European Statistical Meeting
Berlin, 15.02.2019

IQWiG Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
Institute for Quality and Efficiency in Health Care

Meta-analysis with few studies: Bayesian approach

Ralf Bender

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

Outline

- Introduction
  - Bayesian vs. frequentist methods
  - IQWiG methods paper
- Methods for evidence synthesis
  - Fixed-effect (FE) and random effects (RE) model
  - Qualitative and quantitative evidence synthesis
- Meta-analysis with very few studies
  - Estimation of heterogeneity parameter problematic
  - Bayesian approach with informative prior
- Discussion
- Conclusion
- References

Introduction

Definition of Bayesian methods in HTA:

"The explicit quantitative use of external evidence in the design, monitoring, analysis, interpretation, and reporting of a health technology assessment."

(Spiegelhalter et al., 1999)

With this very general definition almost all HTA reports are based upon Bayesian methods, because almost always multiple sources are used, e.g., the main meta-analysis of RCTs for the benefit assessment AND registry data for epidemiological questions.

My understanding

Frequentist methods:

- Point and interval estimation of relevant parameters
- Significance testing
- Output: Point estimates, confidence intervals, *p*-values

Bayesian methods:

- Specification of prior distributions
- Calculation of posteriori distributions from prior distribution and likelihood
- Output: Expected values, credible intervals, Bayes factors

The IQWiG methods paper

- Version 1 (2005):
  Just a note that Bayesian methods exist in the context of model uncertainty.

- Versions 2 (2006) and 3 (2008):
  Bayesian methods mentioned as general alternative to frequentist methods and that IQWiG will apply Bayesian methods "where necessary".

- Versions 4.0 (2011) and 4.1 (2013):
  Designation of indirect comparisons as possible application area for Bayesian methods.

https://www.iqwig.de/de/methoden/methodenpapier.3020.html

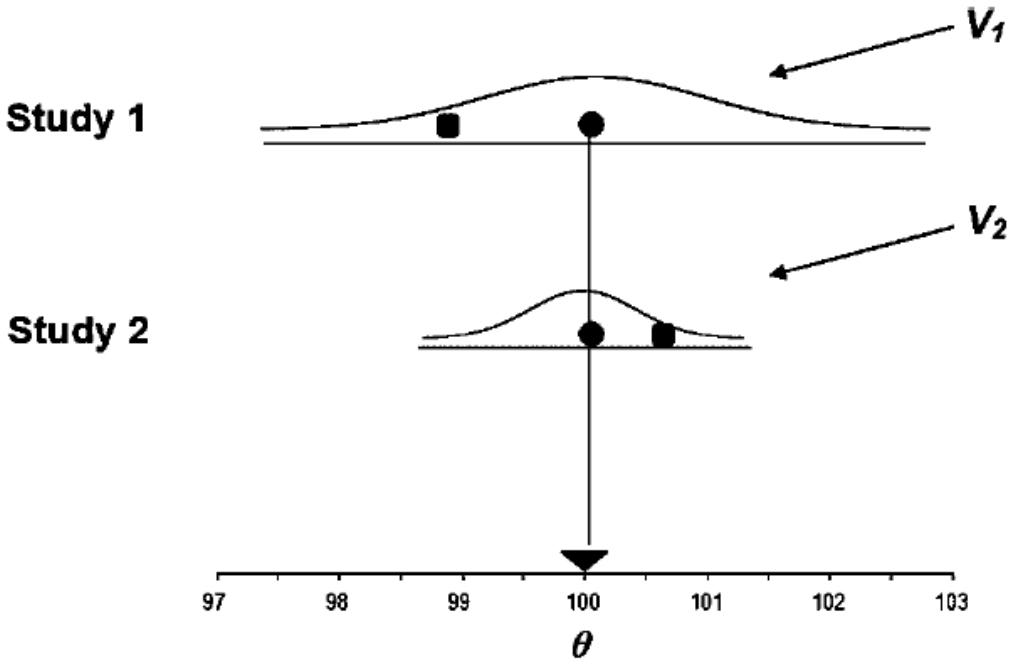
The IQWiG methods paper

- Version 4.2 (2015):
  Use of Bayesian methods mentioned for health economic evaluations and indirect comparisons.
- Version 5.0 (2017):
  Use of Bayesian methods mentioned for health economic evaluations, indirect comparisons, and pairwise meta-analyses with very few studies.

