



Workshop *"Evidenzsynthesemethoden in Situationen mit sehr wenigen und kleinen Studien"*, Jena, 09.09.2025

Challenges and solution approaches for meta-analyses with very few studies

Ralf Bender

IQWiG, Cologne, Germany

Outline

- Introduction
 - Main meta-analytic models
 - Main estimation methods
 - Qualitative summary of study results (QSSR)
- Meta-analysis with very few studies
 - Problems, examples
 - Procedures used in IQWiG
- Alternative approaches
- Summary
- References

Introduction

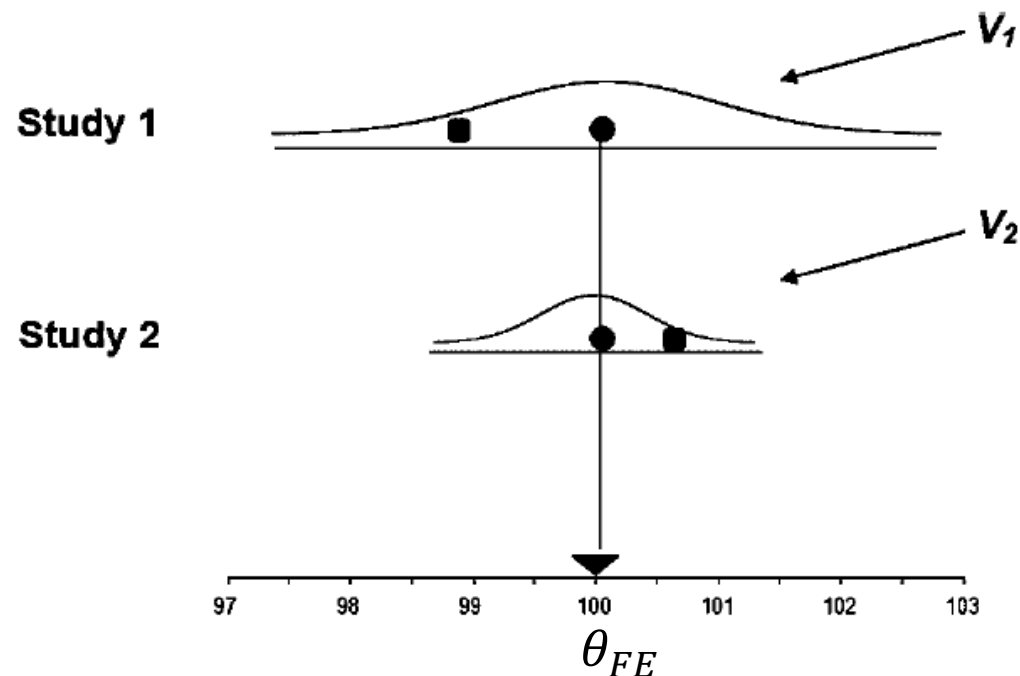
2 main meta-analytic models:

- 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 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., 2018).

