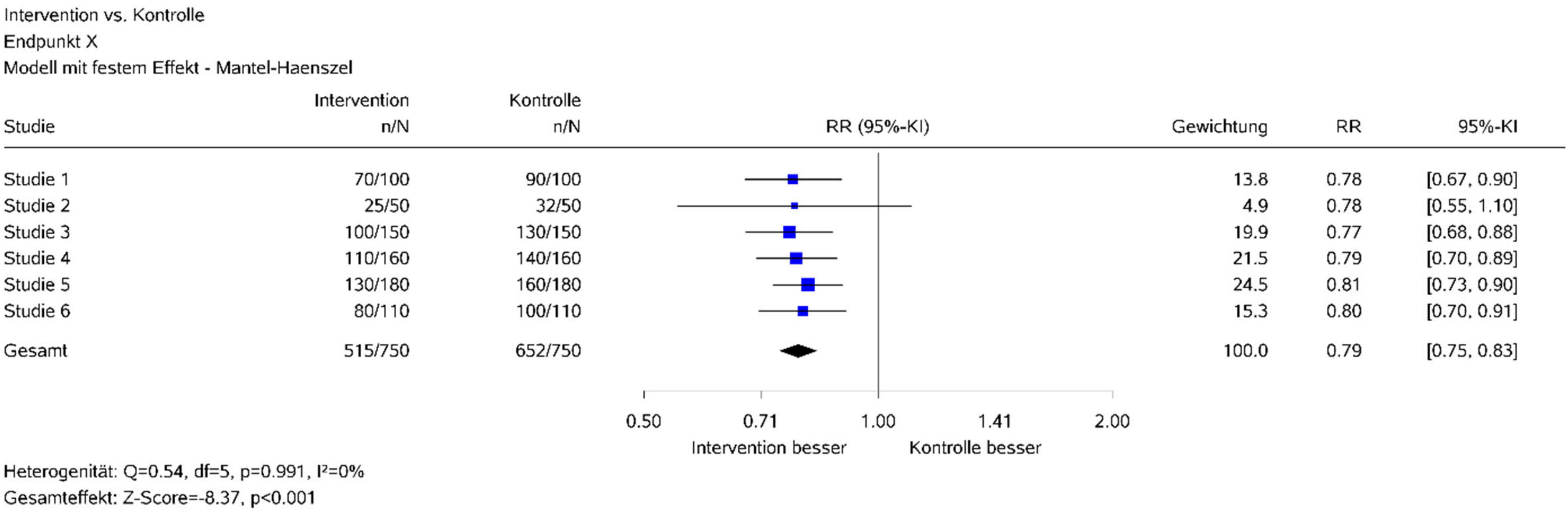
Qualitative summary of results

Concept of conclusive effects (IQWiG, 2022):

- 2 or more estimates are in the same direction
  - Total weight of these studies ≥ 80%
  - ≥ 2 studies are statistically significant
  - Weight of significant studies ≥ 50%
- Moderately and clearly conclusive effects
  - 2 or 3 studies significant ⇒ clearly
  - 2 studies significant, 1 study n.s. ⇒ moderately
  - Conclusive situation with 4 studies:
    all 4 studies significant ⇒ clearly
    Null $\notin$ prediction interval ⇒ clearly
    Null $\in$ prediction interval ⇒ moderately

General examples

Example 1: Clear data situation



Intervention vs. Kontrolle
Endpunkt X
Modell mit festem Effekt - Mantel-Haenszel

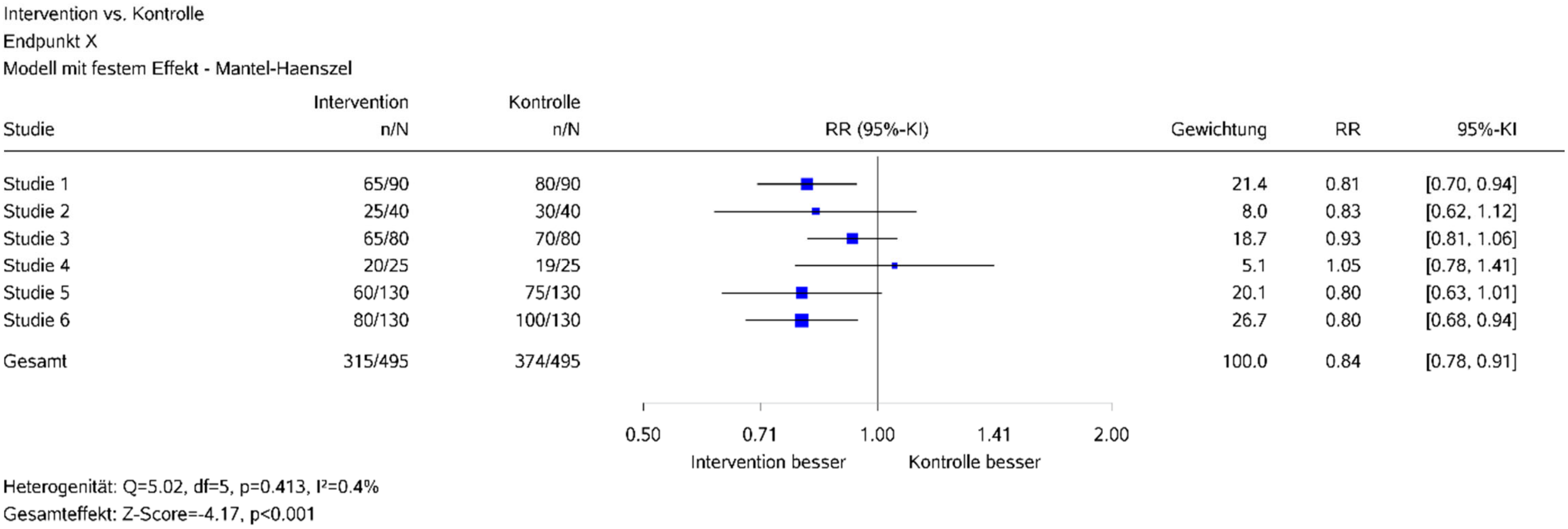
Studie	Intervention n/N	Kontrolle n/N	RR (95%-KI)	Gewichtung	RR	95%-KI
Studie 1	70/100	90/100		13.8	0.78	[0.67, 0.90]
Studie 2	25/50	32/50		4.9	0.78	[0.55, 1.10]
Studie 3	100/150	130/150		19.9	0.77	[0.68, 0.88]
Studie 4	110/160	140/160		21.5	0.79	[0.70, 0.89]
Studie 5	130/180	160/180		24.5	0.81	[0.73, 0.90]
Studie 6	80/110	100/110		15.3	0.80	[0.70, 0.91]
Gesamt	515/750	652/750		100.0	0.79	[0.75, 0.83]

Heterogenität: Q=0.54, df=5, p=0.991, I^2=0%
Gesamteffekt: Z-Score=-8.37, p<0.001

