Recent Advances in Meta-Analysis

Bayesian random-effects meta-analysis with empirical heterogeneity priors generated from IQWiG reports

Sibylle Sturtz, Jona Lilienthal, Ralf Bender

Institute for Quality and Efficiency in Health Care (IQWiG), Köln, Germany

Situation

- IQWiG's (former) approach to evidence synthesis:
  Frequentist methods of meta-analysis combined with qualitative evidence synthesis
- Approach is complex: model choice, qualitative evidence synthesis

IQWiG's former approach of evidence synthesis (in short)

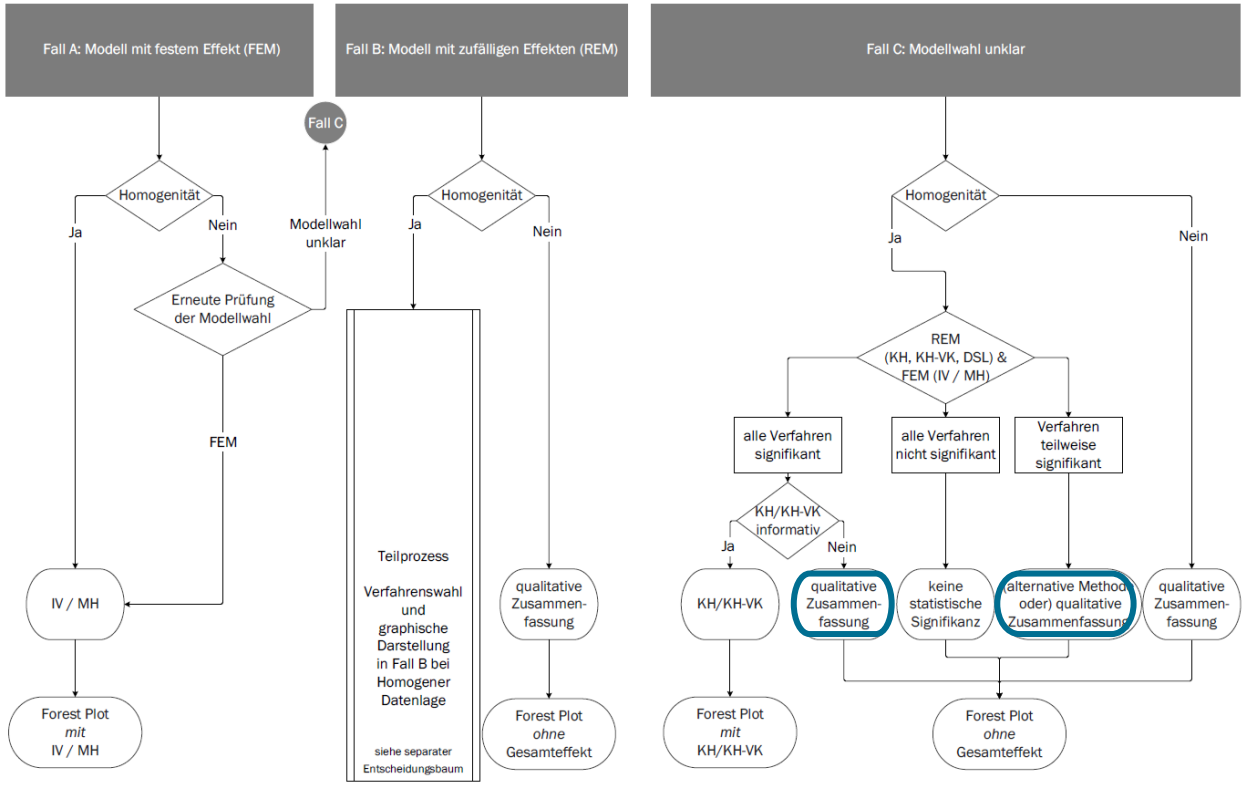
het.

