Model averaging for robust extrapolation in evidence synthesis

Christian Röver[1], Simon Wandel[2], Tim Friede[1]

[1]Department of Medical Statistics,
University Medical Center Göttingen,
Göttingen, Germany

[2]Novartis Pharma AG,
Basel, Switzerland

December 6, 2018

InSPiRe
Innovative methodology for small populations research

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement number FP HEALTH 2013-602144.

Overview

- meta-analysis & extrapolation
- NNHM, example
- informative priors, mixture priors
- example applications
- conclusions

Extrapolation & meta-analysis

- **extrapolation** desirable when evidence sparse
  or relevance unclear:
  paediatric/adult applications, bridging studies,...
- **common situation** in meta-analysis:
  majority of analyses in *Cochrane data base* include ≤ 3 studies[1],
  many overall + subgroup analysis results
- **aims**:
  - formal utilization of related evidence
  - robust procedure
    (no naïve, over-optimistic pooling)

[1] R.M. Turner et al. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. *International Journal of Epidemiology* 41(3):818–827, 2012.
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.

Meta-analysis

The common NNHM (random-effects) model

- k studies
- estimates $y_i \in \mathbb{R}$ $(i = 1, \ldots, k)$
- standard errors $\sigma_i > 0$
- **normal-normal hierarchical model (NNHM)**:

$$y_i|\theta_i, \sigma_i \sim \mathrm{N}(\theta_i, \sigma_i^2), \qquad \theta_i|\mu, \tau \sim \mathrm{N}(\mu, \tau^2)$$

$$\Rightarrow \qquad y_i|\mu, \sigma_i, \tau \sim \mathrm{N}(\mu, \sigma_i^2 + \tau^2)$$

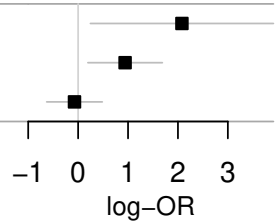
- data: y_i (and σ_i)
- two unknowns:
  - "effect" $\mu \in \mathbb{R}$ (of primary interest)
  - "heterogeneity" $\tau \geq 0$ (between-study variance component)

Migraine example data

Triptans for headache relief in children

- investigation of efficacy of migraine treatments in children (OR > 1 indicates benefit)
- desirable: RCTs with placebo control
- paediatric patients: ethical concerns / feasibility

publication	subjects	triptan	placebo	log–OR	CI
Ueberall (1999)	children	12 / 14	6 / 14	2.079	[0.246, 3.913]
Hämäläinen (2002)	children	38 / 59	24 / 58	0.941	[0.195, 1.688]
Ho (2012)	children	53 / 98	57 / 102	–0.073	[–0.630, 0.485]

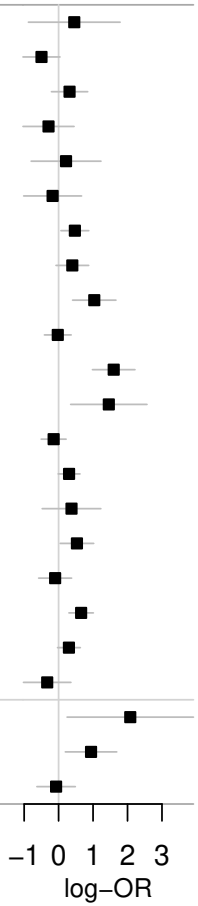


- 3 paediatric studies ($<$ 12yr)

Migraine example data

Triptans for headache relief in children (and adolescents)

publication	subjects	triptan	placebo	log–OR	CI
Hämäläinen (1997b)	adolescents	7 / 23	5 / 23	0.454	[–0.876, 1.785]
Rothner (1997)	adolescents	113 / 226	46 / 74	–0.496	[–1.034, 0.041]
Winner (1997)	adolescents	111 / 222	32 / 76	0.318	[–0.207, 0.844]
Rothner (1999a)	adolescents	96 / 186	20 / 34	–0.292	[–1.033, 0.449]
Rothner (1999b)	adolescents	17 / 62	7 / 30	0.216	[–0.797, 1.230]
Rothner (1999c)	adolescents	23 / 66	14 / 36	–0.174	[–1.014, 0.666]
Winner (2000)	adolescents	243 / 377	69 / 130	0.472	[0.068, 0.876]
Winner (2002)	adolescents	98 / 149	80 / 142	0.398	[–0.076, 0.872]
Ahonen (2004)	adolescents	53 / 83	32 / 83	1.035	[0.406, 1.664]
Visser (2004a)	adolescents	159 / 233	165 / 240	–0.024	[–0.412, 0.364]
Ahonen (2006)	adolescents	71 / 96	35 / 96	1.599	[0.982, 2.216]
Evers (2006)	adolescents	18 / 29	8 / 29	1.458	[0.350, 2.565]
Rothner (2006)	adolescents	262 / 480	93 / 160	–0.144	[–0.506, 0.218]
Winner (2006)	adolescents	316 / 483	141 / 242	0.304	[–0.013, 0.621]
Callenbach (2007)	adolescents	19 / 46	15 / 46	0.375	[–0.477, 1.226]
Lewis (2007)	adolescents	97 / 148	67 / 127	0.533	[0.046, 1.019]
Winner (2007)	adolescents	82 / 144	79 / 133	–0.101	[–0.579, 0.377]
Linder (2008)	adolescents	383 / 544	94 / 170	0.654	[0.300, 1.008]
Ho (2012)	adolescents	167 / 284	147 / 286	0.300	[–0.031, 0.631]
Fujita (2014)	adolescents	23 / 74	27 / 70	–0.331	[–1.019, 0.357]
Ueberall (1999)	children	12 / 14	6 / 14	2.079	[0.246, 3.913]
Hämäläinen (2002)	children	38 / 59	24 / 58	0.941	[0.195, 1.688]
Ho (2012)	children	53 / 98	57 / 102	–0.073	[–0.630, 0.485]



- 3 paediatric studies (< 12yr) + 20 adolescent studies (12–17yr)[2]

[2] L. Richer et al. Drugs for the acute treatment of migraine in children and adolescents. *Cochrane Database of Systematic Reviews*, 4:CD005220, 2016.

Extrapolation

Bayesian framework

- extrapolation:
  - Bayesian methods commonly **suggested**[3]
  - Bayesian methods predominant approach **in practice**[4]
- obvious approaches:
  - via hierarchical models
  - via **informative prior** distribution
- here: Bayesian meta-analysis via `bayesmeta` R package[5]

[3] e.g.: European Medicines Agency (EMEA). Guideline on clinical trials in small populations, July 2006.
Food and Drug Administration (FDA). Leveraging existing clinical data for extrapolation to pediatric uses of medical devices - guidance for industry and food and drug administration staff. Draft guidance, June 2016.

[4] I. Wadswoth, L.V. Hampson, T. Jaki. Extrapolation of efficacy and other data to support the development of new medicines for children: A systematic review of methods. *Statistical Methods in Medical Research*, 2016.

[5] `http://cran.r-project.org/package=bayesmeta`

Informative priors & robustness

- **danger**: posterior as simplistic prior / data "compromise"
- **desirable**: sensible behaviour in case of prior / data conflict; in case of doubt, data should overrule prior
- approach: **robustness** via **heavy-tailed mixture priors**[6]
- here: two parameters–
  - informative priors for effect and/or heterogeneity?
  - include further external information?[7]
- in following (for simplicity): informative joint effect / heterogeneity prior

---

[6] A. O'Hagan L. Pericchi. Bayesian heavy-tailed models and conflict resolution: A review. *Brazilian Journal of Probability and Statistics*, 26(4):372–401, 2012.

H. Schmidli, S. Gsteiger, S. Roychoudhury, A. O'Hagan, D. Spiegelhalter, B. Neuenschwander. Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics*, 70(4):1023–1032, 2014.

[7] R.M. Turner, D. Jackson, Y. Wei, S.G. Thompson, J.P.T. Higgins. Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian meta-analysis. *Statistics in Medicine*, 34(6):984-998, 2015.