⇒ Proof of an intervention effect

General examples

Example 2: Less clear data situation



Intervention vs. Kontrolle
Endpunkt X
Modell mit festem Effekt - Mantel-Haenszel

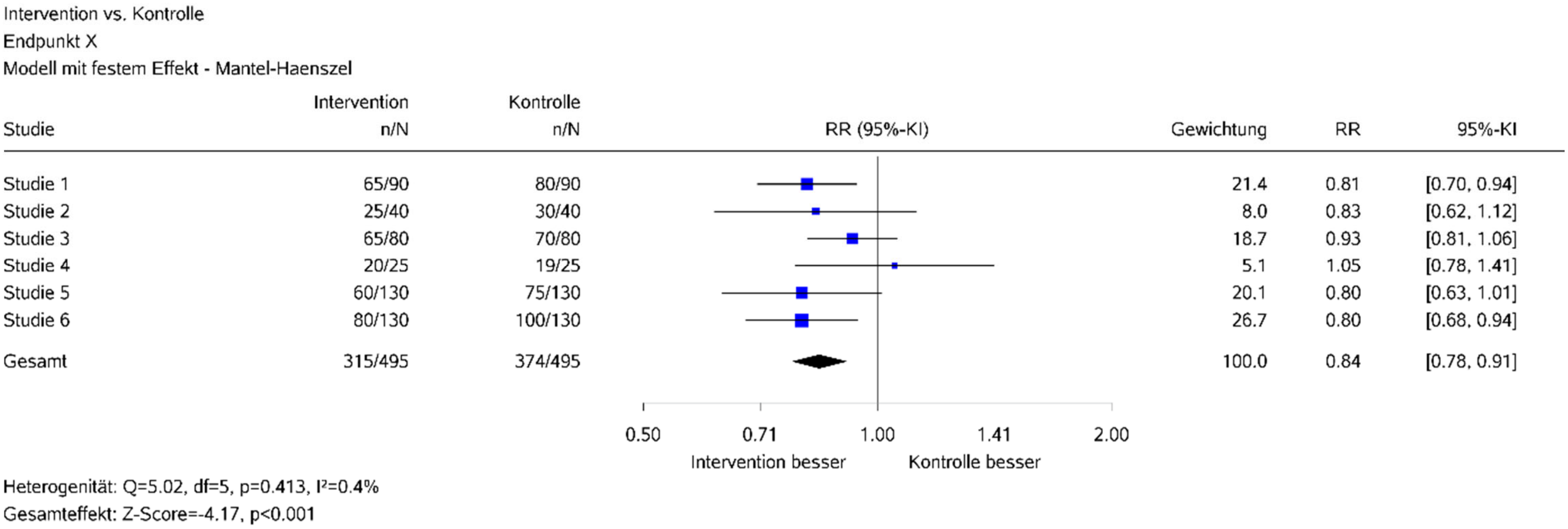
Studie	Intervention n/N	Kontrolle n/N	RR (95%-KI)	Gewichtung	RR	95%-KI
Studie 1	65/90	80/90		21.4	0.81	[0.70, 0.94]
Studie 2	25/40	30/40		8.0	0.83	[0.62, 1.12]
Studie 3	65/80	70/80		18.7	0.93	[0.81, 1.06]
Studie 4	20/25	19/25		5.1	1.05	[0.78, 1.41]
Studie 5	60/130	75/130		20.1	0.80	[0.63, 1.01]
Studie 6	80/130	100/130		26.7	0.80	[0.68, 0.94]
Gesamt	315/495	374/495		100.0	0.84	[0.78, 0.91]

Heterogenität: Q=5.02, df=5, p=0.413, I^2=0.4%
Gesamteffekt: Z-Score=-4.17, p<0.001

Poll 3: Significant effect?

General examples

Example 2: Less clear data situation



Intervention vs. Kontrolle
Endpunkt X
Modell mit festem Effekt - Mantel-Haenszel

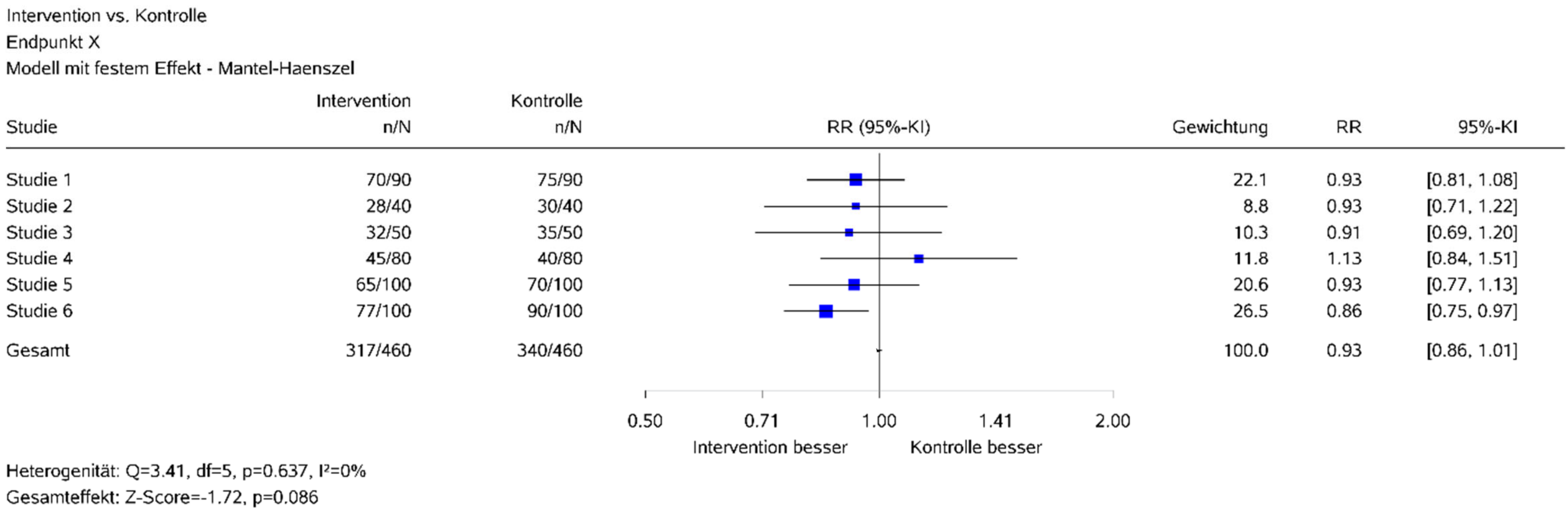
Studie	Intervention n/N	Kontrolle n/N	RR (95%-KI)	Gewichtung	RR	95%-KI
Studie 1	65/90	80/90		21.4	0.81	[0.70, 0.94]
Studie 2	25/40	30/40		8.0	0.83	[0.62, 1.12]
Studie 3	65/80	70/80		18.7	0.93	[0.81, 1.06]
Studie 4	20/25	19/25		5.1	1.05	[0.78, 1.41]
Studie 5	60/130	75/130		20.1	0.80	[0.63, 1.01]
Studie 6	80/130	100/130		26.7	0.80	[0.68, 0.94]
Gesamt	315/495	374/495		100.0	0.84	[0.78, 0.91]

Heterogenität: Q=5.02, df=5, p=0.413, I²=0.4%
Gesamteffekt: Z-Score=-4.17, p<0.001

⇒ Proof of an intervention effect

General examples

Example 3: Unclear data situation



Intervention vs. Kontrolle
Endpunkt X
Modell mit festem Effekt - Mantel-Haenszel

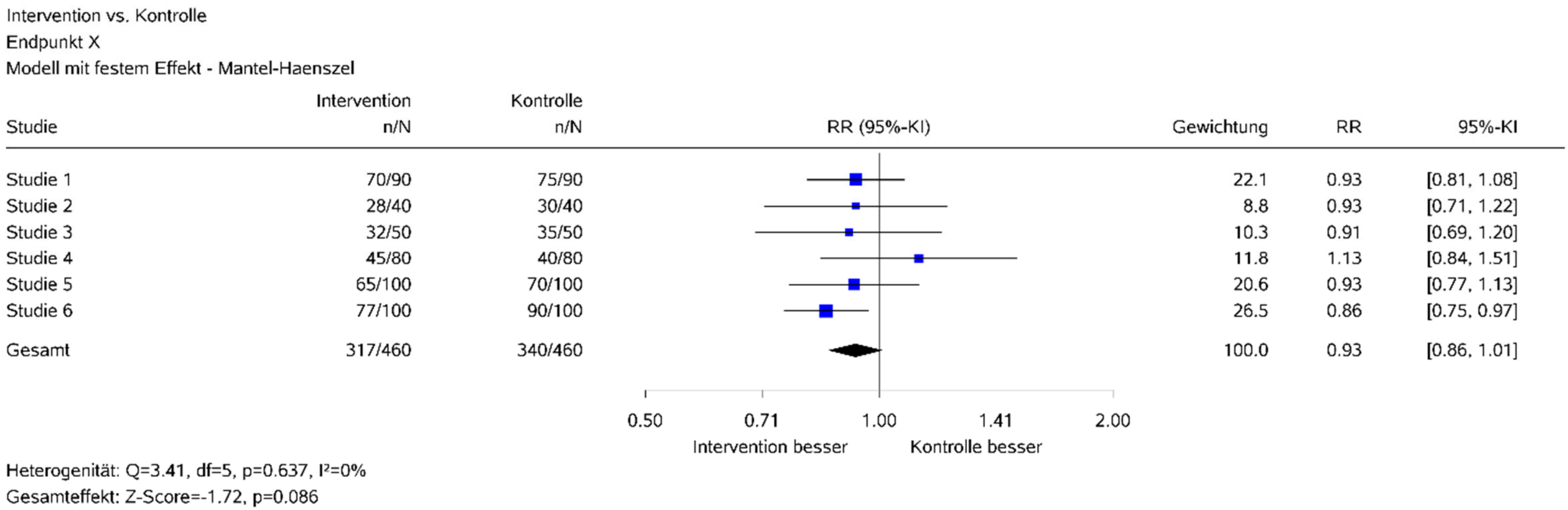
	Intervention	Kontrolle				
Studie	n/N	n/N	RR (95%-KI)	Gewichtung	RR	95%-KI
Studie 1	70/90	75/90		22.1	0.93	[0.81, 1.08]
Studie 2	28/40	30/40		8.8	0.93	[0.71, 1.22]
Studie 3	32/50	35/50		10.3	0.91	[0.69, 1.20]
Studie 4	45/80	40/80		11.8	1.13	[0.84, 1.51]
Studie 5	65/100	70/100		20.6	0.93	[0.77, 1.13]
Studie 6	77/100	90/100		26.5	0.86	[0.75, 0.97]
Gesamt	317/460	340/460		100.0	0.93	[0.86, 1.01]

Heterogenität: Q=3.41, df=5, p=0.637, I²=0%
Gesamteffekt: Z-Score=-1.72, p=0.086

Poll 4: Significant effect?

