The use of historical control data in clinical trials

Christian Röver

Department of Medical Statistics,
University Medical Center Göttingen,
Göttingen, Germany

Symposium “Statistical Planning of Translational Studies”
Göttingen
March 20, 2024

Incorporating historical controls in preclinical trials using meta-analytic models

Christian Röver

Department of Medical Statistics,
University Medical Center Göttingen,
Göttingen, Germany

Symposium “Statistical Planning of Translational Studies”
Göttingen
March 20, 2024

Introduction

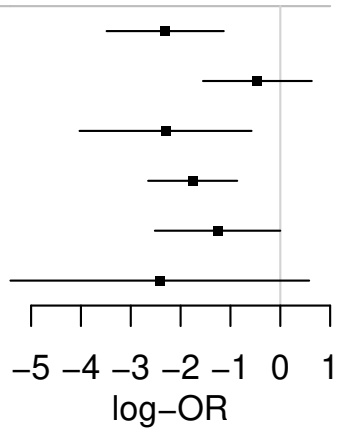
- clinical-trials background
  - meta-analysis, normal-normal model
  - mean effect, shrinkage estimation and prediction
- pre-clinical application
  - cross-checking / matching historical and concurrent data
  - augmenting a control group with historical data
- summary / outlook

The NNHM

The meta-analysis problem: pediatric transplantation example

- meta-analysis problem: want to model/combine measurements (usually: *effects*) that are associated with standard errors
- example: Effect of IL-2RA on *acute rejection* events in pediatric liver transplantation [1]

study	treatment	control	log–OR	95% CI
Heffron (2003)	14 / 61	15 / 20	–2.31	[–3.48, –1.13]
Gibelli (2004)	16 / 28	19 / 28	–0.46	[–1.55, 0.63]
Schuller (2005)	3 / 18	8 / 12	–2.30	[–4.03, –0.58]
Ganschow (2005)	9 / 54	29 / 54	–1.76	[–2.65, –0.86]
Spada (2006)	4 / 36	11 / 36	–1.26	[–2.52, –0.00]
Gras (2008)	0 / 50	3 / 34	–2.42	[–5.41, 0.58]

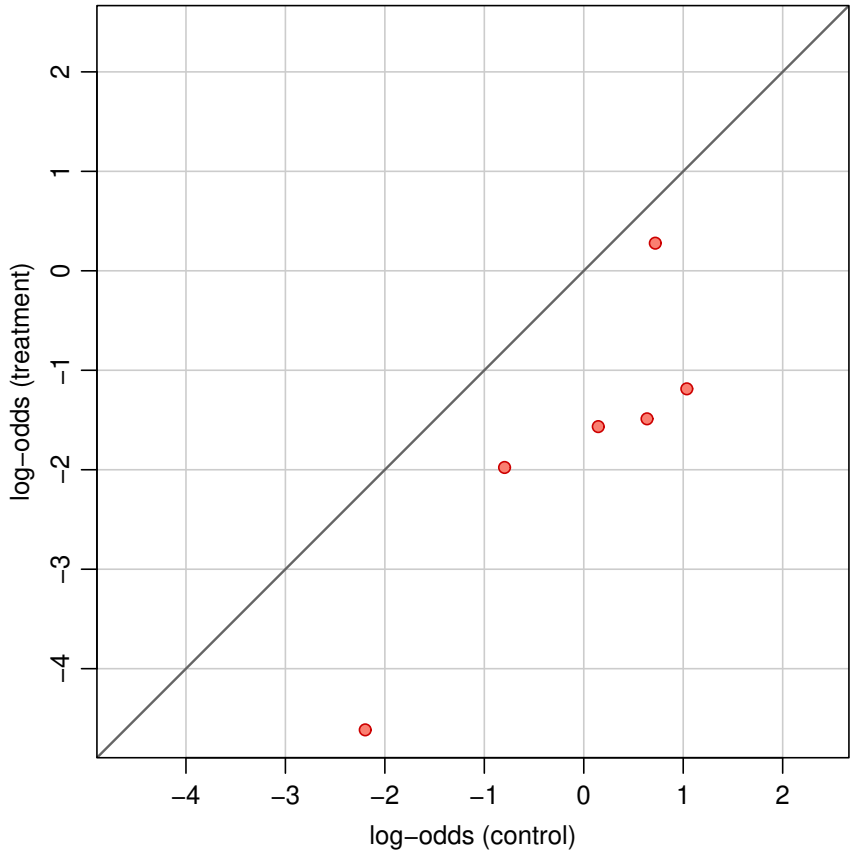


- idea: characterize/model “population” of (log-) ORs

[1] N.D. Crins, C. Röver, A.D. Goralczyk, T. Friede. Interleukin-2 receptor antagonists for pediatric liver transplant recipients: A systematic review and meta-analysis of controlled studies. *Pediatric Transplantation*, **18**(8):839–850, 2014.

