Vorschlag für A-Priori-Verteilungen basierend auf empirischen Untersuchungen

gmds 2023, 19. September, ATF-Workshop

Jona Lilienthal[1]

Inhaltliche Zusammenarbeit mit

Sibylle Sturtz[1], Christoph Schürmann[1], Matthias Maiworm[1,3], Christian Röver[2], Tim Friede[2], Ralf Bender[1]

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

[2] University Medical Center Göttingen, Göttingen, Germany

[3] SALETELLIGENCE GmbH, Bielefeld, Germany

Situation

- IQWiG's current approach to evidence synthesis:
  Frequentist methods of meta-analyses combined with method of qualitative evidence synthesis
- Approach is complex (model choice, qualitative evidence synthesis)
- Difficulties in the case of 2 - 4 studies
  - Estimation of heterogeneity between studies unreliable
  - Knapp-Hartung can lead to inflated confidence intervals of the pooled estimate
- Bayesian meta-analyses as an alternative?

IQWiG's method of evidence synthesis (in short)

homogeneous

- 2 studies: FE meta-analysis
- 3 - 4 studies: RE meta-analysis (Knapp-Hartung method if proper)
  otherwise: qualitative evidence synthesis
- > 4 studies: RE meta-analsis (Knapp-Hartung)

het.

- qualitative evidence synthesis

Qualitative evidence synthesis

There is an overall effect if …

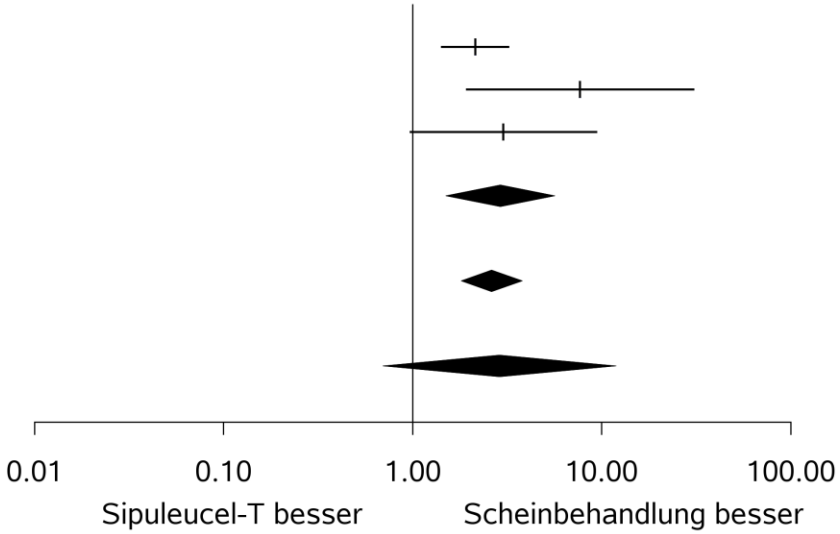
- all studies statistically significant (same direction, 5 % level)
- ≥ 3 studies:
  have a subset of studies with effects pointing towards same direction and
  - studies in subset have cum. weight > 80 %
  - stat. significant studies in subset have cum. weight > 50 %
- ≥ 4 studies:
  95 % prediction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

(Schulz et al. 2022)

Example

Sipuleucel-T vs. Scheinbehandlung
Fieber

Studie	Sipuleucel-T n/N	Scheinbehandlung n/N	Gewichtung	RR	95%-KI
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]
FEM - Mantel-Haenszel				2.62	[1.80, 3.80]
REM - Knapp-Hartung				2.88	[0.70, 11.92]



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

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

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

Prior distributions

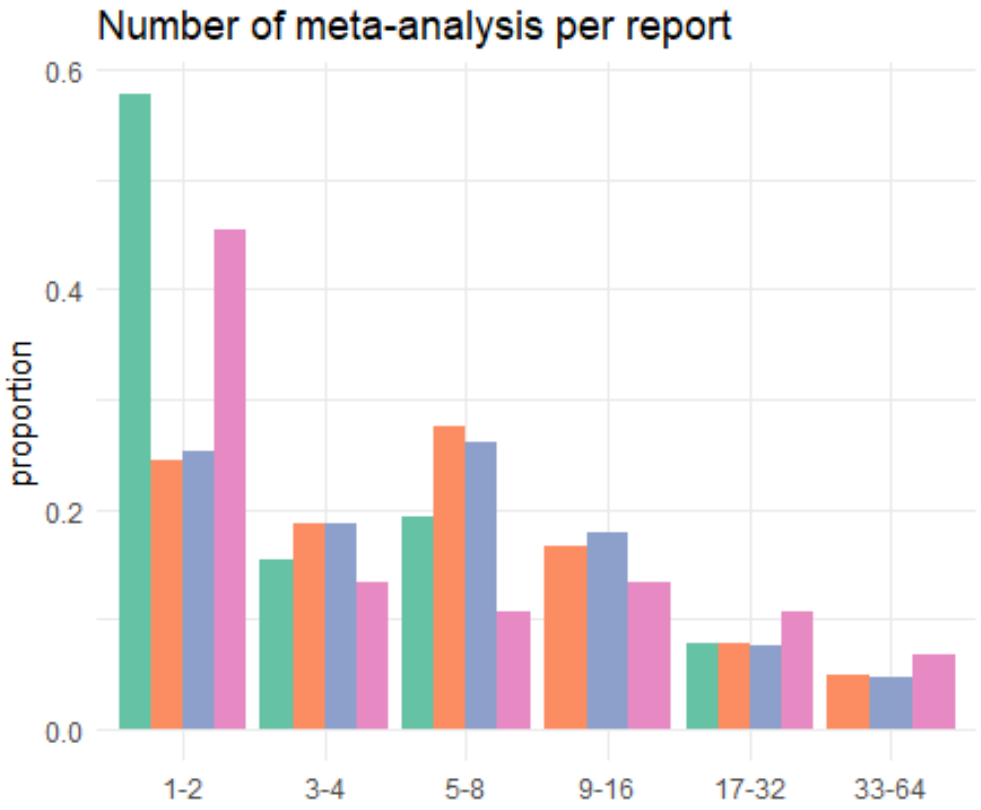
- Prior for effect: improper uniform or normal, e.g. $\mu \sim N(\mu_P, \sigma_P^2)$
- Prior for heterogeneity: $\tau \sim\ ?$

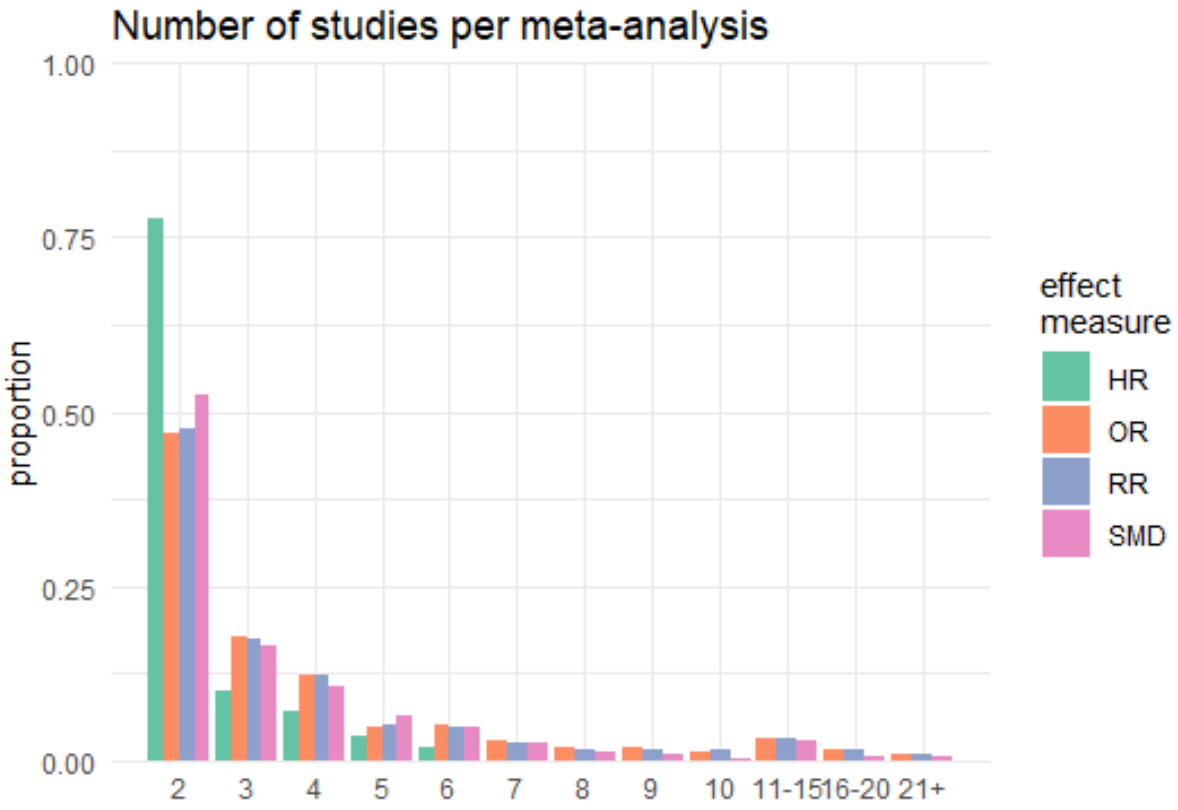
Database of meta-analyses in IQWiG reports