- qualitative evidence synthesis

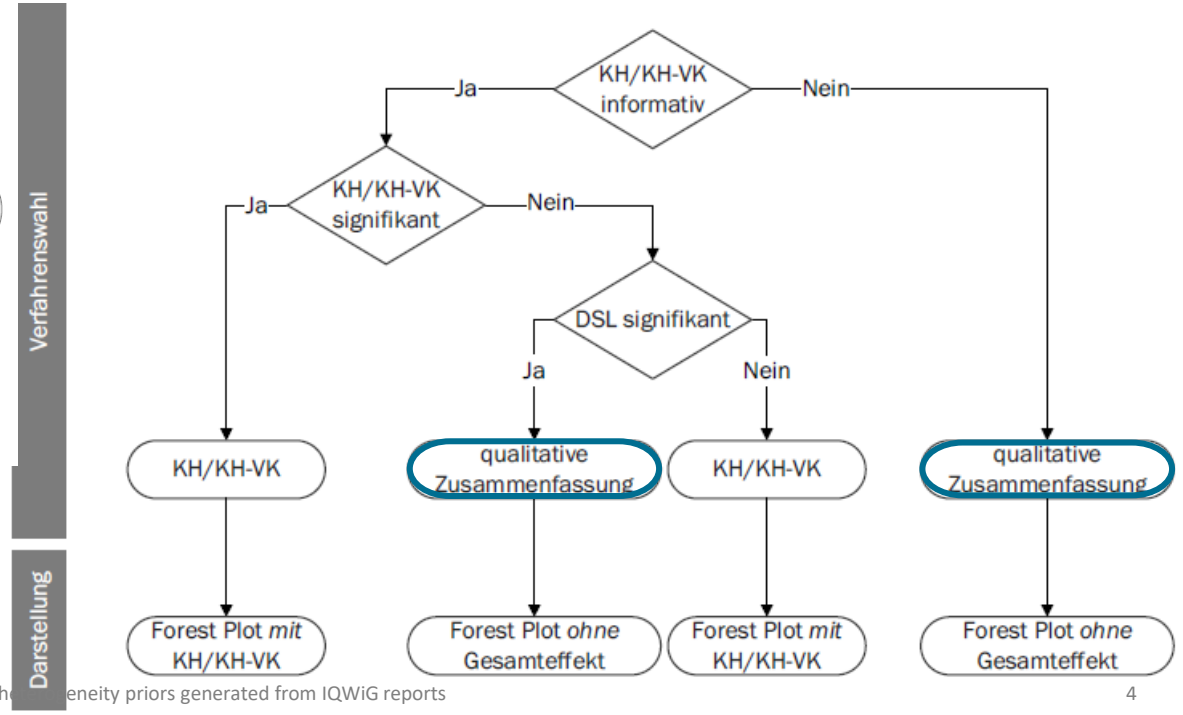
homogeneous

- 2 studies: common/fixed-effect meta-analysis
- 3 or 4 studies:
  random-effects meta-analysis (Knapp-Hartung) if proper, otherwise qualitative evidence synthesis
- ≥ 5 studies: random-effects meta-analysis (Knapp-Hartung)

IQWiG's former approach of evidence synthesis in case of 2-4 studies (in detail)



DSL: Verfahren nach DerSimonian und Laird, FEM: Modell mit festem Effekt, IV: Verfahren mit Inverser Varianz, KH/KH-VK: Verfahren nach Knapp und Hartung ohne/mit Varianzkorrektur, MH: Verfahren nach Mantel und Haenszel, REM: Modell mit zufälligen Effekten



Qualitative evidence synthesis (QES)

There is an overall effect if …

- all studies statistically significant (same direction, 5 % level), or
- ≥ 3 studies:
  have a subset of studies with effects in same direction and
  - studies in subset have cum. weight > 80 %
  - stat. significant studies in subset have cum. weight > 50 %, or
- ≥ 4 studies:
  95 % prediction interval does not include null effect

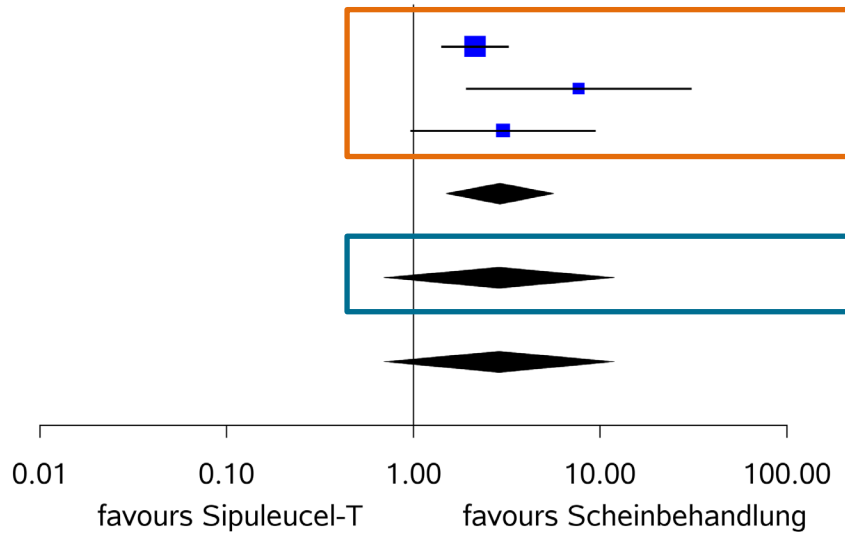
Here, overall effect is non-quantifiable