The NNHM

Reminder: stratification

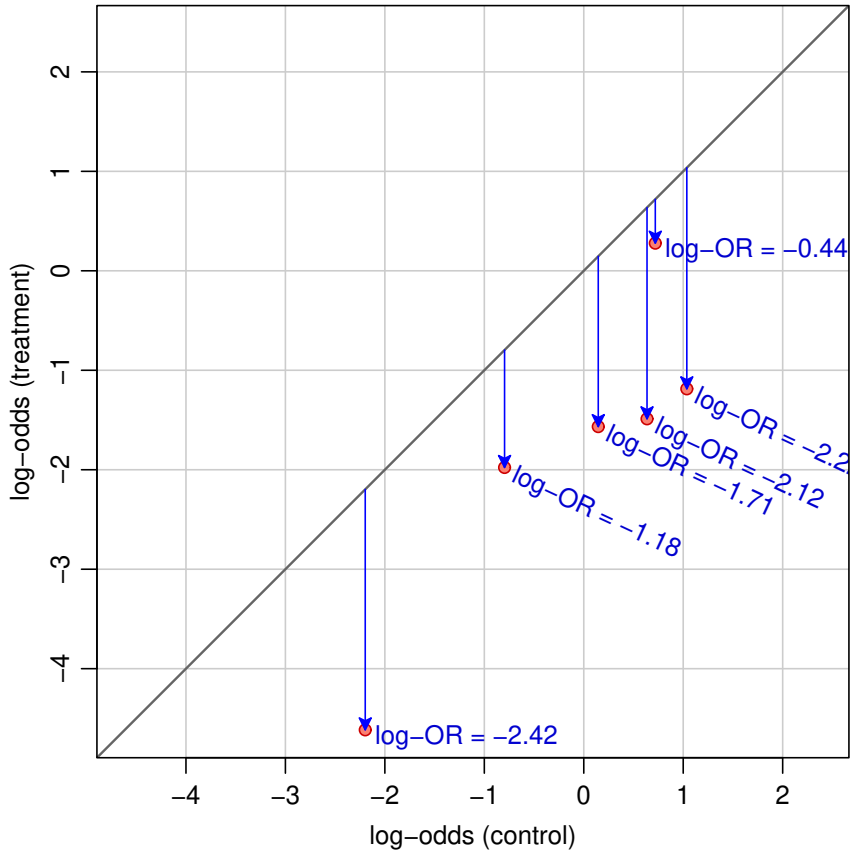


- important principle: **stratified analysis** of ("relative") effects (here: consideration of **log-ORs**)
- pooling of "baselines" (here: odds) may "break randomization" [2]

[2] S. Dias, A.E. Ades. Absolute or relative effects? Arm-based synthesis of trial data. *Research Synthesis Methods*, **7**(1):23–28, 2016.

The NNHM

Reminder: stratification

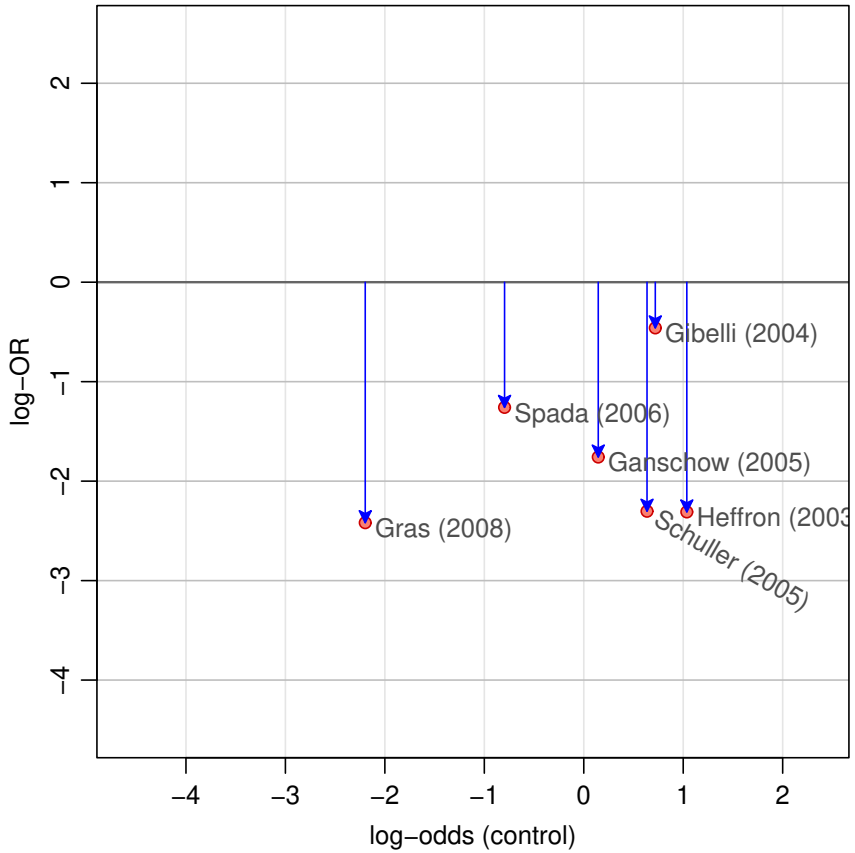


- important principle: **stratified analysis** of ("relative") effects (here: consideration of **log-ORs**)
- pooling of "baselines" (here: odds) may "break randomization" [2]

[2] S. Dias, A.E. Ades. Absolute or relative effects? Arm-based synthesis of trial data. *Research Synthesis Methods*, **7**(1):23–28, 2016.

The NNHM

Reminder: stratification



- important principle: **stratified analysis** of ("relative") effects (here: consideration of **log-ORs**)
- pooling of "baselines" (here: odds) may "break randomization" [2]

[2] S. Dias, A.E. Ades. Absolute or relative effects? Arm-based synthesis of trial data. *Research Synthesis Methods*, **7**(1):23–28, 2016.

- possible aims:
  - to characterize log-ORs,
    infer **location** (mean) and **scale** (variability)
  - mutual support of estimates — e.g., does consideration of previous studies constrain the uncertain estimate from the most recent study?
    → "**shrinkage estimation**"
  - **prediction** of a "future" study
    e.g., to aid in study design[3]

[3] H. Schmidli, B. Neuenschwander, T. Friede. Meta-analytic-predictive use of historical variance data for the design and analysis of clinical trials. *Computational Statistics and Data Analysis*, **113**:100–110, 2017.

The NNHM

The normal-normal hierarchical model (NNHM)

- have:
  - estimates y_i
  - standard errors σ_i

  $(i = 1, \ldots, k)$
- assume:

$$\begin{aligned} y_i|\theta_i &\sim \text{Normal}(\theta_i, \sigma_i^2) \\ \theta_i|\mu, \tau &\sim \text{Normal}(\mu, \tau^2) \end{aligned}$$

  or (marginally):

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

- random-effects (RE) model,
  normal-normal hierarchical model (NNHM)
- "**overall mean**" μ
- "**study-specific effects**" θ_i
- **heterogeneity** parameter $\tau \geq 0$

The NNHM

The technical bits

- frequentist and Bayesian approaches common;
  (focus on Bayesian here)
- implementation straightforward using MCMC (e.g., JAGS, Stan)
- semi-analytically implemented in `bayesmeta` R package [4, 5]
- prior for overall mean μ:
  uniform often sensible; for informative priors: normal [5]
- prior for heterogeneity τ:
  uninformative priors possible (sensible) for large k
  otherwise (small k): (weakly) informative priors necessary. [6]

---

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

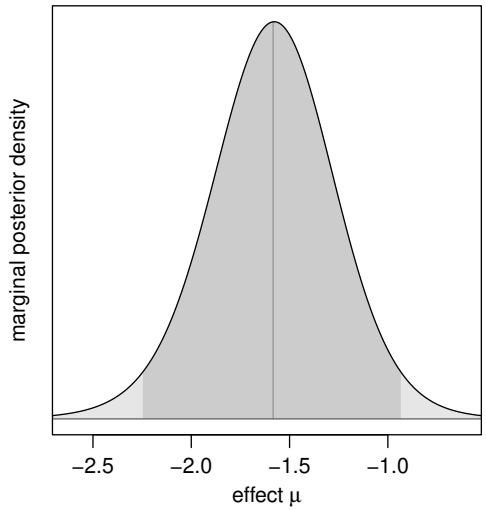
[5] C. Röver. Bayesian random-effects meta-analysis using the bayesmeta R package. *Journal of Statistical Software*, **93**(6):1–51, 2020.

