Methoden zur empirischen Untersuchung der Heterogenität in Metaanalysen

Christian Röver

Institut für Medizinische Statistik
Universitätsmedizin Göttingen

GMDS, Heilbronn, 19. September 2023

- **data**: k studies,
  effect estimates y_i, standard errors σ_i ($i = 1, \ldots, k$),
- **normal-normal hierarchical model (NNHM)**:

$$y_i|\mu,\tau \;\sim\; \mathrm{Normal}(\mu,\, \sigma_i^2 + \tau^2)$$

- **parameters**:
  overall mean μ, heterogeneity τ
- **priors** required for μ and τ
- **aim** (usually): inferring μ
  (sometimes also: prediction, shrinkage estimation, . . .)

Introduction

- Bayesian approach to meta analysis requires **prior distributions**, in particular: between-study **heterogeneity** (τ)
- **improper** or **proper (informative)** distributions may be appropriate [1]
  - **effect prior** (μ):
    (improper) uniform commonly works well, obvious alternative: normal (uncontroversial)
  - **heterogeneity prior** (τ):
    (improper) uniform Ok in case of *many studies*,
    proper distributions necessary/sensible in case of few studies [2]

    no obvious distribution family
    (certain families unsuitable/discouraged, scale parameter sensible)

[1] C. Röver. Bayesian random-effects meta-analysis using the `bayesmeta` R package. *Journal of Statistical Software*, **93**(6), 2020.

[2] R. Bender *et al.* Methods for evidence synthesis in the case of very few studies. *Research Synthesis Methods*, 9(3):381–392, 2018.

- (weakly) informative priors — different **motivations**:
  - epistemic (“classical”, prior *information*)
  - regularisation (“nudging parameters”)
  - pragmatic (operating characteristics)

- guidance on prior choice [3]
  - plausibility considerations
    (effect variability, prior predictive, shrinkage, effective sample sizes, . . .)
  - robustness considerations
    (greater heterogeneity = “more conservative”?
    stochastic ordering of priors?)
  - **empirical** data

[3] C. Röver, R. Bender, S. Dias, C. H. Schmid, H. Schmidli, S. Sturtz, S. Weber, T. Friede. On weakly informative prior distributions for the heterogeneity parameter in Bayesian random-effects meta-analysis. *Research Synthesis Methods*, **12**(4):448–474, 2021. arXiv: 2007.08352

Historical heterogeneity data

Some data sources

- previous empirical investigations,
  e.g., Kontopantelis *et al.* (2013)[4], Steel *et al.* (2015)[5],
  van Erp *et al.* (2017)[6], Seide *et al.* (2019)[7], Günhan *et al.* (2020)[8]
- here: utilize Seide *et al.* (2019) data set [9] as working example
  (40 meta-analyses from IQWiG publications,
  131 studies, log-OR endpoints)

[4] E. Kontopantelis *et al.* A re-analysis of the Cochrane Library data: The dangers of unobserved heterogeneity in meta-analyses. *PLoS ONE*, 8(7):e69930, 2013.

[5] P. Steel *et al.* Improving the meta-analytic assessment of effect size variance with an informed Bayesian prior. *Journal of Management*, 41(2):718–743, 2015.

[6] S. van Erp *et al.* Estimates of between-study heterogeneity for 705 meta-analyses reported in *Psychological Bulletin* from 1990–2013. *Journal of Open Psychology Data*, 5(7), 2017.

[7] S. Seide *et al.* Likelihood-based random-effects meta-analysis with few studies: Empirical and simulation studies. *BMC Medical Research Methodology*, 19:16, 2019.

[8] B.K. Günhan *et al.* Random-effects meta-analysis of few studies involving rare events. *Research Synthesis Methods*, 11(1):74–90, 2020.

[9] S. Seide *et al.* Meta-analysis data extracted from IQWiG publications. *Göttingen Research Online*, DOI: 10.25625/BWYBNK, 2018.