- Database of meta-analyses of IQWiG reports until 31-12-2021
- 132 reports included meta-analyses, a total of 1693 meta-analyses
- Different comparison / outcome / endpoint types
  - 67 % pharmacological (vs. other pharmacological or placebo),
    32 % non-pharmacological (vs. other non-pharmacological or placebo),
    1 % pharmacological vs. non-pharmacological
  - 7 % mortality, 46 % morbidity, 10 % quality of life, 32 % adverse events, 6 % others
- (Recalculated) Effect measures: HR, OR, RR, SMD

Effect measure	Reports	MAs	Studies
HR	25	112	271
OR	98	883	3570
RR	103	917	3689
SMD	69	645	2287

Database of meta-analyses in IQWiG reports





- Most meta-analyses in our data have less than 4 studies

Bayesian meta-analytic approach

Turner et al. 2015, Rhodes et al. 2015:

- Extend Bayesian model to multiple meta-analyses
- Posterior predictive distribution as informative prior for new analyses
- Empirical information from Cochrane Database of Systematic Reviews

  (14 886 binary outcomes, 6492 continuous outcomes)
  - For SMDs: Rhodes et al. (2015): $\log(\tau^2) \sim t_5(-3.44, 2.59)$
  - For ORs: Turner et al. (2015): $\tau^2 \sim LN(-2.56, 1.74^2)$
- Also distributions dependent on outcome and intervention comparison type

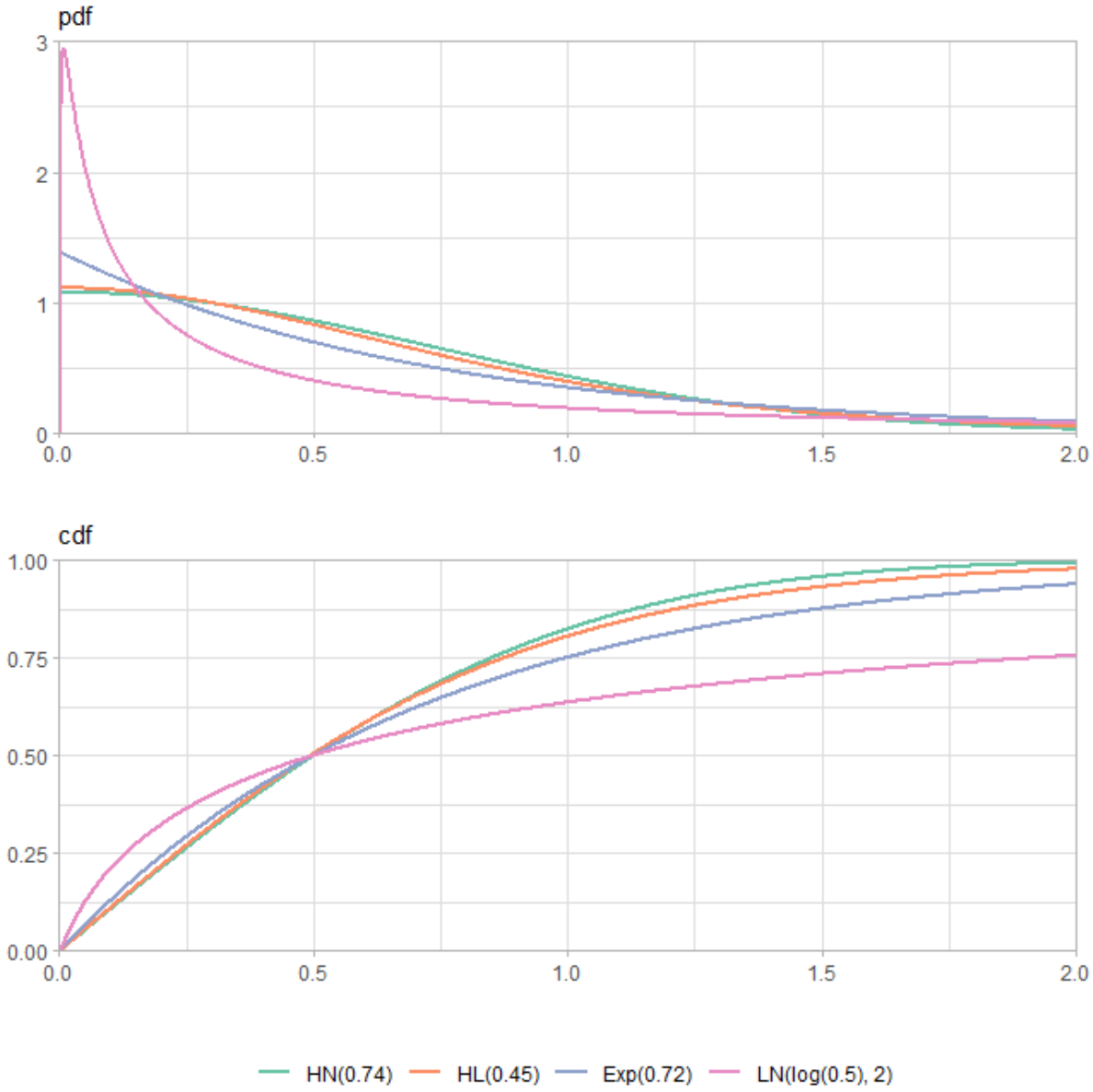
Röver et al. 2023:

$$y_{ij}|\mu_j, \tau_j, \sigma_{ij} \sim N(\mu_j, \sigma_{ij}^2 + \tau_j^2)$$
$$\mu_j \sim N(\mu_P, \sigma_P^2)$$
$$\tau_j|\xi \sim P(\xi), \quad \xi \sim H$$

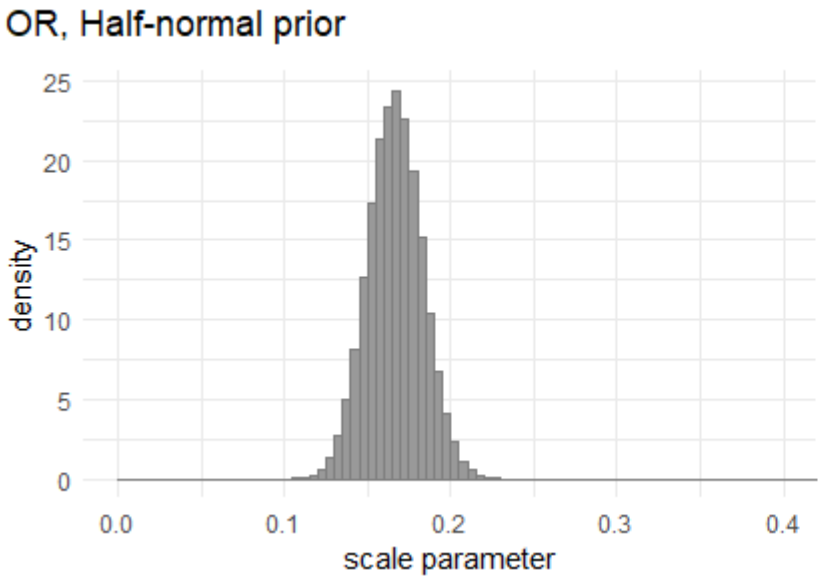
For example: $\tau_j|s \sim HN(s),\ s \sim Unif[0, b]$ with fixed, large b

Application to IQWiG data

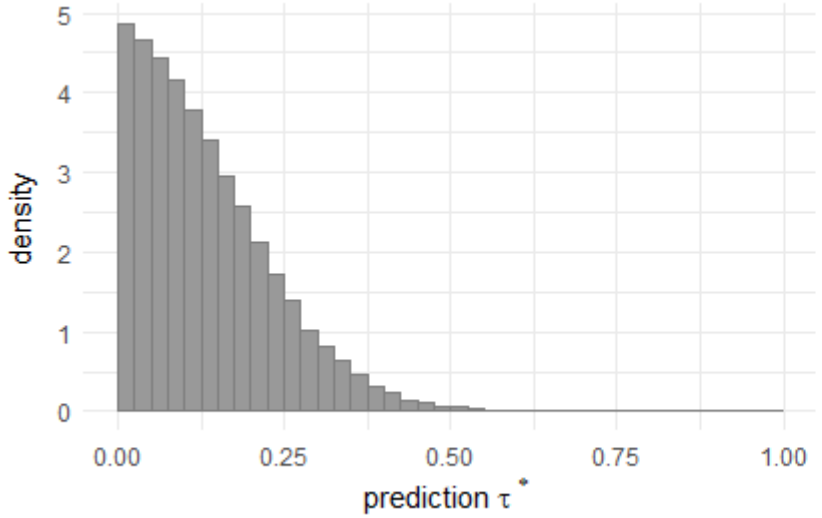
- Separately for effect measures
- Effect prior: $\mu \sim N(0, 100^2)$
- Heterogeneity prior:
  - Half-normal (HN)
  - Half-logistic (HL)
  - Exponential (Exp)
  - Log-normal (LN)