[6] 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.

The NNHM

Pediatric transplantation example

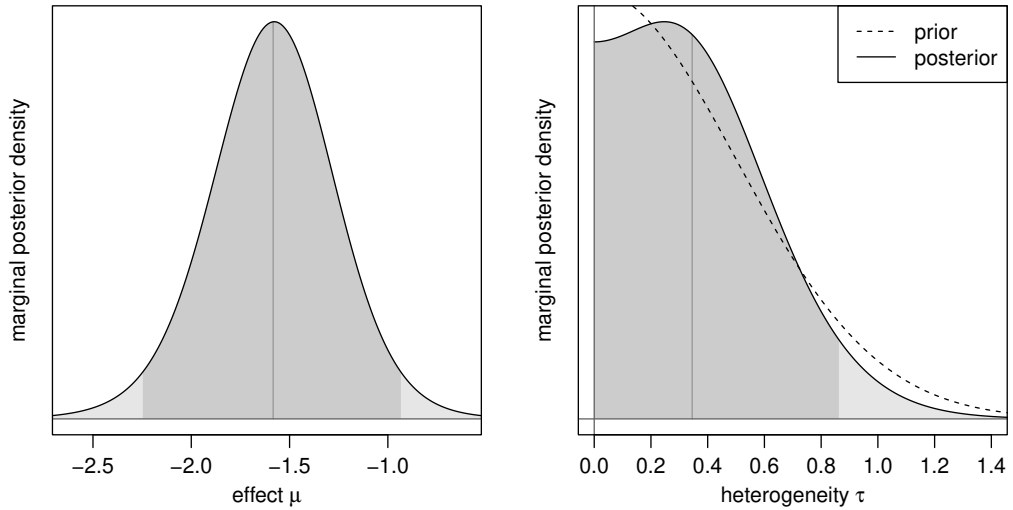
- may infer mean effect (μ)



The NNHM

Pediatric transplantation example

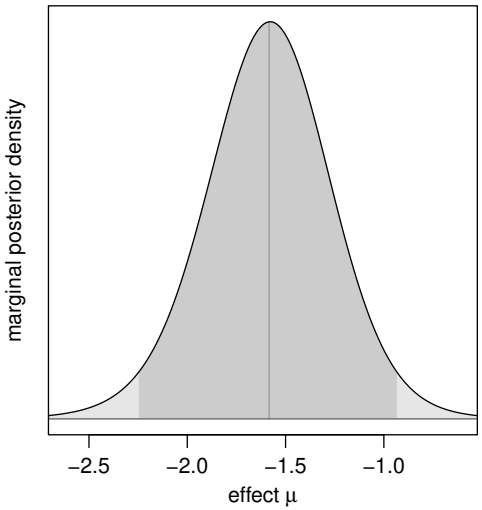
- may infer mean effect (μ), heterogeneity (τ)

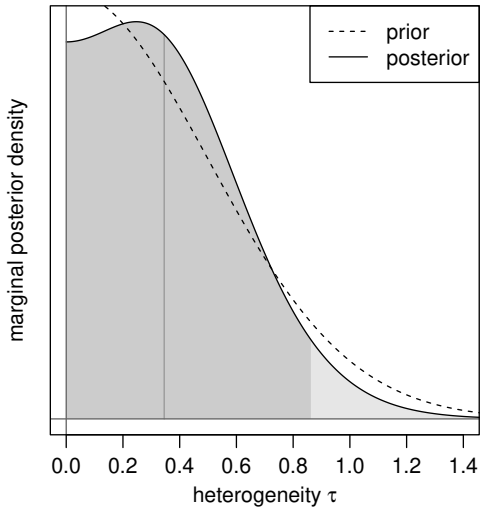


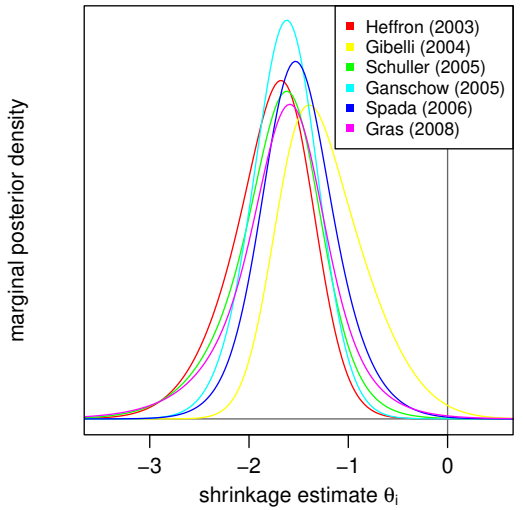
The NNHM

Pediatric transplantation example

- may infer mean effect (μ), heterogeneity (τ), shrinkage estimates (θ_i)