K.M. Rhodes, R.M. Turner, J.P.T. Higgins. 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.

Robust mixture priors

Setup

- idea: prior $p(\theta)$ for children's data as a **mixture**:

$$p(\theta) \;=\; (1-w) \times p_1(\theta) \;+\; w \times p_2(\theta)$$

  where
  - $p_1(\theta)$ is **uninformative / vague**
  - $p_2(\theta)$ is **informative** (based on adolescent data + prior p_1)
  - $w \in [0, 1]$ expresses certainty about external data's relevance

- interpretation: e.g., $w = 50\%$ - -
  - **same effect** with probability $w = 50\%$
  - **separate effects** with probability $(1-w) = 50\%$

- **mixture** setup should lead to **robust** behaviour in case of **prior/data conflict**[8]

[8] A. O'Hagan L. Pericchi. Bayesian heavy-tailed models and conflict resolution: A review. *Brazilian Journal of Probability and Statistics*, 26(4):372–401, 2012.

Robust mixture priors

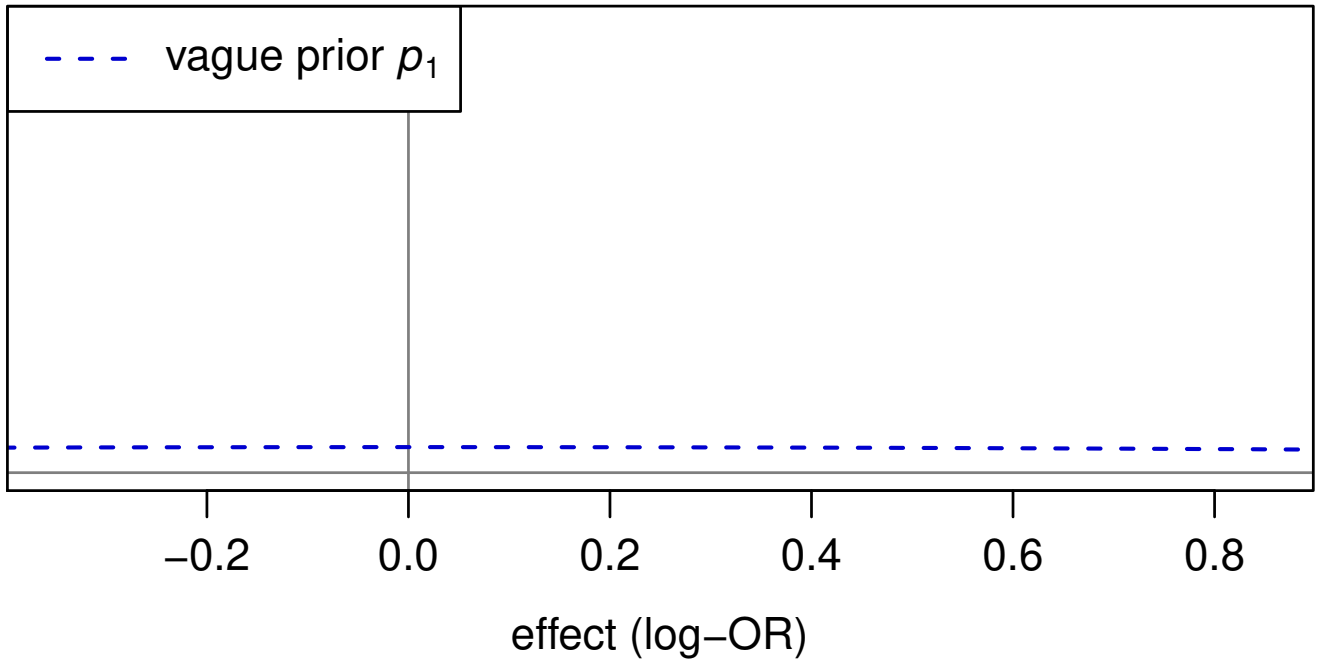
Inference

- technically: **mixture prior** implies **mixture posterior** ($\rightarrow$ model averaging)
- **posterior** again is a **mixture** of (conditional) posteriors under priors p_1 and p_2
- **weighting** of posteriors is given through **marginal likelihoods (Bayes factor)** and weight w
- only need to determine two posteriors and Bayes factor, then re-weight
- equivalence of *meta-analytic-predictive (MAP)* and *meta-analytic-combined (MAC)* approaches simplifies computations[9]

[9] H. Schmidli, S. Gsteiger, S. Roychoudhuri, A. O'Hagan, D. Spiegelhalter, B. Neuenschwander. Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics*, 70(4);1023–1032, 2014.

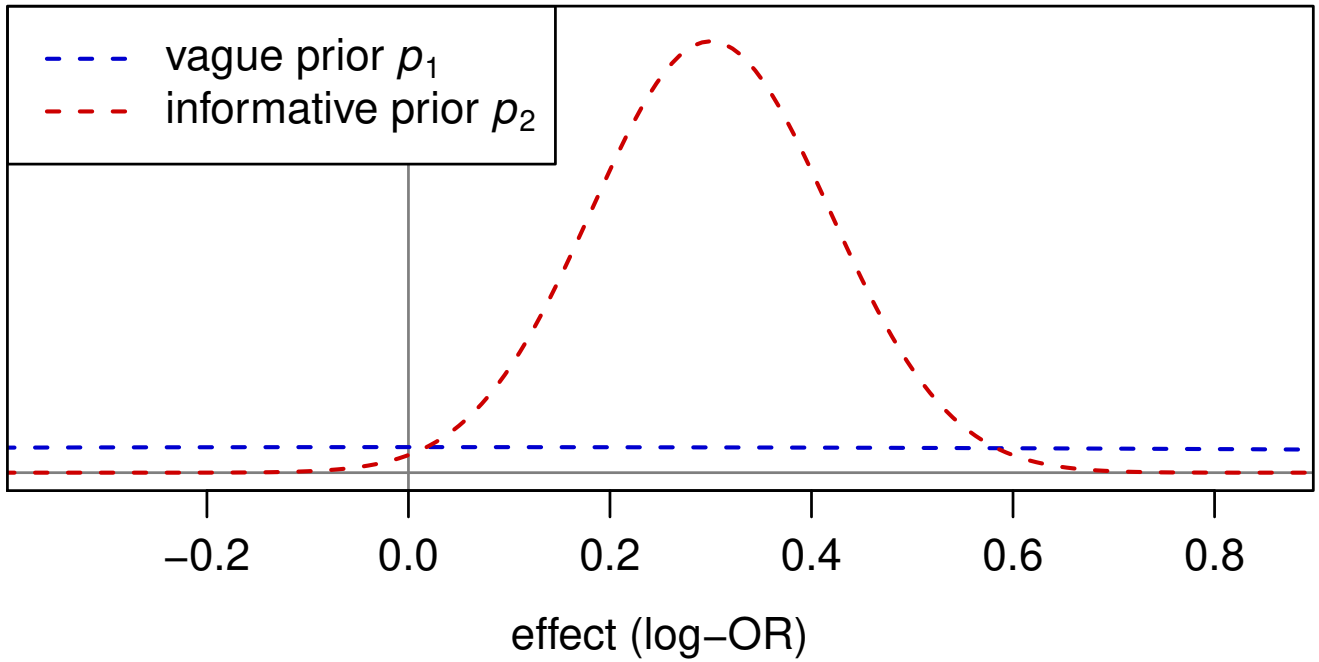
Example: children's effect prior setup

- **vague prior** p_1:
  - effect: $\mu \sim \mathrm{N}(0, 2^2)$
  - heterogeneity: $\tau \sim \mathrm{halfNormal}(0.5)$