(Schulz et al. 2022)

Example

Sipuleucel-T vs. Scheinbehandlung - id_769
Fieber

Study	Sipuleucel-T n/N	Scheinbehandlung n/N	RR (95% CI)	weight	RR	95% CI
D9902B	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]
REM - DerSimonian-Laird	146/485	28/244		100.0	2.91	[1.50, 5.65]
REM - Knapp-Hartung					2.88	[0.70, 11.92]
REM - Knapp-Hartung (variance corr.)					2.88	[0.70, 11.92]



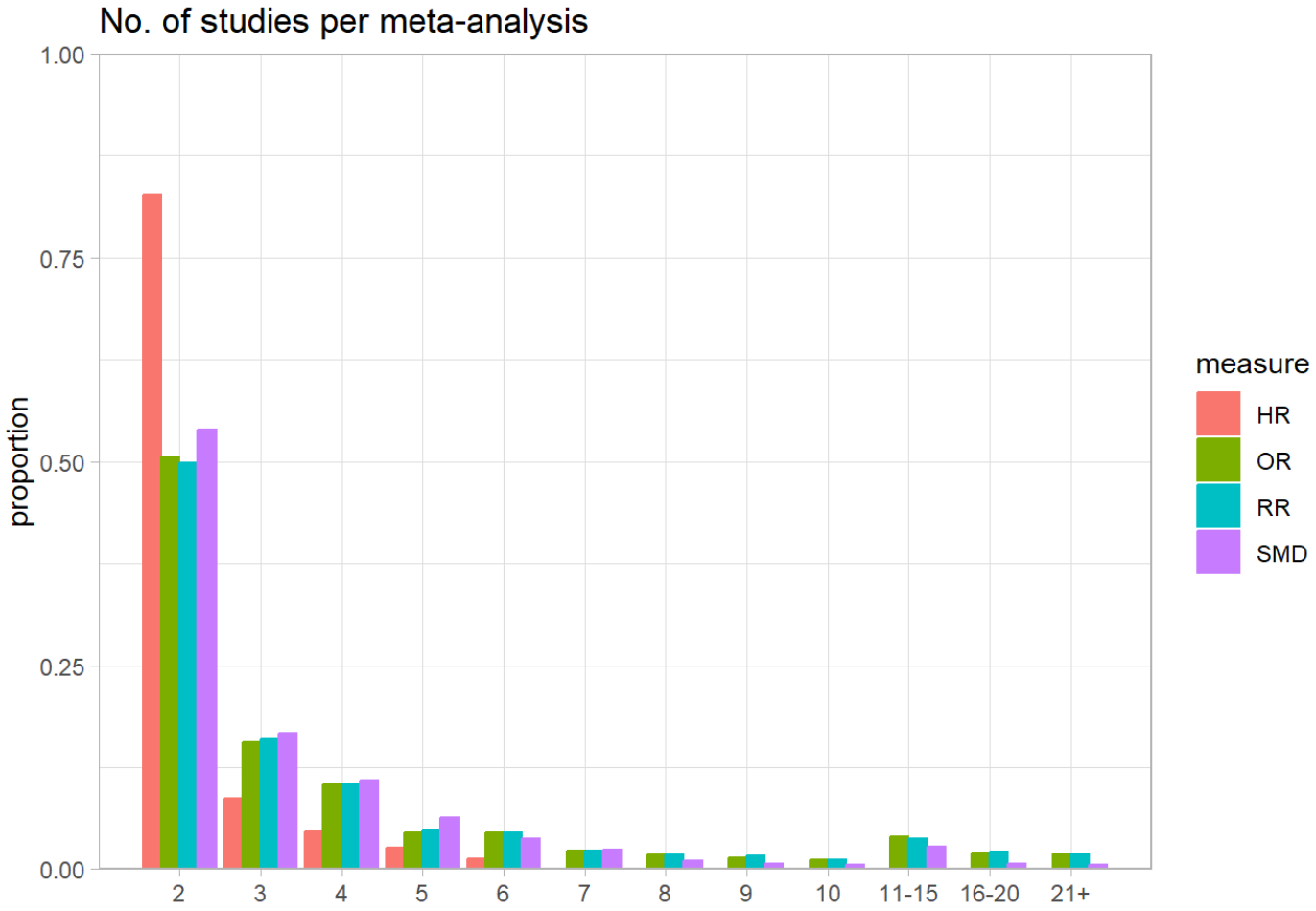
Heterogeneity: Q=3.29, df=2, p=0.193, I^2=39.1%
Overall effect (REM - DerSimonian-Laird): Z-Score=3.15, p=0.002, Tau=0.388