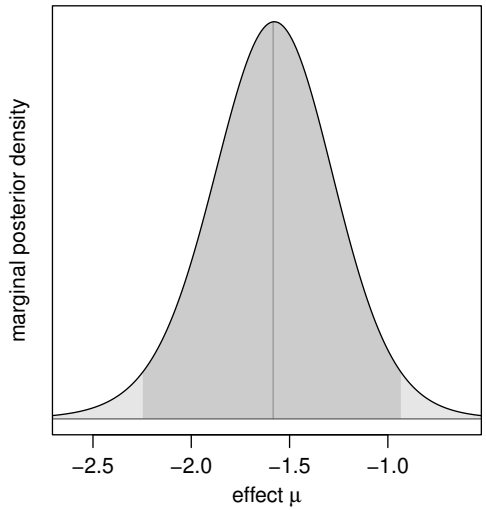


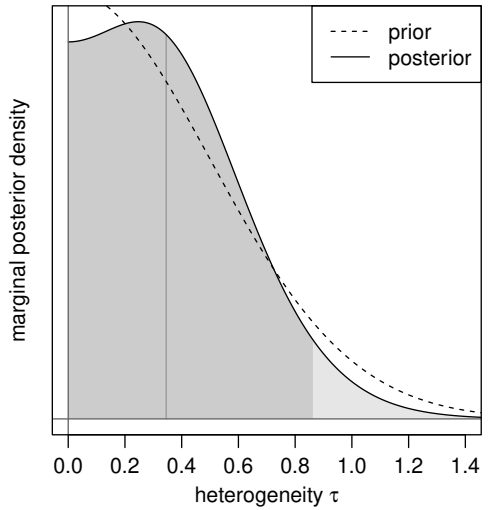


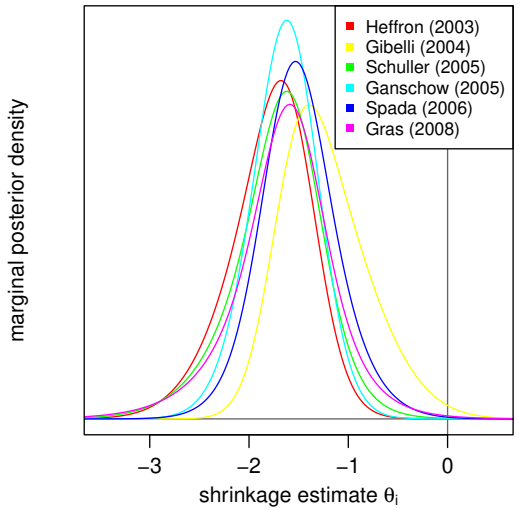
The NNHM

Pediatric transplantation example

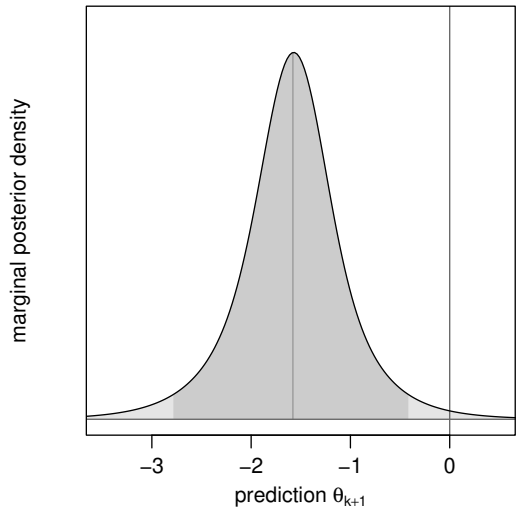
- may infer mean effect (μ), heterogeneity (τ), shrinkage estimates (θ_i)







- and prediction ($\theta_{k+1}|\mu, \tau \sim \text{Normal}(\mu, \tau^2)$)

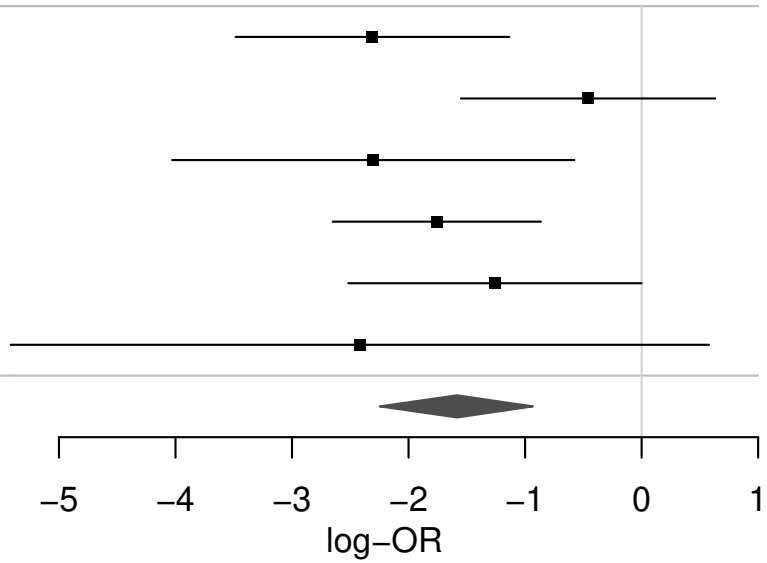


The NNHM

Example

- results commonly illustrated in a “forest plot”

study	estimate	95% CI
Heffron (2003)	–2.31	[–3.48, –1.13]
Gibelli (2004)	–0.46	[–1.55, 0.63]
Schuller (2005)	–2.30	[–4.03, –0.58]
Ganschow (2005)	–1.76	[–2.65, –0.86]
Spada (2006)	–1.26	[–2.52, –0.00]
Gras (2008)	–2.42	[–5.41, 0.58]
mean	**–1.58**	**[–2.25, –0.93]**



Heterogeneity (tau): 0.35 [0.00, 0.86]

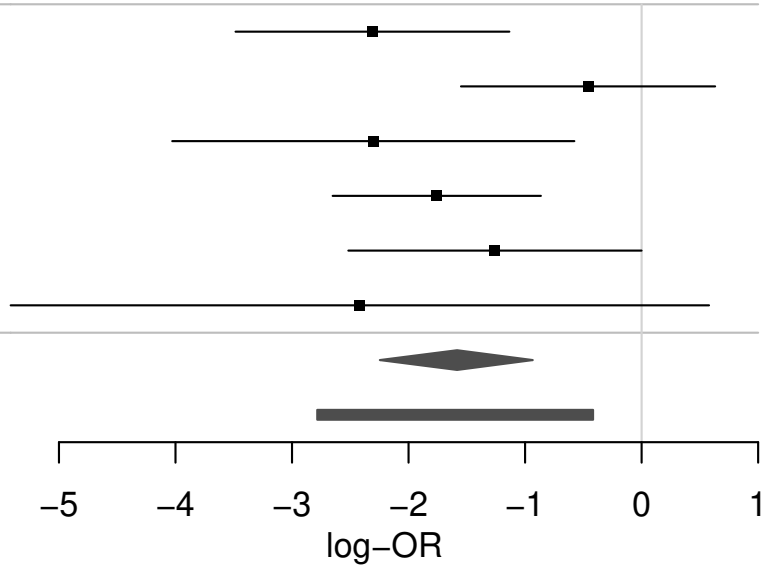
- overall mean (μ) estimate shown as a diamond