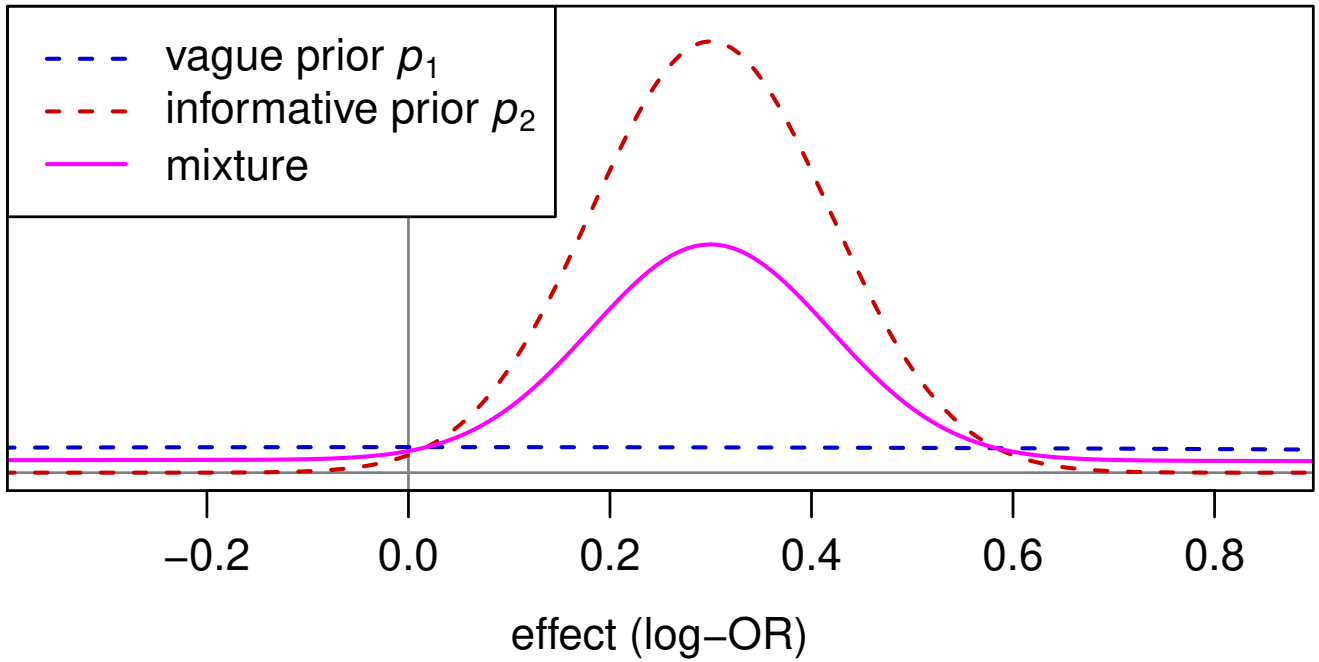
Example: children's effect prior setup

- vague prior p_1:
  - effect: $\mu \sim \mathrm{N}(0, 2^2)$
  - heterogeneity: $\tau \sim \mathrm{halfNormal}(0.5)$
- **informative prior** p_2
  (posterior from adolescent studies):
  - effect: $\mu = 0.30\ [0.07, 0.54]$
  - heterogeneity: $\tau = 0.41\ [0.21, 0.65]$



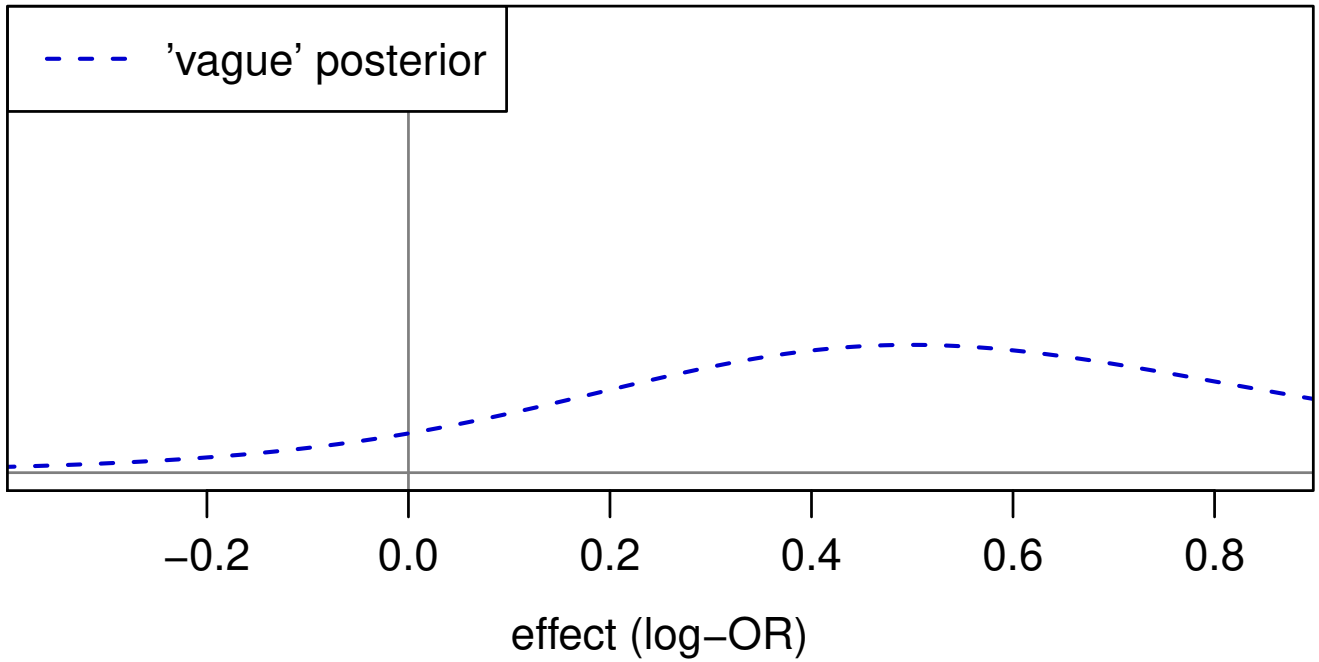
Example: children's effect prior setup

- **vague prior** p_1: $(1-w = 50\%)$
  - effect: $\mu \sim \mathrm{N}(0, 2^2)$
  - heterogeneity: $\tau \sim \mathrm{halfNormal}(0.5)$
- **informative prior** p_2 $(w = 50\%)$
  (posterior from adolescent studies):
  - effect: $\mu = 0.30\ [0.07, 0.54]$
  - heterogeneity: $\tau = 0.41\ [0.21, 0.65]$



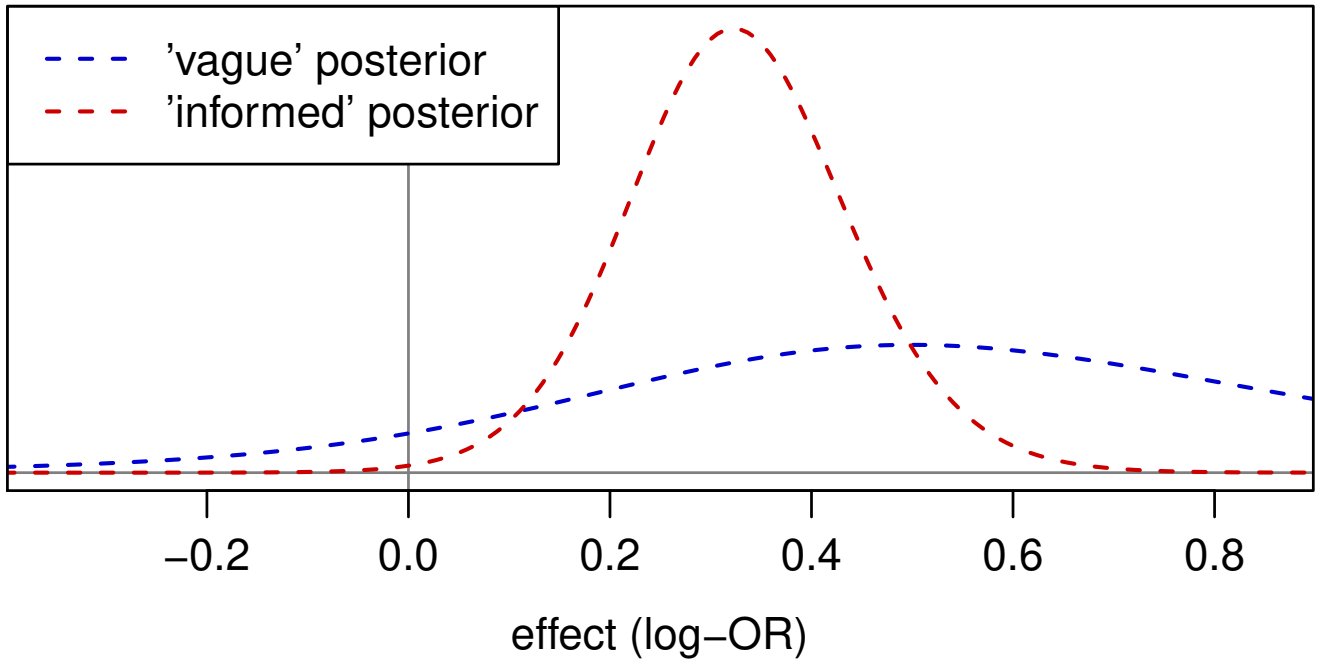
Example: children's effect posterior

- based on **vague** prior p_1 (only children's data):
  - effect: $\mu = 0.55\ [-0.24,\ 1.50]$
  - heterogeneity: $\tau = 0.49\ [0.00,\ 1.04]$