Historical heterogeneity data

General considerations

- sample of heterogeneity **estimates** is not immediately useful (would be overdispersed, commonly includes zeroes)
- need: coherent **meta-analysis** model for "historical" data
- heterogeneity is not easily/accurately quantified via estimate $\pm$ standard error — NNHM not useful here
- may (technically) be thought of as
  - several meta-analyses with an additional "top layer", *or*
  - an IPD meta-analysis where meta-analyses take on the "study role", and studies take the "patient role"
- modeling aspects: heterogeneity distribution family, parametrization, hyperpriors
- **aim**: modeling of (exchangeable) heterogeneity "population", eventually: **predictive distribution** of a "new" (future) heterogeneity value (as a prior for a future analysis)

Earlier work

Utilizing historical heterogeneity data

- idea originally introduced by Higgins and Whitehead (1996) [10] (?)
- derivation of more generally valid priors by Rhodes *et al.* (2015) [11] and Turner *et al.* (2015) [12]
  - applied to data from the *Cochrane Database of Systematic Reviews* (SMD and log-OR endpoints)
  - predictive distributions were summarized (tabulated) for specific outcome types
- aim here: to implement approach flexibly to try out distribution families and ways of summarizing predictive distributions
- eventual goal: to investigate IQWiG data (see following presentation!)

[10] J. P. T. Higgins, A. Whitehead. Borrowing strength from external trials in a meta-analysis. *Statistics in Medicine* 15(24):2733–2749, 1996.

[11] K.M. Rhodes *et al.* Predictive distributions were developed for the extent of heterogeneity in meta-analyses of continuous outcome data. *Journal of Clinical Epidemiology*, 68(1):52–60, 2015.

[12] R.M. Turner *et al.* Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian meta-analysis. *Statistics in Medicine*, 34(6):984–998, 2015.

The extended NNHM

For several meta-analyses; general specification

- **data**: N meta-analyses, each involving k_j studies, effect estimates y_{ij}, standard errors σ_{ij} ($i=1,\ldots,k_j$, $j=1,\ldots,N$),
- **NNHM stage**:
$$y_{ij}|\mu,\tau_j,\sigma_{ij} \;\sim\; \mathrm{Normal}(\mu_j,\, \sigma_{ij}^2+\tau_j^2)$$
- effect stage:
$$\mu_j|\mu_\mathrm{p},\sigma_\mathrm{p} \;\sim\; \mathrm{Normal}(\mu_\mathrm{p},\sigma_\mathrm{p}^2)$$
for fixed “neutral” μ_p and “large” σ_p ($\rightarrow$ stratification, no pooling)
- **heterogeneity stage**:
$$\tau_j|\theta \;\sim\; \mathrm{P}(\theta)$$
for some “heterogeneity distribution” $\mathrm{P}(\theta)$
- **parameters**: N means μ_j and heterogeneities τ_j; “distribution” parameter(s) θ
- (hyper-) **prior** required for θ
- **aim**: prediction $\tau^\star$

The extended NNHM

For several meta-analyses; half-normal example

- **data**: N meta-analyses, each involving k_j studies,
  effect estimates y_{ij}, standard errors σ_{ij} ($i = 1, \ldots, k_j$, $j = 1, \ldots, N$),
- **NNHM stage**:

$$y_{ij}|\mu, \tau_j, \sigma_{ij} \;\sim\; \text{Normal}(\mu_j,\, \sigma_{ij}^2 + \tau_j^2)$$

- effect stage:

$$\mu_j|\mu_\text{p}, \sigma_\text{p} \;\sim\; \text{Normal}(\mu_\text{p}, \sigma_\text{p}^2)$$

  for "neutral" μ_p and "large" σ_p (stratification)
- **heterogeneity stage**:

$$\tau_j|s \;\sim\; \text{half-Normal}(s)$$

- **parameters**:
  N means μ_j and heterogeneities τ_j; "distribution" parameter s
- (hyper-) **prior** required for half-normal scale s
  e.g.: $s \;\sim\; \text{Uniform}(0, 10)$
- **aim**: prediction $\tau^\star$