- Knapp-Hartung not informative
- QES: all studies point in the same direction, sign. studies have weight > 50 %
  => overall effect, non-quantifiable

Situation

- IQWiG's (former) approach to evidence synthesis:
  Frequentist methods of meta-analysis combined with qualitative evidence synthesis
- Approach is complex: model choice, qualitative evidence synthesis
- Difficulties in meta-analyses in situations with ≤ 4 studies
  - Estimation of heterogeneity between studies unreliable
  - Knapp-Hartung can lead to inflated confidence intervals of the pooled estimate
  - Alternative of qualitative evidence synthesis does not permit the quantification of effects
- Adoption: Bayesian random-effects meta-analysis in certain situations

Database of meta-analyses in IQWiG reports



⇨ Most meta-analyses in our data have less than 5 studies

Bayesian meta-analysis

Alternative: Bayesian meta-analysis in situations with few studies
(e.g. Friede et al. 2017, Bender et al. 2018)

Normal-normal hierarchical model

for k studies with observations y_i with standard error σ_i

$$y_i \sim N\left(\theta_i,\ \sigma_i^2\right), i = 1, \ldots, k$$
$$\theta_i \sim N(\mu,\ \tau^2), i = 1, \ldots, k$$
$$(\mu,\ \tau^2) \sim P$$

Prior distributions

- Prior for μ: improper uniform or normal, e.g. $\mu \sim N(0, 10^2)$
- Prior for τ: choice is critical as the number of studies is small