https://www.iqwig.de/de/methoden/methodenpapier.3020.html

Generic models

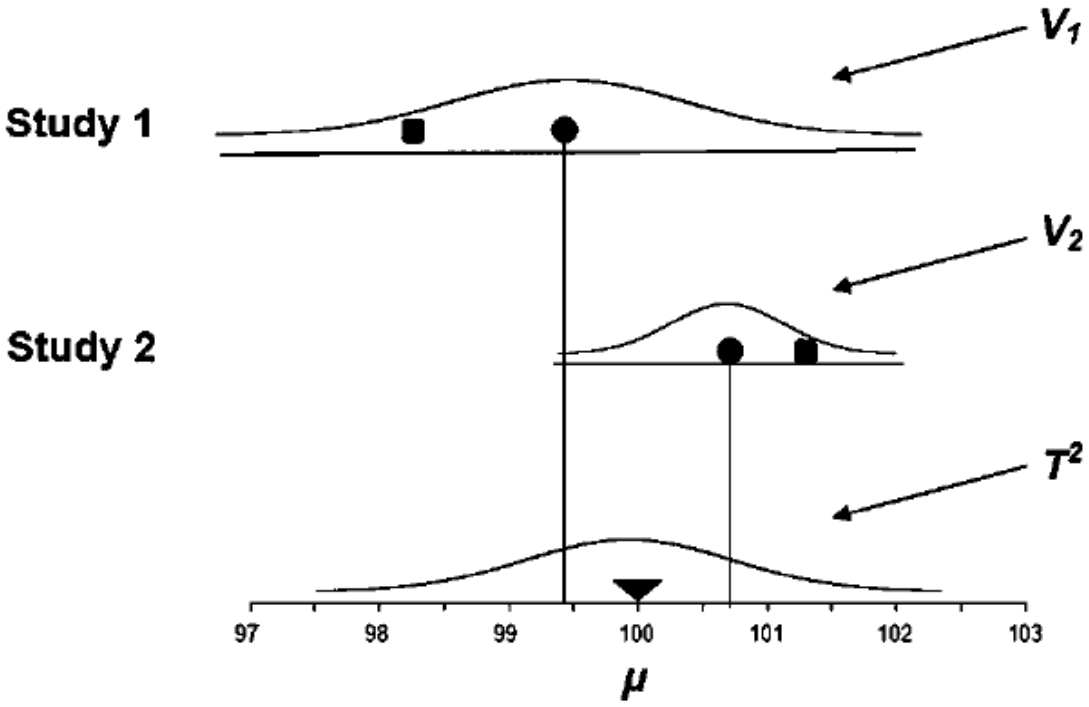
- **Fixed-effect (FE) model**
  - $y_i = \theta_{FE} + \varepsilon_i \ , \ \ \varepsilon_i \sim N(0, v_i) \ , \ \ Var(y_i) = v_i$
  - Assumption: No heterogeneity, same true effect in all studies
  - Parameter of interest: True common treatment effect θ_{FE}



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

Generic models

- **Random-effects (RE) model**
  - $y_i = \theta_i + \varepsilon_i$, $\theta_i = \theta_{RE} + \delta_i$, $\varepsilon_i \sim N(0, v_i)$, $\delta_i \sim N(0, \tau^2)$, $Var(y_i) = v_i + \tau^2$
  - Assumption: Heterogeneity, distribution of true effects
  - Parameter of interest: Mean of the true treatment effects θ_{RE}



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

Methods for evidence synthesis

Qualitative evidence synthesis

- If heterogeneity too large ⇒ no meta-analysis
- Only choice: Qualitative evidence synthesis
- Nevertheless, clear statements possible
- Example:
  2 studies with significant beneficial results in the same direction

  ⇒ Proof of benefit

  But quantification of the effect size is not possible

- Example:
  Category in AMNOG (Germany): "non-quantifiable added benefit"

Meta-analysis with fixed effect

- Inverse variance approach for continuous endpoints
- Effect 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 normal distribution
- For binary endpoints also applicable but not recommended
- For binary data:
  - Mantel-Haenszel method
  - Peto method
  - Beta-binomial model (or other GLMs)

Meta-analysis with random effects

- DerSimonian & Laird (DSL) method criticized (Cornell et al. , 2014)
- DSL ignores estimation uncertainty of τ and v_i
- Effect estimate: $\hat{\theta}_{RE} = \frac{\sum_{i=1}^{k} y_i w_{i,RE}}{\sum_{i=1}^{k} w_{i,RE}}$, with $w_{i,RE} = 1/(\hat{v}_i + \hat{\tau}^2)$
- τ^2 estimated by using the method of moments (DSL)
- 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 can lead to a strongly increased type 1 error

DSL for RE meta-analysis

IQWiG Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
Institute for Quality and Efficiency in Health Care

Table I. Overview of the empirical type I error and power.