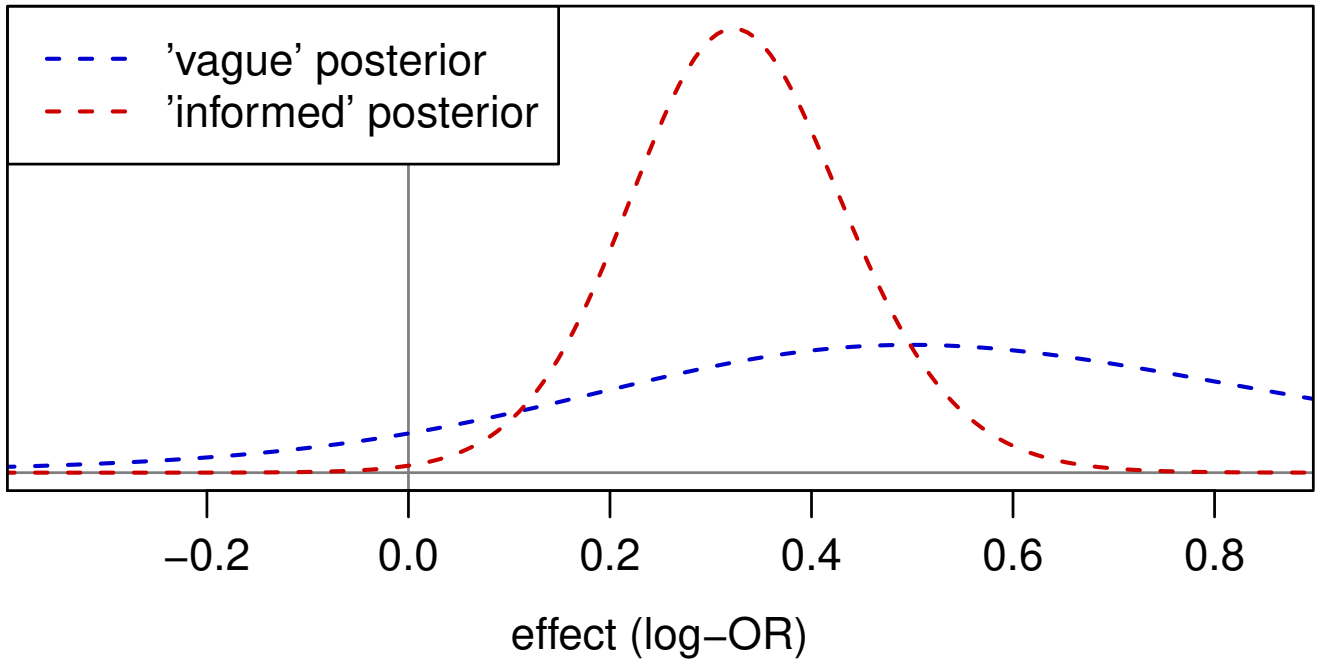
Example: children's effect posterior

- based on vague prior p_1 (only children's data):
  - effect: $\mu = 0.55\ [-0.24, 1.50]$
  - heterogeneity: $\tau = 0.49\ [0.00, 1.04]$
- based on **informative** prior p_2 (adolescents' + children's data):
  - effect: $\mu = 0.33\ [0.10, 0.56]$
  - heterogeneity: $\tau = 0.42\ [0.23, 0.64]$



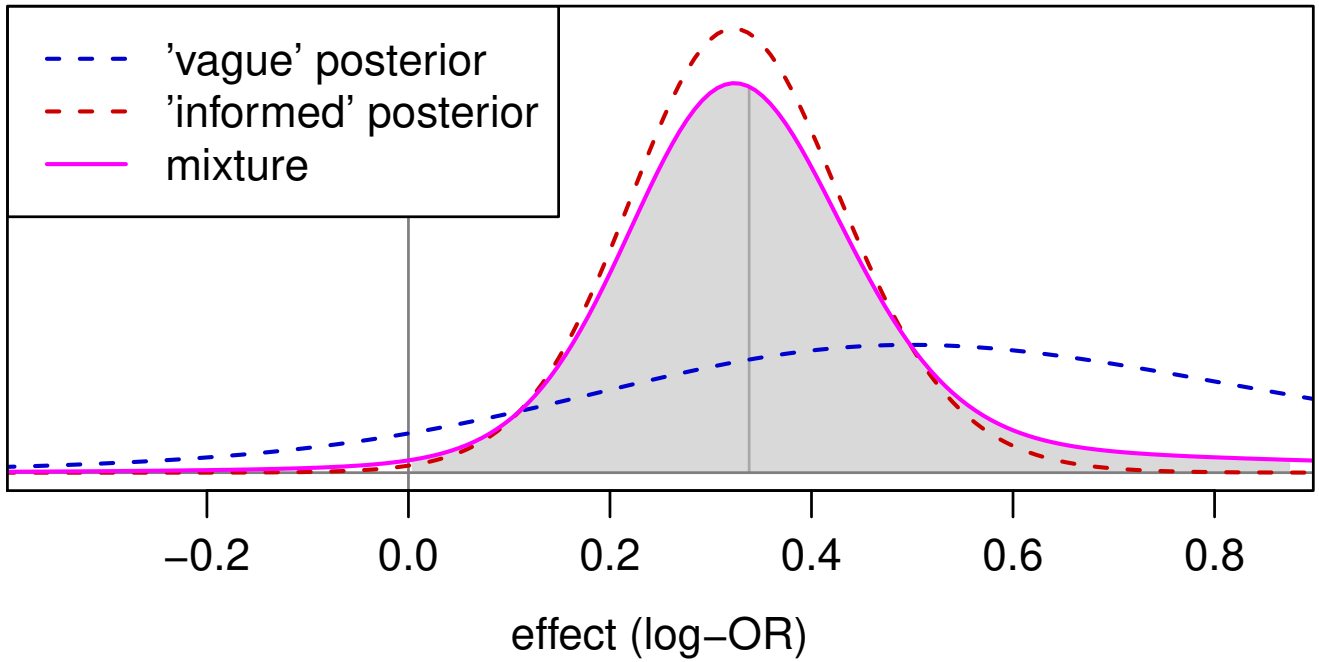
Example: children's effect posterior

- based on **vague** prior p_1 (only children's data): weight: 16.3%
  - effect: $\mu = 0.55\ [-0.24, 1.50]$
  - heterogeneity: $\tau = 0.49\ [0.00, 1.04]$
- based on **informative** prior p_2 weight: 83.7%
  (adolescents' + children's data):
  - effect: $\mu = 0.33\ [0.10, 0.56]$
  - heterogeneity: $\tau = 0.42\ [0.23, 0.64]$
- **Bayes factor**: 5.12