Implementation
Details

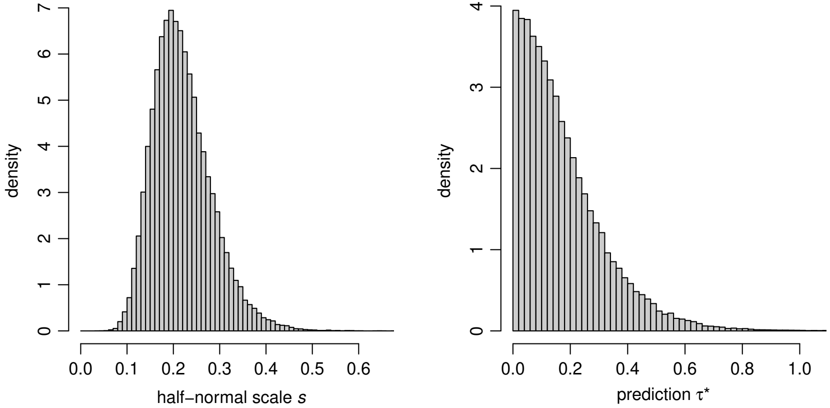
- implementation via MCMC (JAGS)
- several heterogeneity distributions:
  - half-Normal
  - exponential
  - half-Cauchy
  - log-Normal
  - ...

  (all parameterized via a scale parameter;
  only log-Normal with additional shape parameter)
- MCMC returns samples of (scale / shape) parameter(s) θ and predictions $\tau^\star$

Example application

Seide data set (40 meta-analyses), half-normal model

- posterior: scale parameter *s*
- posterior predictive: $\tau^\star | s \; \sim \;$ half-Normal(s)



- **posterior predictive** serves as **prior** for new analysis

Example application

Summarizing and communicating the posterior predictive: approaches

- **have**: MCMC samples
  - scale parameter s
  - $\tau^\star | s \;\sim\;$ half-Normal(s)
- **want**:
  - simple (analytical) form for posterior predictive ($\tau^\star$) distribution (simple implementation / communication)
- different ways of fitting a parametric distribution
  1. derive point estimate $\hat{s}$,
     use half-Normal($\hat{s}$) as an approximation
  2. summarize posterior distribution of s,
     derive/approximate (half-normal scale) mixture distribution
  3. fit a distribution to $\tau^\star$ sample
     (e.g., moment or ML estimation)

Example application

Summarizing and communicating the posterior predictive: application

1. point estimate $\hat{s}$:
   posterior mean $\bar{s} = 0.22 \rightarrow$ half-Normal(0.22) approximation

[13] C. Röver, R. Bender, S. Dias, C. H. Schmid, H. Schmidli, S. Sturtz, S. Weber, T. Friede. On weakly informative prior distributions for the heterogeneity parameter in Bayesian random-effects meta-analysis. *Research Synthesis Methods*, **12**(4):448–474, 2021. arXiv: 2007.08352

Example application

Summarizing and communicating the posterior predictive: application

1. point estimate $\hat{s}$:
   posterior mean $\bar{s} = 0.22 \rightarrow$ half-Normal(0.22) approximation

2. summarizing posterior of s:
   mean 0.22, s.d. 0.064 ($c_v = \frac{0.064}{0.22} = 0.29$)
   $\rightarrow$ approximate via *scaled inverse* χ *dist.* with 8.2 d.f. and scale 0.57
   $\rightarrow$ implies *half-Student-t* with 8.2 d.f. and scale $\frac{0.57}{\sqrt{8.2}} = 0.20$ for $\tau^\star$ [13]

[13] C. Röver, R. Bender, S. Dias, C. H. Schmid, H. Schmidli, S. Sturtz, S. Weber, T. Friede. On weakly informative prior distributions for the heterogeneity parameter in Bayesian random-effects meta-analysis. *Research Synthesis Methods*, **12**(4):448–474, 2021. arXiv: 2007.08352

Example application

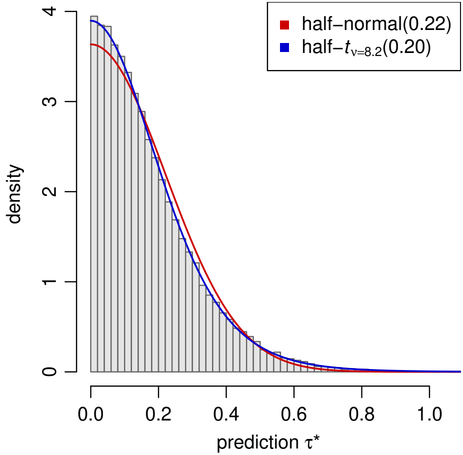
Summarizing and communicating the posterior predictive: application