		Empirical type I error for $p_C = p_T = 0.2$			Empirical power for $p_T = 0.2$ and $p_T = 0.3$		
k	Het (mean I^2)	FE	DL	HK	FE	DL	HK
2	0.15	0.0466	0.0382	0.0481	0.7171	0.6074	0.1487
3	0.15	0.0459	0.0352	0.0477	0.7142	0.6169	0.2976
4	0.14	0.0417	0.0311	0.0473	0.6965	0.6146	0.3999
5	0.13	0.0391	0.0308	0.0473	0.7008	0.6267	0.4720
6	0.12	0.0373	0.0306	0.0447	0.6808	0.6147	0.5015
2	0.25	0.1313	0.0895	0.0469	0.6501	0.4980	0.1137
3	0.25	0.1016	0.0684	0.0525	0.6537	0.5030	0.2115
4	0.25	0.0861	0.0613	0.0467	0.6552	0.5113	0.2762
5	0.25	0.0900	0.0614	0.0467	0.6463	0.5030	0.3148
6	0.25	0.0835	0.0574	0.0423	0.6302	0.4948	0.3395
2	0.50	0.4142	0.2184	0.0489	0.5998	0.3447	0.0706
3	0.50	0.2884	0.1367	0.0493	0.5892	0.3362	0.1131
4	0.50	0.2391	0.1104	0.0467	0.5814	0.3377	0.1476
5	0.50	0.2231	0.0956	0.0443	0.5535	0.3089	0.1611
6	0.50	0.2077	0.0864	0.0421	0.5541	0.3171	0.1767
2	0.75	0.7306	0.2866	0.0639	0.7455	0.3050	0.0567
3	0.75	0.5384	0.1786	0.0509	0.6097	0.2307	0.0695
4	0.75	0.4664	0.1385	0.0501	0.5673	0.2082	0.0747
5	0.75	0.4303	0.1223	0.0466	0.5473	0.1982	0.0853
6	0.75	0.4023	0.1114	0.0468	0.5263	0.1936	0.0900

k, number of studies; Het, heterogeneity; FE, fixed effects approach; DL, DerSimonian and Laird approach; HK, Hartung and Knapp approach; p_c, event rate in control group; p_T, event rate in treatment group.

Gonnermann et al., 2015

Meta-analysis with random effects

- Knapp-Hartung (KH) method recommended (Veroniki et al., 2019)
- Effect estimate: $\hat{\theta}_{RE} = \frac{\sum_{i=1}^{k} y_i w_{i,RE}}{\sum_{i=1}^{k} w_{i,RE}}$, with $w_{i,RE} = 1/(\hat{v}_i + \hat{\tau}^2)$
- τ^2 estimated by using the iterative 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

($z_{0.975}$ =**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**)

- KH in general holds type-1 error

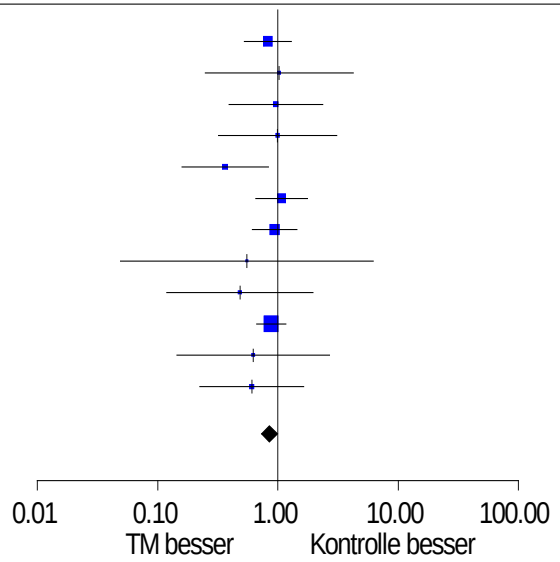
KH method for RE meta-analysis

- 95% CI: $\hat{\theta}_{RE} \pm t_{k-1,1-\frac{\alpha}{2}} \sqrt{\frac{\sum_{i=1}^{k} w_{i,RE}(y_i-\hat{\theta}_{RE})^2}{(k-1)\sum_{i=1}^{k} w_{i,RE}}}$
- If study results very homogenous:
  CI of KH method can be misleadingly narrow
- Ad hoc variance correction (Knapp & Hartung, 2003)
- $Var(\hat{\theta}_{RE}) = max\left[\frac{1}{\sum_{i=1}^{k} w_{i,RE}}, \frac{\sum_{i=1}^{k} w_{i,RE}(y_i-\hat{\theta}_{RE})^2}{(k-1)\sum_{i=1}^{k} w_{i,RE}}\right]$
- Misleadingly narrow CIs avoided
- In practice, results of KH should always be compared with DSL to avoid misleadingly narrow CIs (Jackson et al., 2017)