The NNHM

Example

- results commonly illustrated in a “forest plot”

study	estimate	95% CI
Heffron (2003)	–2.31	[–3.48, –1.13]
Gibelli (2004)	–0.46	[–1.55, 0.63]
Schuller (2005)	–2.30	[–4.03, –0.58]
Ganschow (2005)	–1.76	[–2.65, –0.86]
Spada (2006)	–1.26	[–2.52, –0.00]
Gras (2008)	–2.42	[–5.41, 0.58]
mean	**–1.58**	**[–2.25, –0.93]**
prediction	**–1.58**	**[–2.78, –0.42]**



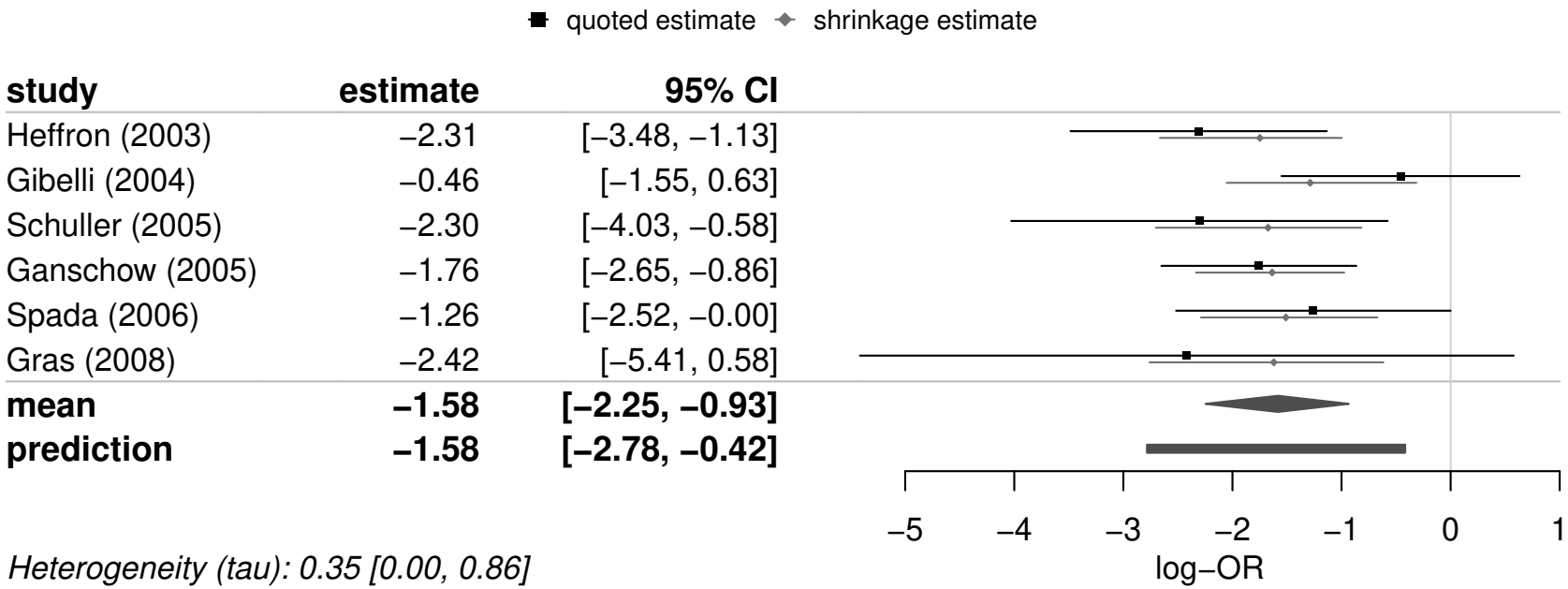
Heterogeneity (tau): 0.35 [0.00, 0.86]

- overall mean (μ) estimate shown as a diamond
- prediction interval

The NNHM

Example

- results commonly illustrated in a “forest plot”



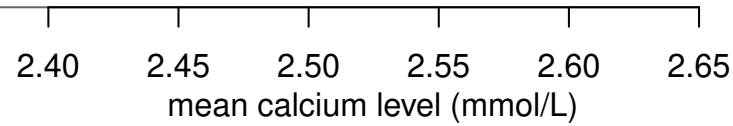
- overall mean (μ) estimate shown as a diamond
- prediction interval
- shrinkage estimates

Historical control data

Serum calcium example data set [7]

- 17 historical, 1 concurrent control group

study	N	mean	se
Study−15	10	2.501	0.018
Study−16	10	2.564	0.020
Study−17	10	2.550	0.019
Study−18	10	2.528	0.029
Study−19	10	2.488	0.012
Study−20	5	2.508	0.022
Study−21	5	2.524	0.031
Study−22	10	2.612	0.016
Study−23	10	2.480	0.013
Study−24	10	2.572	0.019
Study−25	10	2.609	0.008
Study−26	10	2.508	0.015
Study−27	10	2.561	0.025
Study−28	6	2.570	0.018
Study−29	10	2.437	0.019
Study−30	10	2.511	0.012
Study−32	10	2.533	0.031
concurrent	10	2.570	0.019



[7] A. Gurjanov, A. Kreuchwig, T. Steger-Hartmann, L.A.I. Vaas. Hurdles and signposts on the road to virtual control groups — A case study illustrating the influence of anesthesia protocols on electrolyte levels in rats. *Frontiers in Pharmacology* **14**:1142534, 2023.