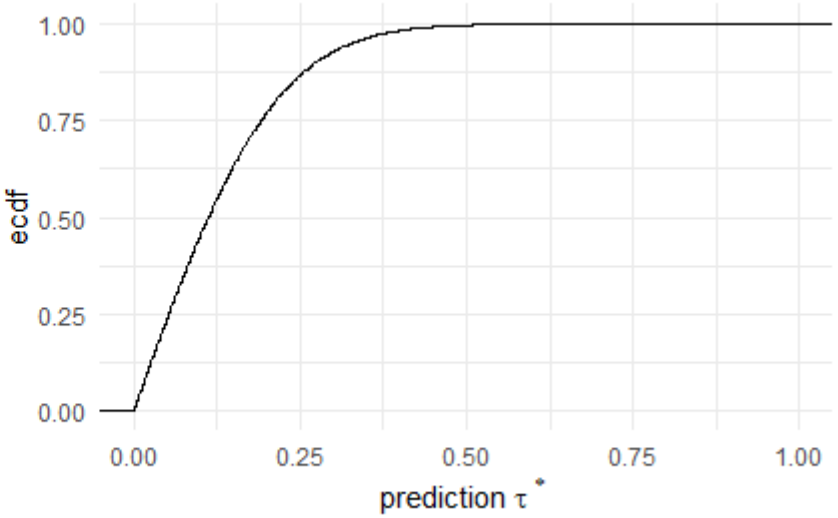


Results

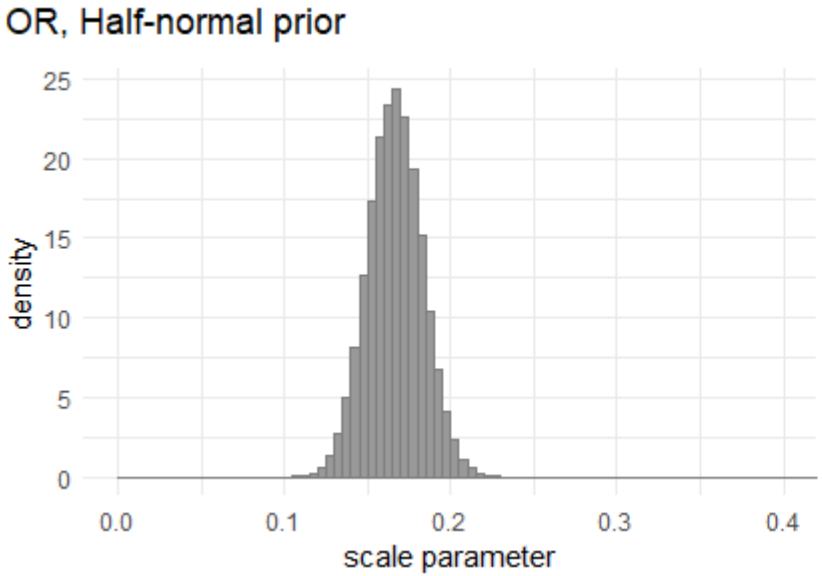


	scale
mean	0.167
sd	0.016
q25	0.155
median	0.166
q75	0.178
q95	0.194
q99	0.205

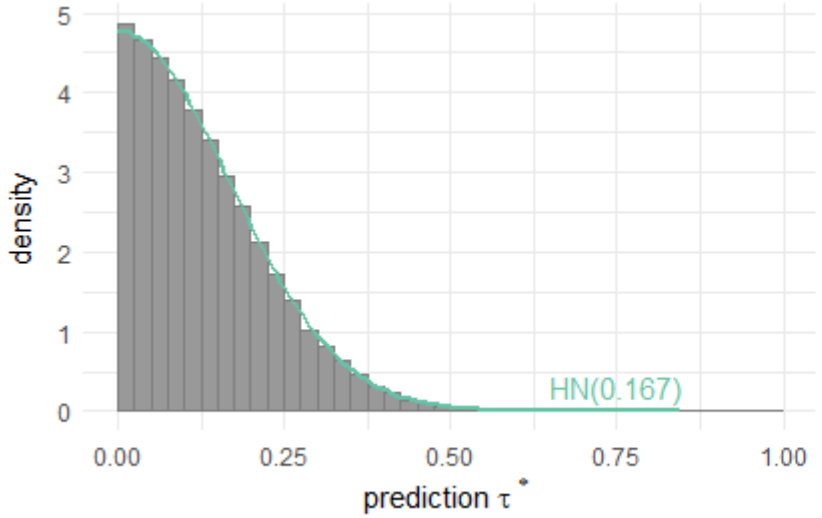


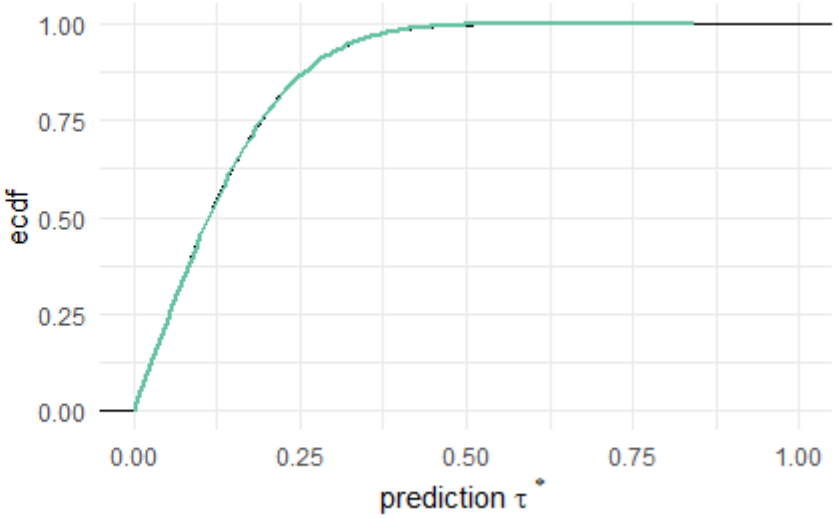


Results

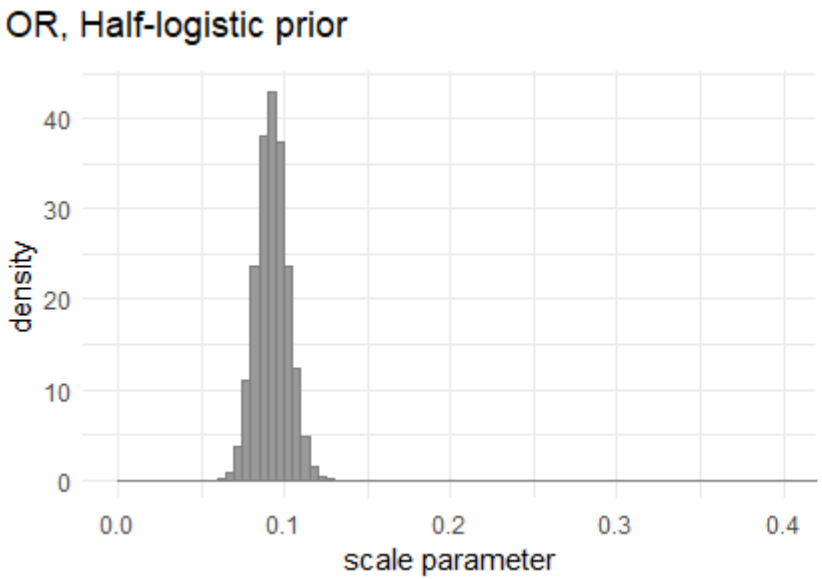


	scale
mean	0.167
sd	0.016
q25	0.155
median	0.166
q75	0.178
q95	0.194
q99	0.205

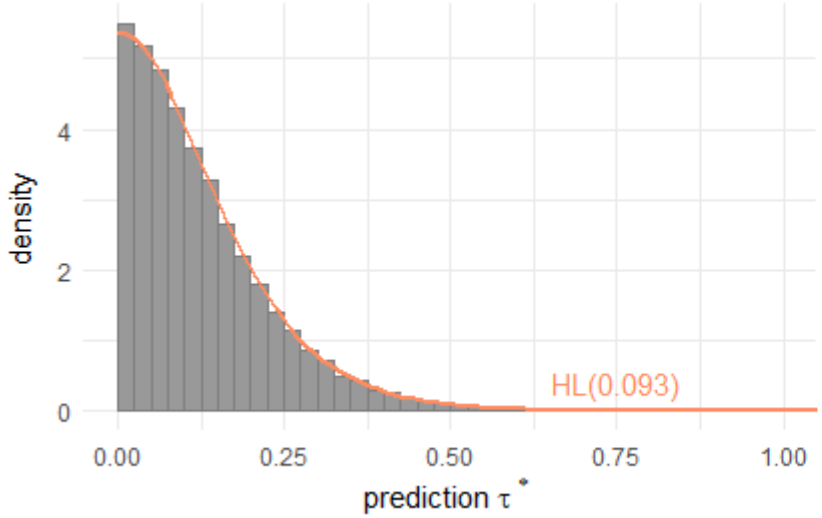