Meta-analysis: FEM

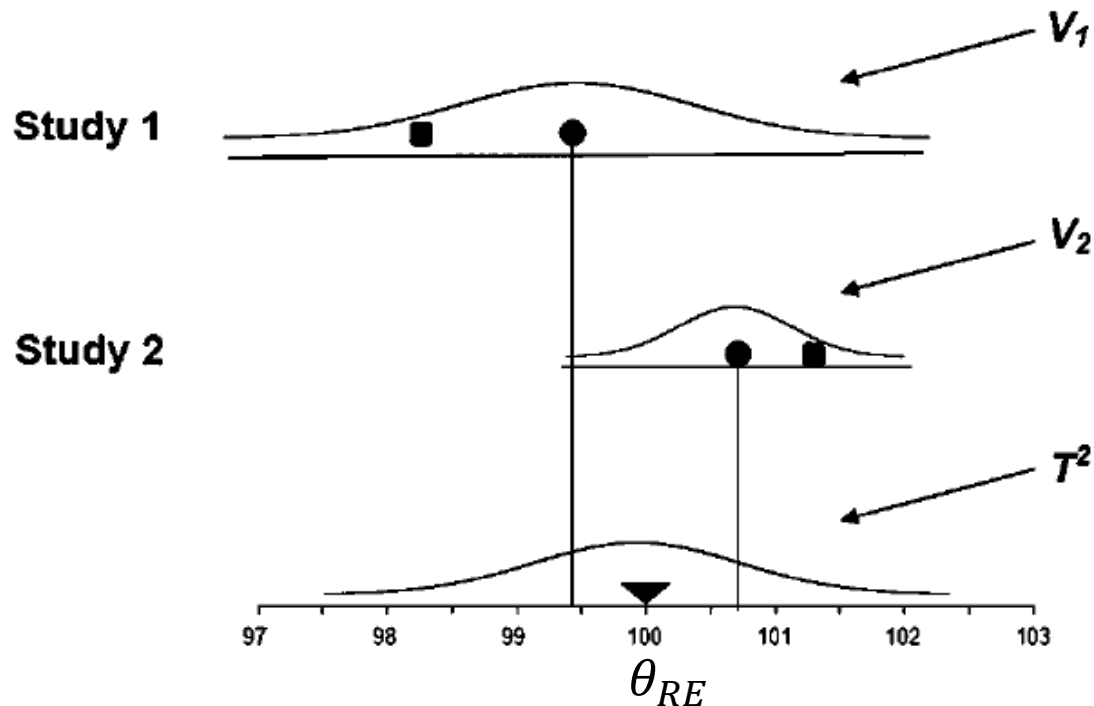
- $y_i = \theta_{FE} + \varepsilon_i$, $\varepsilon_i \sim N(0, v_i)$, $Var(y_i) = v_i$
- Assumption: All studies estimate the same effect.
- Parameter of interest: **Fixed effect** θ_{FE} (better term: "*Common effect*")



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 θ_{RE} of the effects**



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

REM: Prediction interval

- Confidence interval (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

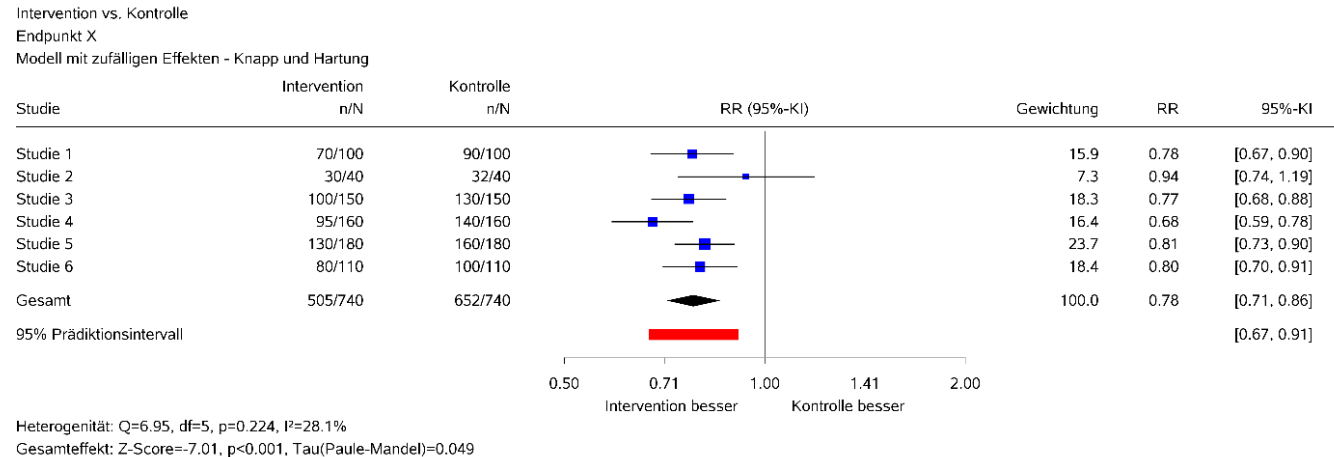
- Prediction interval (PI):