Example: IQWiG report N16-02

Telemonitoring vs. Kontrolle
kardiovaskuläre Mortalität
Modell mit zufälligen Effekten - Knapp und Hartung

Studie	TM n/N	Kontrolle n/N	Gewichtung	OR	95%-KI
CONNECT	36/1014	42/983	14.5	0.82	[0.52, 1.30]
CONNECT-OptiVol	4/87	4/89	1.5	1.02	[0.25, 4.23]
ECOST	10/211	10/203	3.7	0.96	[0.39, 2.36]
effecT	7/88	6/75	2.3	0.99	[0.32, 3.10]
IN-TIME	8/333	21/331	4.4	0.36	[0.16, 0.83]
MORE-CARE	36/437	33/428	12.4	1.07	[0.66, 1.76]
OptiLink HF	46/505	48/497	16.6	0.94	[0.61, 1.43]
QUANTUM	1/70	2/78	0.5	0.55	[0.05, 6.21]
REDUCEhf	3/202	6/198	1.5	0.48	[0.12, 1.96]
REM-HF	107/824	120/826	38.1	0.88	[0.66, 1.16]
TELECART	3/89	5/94	1.4	0.62	[0.14, 2.68]
TRUST	9/908	7/431	3.0	0.61	[0.22, 1.64]
Gesamt	270/4768	304/4233	100.0	0.85	[0.73, 0.99]



Heterogenität: Q=6.77, df=11, p=0.818, I^2=0%
Gesamteffekt: Z Score=-2.32, p=0.040, Tau(Paule-Mandel)=0

- **Results**
  - $\hat{\theta}_{KH} = 0.85$, 95% CI: [0.73, 0.99]
  - $\hat{\theta}_{DSL} = 0.85$, 95% CI: [0.72, 1.01]
  - $\hat{\theta}_{KH-K} = 0.85$, 95% CI: [0.70, 1.03]

Bayesian methods

- Competitive alternative to frequentist methods of meta-analysis is given by Bayesian methods
- Bayesian methodology allows the inclusion of prior knowledge about the unknown parameters in the form of prior distributions
- Inferences about the effects of interest are made by integrating out the unknown parameters from the joint distribution of the prior and the likelihood
- Usually noninformative prior distributions are chosen for the unknown parameters
- In decision-making no difference is made between confidence intervals from frequentist methods and credible intervals from Bayesian methods

Meta-analyses with very few studies

Situation

- Fixed-effect (FE) model
  - Assumption: No true heterogeneity
  - Frequently not adequate
- Random-effects (RE) model
  - Assumption: True heterogeneity (not too large)
  - Knapp-Hartung (KH) method recommended (Veroniki et al., 2019)
  - Problem:
    In the case of very few studies τ cannot be estimated reliably

→ KH method over-conservative in the case of very few (2-4) studies

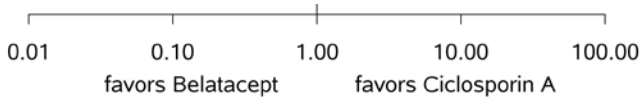
Example

Belatacept after kidney transplant (2 significant studies)

- Belatacept vs ciclosporin A for prophylaxis of graft rejection in adults receiving a renal transplant (IQWiG report A15-25)
- Endpoint "renal insufficiency in chronic kidney disease stage 4/5"

Figure 1
Belatacept vs. Ciclosporin A
Renal insufficiency in chronic kidney disease

Study	log(HR)	SE	HR (95% CI)	weight (DSL)	HR	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]
B-HN(0.5)					0.53	[0.27, 0.98]
B-HN(1.0)					0.52	[0.17, 1.52]