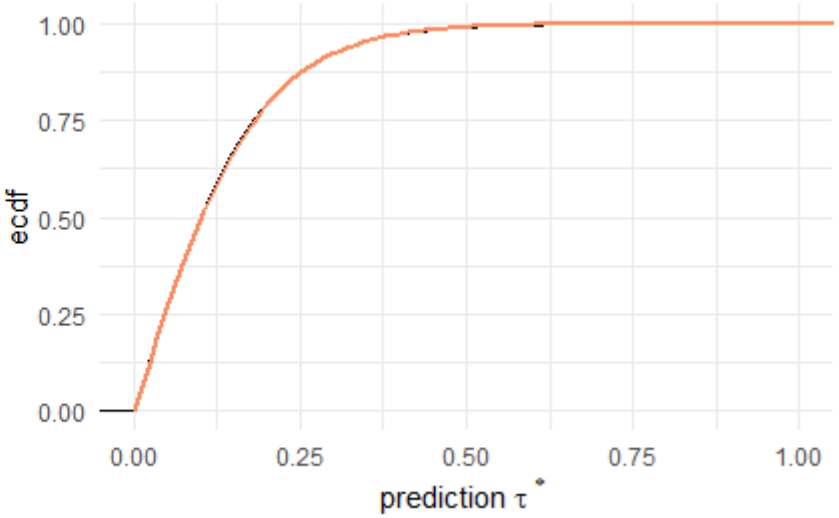


Results

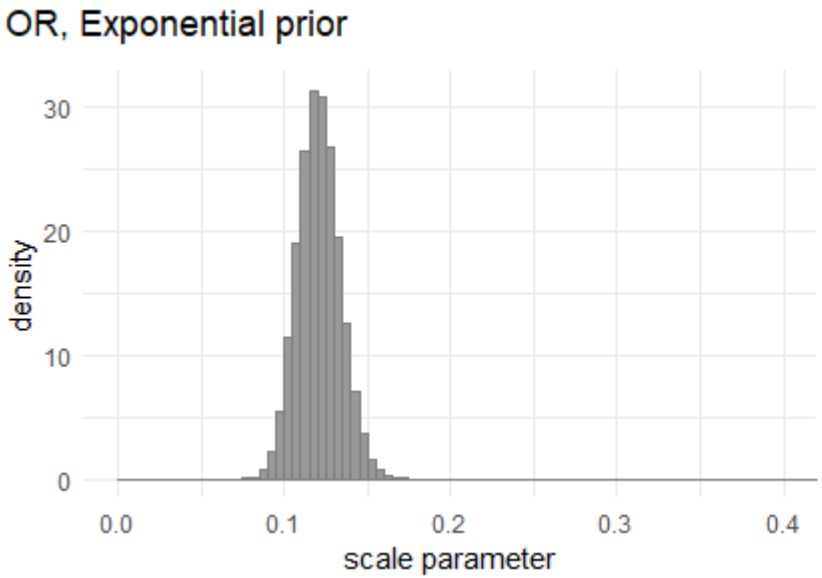


	scale
mean	0.093
sd	0.009
q25	0.087
median	0.093
q75	0.099
q95	0.108
q99	0.115

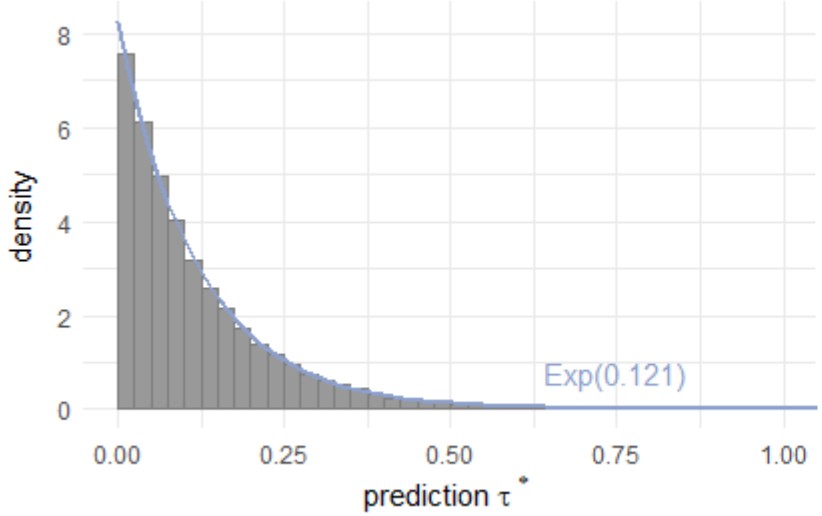


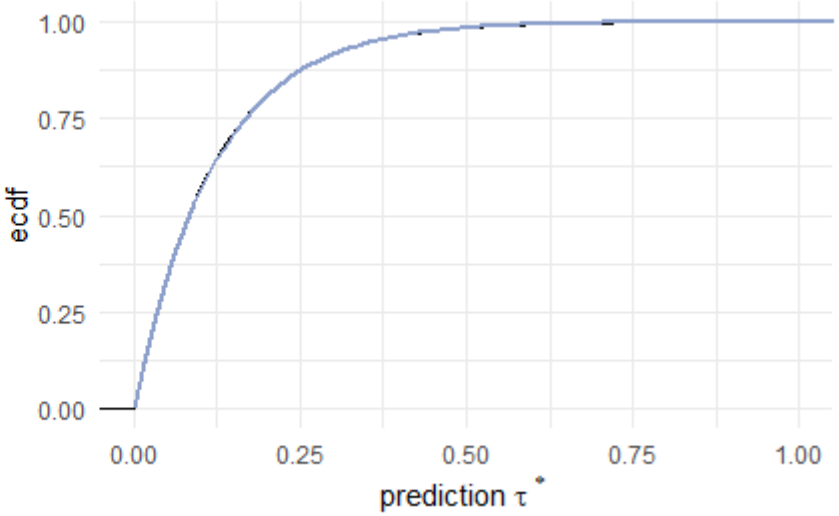


Results

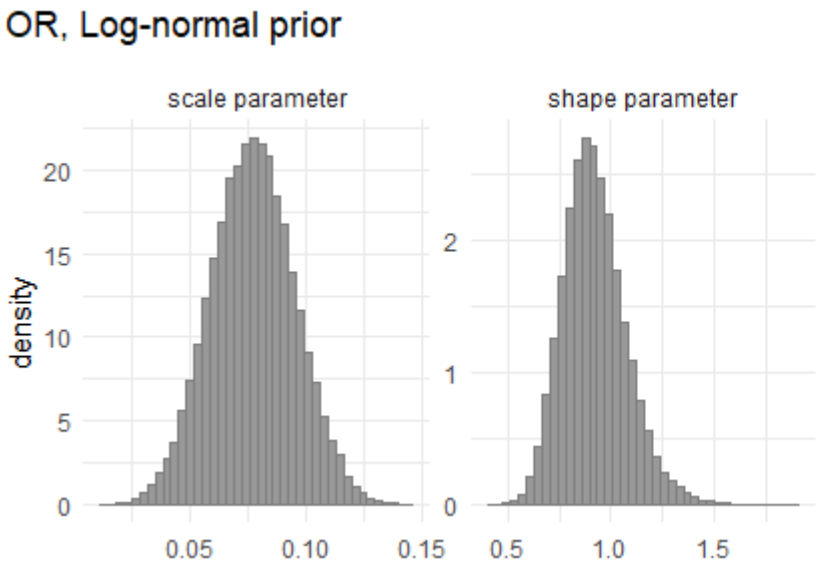


	scale
mean	0.121
sd	0.013
q25	0.112
median	0.120
q75	0.129
q95	0.142
q99	0.152

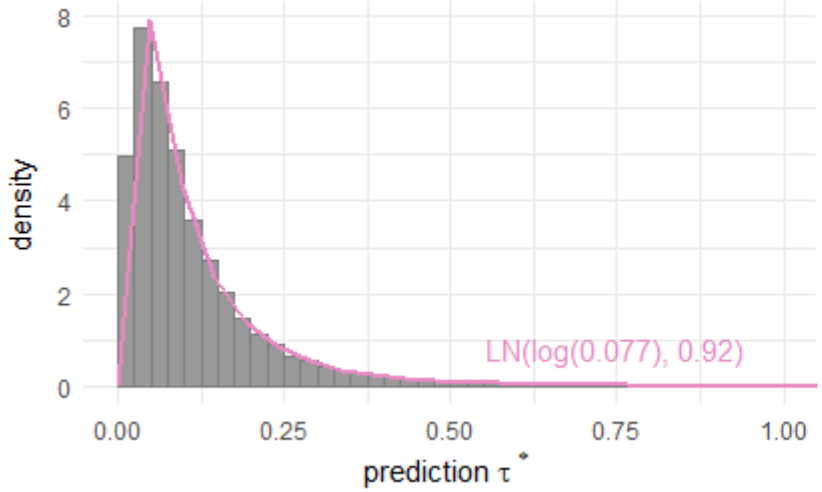


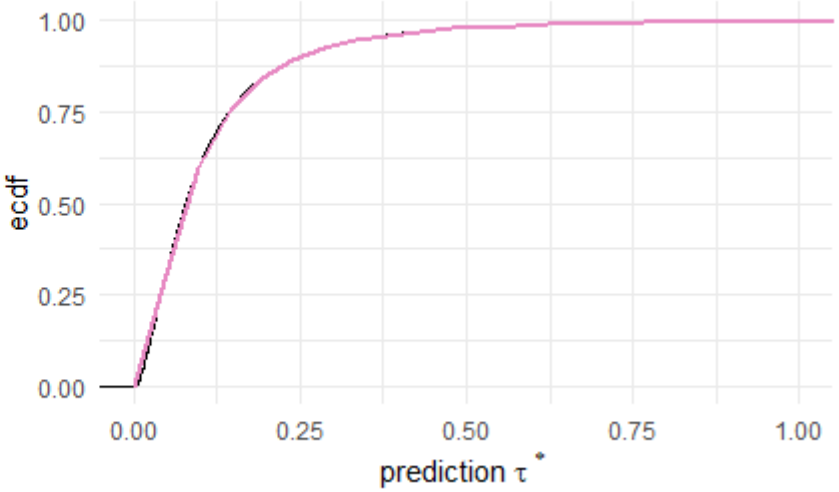