- $\hat{\theta}_{RE} \pm t_{k-1, 1-\frac{\alpha}{2}} \times \sqrt{\tau^2 + Var(\hat{\theta}_{RE})}$

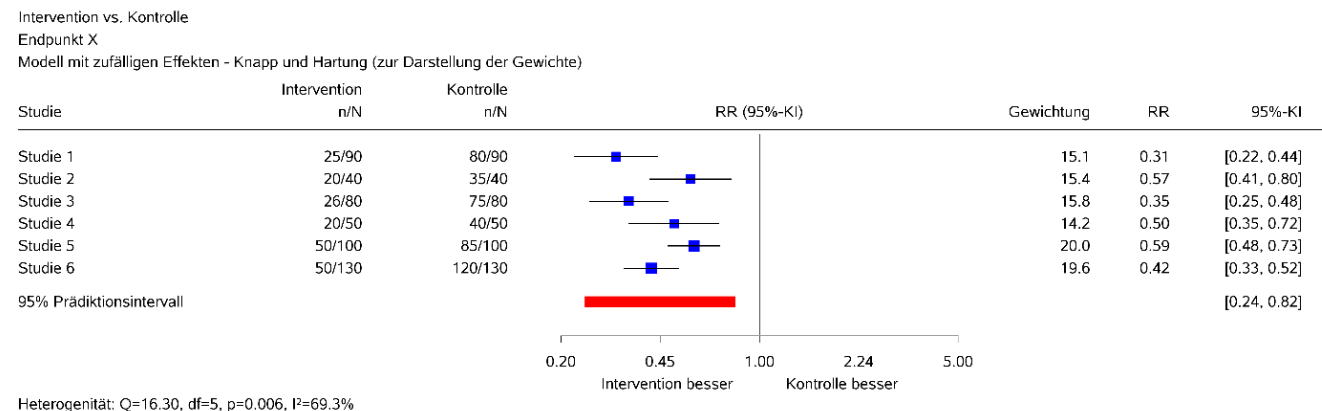
- Range, which includes with high certainty (95%) the true effect of a single study

- Helpful tool for QSSR

PI - graphical illustration of heterogeneity:



No MA meaningful, clearly conclusive effects due to PI



Main estimation methods

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 2×2 tables (formula dependent on effect measure)

Main estimation methods

REM: DerSimonian & Laird (DSL)

- Historically, the standard approach for RE meta-analysis: DSL method (DerSimonian & Laird, 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: $\hat{\theta}_{RE} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{1}{\sum_{i=1}^k w_{i,RE}}}$, z_q : q-quantile of normal distribution
- DSL has been criticized for some time (Cornell et al., 2014)
- DSL ignores the uncertainty of variance estimations
- CIs are frequently too narrow (especially in the case of very few studies)

Main estimation methods

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

- Recommended by the Cochrane Collaboration: HKSJ method (Veroniki et al., 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: $\hat{\theta}_{RE} \pm t_{k-1, 1-\frac{\alpha}{2}} \sqrt{\frac{\sum_{i=1}^k w_{i,RE} (y_i - \hat{\theta}_{RE})^2}{(k-1) \sum_{i=1}^k w_{i,RE}}}$, $t_{m,q}$: q -quantile of t -distribution

- HKSJ holds type 1 error

- CIs frequently very wide (especially in the case of few studies)

- $z_{0.975} = 1.96$, $t_{1;0.975} = 12.7$, $t_{2;0.975} = 4.3$, $t_{3;0.975} = 3.2$, $t_{4;0.975} = 2.8$

Main estimation methods

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

- Problems in homogeneous data situations

- 95% CI: $\hat{\theta}_{RE} \pm t_{k-1, 1-\frac{\alpha}{2}} \sqrt{\frac{\sum_{i=1}^k w_{i,RE} (y_i - \hat{\theta}_{RE})^2}{(k-1) \sum_{i=1}^k w_{i,RE}}}$

- SE may be arbitrarily too small and CI too narrow

- Ad-hoc variance correction (Knapp & Hartung, 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 study results (QSSR)

Concept of conclusive effects (IQWiG, 2023):