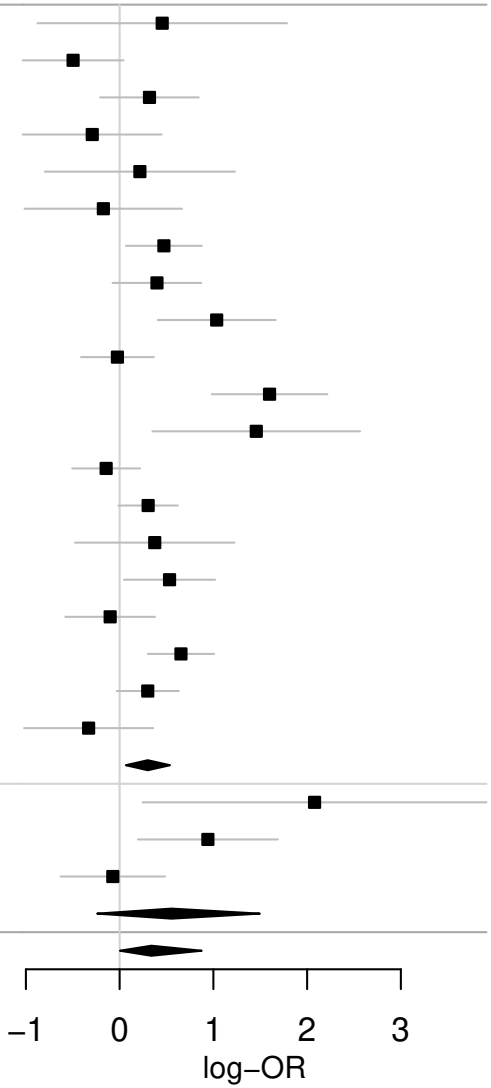
Example: children's effect posterior

- based on **vague** prior p_1 (only children's data): weight: 16.3%
  - effect: $\mu = 0.55\ [-0.24, 1.50]$
  - heterogeneity: $\tau = 0.49\ [0.00, 1.04]$
- based on **informative** prior p_2 weight: 83.7%
  (adolescents' + children's data):
  - effect: $\mu = 0.33\ [0.10, 0.56]$
  - heterogeneity: $\tau = 0.42\ [0.23, 0.64]$
- Bayes factor: 5.12



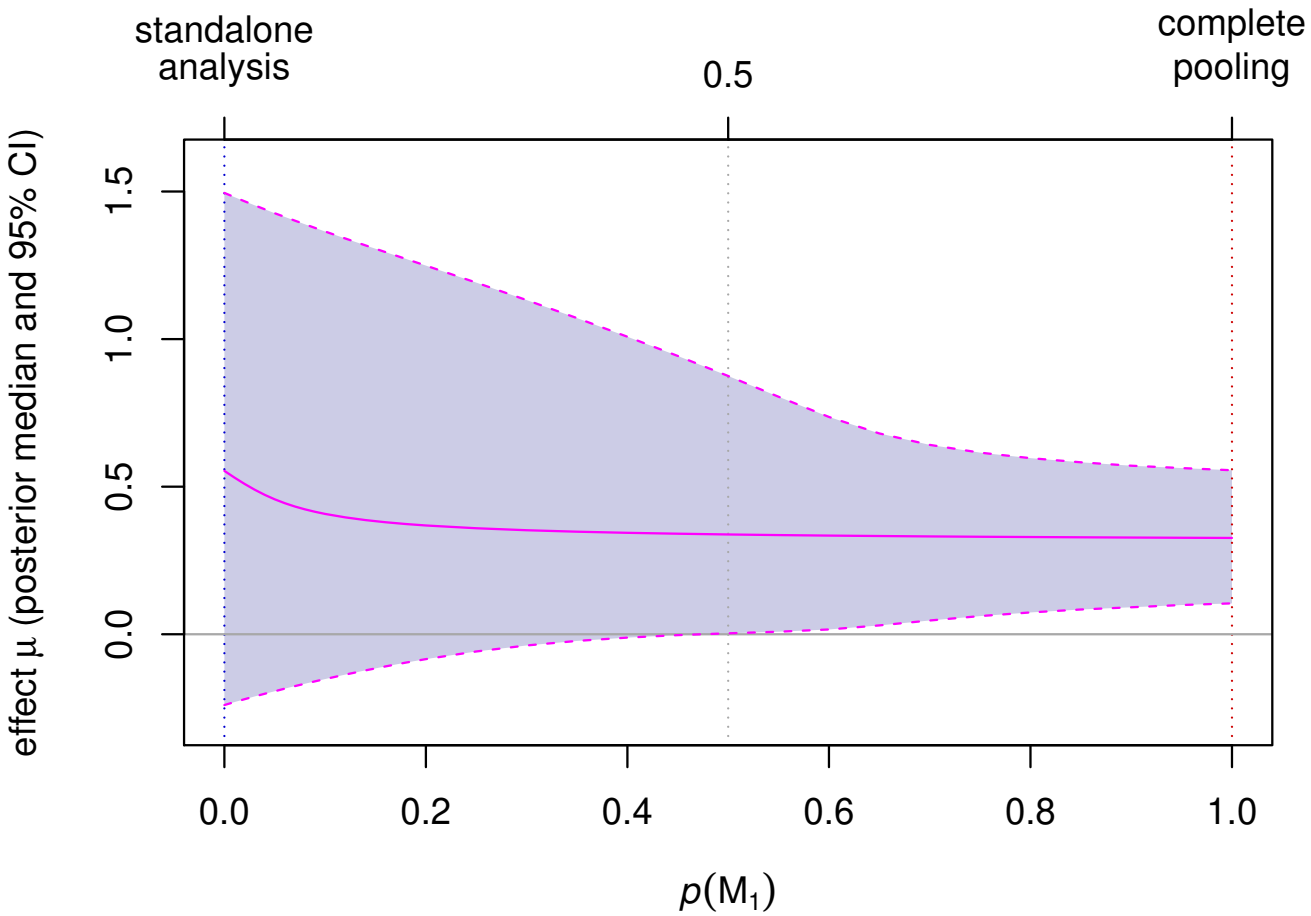
Example: estimates

publication	subjects	log–OR	95% CI
Hämäläinen (1997b)	adolescents	0.454	[–0.876, 1.785]
Rothner (1997)	adolescents	–0.496	[–1.034, 0.041]
Winner (1997)	adolescents	0.318	[–0.207, 0.844]
Rothner (1999a)	adolescents	–0.292	[–1.033, 0.449]
Rothner (1999b)	adolescents	0.216	[–0.797, 1.230]
Rothner (1999c)	adolescents	–0.174	[–1.014, 0.666]
Winner (2000)	adolescents	0.472	[0.068, 0.876]
Winner (2002)	adolescents	0.398	[–0.076, 0.872]
Ahonen (2004)	adolescents	1.035	[0.406, 1.664]
Visser (2004a)	adolescents	–0.024	[–0.412, 0.364]
Ahonen (2006)	adolescents	1.599	[0.982, 2.216]
Evers (2006)	adolescents	1.458	[0.350, 2.565]
Rothner (2006)	adolescents	–0.144	[–0.506, 0.218]
Winner (2006)	adolescents	0.304	[–0.013, 0.621]
Callenbach (2007)	adolescents	0.375	[–0.477, 1.226]
Lewis (2007)	adolescents	0.533	[0.046, 1.019]
Winner (2007)	adolescents	–0.101	[–0.579, 0.377]
Linder (2008)	adolescents	0.654	[0.300, 1.008]
Ho (2012)	adolescents	0.300	[–0.031, 0.631]
Fujita (2014)	adolescents	–0.331	[–1.019, 0.357]
adolescents only		**0.300**	**[0.066, 0.537]**
Ueberall (1999)	children	2.079	[0.246, 3.913]
Hämäläinen (2002)	children	0.941	[0.195, 1.688]
Ho (2012)	children	–0.073	[–0.630, 0.485]
children only		**0.554**	**[–0.240, 1.495]**
children combined		**0.338**	**[0.003, 0.875]**