Results

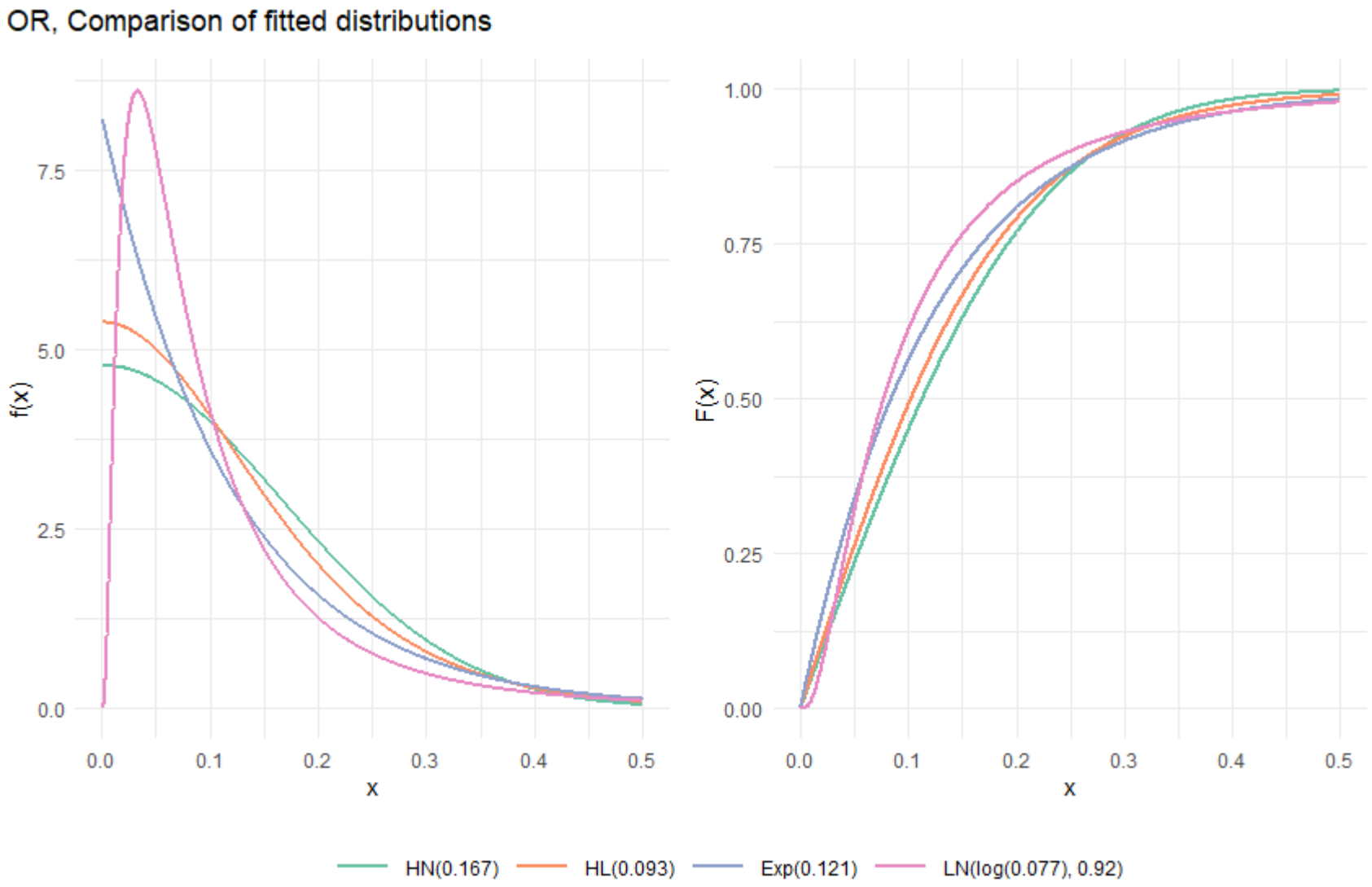


	log(scale)	sigma
mean	0.077	0.920
sd	0.018	0.152
q25	0.065	0.814
median	0.077	0.907
q75	0.090	1.012
q95	0.107	1.187
q99	0.119	1.343



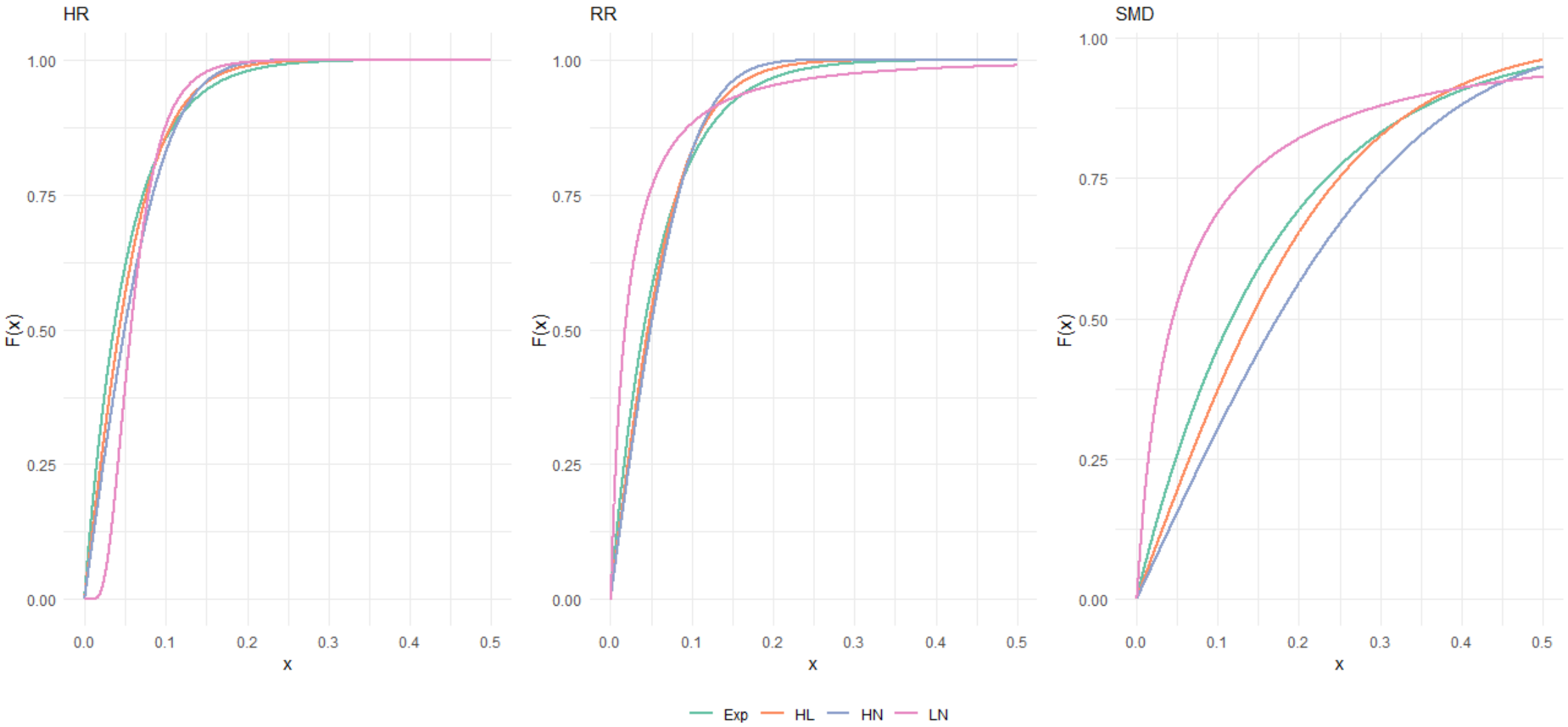


Results


OR, Comparison of fitted distributions
f(x)
7.5
5.0
2.5
0.0
0.0
0.1
0.2
0.3
0.4
0.5
x
F(x)
1.00
0.75
0.50
0.25
0.00
0.0
0.1
0.2
0.3
0.4
0.5
x
HN(0.167)
HL(0.093)
Exp(0.121)
LN(log(0.077), 0.92)

