Applications of Bayesian methods in health technology assessment

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

- Introduction
  - Bayesian vs. frequentist methods
  - IQWiG methods paper

- Bayesian methodology in HTA
  - Clinical trials
  - Economic evaluations
  - (Network) meta-analysis

- Meta-analysis with very few studies

- Discussion

- Conclusion

- References
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.
Introduction

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
Applications in clinical trials:

- Sample size calculation
- Dose-response experiments
- Monitoring of clinical trials
- Use of historical controls
- …

(Spiegelhalter & Freedman, 1994; Ashby, 2006)
Bayesian methods in HTA

Evidence synthesis:

- Pairwise meta-analysis
- Network meta-analysis
- Meta-regression
- Multi-level models

Health economic models:

- Health economic decision models with parameter uncertainty
- Probabilistic methods for Bayesian networks
Bayesian methods in IQWiG reports

Use of frequentist methods:

- Usual methods for parameter estimation and significance testing
- Pairwise meta-analysis, meta-regression

Use of Bayesian methods:

- Network meta-analysis
- Reason: The first complex methods for network meta-analysis were developed in a Bayesian framework (Lu & Ades, 2004)
Example: G09-01: Antidepressants

- Health economic evaluation of venlafaxine, duloxetine, bupropion, and mirtazapine compared to further prescribable pharmaceutical treatments
- Markov model was used for health economic evaluation
- Effect estimates of meta-analyses, indirect comparisons (Bucher method) and network meta-analyses were used as input for the Markov model
- For network meta-analysis Bayesian methods using MCMC and uninformative prior distributions were applied (Sturtz & Bender, 2012)
- Reason: The frequentist methods for network meta-analyses available at this time could not deal with multi-arm trials
Example: A16-70: Rheumatoid arthritis

- Benefit assessment of biotechnologically produced drugs for the treatment of rheumatoid arthritis
- Comparison of 9 drugs
- Network meta-analysis
- Application of R package `netmeta` (Schwarzer et al., 2015)
- Use of frequentist methods now available (even for multi-arm trials)
- Simulation study demonstrated slightly better results for `netmeta` compared to Bayesian methods (Kiefer, 2015)
- No (arbitrary) choice of prior distributions required
Use of Bayesian methods in IQWiG?

- For network meta-analysis, Bayesian methods no longer required.
  - Reason: Application of R package `netmeta`.

- No application of Bayesian health economic models.
  - Reason: Currently no commission for health economic evaluations by the Joint Federal Committee.

→ **No room for Bayesian methods in IQWiG?**
Use of Bayesian methods in IQWiG?

Bayesian methods still play a role:

- For network meta-analysis Bayesian methods no longer required, but nevertheless a valid option (at least for sensitivity analyses etc.)

- Bayesian methods may play a major role for meta-analyses with very few trials in the future
Meta-analyses with very few studies

Situation

- Fixed-effect (FE) model
  - Assumption: No true heterogeneity

- Random-effects (RE) model
  - Assumption: True heterogeneity (not too large)
  - DerSimonian & Laird (DSL) method (DerSimonian & Laird, 1986)
  - DSL ignores estimation uncertainty of $\tau$ (Veroniki et al., 2018)
  - A number of improved methods available
  - Knapp-Hartung (KH) method recommended (Veroniki et al., 2018)
  - Problem:
    In the case of very few studies $\tau$ cannot be estimated reliably

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

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

- It may be possible 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)
Methods for evidence synthesis

Bayesian methods

- However, it cannot be expected that a clear-cut choice for reliable prior information is available for all intervention types and all medical disciplines.

- For binary data, use of half-normal priors with scale 0.5 and 1 for $\tau$ suggested (Friede et al., 2017).

- Even if these values are adequate, a decision is required which of these priors should be used.

- A general scientific agreement is required which distribution for the heterogeneity parameter is valid for which situation.
**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, 0.71]</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: $Q=2.06$, $df=1$, $p=0.151$, $I^2=51.5\%$
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

<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></td>
<td>44.6</td>
<td>[0.32, 0.61]</td>
</tr>
<tr>
<td>BENEFIT-EXT</td>
<td>-0.51</td>
<td>0.13</td>
<td></td>
<td></td>
<td>55.4</td>
<td>[0.46, 0.78]</td>
</tr>
<tr>
<td>DSL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100.0</td>
<td>[0.39, 0.71]</td>
</tr>
<tr>
<td>CE IV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.52</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
**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, $I^2$=51.5%
Overall effect: $Z$ Score=4.21, $p$$<$$0.001$, $ Tau=0.157$

1) KH over-conservative; decision of no added benefit critical
2) Bayesian approach requires the decision of the "right" prior
Discussion

- No satisfactory standard method is currently available to perform meta-analyses in the case of very few studies
- FE model in practice possible, but has limitations (over-confident in the case of true heterogeneity)
- 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
- Thus, in the situation with very few studies, the simple FE model should be applied more frequently (Bender et al., 2018)
Conclusion

- Bayesian methods with informative priors may be a valid 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.

- Can this workshop be a starting point to reach such an agreement?
References