Example: sensitivity check

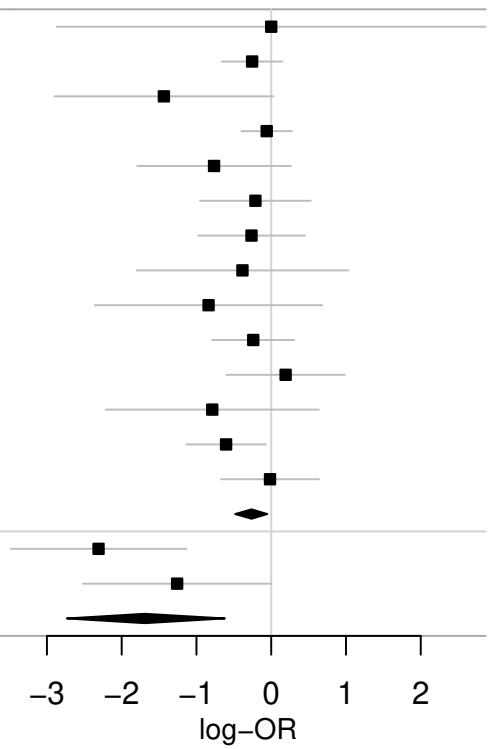
- what role does the specification of **prior weight** w play?
- $w = 0 \quad \Rightarrow \quad$ ignorance of adolescent data
- $w = 1 \quad \Rightarrow \quad$ complete pooling



2nd example: paediatric transplantation

- effect of Interleukin-2 receptor antagonists (IL-2RA) on acute rejection reaction after liver transplantation (OR < 1 indicates benefit)
- 2 RCTs in children[10], 14 in adults[11]. In conflict?

publication	subjects	log–OR	95% CI
Washburn (2001)	adults	0.000	[–2.869, 2.869]
Neuhaus (2002)	adults	–0.256	[–0.663, 0.152]
Yan (2004)	adults	–1.435	[–2.900, 0.030]
Boillot (2005)	adults	–0.060	[–0.399, 0.278]
Fasola (2005)	adults	–0.765	[–1.792, 0.263]
Yoshida (2005)	adults	–0.211	[–0.952, 0.529]
de Simone (2007)	adults	–0.264	[–0.978, 0.450]
Kato, cohort 1 (2007)	adults	–0.385	[–1.801, 1.031]
Kato, cohort 2 (2007)	adults	–0.838	[–2.358, 0.683]
Klintmalm (2007)	adults	–0.241	[–0.789, 0.308]
Schmeding (2007)	adults	0.193	[–0.599, 0.985]
Lupo (2008)	adults	–0.788	[–2.214, 0.637]
Neuberger (2009)	adults	–0.604	[–1.134, –0.074]
Calmus (2010)	adults	–0.016	[–0.671, 0.638]
adults only		**–0.263**	**[–0.482, –0.053]**
Heffron (2003)	children	–2.310	[–3.485, –1.135]
Spada (2006)	children	–1.258	[–2.517, –0.000]
children only		**–1.693**	**[–2.735, –0.620]**

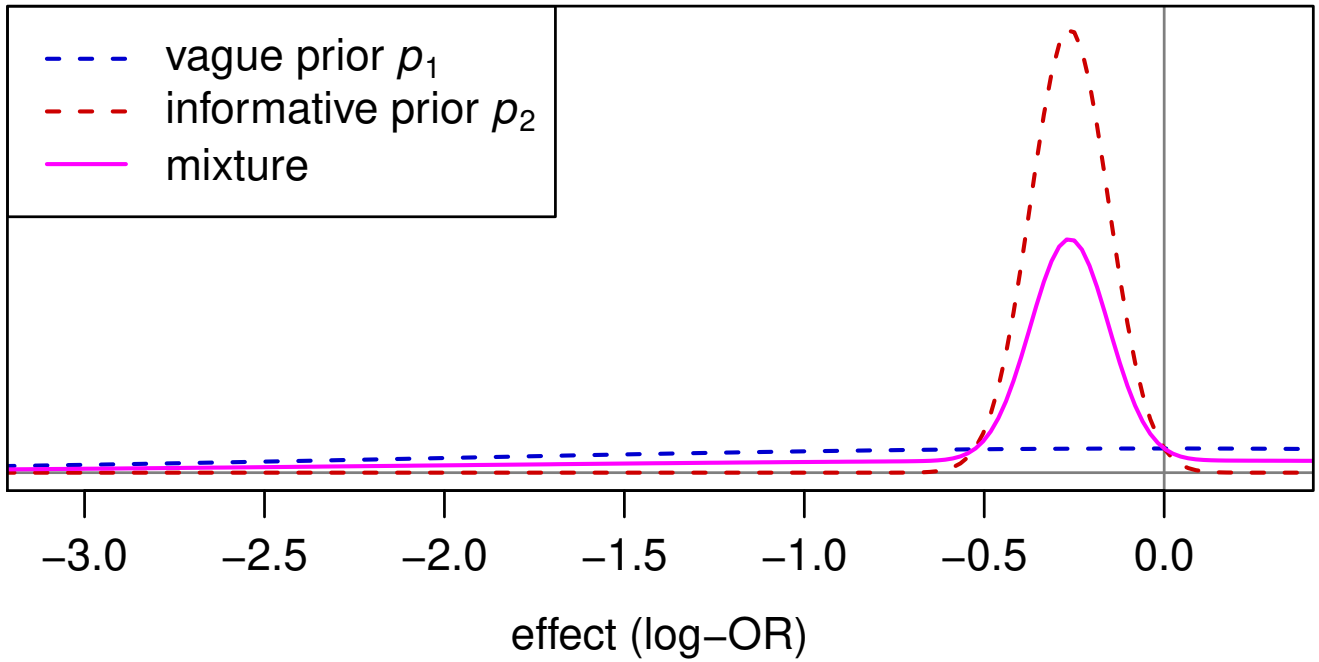


[10] N. Crins et al. Interleukin-2 receptor antagonists for pediatric liver transplant recipients: a systematic review ans meta-analysis of controlled studies. *Pediatric Transplantation*, 18(8):839–850, 2014.

[11] A. Goralczyk et al. Interleukin-2 receptor antagonists for liver transplant recipients: a systematic review ans meta-analysis of controlled studies. *Hepatology*, 54(2):541–554, 2011.

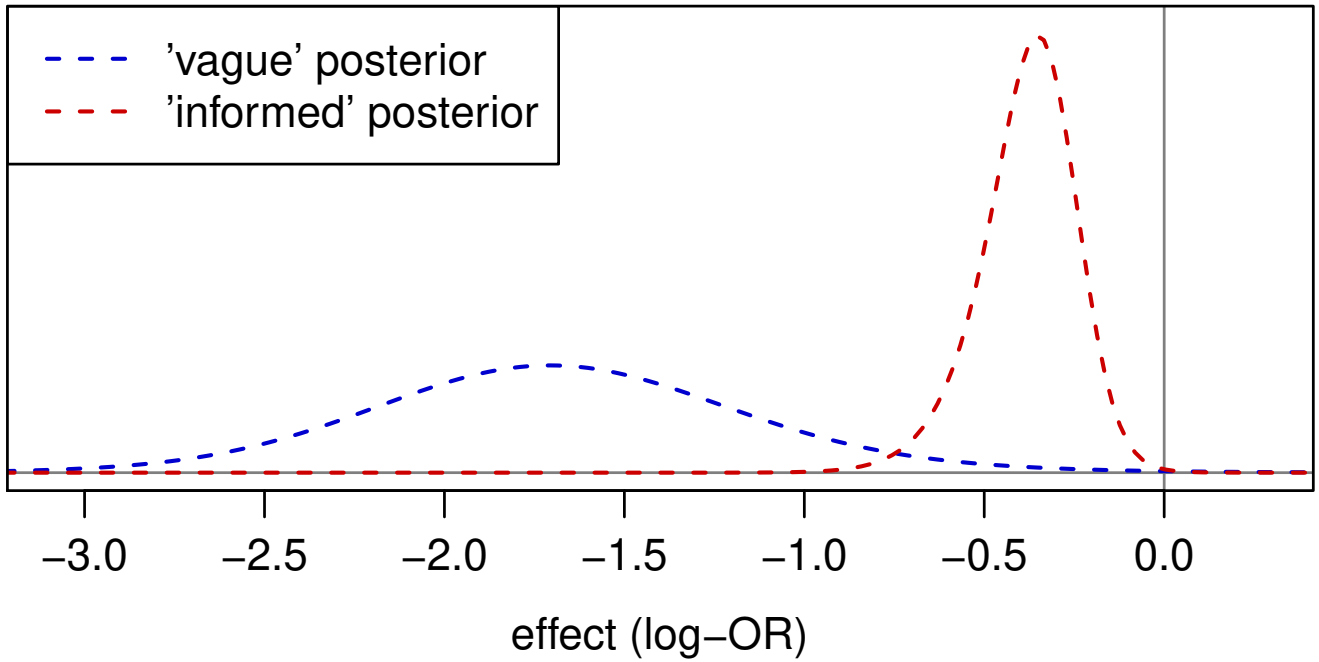
2nd example: children's effect prior setup