Results

Comparison of fitted distributions



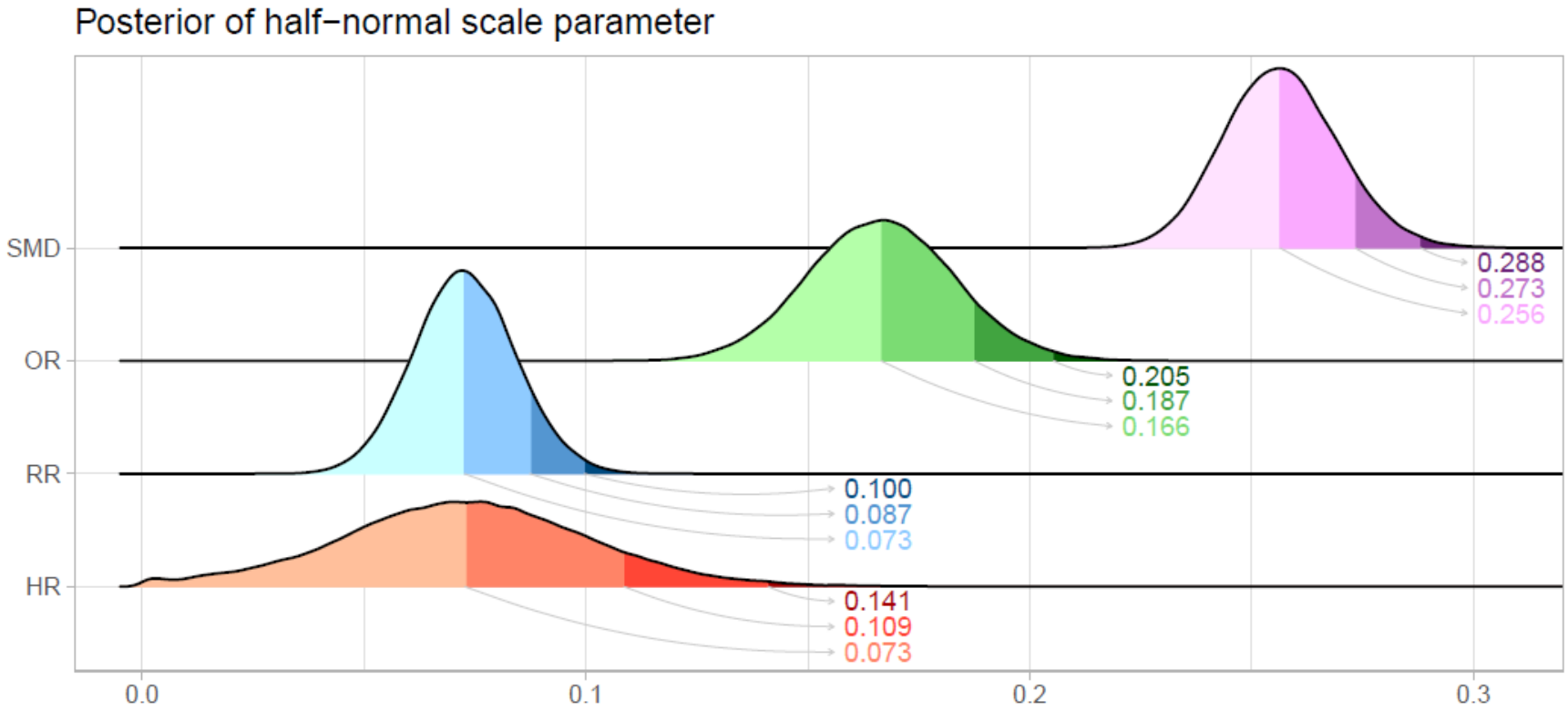
Selection of prior distribution for new analysis

- Good approximation using point estimate

Effect measure	Prior: HN	Prior: HL	Prior: Exp	Prior: LN
OR	HN(0.167)	HL(0.093)	Exp(0.121)	LN(log(0.077), 0.92)
HR	HN(0.073)	HL(0.039)	Exp(0.052)	LN(log(0.057), 0.482)
RR	HN(0.073)	HL(0.042)	Exp(0.059)	LN(log(0.018), 1.445)
SMD	HN(0.257)	HL(0.128)	Exp(0.169)	LN(log(0.045), 1.627)

Selection of prior distribution for new analysis

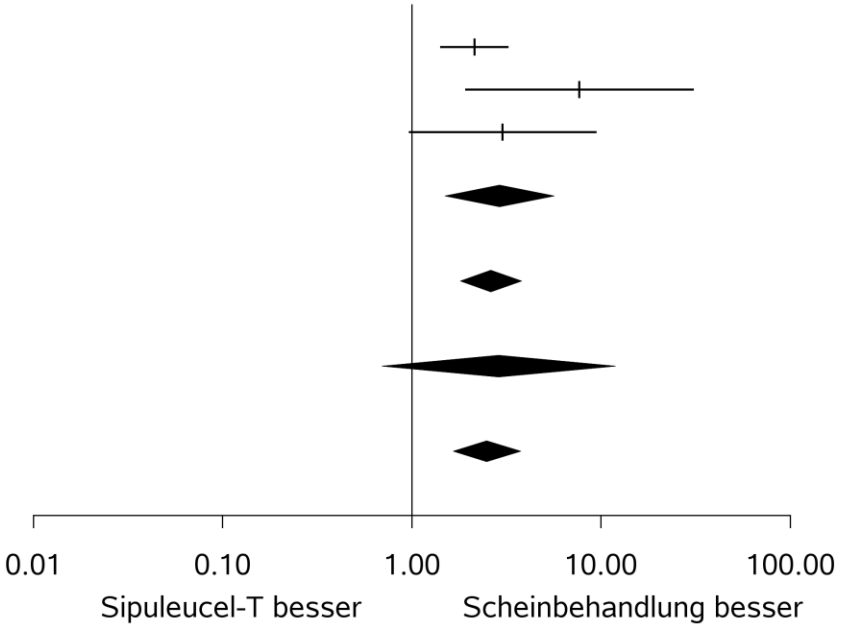
- Good approximation using point estimate
- Half-normal distribution adequate choice
- Use upper tail quantile of posterior distribution of scale parameter
  - Pragmatic suggestion: SMD: HN(0.3), OR: HN(0.2), RR/HR: HN(0.1)



Example (cont.)

Sipuleucel-T vs. Scheinbehandlung
Fieber

Studie	Sipuleucel-T n/N	Scheinbehandlung n/N	RR (95%-KI)	Gewichtung	RR	95%-KI
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]
FEM - Mantel-Haenszel					2.62	[1.80, 3.80]
REM - Knapp-Hartung					2.88	[0.70, 11.92]
Bayes - HN(0.1)					2.48	[1.66, 3.77]



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

Comparison to current IQWiG approach

- Re-run all meta-analyses in our data set
- Methods:
  - current IQWiG approach (FE/RE/QES)
  - Bayesian meta-analysis with HN-prior for heterogeneity:
    SMD: HN(0.3), OR: HN(0.2), RR/HR: HN(0.1)
- Investigate disagreement regarding statistical significance

  OR (883 meta-analyses):

	Bayes HN(0.2)		
IQWiG	No evidence	Evidence	Sum
No evidence	578	51	629
Evidence	46	208	254
Sum	624	259	883

→ Disagreement: (46+51)/883 ≈ 11 %

Proportion of disagreement to current approach

measure	set	No of analyses	% disagreement Bayes HN(prop)	Bayes HN(median)
RR	all	917	10	9
RR	IQWiG proc: no evidence	666	9	9
RR	IQWiG proc: evidence	251	13	9
RR	all studies stat. sig.	65	3	3
HR	all	112	4	4
HR	IQWiG proc: no evidence	80	3	3
HR	IQWiG proc: evidence	32	9	9
HR	all studies stat. sig.	14	0	0
OR	all	883	11	10
OR	IQWiG proc: no evidence	639	8	9
OR	IQWiG proc: evidence	244	16	15
OR	all studies stat. sig.	64	3	3
SMD	all	645	18	17
SMD	IQWiG proc: no evidence	427	8	10
SMD	IQWiG proc: evidence	218	36	32
SMD	all studies stat. sig.	83	12	10

In MAs in which IQWiG states no evidence, Bayes-MA with proposed HN parameter claims evidence in 9 % of cases

In MAs in which IQWiG states evidence, Bayes-MA with proposed HN parameter claims no evidence in 13 % of cases

- Disagreement is higher for analyses in which IQWiG claims evidence
- If all studies are stat. sig.: Bayes mostly agrees

Conclusion/Outlook

- Bayesian meta-analysis as alternative in situations with few studies
  - Frequentist methods often fail
  - Avoids complex model choice, makes quantification possible
  - Specification of heterogeneity prior distribution necessary