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

Qualitative summary of study results (QSSR)

Concept of conclusive effects (IQWiG, 2023):

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

Meta-analyses with very few studies

Meta-analytic problems in the case of very few (< 5) studies:

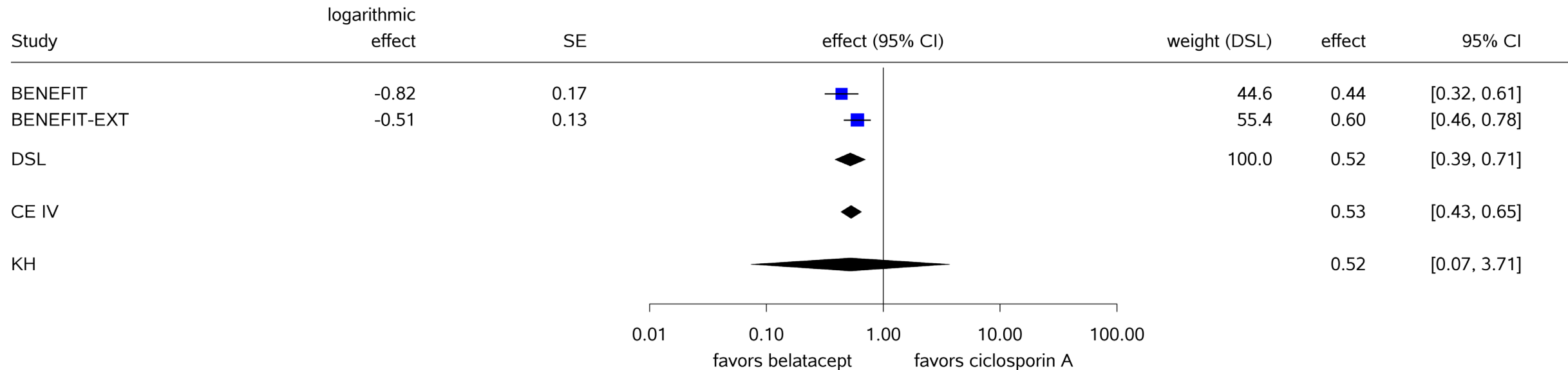
- Choice between FEM and REM difficult
- Heterogeneity parameter cannot be estimated adequately
- Confidence intervals of DSL are too narrow
- Confidence intervals of HKSJ are wide or even non-informative
- In homogeneous situations CIs of HKSJ are sometimes too narrow
(For this situation the ad hoc variance correction is required)

Reference: Bender et al. (2018)

Meta-analyses with very studies

Example: 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"



Heterogeneity: $Q=2.06$, $df=1$, $p=0.151$, $I^2=51.5\%$
 Overall effect: Z Score=-4.21, $p<0.001$, Tau=0.157