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

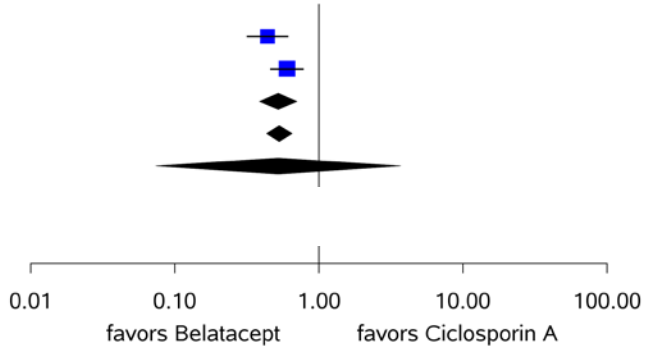
Example

Belatacept after kidney transplant (2 significant studies)

- Belatacept vs ciclosporin A for prophylaxis of graft rejection in adults receiving a renal transplant (IQWiG report A15-25)
- Endpoint "renal insufficiency in chronic kidney disease stage 4/5"

Figure 1
Belatacept vs. Ciclosporin A
Renal insufficiency in chronic kidney disease

Study	log(HR)	SE	HR (95% CI)	weight (DSL)	HR	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]
B-HN(0.5)					0.53	[0.27, 0.98]
B-HN(1.0)					0.52	[0.17, 1.52]



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) KH over-conservative

2) Decision of no added benefit is critical

Current practice in IQWiG

- If heterogeneity is too large for a meaningful pooling of the available study results, apply **qualitative evidence synthesis**
- If the FE assumption seems to be not violated too strongly, apply **FE meta-analysis**
- In the case where the pooling of study results seems to be meaningful despite of heterogeneity, apply **RE meta-analysis** by using the **KH method**; but compare the results with the qualitative evidence synthesis (see Bender et al., 2018)
- If the qualitative evidence synthesis yields a proof of added benefit, this result overrules a non-significant result of the KH method (proof of added benefit, but with non-quantifiable extent)

Bayesian methods

- Bayesian methodology allows the inclusion of prior knowledge about the heterogeneity parameter in the form of (weakly) informative prior distributions (Friede et al., 2017)
- Compromise between over-confident FE meta-analysis and over-conservative RE meta-analysis based upon KH method ?
- Reliable information on the prior distribution of the unknown parameters is required

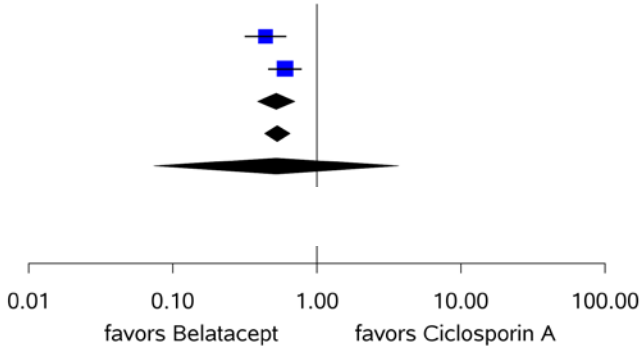
Example (continued)

Belatacept after kidney transplant (2 significant studies)

- Belatacept vs ciclosporin A for prophylaxis of graft rejection in adults receiving a renal transplant (IQWiG report A15-25)
- Endpoint "renal insufficiency in chronic kidney disease stage 4/5"

Figure 1
Belatacept vs. Ciclosporin A
Renal insufficiency in chronic kidney disease

Study	log(HR)	SE	HR (95% CI)	weight (DSL)	HR	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]
B-HN(0.5)					0.53	[0.27, 0.98]
B-HN(1.0)					0.52	[0.17, 1.52]



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

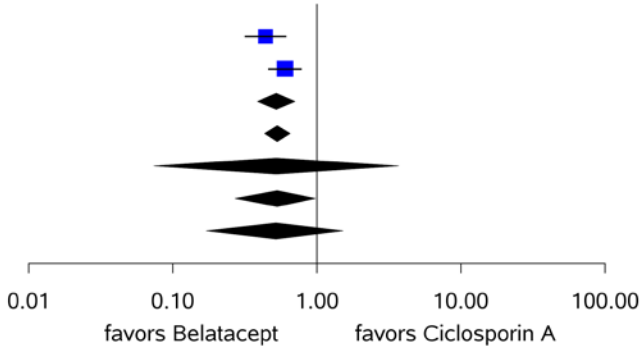
Example (continued)

Belatacept after kidney transplant (2 significant studies)

- Belatacept vs ciclosporin A for prophylaxis of graft rejection in adults receiving a renal transplant (IQWiG report A15-25)
- Endpoint "renal insufficiency in chronic kidney disease stage 4/5"