Historical control data

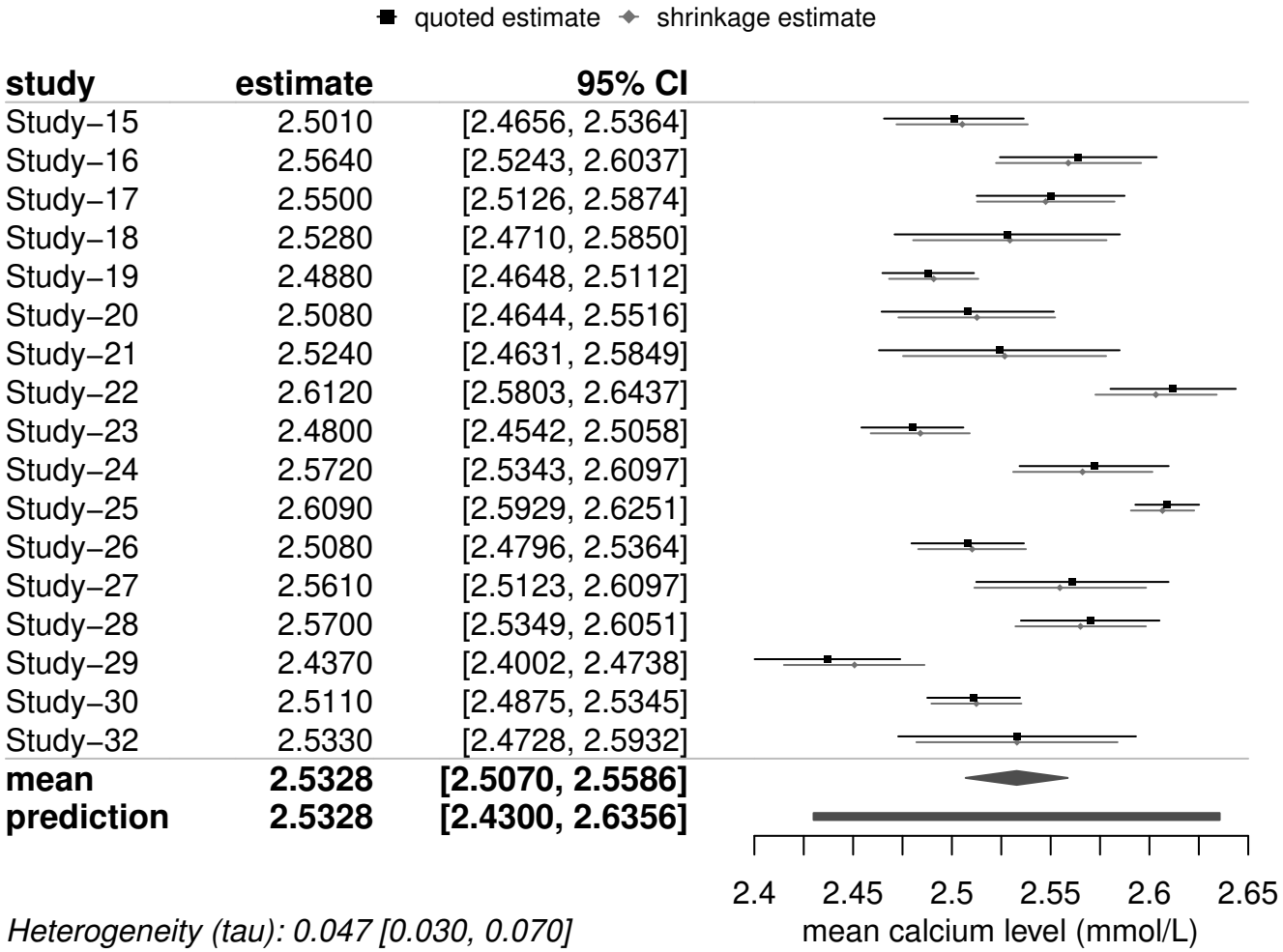
Matching historical and concurrent data

- are new data "**consistent**" with historical data?
- link between historical and new data: **predictive distribution**;
  in particular:
  - (posterior) predictive distribution $p(\theta_{18}|y_1, \ldots, y_{17})$
  - new observation y_{18} (with s.e. σ_{18})
  - link: $y_{18}|\theta_{18}, \sigma_{18} \sim \text{Normal}(\theta_{18}, \sigma_{18}^2)$
- predictive distribution of y_{18} (rather than θ_{18})
  (i.e.: $p(y_{18}|y_1, \ldots, y_{17})$)
  results as a sum of predictive distribution and an additional normal term.

Historical control data

Matching historical and concurrent data

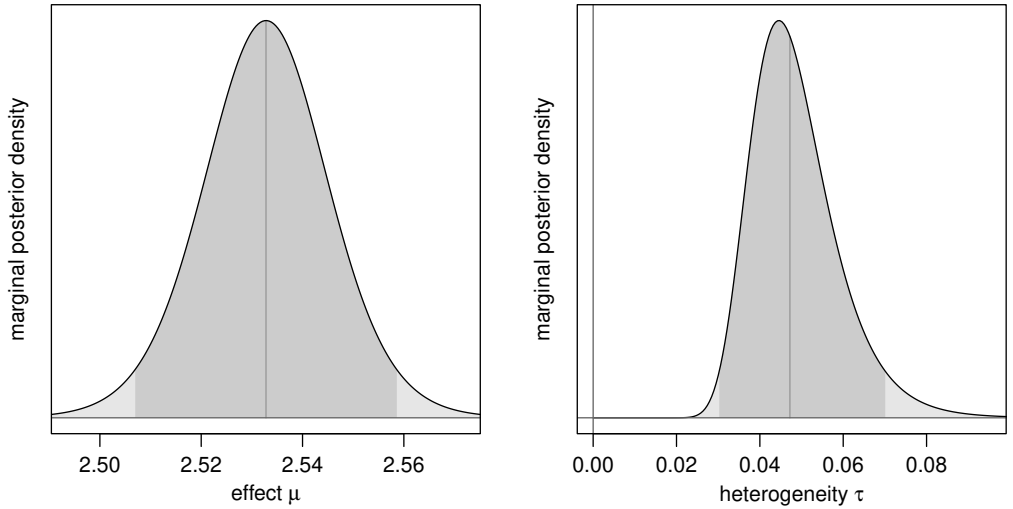
- perform meta analysis of 17 historical estimates. . .