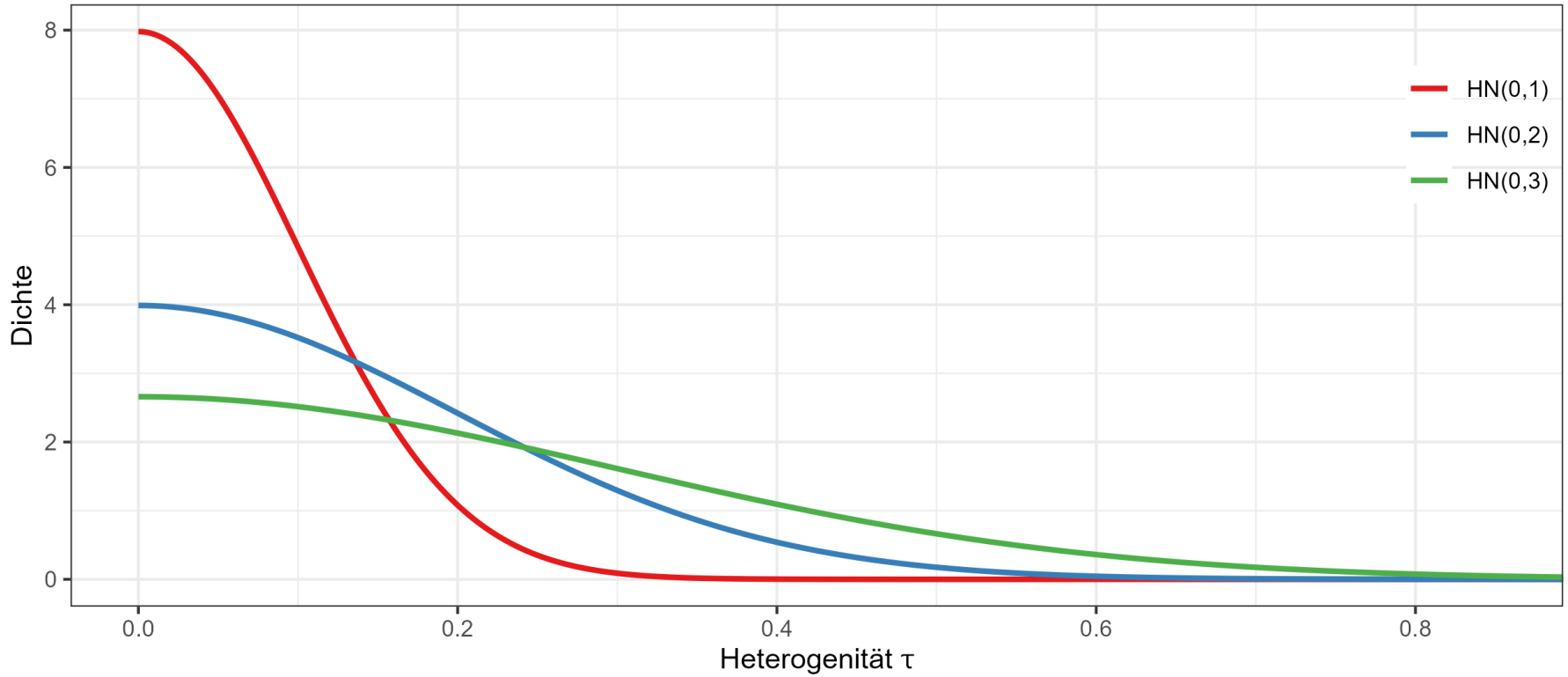
known parameters, unknown parameters

Prior distribution for heterogeneity parameter

- Informative prior distributions based on empirical information
- Turner et al. (2015) and Rhodes et al. (2015):
  - Empirical information from Cochrane Database of Systematic Reviews
  - Bayesian model to derive posterior predictive distribution
    - SMDs: $\log(\tau^2) \sim t_5(-3.44, 2.59^2)$
    - ORs: $\tau^2 \sim LN(-2.56, 1.74^2)$
- Röver et al. (2023): extended normal-normal hierarchical model for Bayesian random effects meta-analysis to estimate the distribution of τ^2
- Lilienthal et al. (2024):
  - Application of Röver's model to meta-analyses from IQWiG reports
  - Distribution families: half normal (HN), log normal, exponential, half Cauchy, half logistic
  - Effect measures: OR, RR, HR, SMD

Prior distribution for heterogeneity parameter

- Recommendations for prior distributions based on IQWiG meta-analyses database (Lilienthal et al., 2024)
  - RR, HR: **HN(0.1)**
  - OR: **HN(0.2)**
  - SMD: **HN(0.3)**



Comparison

- Former approach (KH+QES) vs. Bayesian meta-analysis using the proposed priors

	RR	HR	OR	SMD
P(Bayes result = KH+QES result)	90%	96%	90%	82%
P(Bayes effect | KH+QES no effect)	9%	3%	9%	8%
P(Bayes no effect | KH+QES effect)	13%	9%	16%	36%

- To avoid a large difference regarding effect decision:
  Combination of Bayesian meta-analysis and qualitative evidence synthesis

Lilienthal et al., 2024

IQWiG's former approach of evidence synthesis (in short)

het.

- qualitative evidence synthesis

homogeneous

- 2 studies: common/fixed-effect meta-analysis
- 3 or 4 studies:
  random-effects meta-analysis (Knapp-Hartung) if proper, otherwise qualitative evidence synthesis
- ≥ 5 studies: random-effects meta-analysis (Knapp-Hartung)

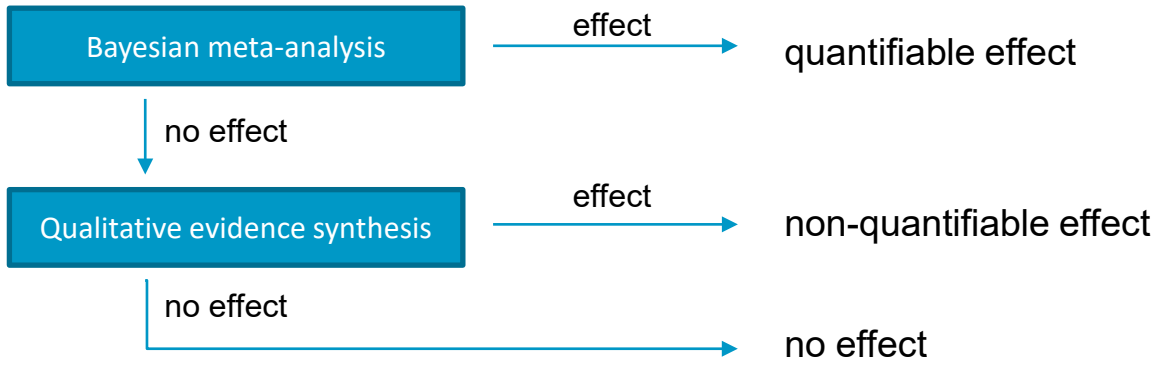
New approach

het.