General examples

Example 3: Unclear data situation



Intervention vs. Kontrolle
Endpunkt X
Modell mit festem Effekt - Mantel-Haenszel

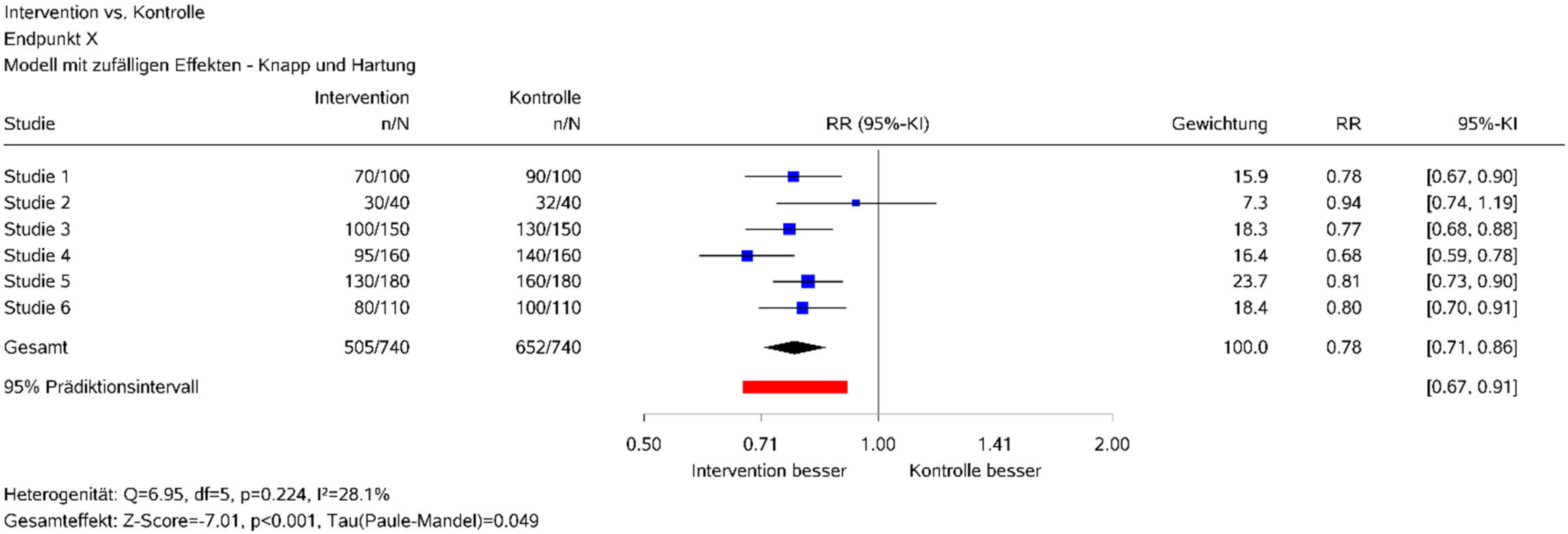
Studie	Intervention n/N	Kontrolle n/N	RR (95%-KI)	Gewichtung	RR	95%-KI
Studie 1	70/90	75/90		22.1	0.93	[0.81, 1.08]
Studie 2	28/40	30/40		8.8	0.93	[0.71, 1.22]
Studie 3	32/50	35/50		10.3	0.91	[0.69, 1.20]
Studie 4	45/80	40/80		11.8	1.13	[0.84, 1.51]
Studie 5	65/100	70/100		20.6	0.93	[0.77, 1.13]
Studie 6	77/100	90/100		26.5	0.86	[0.75, 0.97]
Gesamt	317/460	340/460		100.0	0.93	[0.86, 1.01]

Heterogenität: Q=3.41, df=5, p=0.637, I²=0%
Gesamteffekt: Z-Score=-1.72, p=0.086

⇒ No proof of an intervention effect

General examples

Example 4: REM in clear data situation



Intervention vs. Kontrolle
Endpunkt X
Modell mit zufälligen Effekten - Knapp und Hartung

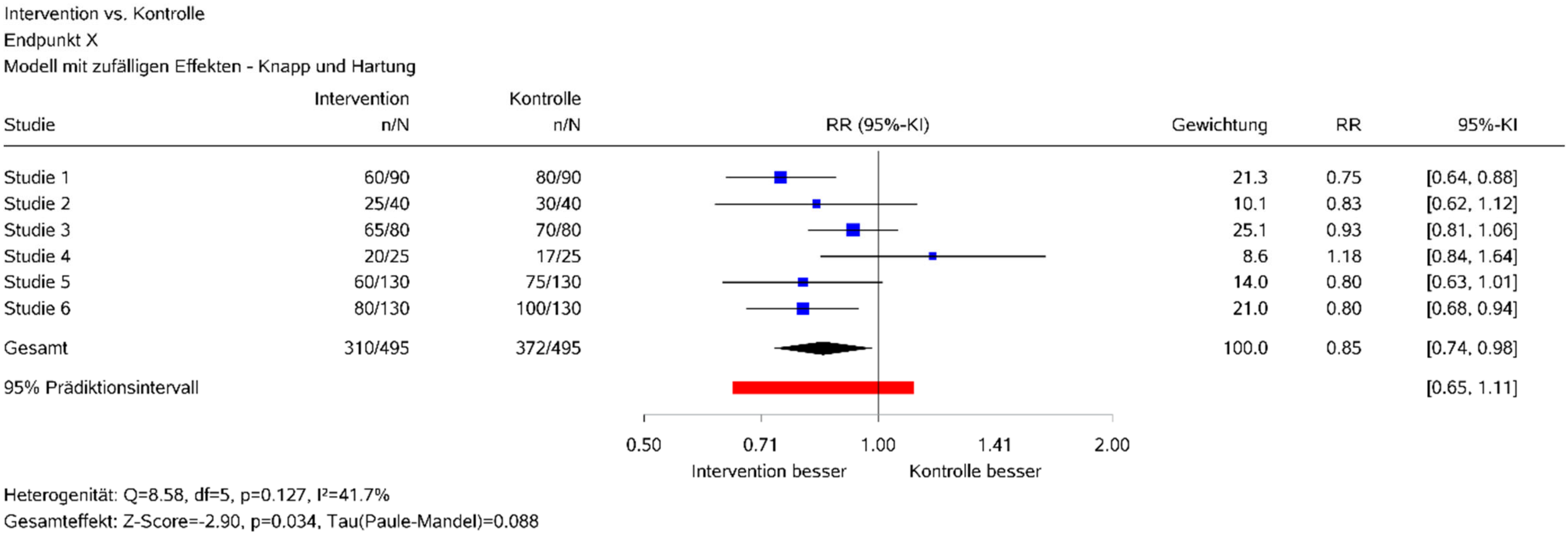
Studie	Intervention n/N	Kontrolle n/N	RR (95%-KI)	Gewichtung	RR	95%-KI
Studie 1	70/100	90/100		15.9	0.78	[0.67, 0.90]
Studie 2	30/40	32/40		7.3	0.94	[0.74, 1.19]
Studie 3	100/150	130/150		18.3	0.77	[0.68, 0.88]
Studie 4	95/160	140/160		16.4	0.68	[0.59, 0.78]
Studie 5	130/180	160/180		23.7	0.81	[0.73, 0.90]
Studie 6	80/110	100/110		18.4	0.80	[0.70, 0.91]
Gesamt	505/740	652/740		100.0	0.78	[0.71, 0.86]
95% Prädiktionsintervall						[0.67, 0.91]