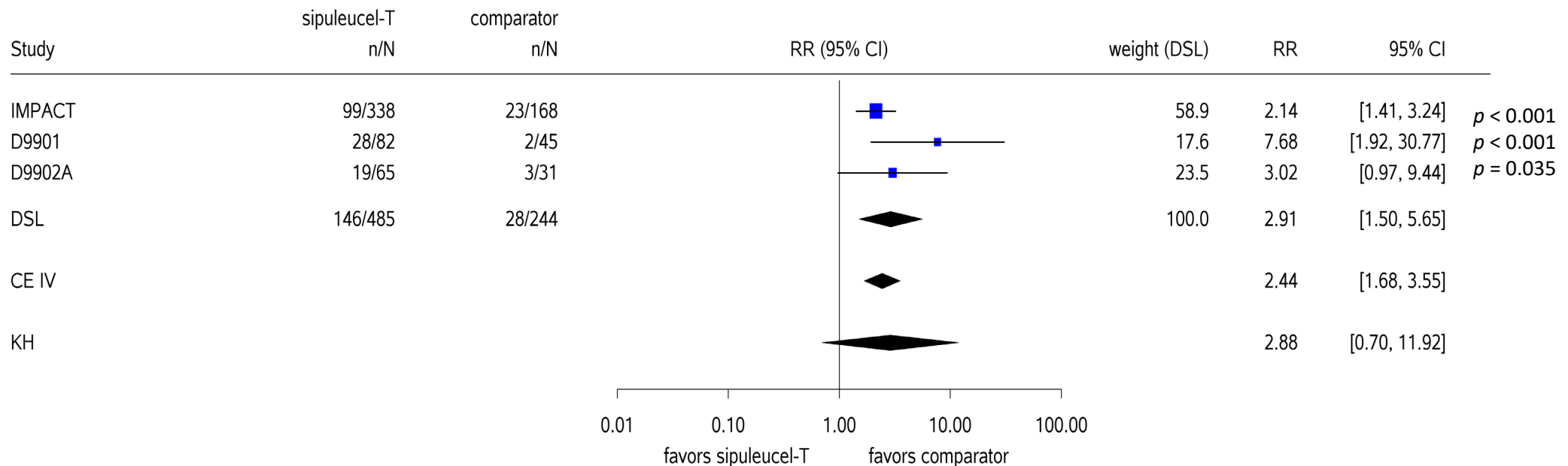


- 1) HKSJ over-conservative
- 2) Decision of no added benefit would be critical

Meta-analyses with very studies

Example: 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



Heterogeneity: $Q=3.29$, $df=2$, $p=0.193$, $I^2=39.1\%$

Overall effect: Z Score=3.15, $p=0.002$, Tau=0.388



Despite 3 significant studies, no significant effect with HKSJ

Alternative approach: Bayesian REM

- Up to November 2024, IQWiG used a circumstantial procedure
 - Step 1: Preliminary model choice (in general: REM)
 - Step 2: Evaluation of heterogeneity (may lead to QSSR)
 - Step 3: Final model and method choice (frequently leads to QSSR)
- Bayesian methods may have advantages
(Bender et al. *RSM* 2018, Röver et al. *RSM* 2021)
- Priors for a Bayesian REM meta-analysis:
 - Treatment effect: **Non-informative** prior distribution
 - Heterogeneity: Advantages with **informative** prior distributions
- Suitable prior distributions obtained from meta-analyses performed in IQWiG 2005-2021 (Lilienthal et al. *RSM* 2024, GMDS Heilbronn 2023)

New approach applied in IQWiG

- **Combination of Bayes with QSSR for 3-4 studies:**
 - 2 studies: FEM (as before)
 - 3-4 studies: Bayes and comparison with QSSR
 - ≥ 5 studies: REM by using HKSJ (as before)
- Advantages of the new approach:
 - Less effort (choice of model and estimation method less complicated)
 - Improved option for the determination of the extent of added benefit