- qualitative evidence synthesis

homogeneous

- 2 studies: common/fixed-effect meta-analysis
- **3 or 4 studies: Bayesian meta-analysis + qualitative evidence synthesis**



- 5 studies: random-effects meta-analysis (Knapp-Hartung)

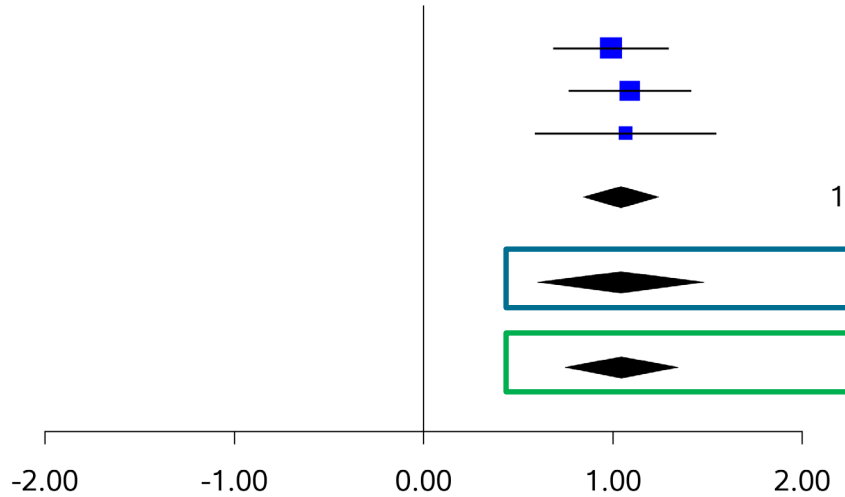
Retrospective evaluation of the new approach

- Comparison of
  - former approach (Knapp-Hartung +/- qualitative evidence synthesis)
  - new approach (Bayesian meta-analysis +/- qualitative evidence synthesis)

- Re-calculation of all meta-analyses of IQWiG reports
  - 3 or 4 studies
  - not heterogeneous (p > 5%)
  - resulting dataset: N = 437 meta-analyses

Rosiglitazone 8mg + Metformin vs. Placebo + Metformin - id_94
Gewichtsänderung

Study	Rosiglitazone 8mg + Metformin n	mean	SD	Placebo + Metformin n	mean	SD	Hedges' g (95% CI)	weight	Hedges' g	95% CI
49653/094	94	1.90	3.70	94	-1.20	2.40		43.7	0.99	[0.69, 1.29]
49653/093	90	2.30	3.60	81	-1.30	2.90		38.7	1.09	[0.77, 1.41]
49653/044	43	2.30	3.30	35	-0.90	2.50		17.6	1.07	[0.59, 1.55]
REM - DerSimonian-Laird	227			210				100.0	1.04	[0.84, 1.24]
REM - Knapp-Hartung (variance corr.)									1.04	[0.60, 1.48]
Bayesian HN(0.3)									1.04	[0.75, 1.35]

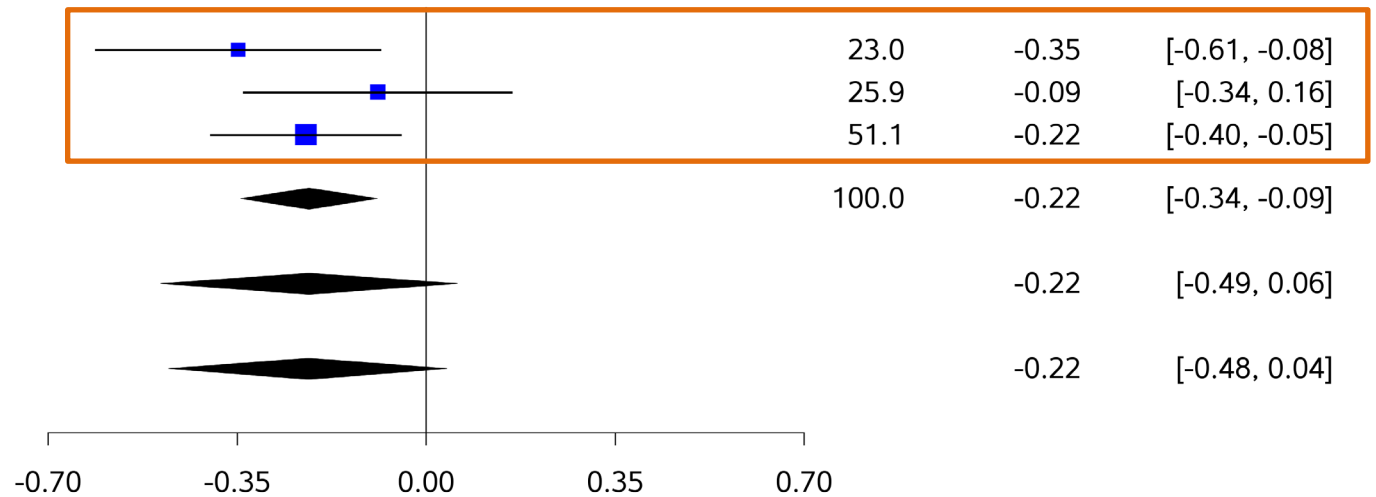


Heterogeneity: Q=0.21, df=2, p=0.901, I^2=0%
Overall effect (REM - DerSimonian-Laird): Z-Score=10.19, p<0.001, Tau=0