Heterogenität: Q=6.95, df=5, p=0.224, I²=28.1%
Gesamteffekt: Z-Score=-7.01, p<0.001, Tau(Paule-Mandel)=0.049

⇒ Proof of an intervention effect

General examples

Example 5: REM in less clear data situation



Intervention vs. Kontrolle
Endpunkt X
Modell mit zufälligen Effekten - Knapp und Hartung

Studie	Intervention n/N	Kontrolle n/N	RR (95%-KI)	Gewichtung	RR	95%-KI
Studie 1	60/90	80/90		21.3	0.75	[0.64, 0.88]
Studie 2	25/40	30/40		10.1	0.83	[0.62, 1.12]
Studie 3	65/80	70/80		25.1	0.93	[0.81, 1.06]
Studie 4	20/25	17/25		8.6	1.18	[0.84, 1.64]
Studie 5	60/130	75/130		14.0	0.80	[0.63, 1.01]
Studie 6	80/130	100/130		21.0	0.80	[0.68, 0.94]
Gesamt	310/495	372/495		100.0	0.85	[0.74, 0.98]
95% Prädiktionsintervall						[0.65, 1.11]

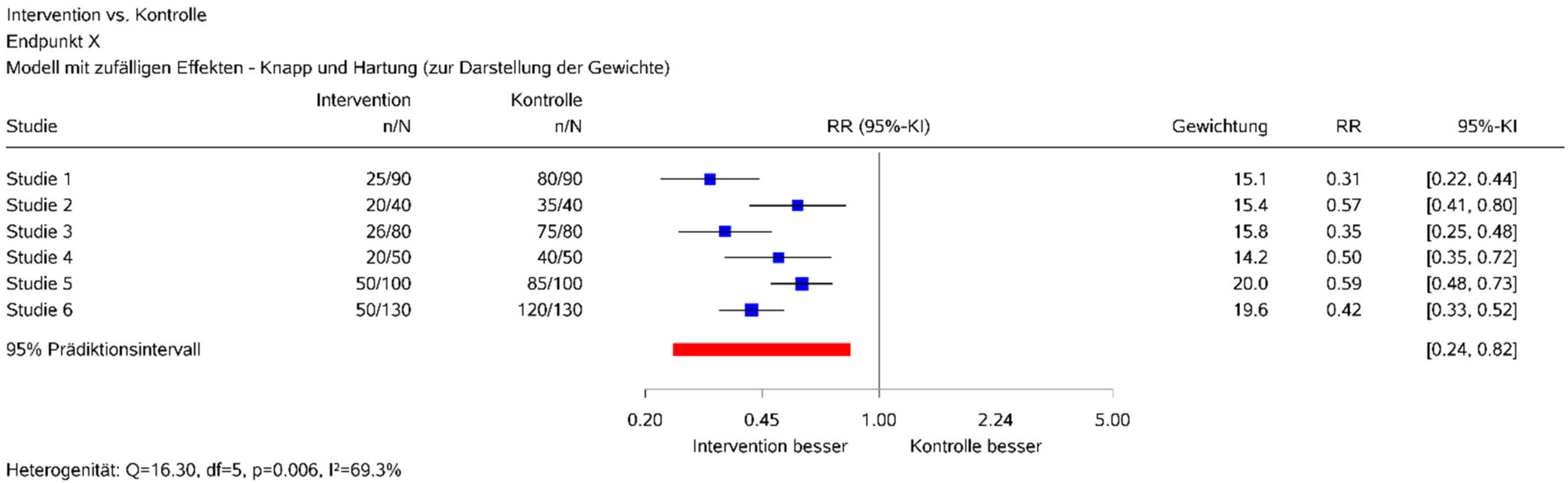
Heterogenität: Q=8.58, df=5, p=0.127, I²=41.7%
Gesamteffekt: Z-Score=-2.90, p=0.034, Tau(Paule-Mandel)=0.088

Provided there is sufficient certainty of the study results, the pooled effect estimate indicates **proof of an intervention effect** (on average!).

However, due to heterogeneity, study situations can be expected, in which the intervention has no effect.

General examples

Example 6: Clearly conclusive effects



Intervention vs. Kontrolle
Endpunkt X
Modell mit zufälligen Effekten - Knapp und Hartung (zur Darstellung der Gewichte)

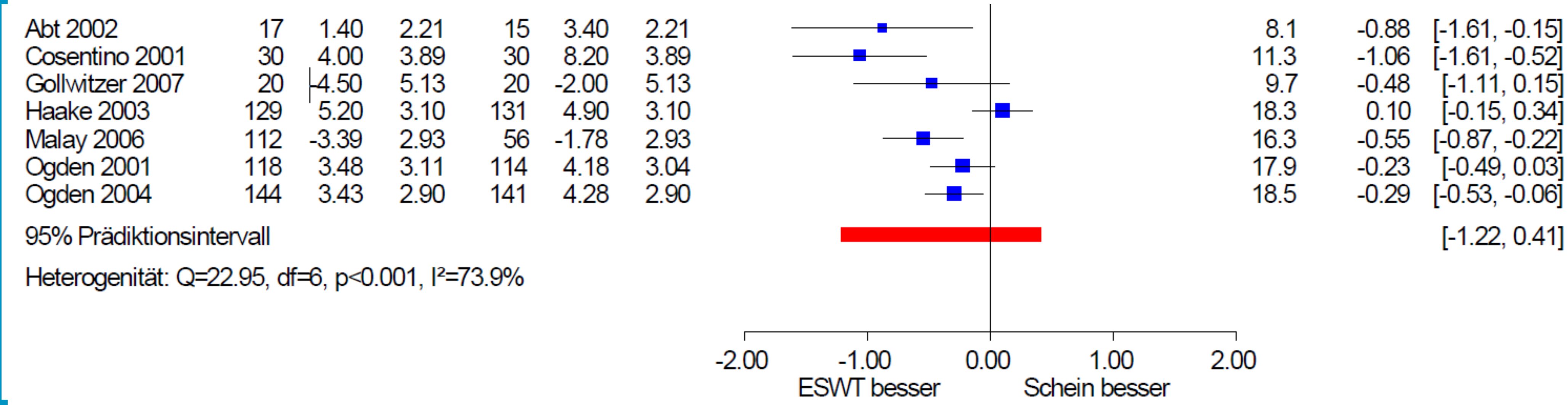
Studie	Intervention n/N	Kontrolle n/N	RR (95%-KI)	Gewichtung	RR	95%-KI
Studie 1	25/90	80/90		15.1	0.31	[0.22, 0.44]
Studie 2	20/40	35/40		15.4	0.57	[0.41, 0.80]
Studie 3	26/80	75/80		15.8	0.35	[0.25, 0.48]
Studie 4	20/50	40/50		14.2	0.50	[0.35, 0.72]
Studie 5	50/100	85/100		20.0	0.59	[0.48, 0.73]
Studie 6	50/130	120/130		19.6	0.42	[0.33, 0.52]
95% Prädiktionsintervall						[0.24, 0.82]

Heterogenität: Q=16.30, df=5, p=0.006, I²=69.3%

Provided there is sufficient certainty of the study results, the clearly conclusive effects indicate **proof of an intervention effect** (but with an unclear effect size).

General examples

Example 7: Moderately conclusive effects



Abt 2002	17	1.40	2.21	15	3.40	2.21	8.1	-0.88	[-1.61, -0.15]
Cosentino 2001	30	4.00	3.89	30	8.20	3.89	11.3	-1.06	[-1.61, -0.52]
Gollwitzer 2007	20	-4.50	5.13	20	-2.00	5.13	9.7	-0.48	[-1.11, 0.15]
Haake 2003	129	5.20	3.10	131	4.90	3.10	18.3	0.10	[-0.15, 0.34]
Malay 2006	112	-3.39	2.93	56	-1.78	2.93	16.3	-0.55	[-0.87, -0.22]
Ogden 2001	118	3.48	3.11	114	4.18	3.04	17.9	-0.23	[-0.49, 0.03]
Ogden 2004	144	3.43	2.90	141	4.28	2.90	18.5	-0.29	[-0.53, -0.06]
95% Prädiktionsintervall									[-1.22, 0.41]