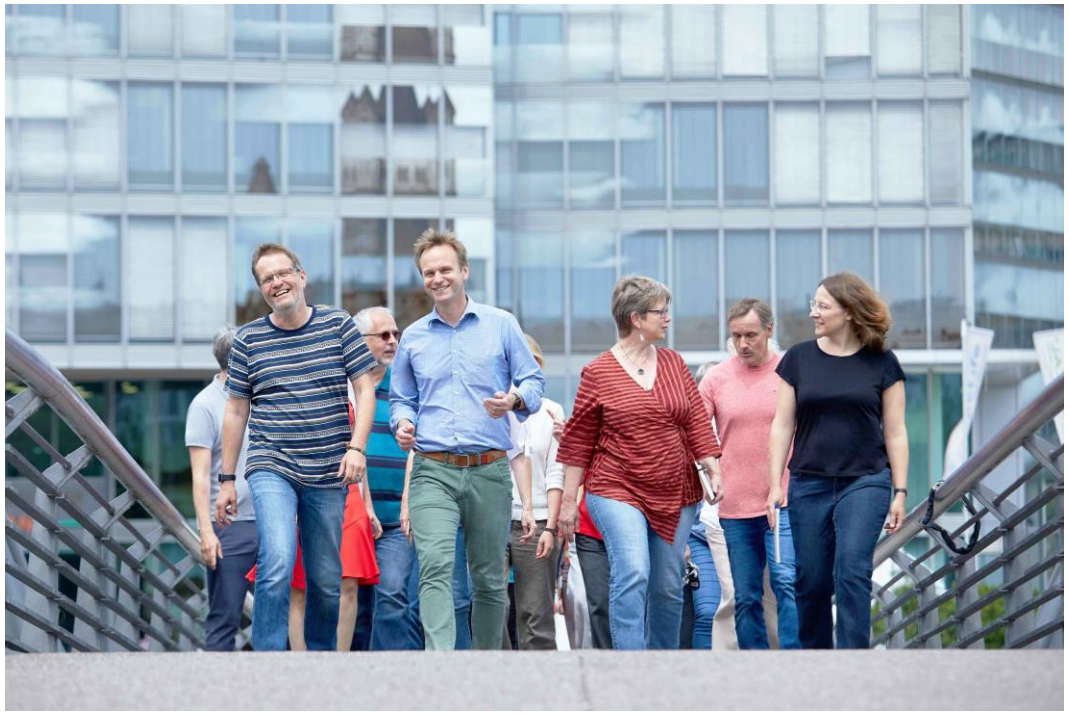
- Derivation of informative prior using meta-analysis data of IQWiG reports
  - Different results for different effect measures
  - Suggestion: Use upper tail quantile of scale parameter distribution for conservative estimation

- Comparison to current IQWiG approach
  - Differences in decisions; Disagreement is higher if IQWiG approach yields evidence
  - Combination of QES and Bayes promising

Literatur

- 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.
- Friede T., Röver C., Wandel S., Neuenschwander B. (2017). Meta-analysis of few small studies in orphan diseases. Research Synthesis Methods 8: 79–91.
- 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.
- Veroniki, A. A., Jackson, D., Bender, R., Kuss, O., Langan, D., Higgins, J. P., ... & Salanti, G. (2019). Methods to calculate uncertainty in the estimated overall effect size from a random-effects meta-analysis. Research synthesis methods, 10(1), 23-43.

Institute for Quality and Efficiency in Health Care (IQWiG)



Im Mediapark 8
D-50670 Cologne

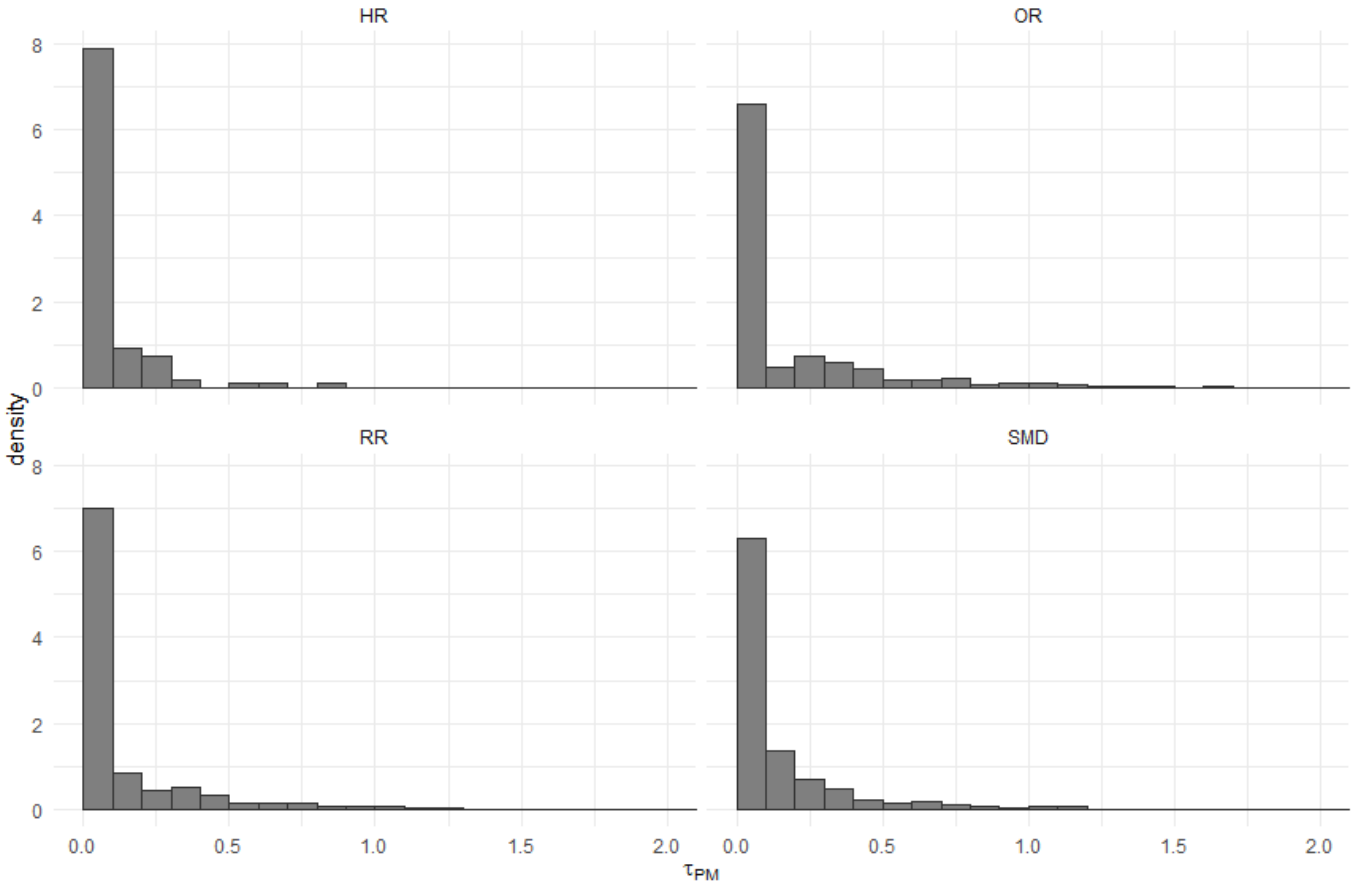
Phone +49 221 35685-0
Fax +49 221 35685-1

info@iqwig.de

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

Empirical distribution of heterogeneity estimates

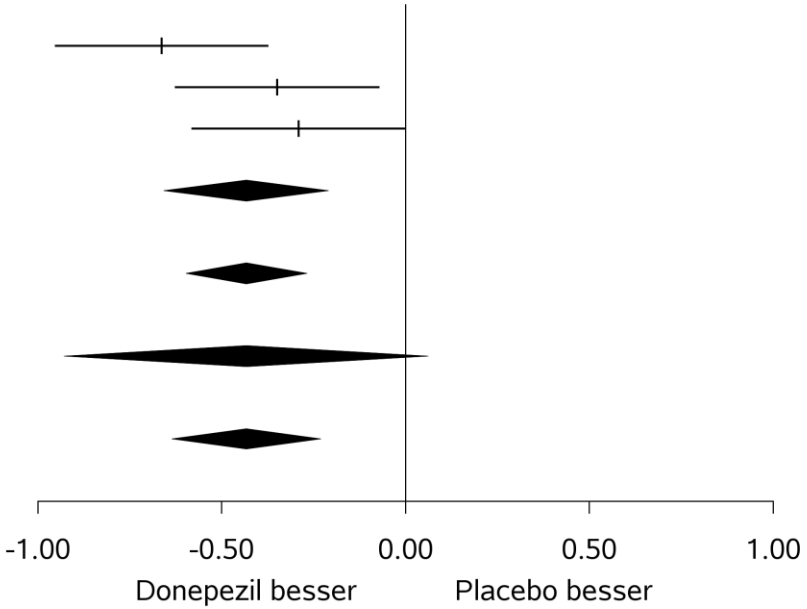
- Paule-Mandel estimator calculated for every meta-analysis
- 61% (HR: 74%, OR: 64%, RR: 63%, SMD: 53%) of estimates are 0



Example 2

Donepezil vs. Placebo
DAD, CMCS, PDS

Studie	logarithmierter RR	SE	Gewichtung	RR	95%-KI
Gauthier 2002	-0.66	0.15	32.9	-0.66	[-0.95, -0.37]
Homma 2000	-0.35	0.14	34.4	-0.35	[-0.63, -0.07]
Winblad 2001	-0.29	0.15	32.7	-0.29	[-0.58, 0.00]
REM - DerSimonian-Laird			100.0	-0.43	[-0.66, -0.21]
FEM - Inverse Varianz				-0.43	[-0.60, -0.27]
REM - Knapp-Hartung				-0.43	[-0.93, 0.06]
Bayes - HN(0.1)				-0.43	[-0.64, -0.23]