former approach: effect, quantifiable
new approach: effect, quantifiable

Exenatide 5mug vs. Placebo - id_521
Gewichtsänderung

Study	Exenatide 5mug n	Exenatide 5mug mean	Exenatide 5mug SD	Placebo n	Placebo mean	Placebo SD	Hedges' g (95% CI)	weight	Hedges' g	95% CI
2993-112 DeFronzo 2005	110	-1.60	4.20	113	-0.30	3.19		23.0	-0.35	[-0.61, -0.08]
2993-113 Buse 2004	125	-0.90	3.35	123	-0.60	3.33		25.9	-0.09	[-0.34, 0.16]
2993-115 Kendall 2005	245	-1.60	3.13	247	-0.90	3.14		51.1	-0.22	[-0.40, -0.05]
REM - DerSimonian-Laird	480			483				100.0	-0.22	[-0.34, -0.09]
REM - Knapp-Hartung									-0.22	[-0.49, 0.06]
Bayesian HN(0.3)									-0.22	[-0.48, 0.04]

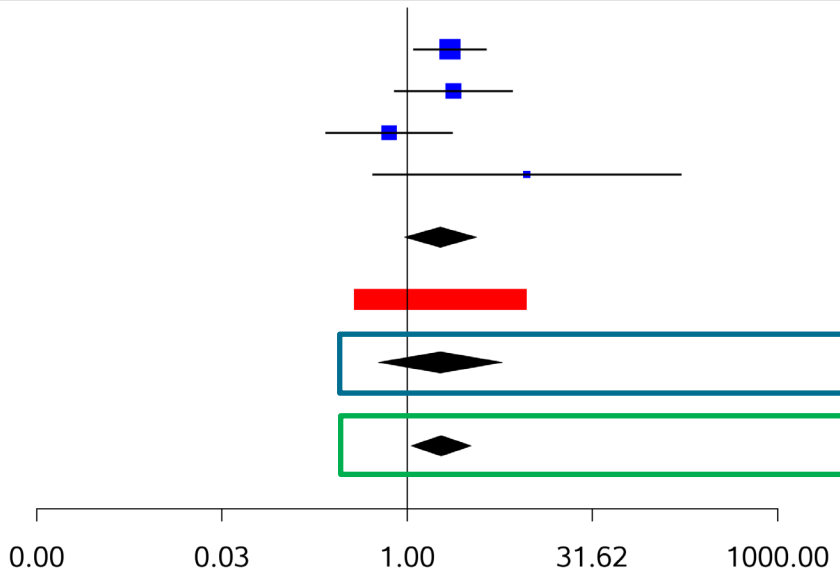


Heterogeneity: Q=1.95, df=2, p=0.377, I^2=0%
Overall effect (REM - DerSimonian-Laird): Z-Score=-3.36, p<0.001, Tau=0

former approach: effect, non-quantifiable
new approach: effect, non-quantifiable

Certolizumab + MTX vs. Placebo + MTX - id_612
SUE

Study	Certolizumab + MTX n/N	Placebo + MTX n/N	OR (95% CI)	weight	OR	95% CI
C87027	45/392	11/199		45.5	2.22	[1.12, 4.39]
C87050	18/248	4/125		25.9	2.37	[0.78, 7.15]
C870-076	5/96	7/98		23.5	0.71	[0.22, 2.33]
RA0025	8/85	0/42		5.1	9.32	[0.53, 165.51]
REM - DerSimonian-Laird	76/821	22/464		100.0	1.86	[0.95, 3.65]
95% prediction interval						[0.37, 9.29]
REM - Knapp-Hartung					1.86	[0.58, 5.93]
Bayesian HN(0.2)					1.89	[1.07, 3.32]