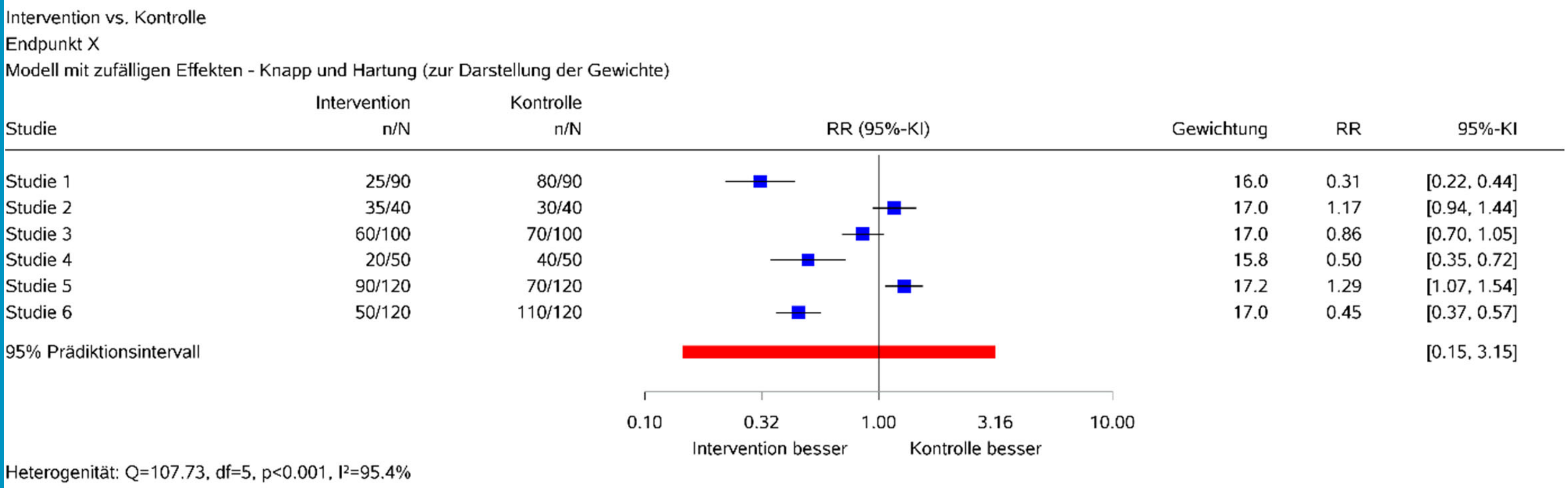
Heterogenität: Q=22.95, df=6, p<0.001, I²=73.9%

The decision, whether the intervention is beneficial depends on the certainty of the study results.

(RCTs with low risk of bias or non-RCTs with high or unclear risk of bias?)

General examples

Example 8: No conclusive effects



Intervention vs. Kontrolle
Endpunkt X
Modell mit zufälligen Effekten - Knapp und Hartung (zur Darstellung der Gewichte)

	Intervention	Kontrolle				
Studie	n/N	n/N	RR (95%-KI)	Gewichtung	RR	95%-KI
Studie 1	25/90	80/90		16.0	0.31	[0.22, 0.44]
Studie 2	35/40	30/40		17.0	1.17	[0.94, 1.44]
Studie 3	60/100	70/100		17.0	0.86	[0.70, 1.05]
Studie 4	20/50	40/50		15.8	0.50	[0.35, 0.72]
Studie 5	90/120	70/120		17.2	1.29	[1.07, 1.54]
Studie 6	50/120	110/120		17.0	0.45	[0.37, 0.57]
95% Prädiktionsintervall						[0.15, 3.15]

Heterogenität: Q=107.73, df=5, p<0.001, I²=95.4%

⇒ No proof of an intervention effect

Very few studies (k<5)

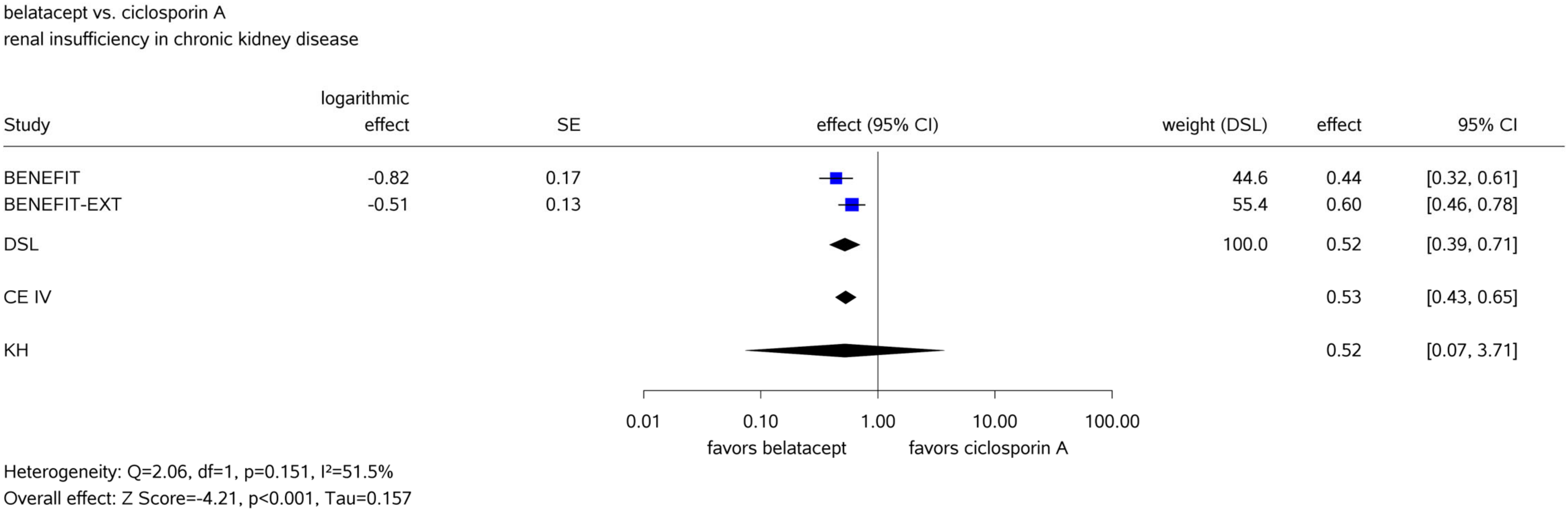
Problems with meta-analyses with very few studies (Bender et al., 2018):

- Choice between FEM and REM difficult
- τ cannot be adequately estimated
- DSL-CIs are too narrow
- HKSJ-CIs are wide or even non-informative
- In homogeneous data situations HKSJ-CIs are sometimes too narrow

Example: IQWiG Report A15-25

Belatacept after kidney transplant (2 significant studies)

- Belatacept vs ciclosporin A for prophylaxis of graft rejection in adults receiving a renal transplant
- Endpoint "renal insufficiency in chronic kidney disease stage 4/5"



belatacept vs. ciclosporin A
renal insufficiency in chronic kidney disease

Study	logarithmic effect	SE	effect (95% CI)	weight (DSL)	effect	95% CI
BENEFIT	-0.82	0.17		44.6	0.44	[0.32, 0.61]
BENEFIT-EXT	-0.51	0.13		55.4	0.60	[0.46, 0.78]
DSL				100.0	0.52	[0.39, 0.71]
CE IV					0.53	[0.43, 0.65]
KH					0.52	[0.07, 3.71]

Heterogeneity: Q=2.06, df=1, p=0.151, I²=51.5%
Overall effect: Z Score=-4.21, p<0.001, Tau=0.157