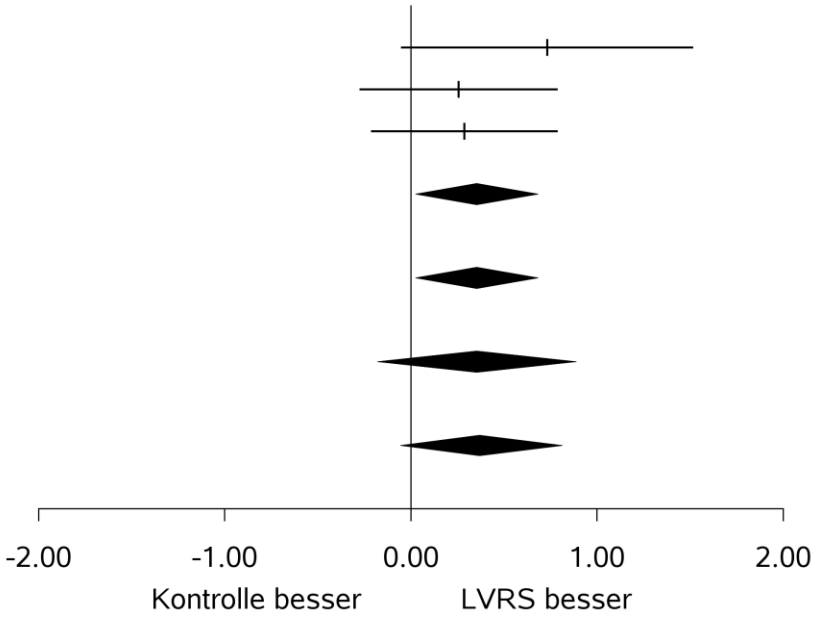


Heterogenität: Q=3.70, df=2, p=0.157, I^2=45.9%
Gesamteffekt (REM - DerSimonian-Laird): Z-Score=-3.78, p<0.001, Tau=0.135

Example 3

LVRS vs. Kontrolle
6-Minuten-Gehtest

Studie	Effekt	SE	Effekt (95%-KI)	Gewichtung	Effekt	95%-KI
Clarenbach 2015	0.73	0.40		17.8	0.73	[-0.05, 1.52]
Goldstein 2003	0.26	0.27		38.7	0.26	[-0.28, 0.79]
Miller 2006 (CLVR)	0.29	0.26		43.5	0.29	[-0.21, 0.79]
REM - DerSimonian-Laird				100.0	0.35	[0.02, 0.68]
FEM - Inverse Varianz					0.35	[0.02, 0.68]
REM - Knapp-Hartung					0.35	[-0.18, 0.89]
Bayes - HN(0.3)					0.37	[-0.06, 0.81]

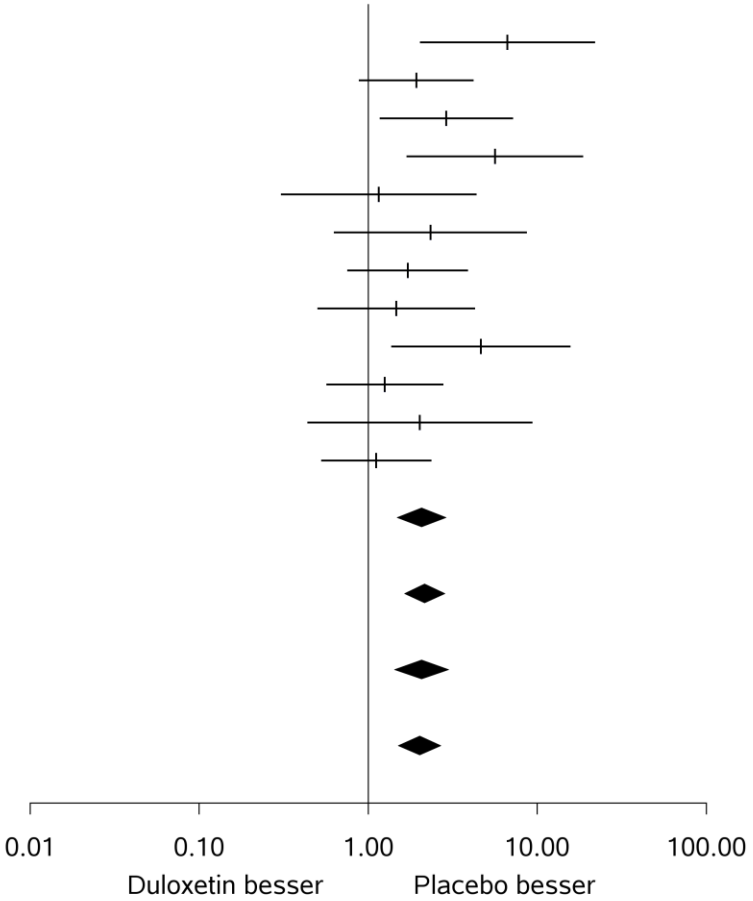


Heterogenität: Q=1.09, df=2, p=0.580, I^2=0%
Gesamteffekt (REM - DerSimonian-Laird): Z-Score=2.10, p=0.035, Tau=0

Example 4

Duloxetin vs. Placebo
Abbruch wegen UE

Studie	Duloxetin n/N	Placebo n/N	Gewichtung	RR	95%-KI
Brannan 2005	20/141	3/141	6.5	6.67	[2.03, 21.93]
Brecht 2007	17/162	9/165	12.2	1.92	[0.88, 4.19]
Detke 2002a	16/128	6/139	9.9	2.90	[1.17, 7.17]
Detke 2002b	17/123	3/122	6.4	5.62	[1.69, 18.69]
Detke 2004	7/188	3/93	5.4	1.15	[0.31, 4.36]
Goldstein 2002	7/70	3/70	5.5	2.33	[0.63, 8.66]
Goldstein 2004	14/91	8/89	11.4	1.71	[0.76, 3.88]
HMAQ-B	8/82	5/75	7.6	1.46	[0.50, 4.28]
HMAT-A	13/84	3/90	6.2	4.64	[1.37, 15.72]
Nierenberg 2007	20/273	8/137	11.9	1.25	[0.57, 2.77]
Perahia 2006b	8/196	2/99	4.2	2.02	[0.44, 9.34]
Raskin 2007	20/207	9/104	12.8	1.12	[0.53, 2.37]
REM - DerSimonian-Laird	167/1745	62/1324	100.0	2.07	[1.48, 2.89]
FEM - Mantel-Haenszel				2.16	[1.64, 2.85]
REM - Knapp-Hartung				2.07	[1.43, 3.00]
Bayes - HN(0.1)				2.01	[1.50, 2.70]



Heterogenität: Q=14.33, df=11, p=0.215, I^2=23.2%
Gesamteffekt (REM - DerSimonian-Laird): Z-Score=4.28, p<0.001, Tau=0.281

Comparison to distributions of Turner and Rhodes

	quantile				
distribution	0.05	0.25	0.5	0.75	0.95
effect measure: OR					
HN(0.2)	0.0125	0.0637	0.1349	0.2301	0.3920
$LN(-1.28, 0.87^2)$	0.0665	0.1546	0.2780	0.5000	1.1630
effect measure: SMD					
HN(0.3)	0.0188	0.0956	0.2023	0.3451	0.5880
$\log t_5(-1.72, 1.295^2)$	0.0132	0.0699	0.1791	0.4589	2.4339

Proportion of disagreement to current approach (k=2)

measure	set	No of analyses	Bayes HN(prop)	Bayes HN(median)	DSL
RR	all	435	8	6	5
RR	IQWiG proc: no evidence	312	1	1	0
RR	IQWiG proc: evidence	123	24	16	17
RR	all studies stat. sig.	42	5	5	5
HR	all	87	3	3	1
HR	IQWiG proc: no evidence	59	0	0	0
HR	IQWiG proc: evidence	28	11	11	4
HR	all studies stat. sig.	13	0	0	0
OR	all	413	9	8	5
OR	IQWiG proc: no evidence	296	2	2	0
OR	IQWiG proc: evidence	117	27	25	18
OR	all studies stat. sig.	38	5	5	8
SMD	all	340	21	19	4
SMD	IQWiG proc: no evidence	239	4	5	1
SMD	IQWiG proc: evidence	101	60	52	11
SMD	all studies stat. sig.	49	20	16	0

Proportion of disagreement to current approach (k=2-4)

measure	set	No of analyses	Bayes HN(prop)	Bayes HN(median)	DSL
RR	all	706	10	8	7
RR	IQWiG proc: no evidence	516	7	7	5
RR	IQWiG proc: evidence	190	17	12	12
RR	all studies stat. sig.	61	3	3	3
HR	all	106	4	4	2
HR	IQWiG proc: no evidence	75	1	1	1
HR	IQWiG proc: evidence	31	10	10	3
HR	all studies stat. sig.	14	0	0	0
OR	all	679	11	10	8
OR	IQWiG proc: no evidence	493	7	8	5
OR	IQWiG proc: evidence	186	19	18	13
OR	all studies stat. sig.	60	3	3	5
SMD	all	514	19	18	7
SMD	IQWiG proc: no evidence	353	6	8	6
SMD	IQWiG proc: evidence	161	46	40	9
SMD	all studies stat. sig.	72	14	11	0

Selection of prior distribution for new analysis

- Good approximation using point estimate
- Half-normal distribution adequate choice
- Use upper tail quantile of posterior distribution of scale parameter
  - Pragmatic suggestion: SMD: HN(0.3), OR: HN(0.2), RR/HR: HN(0.1)

Prior	Effect measure	Half normal scale parameter			
		median	q90	q95	q99
HN	OR	0.1665	0.1874	0.1936	0.2053
HN	HR	0.0727	0.1085	0.1191	0.1410
HN	RR	0.0726	0.0873	0.0916	0.0995
HN	SMD	0.2564	0.2733	0.2783	0.2881