The empirical distribution of $\tau$ from IQWiG reports for the application in Bayesian random-effects meta-analyses

Ralf Bender,
Matthias Maiworm, Sibylle Sturtz
(IQWiG, Germany)
Outline

- Introduction
  - Meta-analysis with very few studies
  - Example
  - Bayesian methods

- Methods
  - Suggestions for prior distributions
  - Meta-analyses from IQWiG reports

- Results
- Interim conclusion
- Outlook
- References
Introduction

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 (2-4) studies $\tau$ cannot be estimated reliably (Bender et al., 2018)

KH method is over-conservative in the case of very few studies

Currently we apply FEM or a qualitative evidence synthesis, but this is circumstantial …
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

<table>
<thead>
<tr>
<th>Study</th>
<th>log(HR)</th>
<th>SE</th>
<th>HR (95% CI)</th>
<th>weight (DSL)</th>
<th>HR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>BENEFIT</td>
<td>-0.82</td>
<td>0.17</td>
<td></td>
<td>44.6</td>
<td>0.44</td>
<td>[0.32, 0.61]</td>
</tr>
<tr>
<td>BENEFIT-EXT</td>
<td>-0.51</td>
<td>0.13</td>
<td></td>
<td>55.4</td>
<td>0.60</td>
<td>[0.46, 0.78]</td>
</tr>
<tr>
<td>DSL</td>
<td></td>
<td></td>
<td></td>
<td>100.0</td>
<td>0.52</td>
<td>[0.39, 0.71]</td>
</tr>
<tr>
<td>CE IV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.53</td>
<td>[0.43, 0.65]</td>
</tr>
<tr>
<td>KH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.52</td>
<td>[0.07, 3.71]</td>
</tr>
<tr>
<td>B-HN(0.5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.53</td>
<td>[0.27, 0.98]</td>
</tr>
<tr>
<td>B-HN(1.0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.52</td>
<td>[0.17, 1.52]</td>
</tr>
</tbody>
</table>

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) Knapp-Hartung is over-conservative
2) Decision of no significant overall effect is critical
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

<table>
<thead>
<tr>
<th>Study</th>
<th>log(HR)</th>
<th>SE</th>
<th>HR (95% CI)</th>
<th>weight (DSL)</th>
<th>HR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>BENEFIT</td>
<td>-0.82</td>
<td>0.17</td>
<td></td>
<td>44.6</td>
<td>0.44</td>
<td>[0.32, 0.61]</td>
</tr>
<tr>
<td>BENEFIT-EXT</td>
<td>-0.51</td>
<td>0.13</td>
<td></td>
<td>55.4</td>
<td>0.60</td>
<td>[0.46, 0.78]</td>
</tr>
<tr>
<td>DSL</td>
<td></td>
<td></td>
<td></td>
<td>100.0</td>
<td>0.52</td>
<td>[0.39, 0.71]</td>
</tr>
<tr>
<td>CE IV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.53</td>
<td>[0.43, 0.65]</td>
</tr>
<tr>
<td>KH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.52</td>
<td>[0.07, 3.71]</td>
</tr>
<tr>
<td>B-HN(0.5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.53</td>
<td>[0.27, 0.98]</td>
</tr>
<tr>
<td>B-HN(1.0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.52</td>
<td>[0.17, 1.52]</td>
</tr>
</tbody>
</table>

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) Bayesian approach = Compromise between DSL and KH  
2) But the final result depends on the prior distribution
Prior distributions

- Bayes: Posterior $\propto$ prior $\times$ likelihood

- Random-effects meta-analysis:
  
  $y_i = \theta_i + \varepsilon_i$, $\theta_i = \theta_{RE} + \delta_i$  
  $\varepsilon_i \sim N(0, \nu_i)$, $\delta_i \sim N(0, \tau^2)$, $Var(y_i) = \nu_i + \tau^2$

- $P((\theta_{RE}, \tau^2) | \text{data}) \propto P((\theta_{RE}, \tau^2)) \times P(\text{data} |(\theta_{RE}, \tau^2))$

- For overall mean effect $\theta_{RE}$: Non-informative prior

- For heterogeneity parameter $\tau$: Weakly informative prior to overcome limitations in the case of few studies  
  (Friede et al., 2017; Röver et al., 2021)
Potential prior distributions for $\tau$:

See Röver et al. (2021)
Prior distributions

- For pragmatic reasons we concentrate at first on half-normal distribution (Röver et al., 2021)

Comparison of HN(0.5) and HN(1.0) with the lognormal distribution proposed by Turner et al. (2015)

Which distribution is suitable in the HTA framework?
Methods

- Collection of all meta-analyses of IQWiG reports from 2005 to June 2020
- Random-effects meta-analysis by means of Knapp-Hartung (IQWiG, 2020)
- Estimation of $\tau$ by means of Paule-Mandel
- Conditions:
  - No meta-analyses for sensitivity/specificity
  - No subgroup analyses
  - No sensitivity analyses
  - Fourfold table available: Calculation of OR and RR
- Histograms to illustrate the empirical distribution of $\tau$
- Comparison with HN(0.5) and HN(1.0)
Results

- Data basis:
  - 653 IQWiG reports
  - 118 reports with meta-analyses (forest plot)
  - 1653 meta-analyses

- Effect measures: OR, RR, SMD, (HR)

- In more than 75% of meta-analyses the number of studies is smaller than 5!

- Restrictions:
  - Only estimates of \( \tau \) larger than zero
  - Only meta-analyses without substantial heterogeneity (Q-test not significant)
Results

Problem:
In about 60% of meta-analyses zero estimates for $\tau$ are obtained (similar to others).

Further restriction:
It makes sense to include only meta-analyses where heterogeneity is not too large for a meaningful pooled effect estimation.

Number of meta-analyses with non-zero estimates for $\tau$ and no substantial heterogeneity:

OR: 243 meta-analyses
RR: 260 meta-analyses
SMD: 166 meta-analyses
(HR: 21 meta-analyses)
The empirical distribution of $\tau$ from IQWiG reports seems to be suitable for OR.

$\text{HN}(0.5)$ distribution seems to be suitable for OR.
The empirical distribution of $\tau$ from IQWiG reports for the application in Bayesian random-effects meta-analyses.

Results

$\text{HN}(0.5)$ distribution seems to be suitable for RR.
Results

Distribution with smaller scale than $\text{HN}(0.5)$ for SMD?

Density

$\tau$
Interim conclusion

- First results are promising
- **HN(0.5)** seems to be suitable for OR and RR (and HR)
- For SMD a distribution with smaller scale parameter seems to be possible
- Pragmatic approach: Use of the same prior distribution for all effect measures, e.g., HN(0.5)
Outlook

- Application of various prior distributions (e.g., HN(0.5), HN(1.0), lognormal, Cauchy) to the IQWiG database of meta-analyses

- Key question: Can the use of qualitative evidence synthesis be avoided by means of Bayesian meta-analysis?

- If possible, decision for a suitable standard prior distribution (together with experts from biometric societies in Germany)

- Application of Bayesian meta-analyses with the chosen standard prior distribution for $\tau$ in the case of very few studies in the future
The empirical distribution of $\tau$ from IQWiG reports for the application in Bayesian random-effects meta-analyses

References