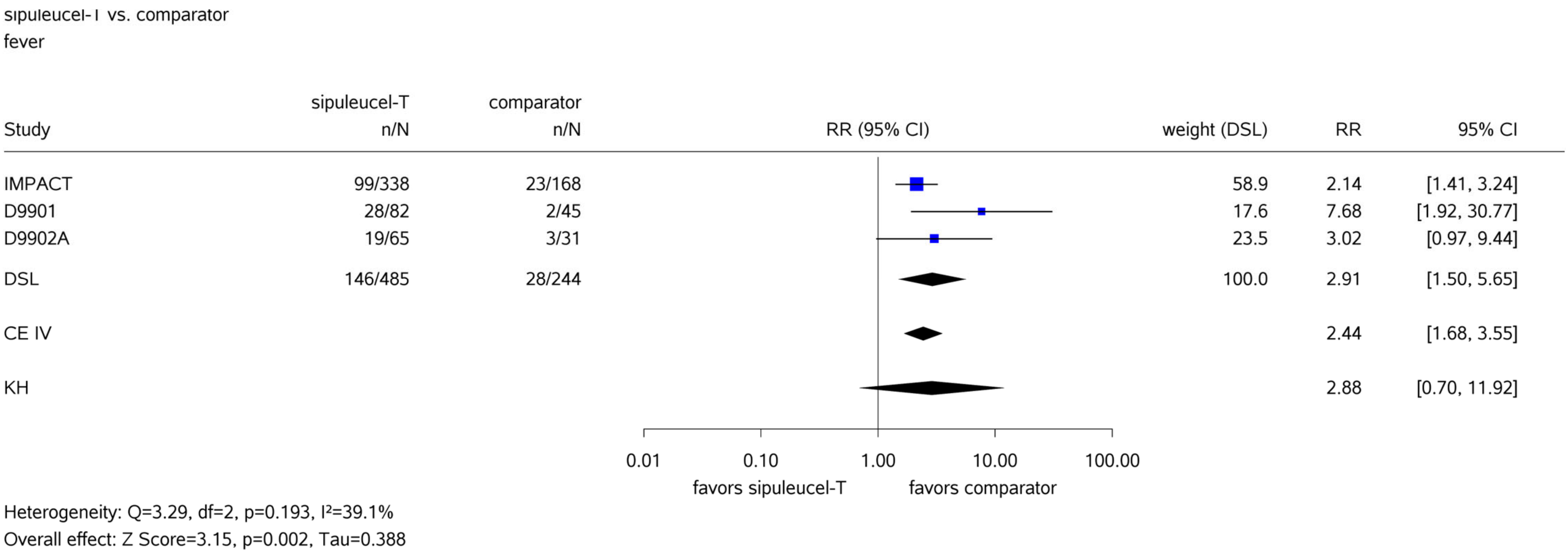
→

1) HKSJ over-conservative
2) Decision of no added benefit would be critical

Example: IQWiG Report A14-38

Sipuleucel-T in prostate cancer (3 significant studies)

- Sipuleucel-T vs appropriate comparator for asymptomatic or minimally symptomatic metastatic prostate cancer in males
- Endpoint fever



sipuleucel-T vs. comparator
fever

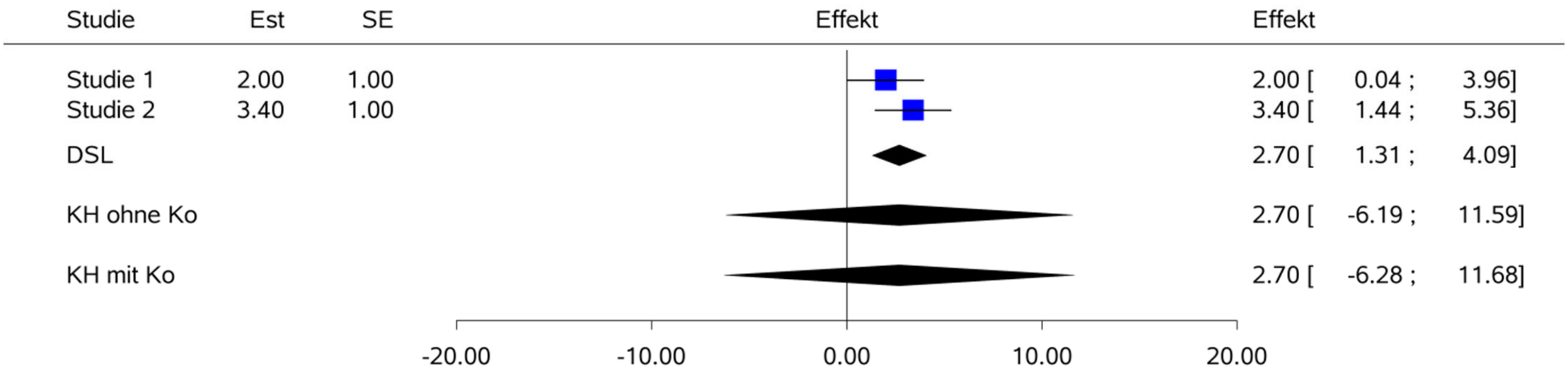
Study	sipuleucel-T n/N	comparator n/N	RR (95% CI)	weight (DSL)	RR	95% CI
IMPACT	99/338	23/168		58.9	2.14	[1.41, 3.24]
D9901	28/82	2/45		17.6	7.68	[1.92, 30.77]
D9902A	19/65	3/31		23.5	3.02	[0.97, 9.44]
DSL	146/485	28/244		100.0	2.91	[1.50, 5.65]
CE IV					2.44	[1.68, 3.55]
KH					2.88	[0.70, 11.92]

Heterogeneity: Q=3.29, df=2, p=0.193, I²=39.1%
Overall effect: Z Score=3.15, p=0.002, Tau=0.388

→ Even in the case of 3 studies HKSJ method over-conservative

Artificial examples

Ad-hoc variance correction (VC) for HKSJ



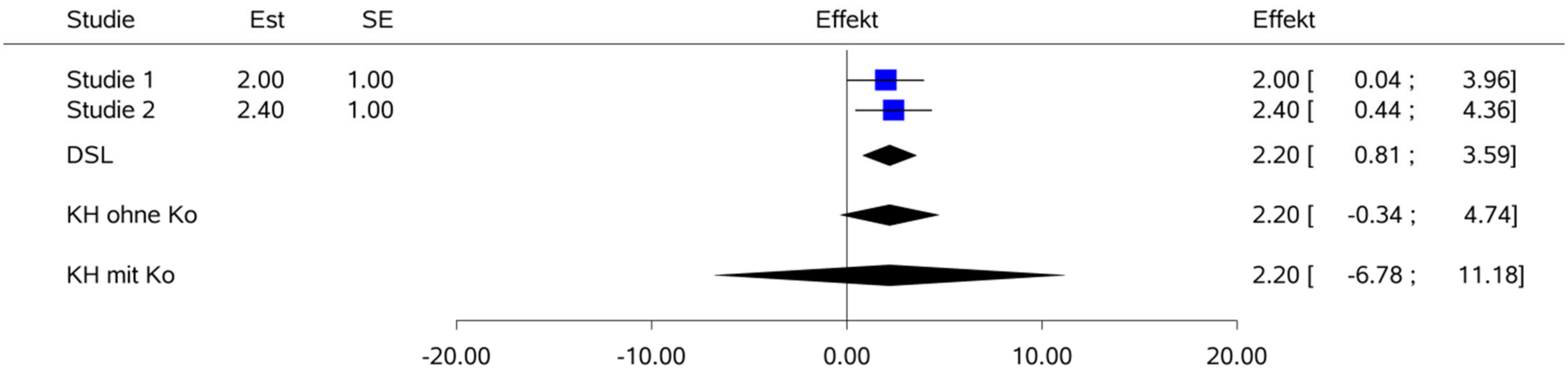
Studie	Est	SE	Effekt
Studie 1	2.00	1.00	2.00 [0.04 ; 3.96]
Studie 2	3.40	1.00	3.40 [1.44 ; 5.36]
DSL			2.70 [1.31 ; 4.09]
KH ohne Ko			2.70 [-6.19 ; 11.59]
KH mit Ko			2.70 [-6.28 ; 11.68]

tau^2 PM: 0.000

→ HKSJ over-conservative
Ad-hoc VC not required

Artificial examples

Ad-hoc variance correction (VC) for HKSJ



Studie	Est	SE	Effekt
Studie 1	2.00	1.00	2.00 [0.04 ; 3.96]
Studie 2	2.40	1.00	2.40 [0.44 ; 4.36]
DSL			2.20 [0.81 ; 3.59]
KH ohne Ko			2.20 [-0.34 ; 4.74]
KH mit Ko			2.20 [-6.78 ; 11.18]

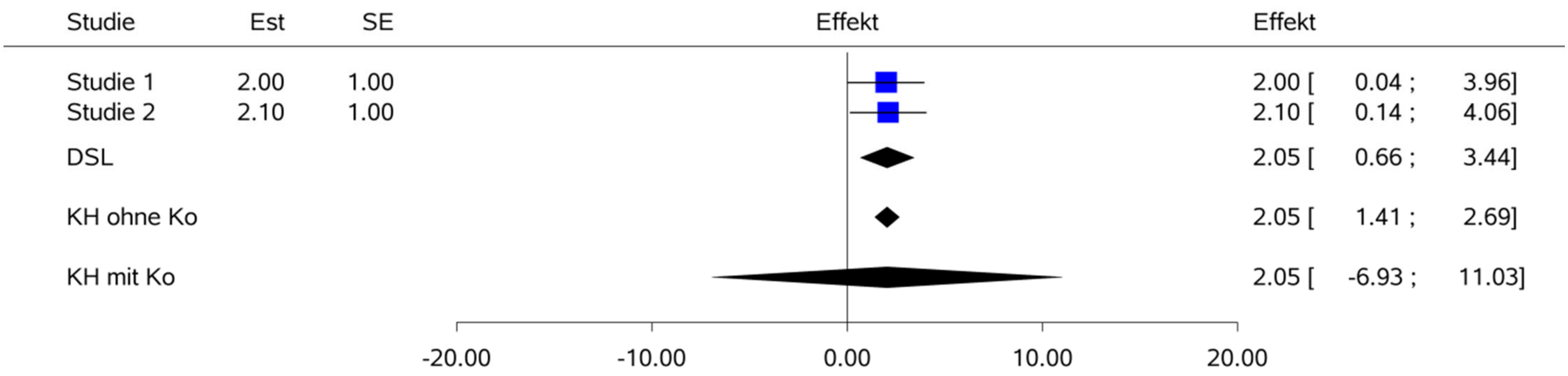
tau^2 PM: 0.000

→ HKSJ CI-width decreases with increasing homogeneity

Is the use of ad-hoc VC required?