- **vague prior** p_1: $(1-w = 50\%)$
  - effect: $\mu \sim \mathrm{N}(0, 2^2)$
  - heterogeneity: $\tau \sim \mathrm{halfNormal}(0.5)$
- **informative prior** p_2 $(w = 50\%)$
  (posterior from adult studies):
  - effect: $\mu = -0.26\ [-0.48, 0.05]$
  - heterogeneity: $\tau = 0.11\ [0.00, 0.34]$



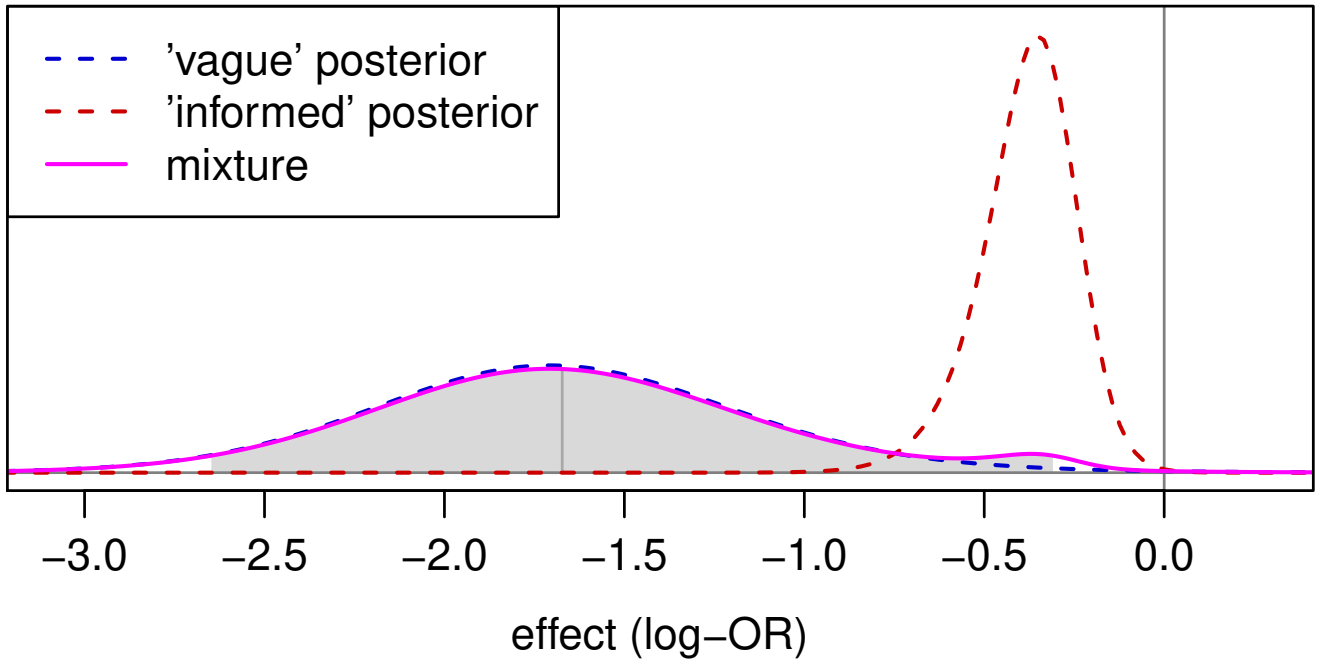
2nd example: children's effect posterior

- based on **vague** prior p_1 (only children's data): weight: 96.9%
  - effect: $\mu = -1.71$ $[-2.73, -0.62]$
  - heterogeneity: $\tau = 0.33$ $[0.00, 0.94]$
- based on **informative** prior p_2 weight: 3.1%
  (adolescents' + children's data):
  - effect: $\mu = -0.37$ $[-0.66, -0.13]$
  - heterogeneity: $\tau = 0.22$ $[0.00, 0.55]$
- **Bayes factor**: 0.032



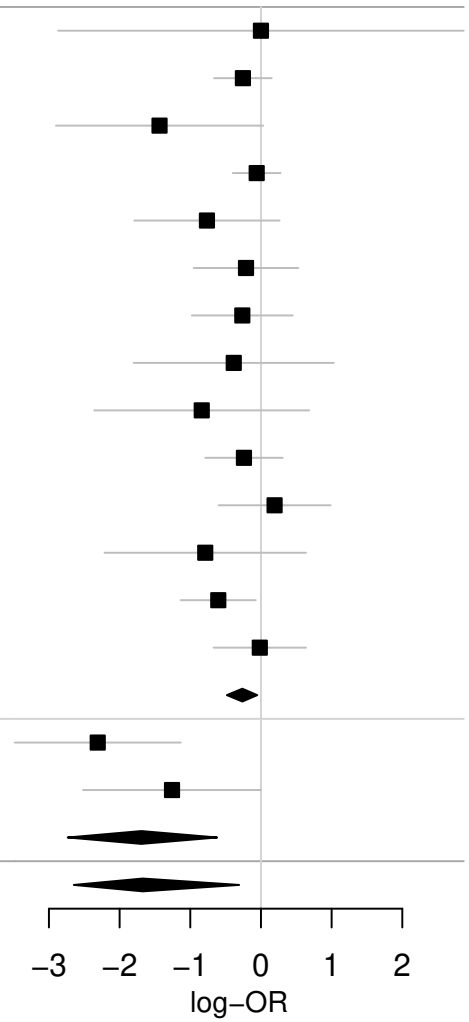
2nd example: children's effect posterior

- based on **vague** prior p_1 (only children's data): weight: 96.9%
  - effect: $\mu = -1.71$ $[-2.73, -0.62]$
  - heterogeneity: $\tau = 0.33$ $[0.00, 0.94]$
- based on **informative** prior p_2 weight: 3.1%
  (adolescents' + children's data):
  - effect: $\mu = -0.37$ $[-0.66, -0.13]$
  - heterogeneity: $\tau = 0.22$ $[0.00, 0.55]$
- **Bayes factor**: 0.032