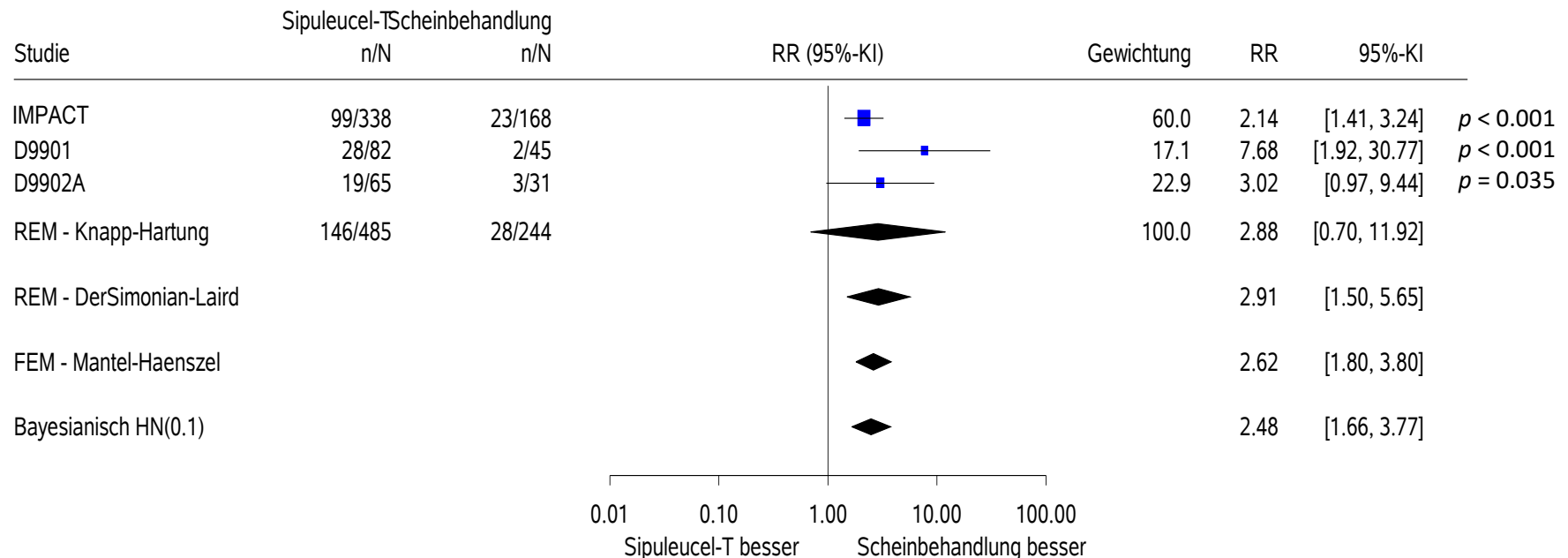
Determination of the extent of added benefit with 3 or 4 studies:

- Former approach: Extent in **23 %** of cases not quantifiable
- Bayes + QSSR: Extent in only **6 %** of cases not quantifiable

Example

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



Heterogenität: $Q=3.29$, $df=2$, $p=0.193$, $I^2=39.1\%$

Gesamteffekt (REM - Knapp-Hartung): $Z\text{-Score}=3.20$, $p=0.085$, $\text{Tau(Paule-Mandel)}=0.370$



Bayes REM is a good compromise between FEM and REM with HKSJ

Alternative approaches

- For binary data:
 - Generalized mixed models
 - Marginal models with generalized estimation equations (GEE)
 - Beta-binomial models (Felsch et al. *BMC-MRM* 2022)
- Combined p -value functions :
 - Talk by Leo Held: ROeS, Graz, 18.09.2025
 - Held, Hofmann & Pawel: <https://arxiv.org/abs/2408.08135>

Summary

- Meta-analysis with very studies is challenging
- Frequently, none of the standard methods leads to useful results
- In qualitative summaries no quantification of the pooled effect
- **Alternative meta-analytic methods for very few studies are needed**
- Promising alternative approaches:
 - Bayesian REM with informative priors for the heterogeneity parameter
(currently applied in IQWiG for 3-4 studies)
 - Beta-binomial model
(especially for double-zero studies)
 - Combined p -value functions
(submitted paper by Held, Hofman & Pawel)

References

- Bender, R. et al. (2018): Methods for evidence synthesis in the case of very few studies. *Res. Syn. Methods* **9**, 382-392.
- Borenstein, M. et al. (2010): A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res. Syn. Methods* **1**, 97-111.
- Cornell, J.E. et al. (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.
- Felsch, M. et al. (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.
- Held, L., Hofman, F. & Pawel, S. (2024): A comparison of combined p -value functions for meta-analysis. arXiv:2408.08135v2.
- IQWiG (2023): *General Methods, Version 7.0 of 19.09.2023*. 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.
- Lilienthal, J. et al. (2024): Bayesian random-effects meta-analysis with empirical heterogeneity priors for application in health technology assessment with very few studies. *Res. Syn. Methods* **15**, 275-287.
- Röver C. et al. (2021): On weakly informative prior distributions for the heterogeneity parameter in Bayesian random-effects meta-analysis. *Res. Syn. Methods* **12**, 448-474.
- Veroniki, A.A. et al. (2019): Methods to calculate uncertainty in the estimated overall effect size from a random-effects meta-analysis. *Res. Syn. Methods* **10**, 23-43.