Artificial examples

Ad-hoc variance correction (VC) for HKSJ



Studie	Est	SE	Effekt
Studie 1	2.00	1.00	2.00 [0.04 ; 3.96]
Studie 2	2.10	1.00	2.10 [0.14 ; 4.06]
DSL			2.05 [0.66 ; 3.44]
KH ohne Ko			2.05 [1.41 ; 2.69]
KH mit Ko			2.05 [-6.93 ; 11.03]

tau^2 PM: 0.000

→ HKSJ-CI clearly too narrow
Variance correction required, but over-conservative

→ Comparison with DSL to decide whether
ad-hoc VC should be used (Schulz et al., 2022)

Procedure in the case of very few studies

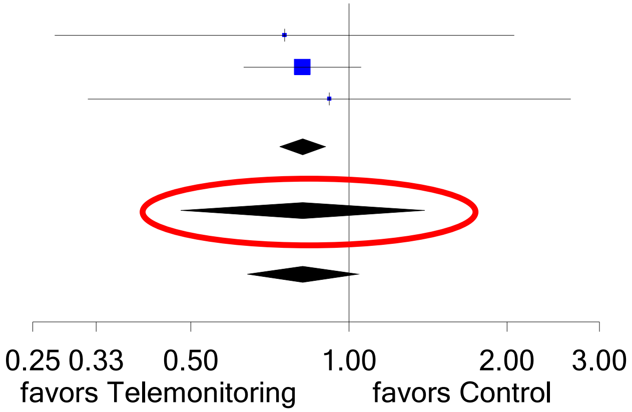
- Step 1: Preliminary model choice
  - PICOS framework
  - In general: RE model
  - 2 studies: FE model (studies with identical design)
- Step 2: Evaluation of heterogeneity
  - Too large, unexplained heterogeneity: MA not useful
  - Q-Test, I^2, visual inspection of forest plot
  - If this is the case: Qualitative summary (QS)
- Step 3: Final model and method choice
  - Strong heterogeneity: Reconsider preliminary choice
  - FE model: IV (continuous) or MH (binary)
  - RE model: HKSJ (if required VC) or QS (comparison with DSL and comparison with QS)

Example: IQWiG Report N16-02

- Use of ad-hoc VC required?
  - Comparison of CIs from DSL and HKSJ
  - HKSJ-CI narrower than DSL-CI ⇒ Use VC

Telemonitoring vs. Control
Mortality

Study	Telemonitoring n/N	Control n/N	OR (95% CI)	weight	OR	95% CI
REDUCEhf	7/202	9/198		5.8	0.75	[0.28, 2.07]
REM-HF	128/824	152/826		88.9	0.82	[0.63, 1.06]
TELECART	7/89	8/94		5.3	0.92	[0.32, 2.64]
REM - HKSJ	142/1115	169/1118		100.0	0.82	[0.74, 0.90]
REM - HKSJ (variance corr.)					0.82	[0.48, 1.39]
REM - DerSimonian-Laird					0.82	[0.64, 1.04]

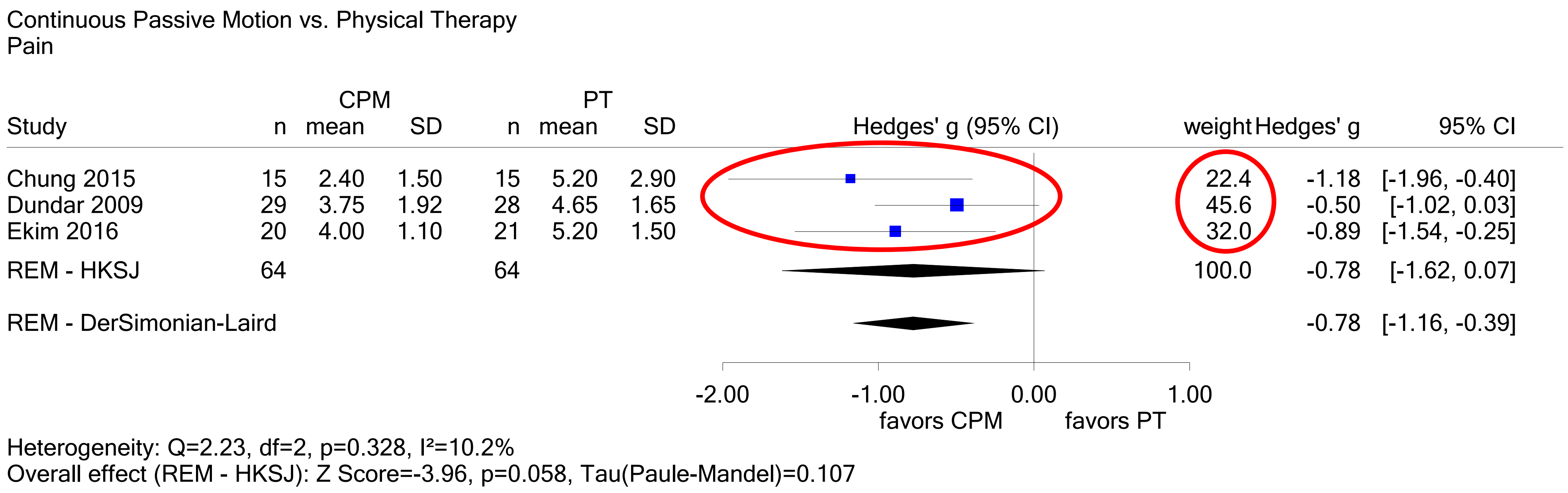


Heterogeneity: Q=0.07, df=2, p=0.965, I²=0%
Overall effect (REM - HKSJ): Z Score=-8.66, p=0.013, Tau(Paule-Mandel)=0

→ HKSJ (VC) ⇒ No proof of an effect

Example: IQWiG Report N16-03

- Is HKSJ informative? Significance of HKSJ vs DSL?
  - HKSJ-CI wider than the union of study CIs?
  - HKSJ informative, but n.s., DSL stat. sign. ⇒ QS