Figure 1
Belatacept vs. Ciclosporin A
Renal insufficiency in chronic kidney disease

Study	log(HR)	SE	HR (95% CI)	weight (DSL)	HR	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]
B-HN(0.5)					0.53	[0.27, 0.98]
B-HN(1.0)					0.52	[0.17, 1.52]



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) Bayesian approach a compromise between FE and RE
2) But the final result is dependent on the prior distribution

Choice of prior for τ

- For binary data, use of half-normal priors with scale 0.5 and 1 for τ are suggested (Friede et al., 2017)
- Pullenayegum (2011) proposed to use the lognormal distribution as prior
- Bodnar et al. (2016) proposed to apply the Berger & Bernado reference prior principle (Jeffreys prior)
- Another suggestion is to use empirical data from the *Cochrane Database of Systematic Reviews* (Turner et al., 2015; Rhodes et al., 2015)
- Alternative: Use of expert beliefs (Ren et al., 2018)
- The choice of the prior distributions is important, especially in the case of sparse data (Bodnar et al., 2016; Weber et al., 2018)

Choice of prior for τ

- Informative prior distributions may determine the final conclusions in the case of sparse data
- It cannot be expected that a clear-cut choice for reliable prior information is available for all intervention types and all medical disciplines
- Additional suggestions for prior distributions in the case of continuous data are required

→ A general scientific agreement is needed which distribution for the heterogeneity parameter is valid for which situation

Discussion

- In general, whenever heterogeneity cannot be excluded, the FE model should not be used
- However, in situations with only 1 single study, results of this study are interpreted and conclusions are made for the considered population
- In the case of 2 or more studies we can technically investigate heterogeneity and we try to assess heterogeneity even if heterogeneity cannot reliably estimated
- To avoid a break in the assessment of study results between the situations with 1 and 2 studies, the simple FE model should be applied more frequently (Bender et al., 2018)
- Especially in the case of 2 studies, the situation of twin studies (i.e., 2 studies with identical design) occurs in practice, which justifies the use of the FE model

Discussion

- If the FE model is clearly not adequate, the RE model should be used for meta-analysis
- In general, the KH method should be used for RE meta-analyses
- If there is a conflict between the results of the KH method and a qualitative evidence synthesis, alternative methods are required
- One option is to use generalized linear models, e.g. the beta-binomial model
- Another option is to use Bayesian methods with informative prior distribution for the heterogeneity parameter

Conclusion

- No satisfactory standard method is currently available to perform meta-analyses in the case of very few studies
- Bayesian methods with informative priors represent a compromise between over-confident FE meta-analysis and over-conservative RE meta-analysis
- A general scientific agreement is required which prior distribution for the heterogeneity parameter is valid for which situation

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**.**
- Bodnar, O., Link, A., Arendacka, B., Possolo, A. & Elster, C. (2017): Bayesian estimation in random effects meta-analysis using a non-informative prior. *Stat. Med.* **36**, 378-399.
- Friede, T., Röver, C., Wandel, S. & Neuenschwander, B. (2017): Meta-analysis of few small studies in orphan diseases. *Res. Syn. Methods* **8**, 79-91.
- Jackson, D., Law, M., Rücker, G. & Schwarzer, G. (2017): The Hartung-Knapp modification for random-effects meta-analysis: A useful refinement but are there any residual concerns? *Stat. Med.* **36**, 3923-3934.
- Knapp, G. & Hartung, J. (2003): Improved tests for a random effects meta-regression with a single covariate. *Stat. Med.* **22**, 2693-2710.
- Pullenayegum, E.M. (2011): An informed reference prior for between-study heterogeneity in meta-analyses of binary outcomes. *Stat. Med.* **30**, 3082-3094.
- 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. *J. Clin. Epidemiol.* **68**, 52-60.
- Spiegelhalter, D.J., Myles, J.P., Jones, D.R. & Abrams, K.R. (1999): An introduction to Bayesian methods in health technology assessment. *BMJ* **319**, 508-512.
- Ren, S., Oakley, J.E. & Stevens, J.W. (2018): Incorporating genuine prior information about between-study heterogeneity in random effects pairwise and network meta-analyses. *Med. Decis. Making* **38**, 531-542.
- 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. *Stat. Med.* **34**, 984-998.
- 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** (in press)*.*
- Weber, K., Hemmings, R. & Koch, A. (2018): How to use prior knowledge and still give new data a chance? *Pharm. Stat.* **17**, 329-341.