2nd example: estimates

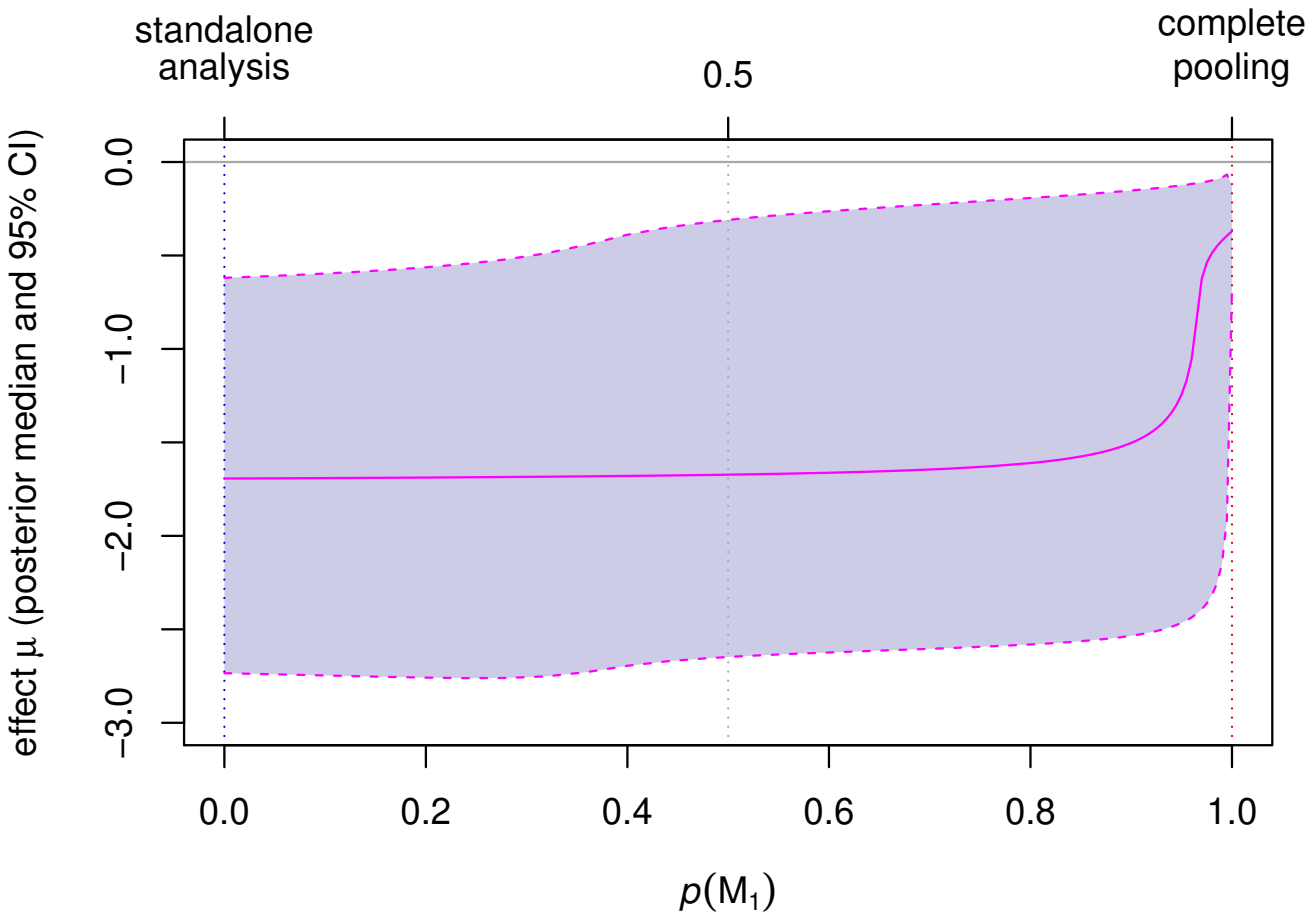
publication	subjects	log–OR	95% CI
Washburn (2001)	adults	0.000	[–2.869, 2.869]
Neuhaus (2002)	adults	–0.256	[–0.663, 0.152]
Yan (2004)	adults	–1.435	[–2.900, 0.030]
Boillot (2005)	adults	–0.060	[–0.399, 0.278]
Fasola (2005)	adults	–0.765	[–1.792, 0.263]
Yoshida (2005)	adults	–0.211	[–0.952, 0.529]
de Simone (2007)	adults	–0.264	[–0.978, 0.450]
Kato, cohort 1 (2007)	adults	–0.385	[–1.801, 1.031]
Kato, cohort 2 (2007)	adults	–0.838	[–2.358, 0.683]
Klintmalm (2007)	adults	–0.241	[–0.789, 0.308]
Schmeding (2007)	adults	0.193	[–0.599, 0.985]
Lupo (2008)	adults	–0.788	[–2.214, 0.637]
Neuberger (2009)	adults	–0.604	[–1.134, –0.074]
Calmus (2010)	adults	–0.016	[–0.671, 0.638]
adults only		**–0.263**	**[–0.482, –0.053]**
Heffron (2003)	children	–2.310	[–3.485, –1.135]
Spada (2006)	children	–1.258	[–2.517, –0.000]
children only		**–1.693**	**[–2.735, –0.620]**
children combined		**–1.673**	**[–2.648, –0.309]**



- prior/data conflict reflected in results

2nd example: sensitivity check

- check: effect of **prior weight** w
- $w = 0 \quad \Rightarrow \quad$ ignorance of adolescent data
- $w = 1 \quad \Rightarrow \quad$ complete pooling



Variations / extensions

More than two prior components

- choice of "vague" standard deviation (here: $\sigma = 2$) is relevant (affects Bayes factor: *Lindley's paradox*)

- may consider more than 2 prior components, e.g.:
  - common effect μ and heterogeneity τ ("complete pooling")
  - common heterogeneity τ only ("heterogeneity pooling")
  - common effect μ only ("effect pooling")
  - no common parameters ("standalone analyses")
- plausible?
- complex models may be barely distinguishable based on little data
- sparser models may be more desirable (*Ockham's razor*)

Conclusions

- meta analyses often based on few studies
  (especially subgroup analyses)
- Bayesian approach formalizes otherwise often informal extrapolation / model choice; incorporates uncertainty
- transparent information flow (prior distribution, Bayes factor,...)
- computations relatively easy using `bayesmeta` R package
- prior settings need to be chosen carefully
- may check sensitivity to model specifications
- model variations: mixtures of > 2 components
- many generalizations possible
  (other sources of external information,
  "main" analysis not a meta-analysis,...)

- C. Röver, S. Wandel, T. Friede.
  Model averaging for robust extrapolation in evidence synthesis.
  Statistics in Medicine (in press), 2018. `arXiv:1805.10890`

+++ additional slides +++

Example R code

Three meta analyses

```
########################
# main MA computations:

require("bayesmeta")
vaguepriorsd <- 2

# meta analysis for adolescents only:
bma.adol <- bayesmeta(y=logOR.adol, sigma=stdErr.adol,
                      mu.prior.mean=0, mu.prior.sd=vaguepriorsd,
                      tau.prior=function(t){dhalfnormal(t,scale=0.5)})

# meta analysis for children only:
bma.child <- bayesmeta(y=logOR.children, sigma=stdErr.children,
                       mu.prior.mean=0, mu.prior.sd=vaguepriorsd,
                       tau.prior=function(t){dhalfnormal(t,scale=0.5)})

# joint meta analysis for all patients:
bma.joint <- bayesmeta(y=c(logOR.adol, logOR.children),
                       sigma=c(stdErr.adol, stdErr.children),
                       mu.prior.mean=0, mu.prior.sd=vaguepriorsd,
                       tau.prior=function(t){dhalfnormal(t,scale=0.5)})
```

Example R code

Posterior

```
bayesfactor <- bma.joint$marginal / (bma.adol$marginal * bma.child$marginal)

prior.odds <- 0.5 / (1 - 0.5)

# determine posterior:
post.odds <- prior.odds * bayesfactor
post.prob <- post.odds / (post.odds + 1)
print(post.prob)  # =0.837

# plot posterior density:
x <- seq(-0.5, 1.0, le=200)
plot(x, (1-post.prob) * bma.child$dposterior(mu=x)
        + post.prob * bma.joint$dposterior(mu=x),
     type="l", xlab="log-OR", ylab="posterior density")

# (...)
```

Simulations

Coverage, CI width; homogeneous and heterogeneous scenarios

	weight	homogeneous		heterogeneous	
# studies	w (%)	coverage	width	coverage	width
$10 + 3$	0	98.7	(1.59)	94.5	(1.67)
	25	99.6	(1.29)	89.7	(1.59)
	50	99.5	(1.06)	81.9	(1.50)
	75	98.8	(0.86)	70.4	(1.38)
	100	97.1	(0.66)	15.6	(0.77)
$3 + 3$	0	98.7	(1.59)	94.5	(1.67)
	25	99.1	(1.42)	92.8	(1.61)
	50	99.1	(1.29)	90.0	(1.56)
	75	98.7	(1.18)	86.2	(1.48)
	100	98.1	(1.06)	74.9	(1.22)