Continuous Passive Motion vs. Physical Therapy
Pain

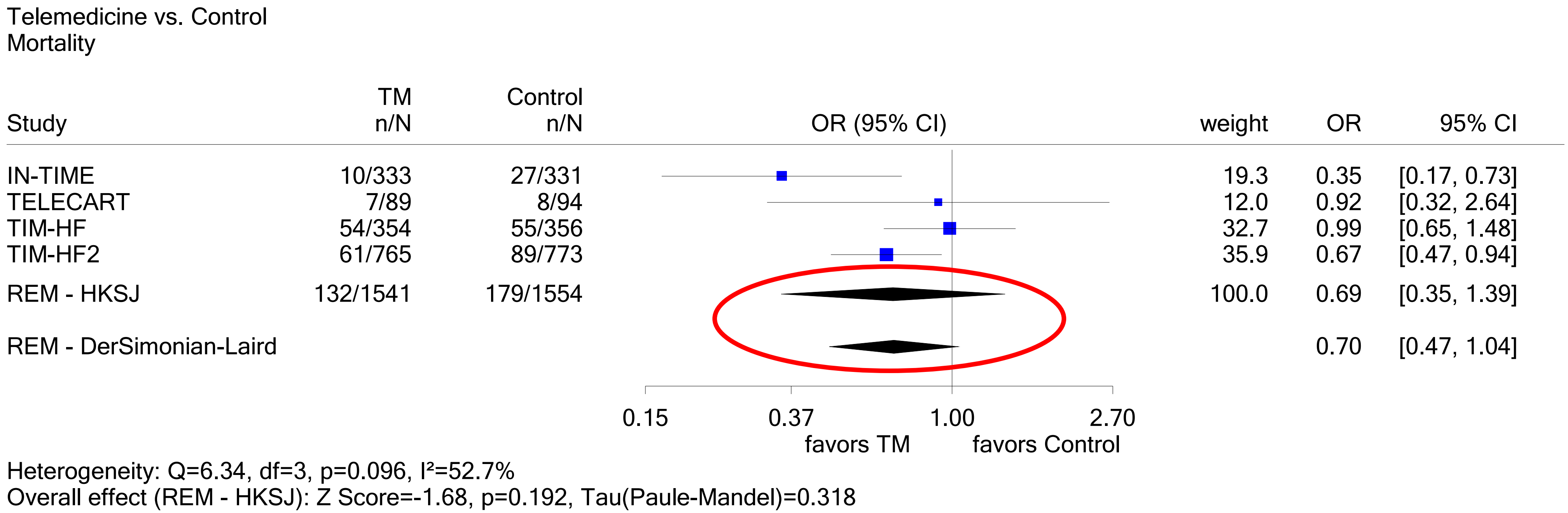
	CPM			PT						
Study	n	mean	SD	n	mean	SD	Hedges' g (95% CI)	weight	Hedges' g	95% CI
Chung 2015	15	2.40	1.50	15	5.20	2.90		22.4	-1.18	[-1.96, -0.40]
Dundar 2009	29	3.75	1.92	28	4.65	1.65		45.6	-0.50	[-1.02, 0.03]
Ekim 2016	20	4.00	1.10	21	5.20	1.50		32.0	-0.89	[-1.54, -0.25]
REM - HKSJ	64			64				100.0	-0.78	[-1.62, 0.07]
REM - DerSimonian-Laird									-0.78	[-1.16, -0.39]

Heterogeneity: Q=2.23, df=2, p=0.328, I²=10.2%
Overall effect (REM - HKSJ): Z Score=-3.96, p=0.058, Tau(Paule-Mandel)=0.107

→ QS ⇒ Benefit of the intervention
(but effect size is unclear)

Example: IQWiG Report N19-01

- Is HKSJ informative? Significance of HKSJ vs DSL?
  - HKSJ-CI wider than the union of study CIs?
  - HKSJ informative, but n.s., DSL n.s. ⇒ HKSJ & DSL



Telemedicine vs. Control
Mortality

Study	TM n/N	Control n/N	OR (95% CI)	weight	OR	95% CI
IN-TIME	10/333	27/331		19.3	0.35	[0.17, 0.73]
TELECART	7/89	8/94		12.0	0.92	[0.32, 2.64]
TIM-HF	54/354	55/356		32.7	0.99	[0.65, 1.48]
TIM-HF2	61/765	89/773		35.9	0.67	[0.47, 0.94]
REM - HKSJ	132/1541	179/1554		100.0	0.69	[0.35, 1.39]
REM - DerSimonian-Laird					0.70	[0.47, 1.04]

Heterogeneity: Q=6.34, df=3, p=0.096, I²=52.7%
Overall effect (REM - HKSJ): Z Score=-1.68, p=0.192, Tau(Paule-Mandel)=0.318

→ HKSJ & DSL ⇒ No proof of an effect

Discussion

- No satisfactory standard method is currently available to perform meta-analyses in the case of very few studies
- FEM possible in practice, but has limitations
- **Therefore, in general, the REM should be used** (unless there are clear reasons to justify the use of the FEM)
- Problem: In the case of very few studies, REM frequently has low power and does not yield informative results
- **In the case of only 2 studies, the FEM should be used** (despite of the general recommendation) unless there are clear reasons against the use of the FEM
- Reason: In situations with only 1 single study, results of this study are interpreted and conclusions are made (in principle, application of the FEM)

Discussion

- In the case of **3-4 studies**: **REM** should be used (unless there are clear reasons to justify the use of the FEM)
- Use of HKSJ (with checks regarding VC and whether the result is informative)
- Application of **HKSJ** or **HKSJ-VC** or **QS**
- For QS:
  - Concept of conclusive effects
  - Prediction intervals
- Other promising possibilities:
  - Beta-binomial model (Felsch et al., *BMC-MRM* 2022)
  - Bayesian meta-analysis with informative prior for τ (Röver et al., *RSM* 2021; Lilienthal et al., work in progress)

Outlook

Beta-binomial model (BBM)

- Suitable for binary data
- Simulation study by IQWiG in collaboration with Tim Mathes (Göttingen) and Oliver Kuß (Düsseldorf)
- Results (Felsch et al., *BMC-MRM* 2022):
  - No advantages in the case of 2 studies
  - More power than HKSJ in the case of 3-4 studies

→ Consideration of inclusion of the BBM in the procedure described before

Outlook

Bayesian meta-analysis

- Required: Slightly informative prior for τ
- Good compromise between DSL und HKSJ
- IQWiG-project in collaboration with Tim Friede and Christian Röver (Göttingen):
  - Derivation of empirical priors for τ from meta-analyses of IQWiG reports (see *"A Day with … SMG"* 11.05.2021: https://training.cochrane.org/learning-events/learning-live/day/day-smg)
  - Currently: Estimation of empirical priors for τ by means of the hierarchical Bayes model according to Röver et al. (*Stat. Med.* 2023, under review)
  - Manuscript in preparation with suggestion of priors for τ for the effect measures RR, OR, HR, SMD (suitable for HTA) (Lilienthal et al., 2023, work in progress)

Summary

Evidence synthesis in the case of very few studies:

- Too large, unexplained heterogeneity: **QS**
- 2 studies:
  Standard model **FEM** (IV or MH)
- 3-4 studies:
  - **REM** with HKSJ or HKSJ-VC (if HKSJ yields useful information)
  - **QS** (if HKSJ yields no useful information or when DSL stat. sign.)
- 5 studies or more: REM with HKSJ or HKSJ-VC
- <u>Future:</u> BBM and Bayes (with informative prior for τ)

Conclusion

- No satisfactory universal standard method is currently available to perform meta-analyses in the case of very few studies

- Additional approaches (beta-binomial model, Bayes) are under consideration

- The procedure currently used by IQWiG (combination of FEM, REM, QS) represents a feasible approach to perform evidence syntheses with very few studies in practice

References

- Bender, R., Friede, T., Koch, A., Kuss, O., Schlattmann, P., Schwarzer, G. & Skipka, G. (2018): Methods for evidence synthesis in the case of very few studies. *Res. Syn. Methods* **9**, 382–392.
- Cornell, J.E., Mulrow, C.D., Localio, R., Stack, C.B., Meibohm, A.R., Guallar, E. & Goodman, S.N. (2014): Random-effects meta-analysis of inconsistent effects: A time for change. *Ann. Intern. Med.* **160**, 267-270.
- DerSimonian, R. & Laird, N. (1986): Meta-analysis in clinical trials. *Control. Clin. Trials* **7**, 177-188.
- IQWiG (2022): *General Methods, Version 6.1 of 24.01.2022*. IQWiG, Cologne, Germany.
- Knapp, G. & Hartung, J. (2003): Improved tests for a random effects meta-regression with a single covariate. *Stat. Med.* **22**, 2693-2710.
- Felsch, M., Beckmann, L., Bender, R., Kuss, O., Skipka, G. & Mathes, T. (2022): Performance of several types of beta-binomial models in comparison to standard approaches for meta-analyses with very few studies. *BMC Med. Res. Methodol.* **22**, 319.
- Röver, C., Bender, R., Dias, S., Schmid, C.H., Schmidli, H., Sturtz, S., Weber, S. & Friede, T. (2021): On weakly informative prior distributions for the heterogeneity parameter in Bayesian random-effects meta-analysis. *Res. Syn. Methods* **12**, 448-474.
- Schulz, A., Schürmann, C., Skipka, G. & Bender, R. (2022): Performing meta-analyses with very few studies. In: Evangelou, E. & Veroniki, A.A., Eds.: *Meta-Research: Methods and Protocols,* pp. 91-102. Humana, New York.
- Veroniki, A.A., Jackson, D., Bender, R., Kuss, O., Langan, D., Higgins, J.P.T., Knapp, G. & Salanti, G. (2019): Methods to calculate uncertainty in the estimated overall effect size from a random-effects meta-analysis. *Res. Syn. Methods* **10**, 23-43.