Historical control data

Matching historical and concurrent data

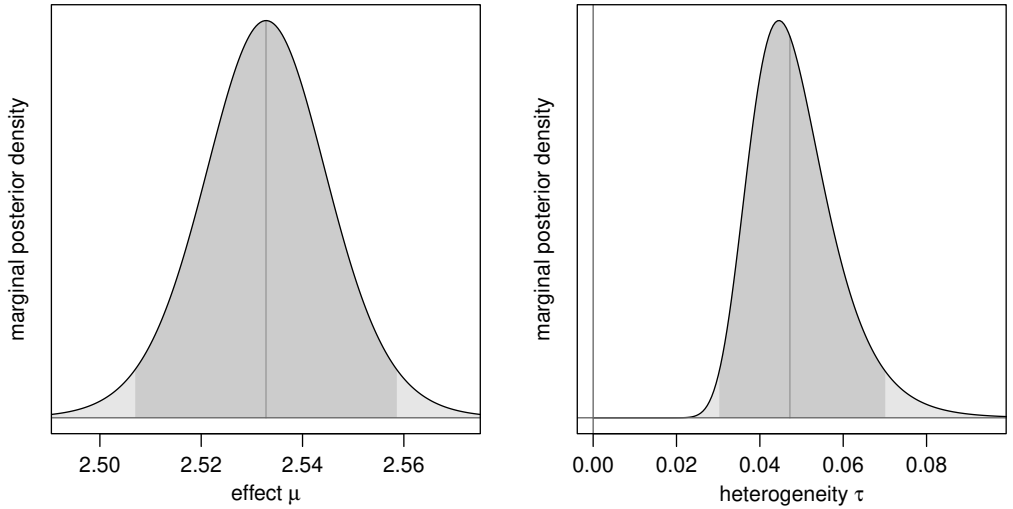
- infer mean effect (μ) and heterogeneity (τ)



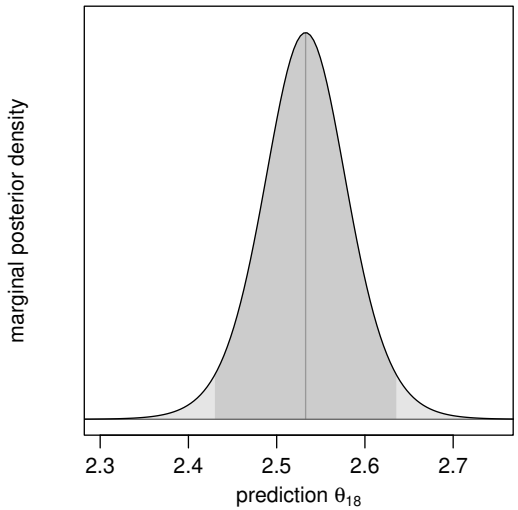
Historical control data

Matching historical and concurrent data

- infer mean effect (μ) and heterogeneity (τ)



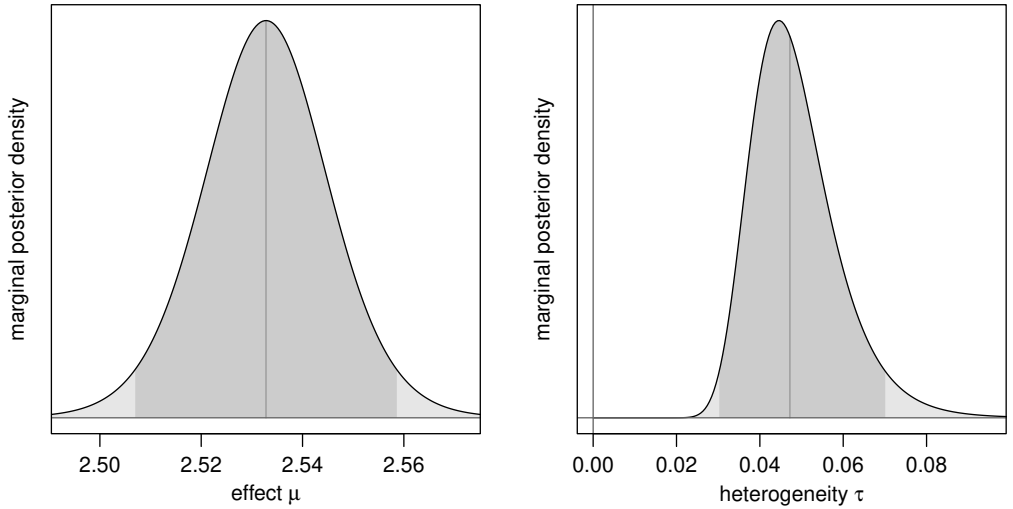
- derive prediction (parameter θ_{18})



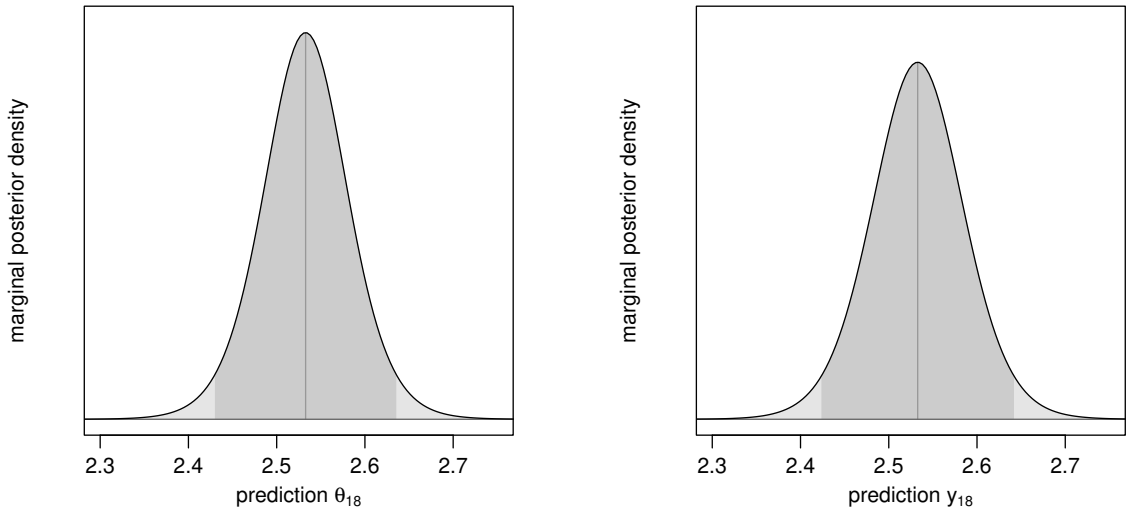
Historical control data

Matching historical and concurrent data

- infer mean effect (μ) and heterogeneity (τ)