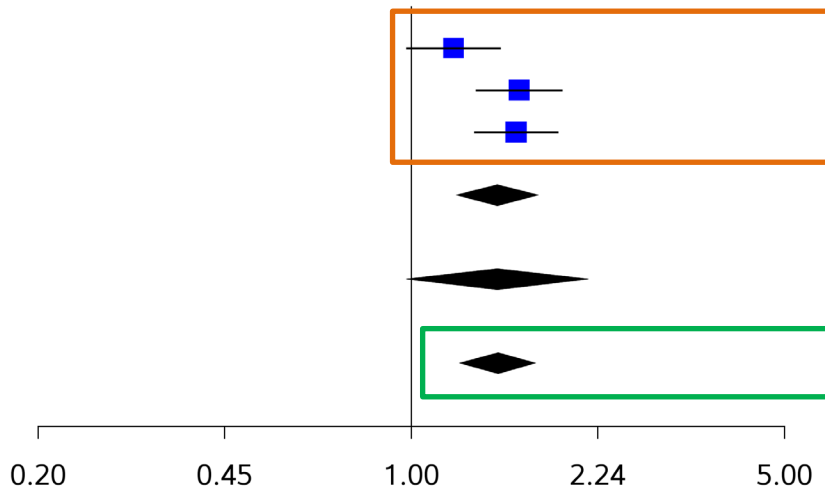


Heterogeneity: Q=4.20, df=3, p=0.241, I^2=28.6%
Overall effect (REM - DerSimonian-Laird): Z-Score=1.81, p=0.070, Tau=0.371

former approach: no effect
new approach: effect, quantifiable

SIM + PEG +RBV vs. PLA +PEG + RBV - id_736
SVR 72

Study	SIM + PEG +RBV n/N	PLA +PEG + RBV n/N	weight	RR	95% CI
PILLAR	60/77	50/77	31.6	1.20	[0.98, 1.47]
QUEST-1	207/264	64/130	33.9	1.59	[1.32, 1.92]
QUEST-2	202/257	67/134	34.5	1.57	[1.31, 1.88]
REM - DerSimonian-Laird	469/598	181/341	100.0	1.45	[1.21, 1.73]
REM - Knapp-Hartung				1.45	[0.98, 2.14]
Bayesian HN(0.1)				1.45	[1.23, 1.71]



Heterogeneity: Q=5.28, df=2, p=0.071, I^2=62.1%
Overall effect (REM - DerSimonian-Laird): Z-Score=4.09, p<0.001, Tau=0.124

former approach: effect, non-quantifiable
new approach: effect, quantifiable

no change regarding effect and quantification:
363/437 (83%)
282 no effect
77 quantifiable effect
4 non-quantifiable effect

change regarding effect:
47/437 (11%)
41 no effect → effect
6 effect → no effect

change regarding quantification:
27/437 (6%)
27 non-quantifiable → quantifiable
0 quantifiable → non-quantifiable

Conclusion

- IQWiG's former approach for meta-analysis with few studies
  - complex and elaborate
  - resulting effects often non-quantifiable
- To use Bayesian meta-analysis, informative prior distributions for the heterogeneity parameter were derived from a database of meta-analyses in IQWiG reports
- New approach combining Bayesian meta-analysis and qualitative evidence synthesis
  - less complex
  - higher rate of quantifiable effects: 83/114 -> 145/149
  - applied since November 2024
    - since then no report published with this data situation

Thanks for joint work to

- Christian Röver & Tim Friede
- Former IQWiG employees Christoph Schürmann & Matthias Maiworm
- & all colleagues from Medical Biometry Department at IQWiG

Literature

- 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. Research Synthesis Methods 9: 382-392.
- Lilienthal J., Sturtz S., Schürmann C., Maiworm M., Röver C., Friede T., Bender R. (2024). Bayesian random-effects meta-analysis with empirical heterogeneity priors for application in health technology assessment with very few studies. Research Synthesis Methods 15(2): 275-287.
- Rhodes, K. M., Turner, R. M., Higgins, J. P. (2015). Predictive distributions were developed for the extent of heterogeneity in meta-analyses of continuous outcome data. Journal of clinical epidemiology, 68(1), 52-60.
- Röver C., Bender R., Dias S., Schmid C., Schmidli H., Sturtz S., Weber S., Friede T. (2021). On weakly informative prior distributions for the heterogeneity parameter in Bayesian random-effects meta-analysis. Research Synthesis Methods 12: 448-474.
- Röver, C., Sturtz, S., Lilienthal, J., Bender, R., Friede, T. (2023). Summarizing empirical information on between-study heterogeneity for Bayesian random-effects meta-analysis. Statistics in Medicine, 42(14), 2439-2354.
- Schulz, A., Schürmann, C., Skipka, G., Bender, R. (2022). Performing meta-analyses with very few studies. In: Evangelou, V. & Veroniki, A.A. (Eds.): Meta-Research, pp. 91-102. Humana, New York.
- Turner, R. M., Jackson, D., Wei, Y., Thompson, S. G., Higgins, J. P. (2015). Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian meta-analysis. Statistics in Medicine, 34(6), 984-998.

Institute for Quality and Efficiency in Health Care (IQWiG)

Siegburger Str. 237
50679 Köln

Telefon	**+49 221 35685 - 0**
Telefax	**+49 221 35685 – 1**
Mail:	**info@iqwig.de**

www.iqwig.de
www.gesundheitsinformation.de
www.themencheck-medizin.de