1. point estimate $\hat{s}$:
   posterior mean $\bar{s} = 0.22 \rightarrow$ half-Normal(0.22) approximation
2. summarizing posterior of s:
   mean 0.22, s.d. 0.064 ($c_v = \frac{0.064}{0.22} = 0.29$)
   $\rightarrow$ approximate via *scaled inverse* χ *dist.* with 8.2 d.f. and scale 0.57
   $\rightarrow$ implies *half-Student-t* with 8.2 d.f. and scale $\frac{0.57}{\sqrt{8.2}} = 0.20$ for $\tau^\star$ [13]
3. fit a distribution to $\tau^\star$ sample
   motivation: use half-Student-t for a half-normal scale mixture
   moment or ML estimates: 8.2 d.f., scale 0.20 (again)

[13] C. Röver, R. Bender, S. Dias, C. H. Schmid, H. Schmidli, S. Sturtz, S. Weber, T. Friede. On weakly informative prior distributions for the heterogeneity parameter in Bayesian random-effects meta-analysis. *Research Synthesis Methods*, **12**(4):448–474, 2021. arXiv: 2007.08352

Example application

Summarizing and communicating the posterior predictive: illustration



- analytic approximations to posterior predictive ($\tau^\star$)
- differences/similarities between approaches (esp.: HN vs. HN-mixture) depend on uncertainty in scale parameter *s*

Example application

Model selection

- comparison of **alternative models**
  (for heterogeneity stage: half-normal, exponential, . . .)

	predictive distribution ($\tau^\star$)				
model	mean	st.dev.	50%	95%	99%
half-normal	0.17	0.15	0.14	0.46	0.66
exponential					
log-normal					
half-Cauchy					

Example application

Model selection

- comparison of **alternative models**
  (for heterogeneity stage: half-normal, exponential, . . .)

	predictive distribution ($\tau^\star$)				
model	mean	st.dev.	50%	95%	99%
half-normal	0.17	0.15	0.14	0.46	0.66
exponential	0.17	0.19	0.11	0.54	0.91
log-normal	0.26	3.46	0.11	0.61	1.90
half-Cauchy			0.08	1.10	5.46

- comparison of **alternative models**
  (for heterogeneity stage: half-normal, exponential, ...)
- **model selection** e.g. via deviance information criterion (DIC)
  (NB: "smaller = better")

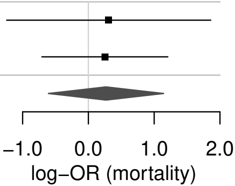
		predictive distribution ($\tau^\star$)				
model	DIC	mean	st.dev.	50%	95%	99%
half-normal	163.8	0.17	0.15	0.14	0.46	0.66
exponential	167.7	0.17	0.19	0.11	0.54	0.91
log-normal	178.0	0.26	3.46	0.11	0.61	1.90
half-Cauchy	212.8			0.08	1.10	5.46

Example application

Applying the derived heterogeneity prior

- consider "new" meta analysis ("41st study"), use half-Student-$t_{\nu=8.2}(0.20)$ distribution as "empirically motivated" prior distribution

study	estimate	95% CI
REVOLENS (Deslee; 2016)	0.31	[–1.24, 1.86]
RENEW (Sciurba; 2016)	0.25	[–0.71, 1.21]
mean	**0.27**	**[–0.61, 1.14]**



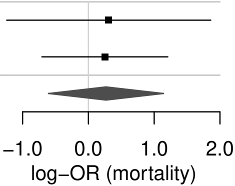
Heterogeneity (tau): 0.13 [0.00, 0.43]

Example application

Applying the derived heterogeneity prior

- consider "new" meta analysis ("41st study"), use half-Student-$t_{\nu=8.2}(0.20)$ distribution as "empirically motivated" prior distribution

study	estimate	95% CI
REVOLENS (Deslee; 2016)	0.31	[–1.24, 1.86]
RENEW (Sciurba; 2016)	0.25	[–0.71, 1.21]
mean	**0.27**	**[–0.61, 1.14]**



Heterogeneity (tau): 0.13 [0.00, 0.43]

- original (frequentist) analysis would lead to zero heterogeneity estimate (effectively a common-effect analysis)

Application

Additional considerations

- how to categorize meta-analyses?
  - need to distinguish between effect measures[14]
  - Rhodes *at al.* (2015) and Turner *at al.* (2015) considered subgroups by *outcome* and *intervention comparison*
  - need to find balance between generality / particularity / precision / complexity / simplicity
- How relevant are prior details (scale, family, . . .) in practice? [15]
- what if assumptions (exchangeability,. . .) are violated?
  - may include "robustification"
  - inclusion of a "vague" mixture component
  - deliberate choice of overdispersed / heavy-tailed prior (upper *and* lower tail?)
  - choice of stochastically larger prior (assuming that heterogeneity overestimation is a conservative form of bias)

[14] J.J. Deeks, D. Altman. Effect measures for meta-analysis of trials with binary outcomes. In: M. Egger, G. Davey Smith, D. Altman, eds. *Systematic reviews in health care: Meta-analysis in context*. London: BMJ Publishing. 2nd ed. 2001 (pp. 313-335).

[15] C. Röver, R. Bender, S. Dias, C. H. Schmid, H. Schmidli, S. Sturtz, S. Weber, T. Friede. On weakly informative prior distributions for the heterogeneity parameter in Bayesian random-effects meta-analysis. *Research Synthesis Methods*, **12**(4):448–474, 2021. Appendix D.4.

Summary

- in general: empirical information **one among several** motivations
- may also be considered as **complementary**
  or for "**sanity check**" / "**reality check**"
- additional considerations: e.g., operating characteristics, robustness, conservatism, ...
- eventual aim: to investigate / characterize "historical" IQWiG data, providing guidance for more general prior recommendations

Summary

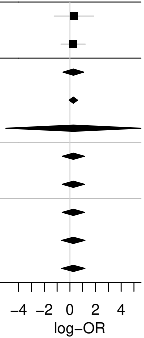
- in general: empirical information **one among several** motivations
- may also be considered as **complementary**
  or for "**sanity check**" / "**reality check**"
- additional considerations: e.g., operating characteristics, robustness, conservatism, ...
- eventual aim: to investigate / characterize "historical" IQWiG data, providing guidance for more general prior recommendations

- C. Röver, S. Sturtz, J. Lilienthal, R. Bender, T. Friede. Summarizing empirical information on between-study heterogeneity for Bayesian random-effects meta-analysis.
  Statistics in Medicine, **42**(14):2439–2454, 2023.
  `arXiv:2202.12538`

+++ additional slides +++

- 41st meta-analysis: several approaches (frequentist, and alternative priors)

study	treatment	control	log–OR	95% CI
REVOLENS	4 / 50	3 / 50	0.3092	[–1.2420, 1.8603]
RENEW	10 / 155	8 / 157	0.2504	[–0.7070, 1.2077]
Wald			**0.2666**	**[–0.5481, 1.0813]**
HKSJ			**0.2666**	**[–0.0675, 0.6007]**
mKH			**0.2666**	**[–5.0151, 5.5482]**
half–Normal(0.22)			**0.2675**	**[–0.6062, 1.1416]**
half–Student–t(8.2; 0.20)			**0.2675**	**[–0.6083, 1.1436]**
Lomax(9.9; 1.5)			**0.2674**	**[–0.6092, 1.1443]**
log–Normal(–2.6; 1.7)			**0.2673**	**[–0.6462, 1.1813]**
half–Cauchy(0.10)			**0.2674**	**[–0.6437, 1.1789]**



Example application

Predictions vs. estimates

- comparison of heterogeneity *estimates* ($\hat{\tau}$) vs. predictive distribution ($\tau^\star$): overdispersed estimates due to estimation uncertainty