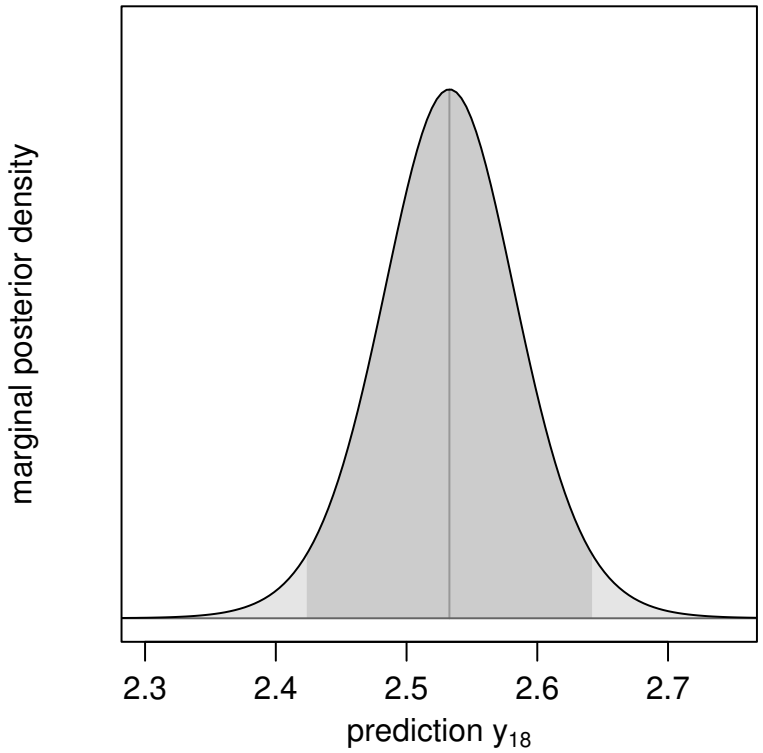
- derive prediction (parameter θ_{18}) **and data (average y_{18})**



Historical control data

Matching historical and concurrent data

- prediction (average y_{18})

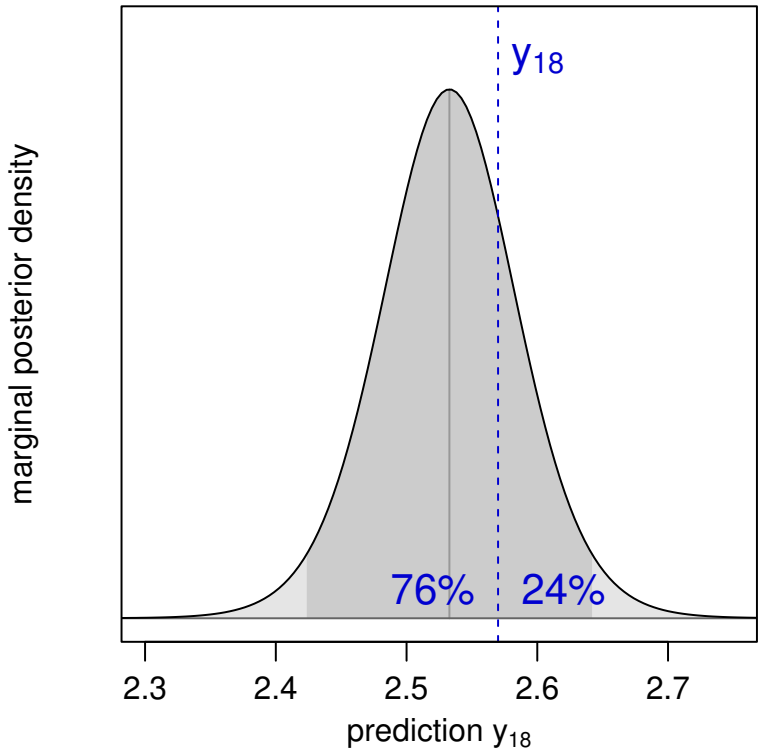


[8] T. Gneiting, F. Balabdaoui, A.E. Raftery. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society B*, **69**(2), 243–268, 2007.

Historical control data

Matching historical and concurrent data

- prediction (average y_{18})



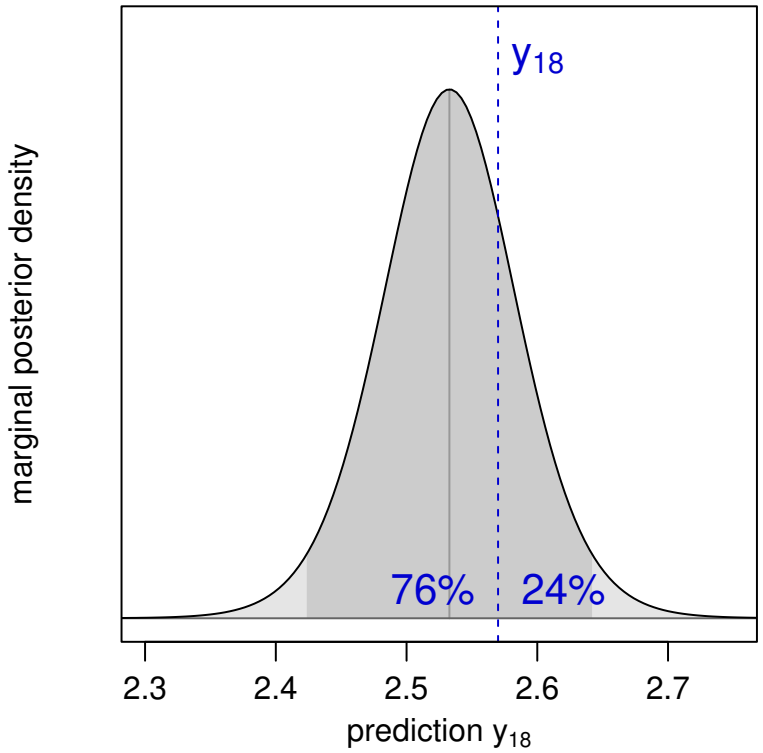
- actual measurement (average $y_{18} = 2.57$) is at 76% quantile,
  $\rightarrow$ consistent with historical means

[8] T. Gneiting, F. Balabdaoui, A.E. Raftery. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society B*, **69**(2), 243–268, 2007.

Historical control data

Matching historical and concurrent data

- prediction (average y_{18})



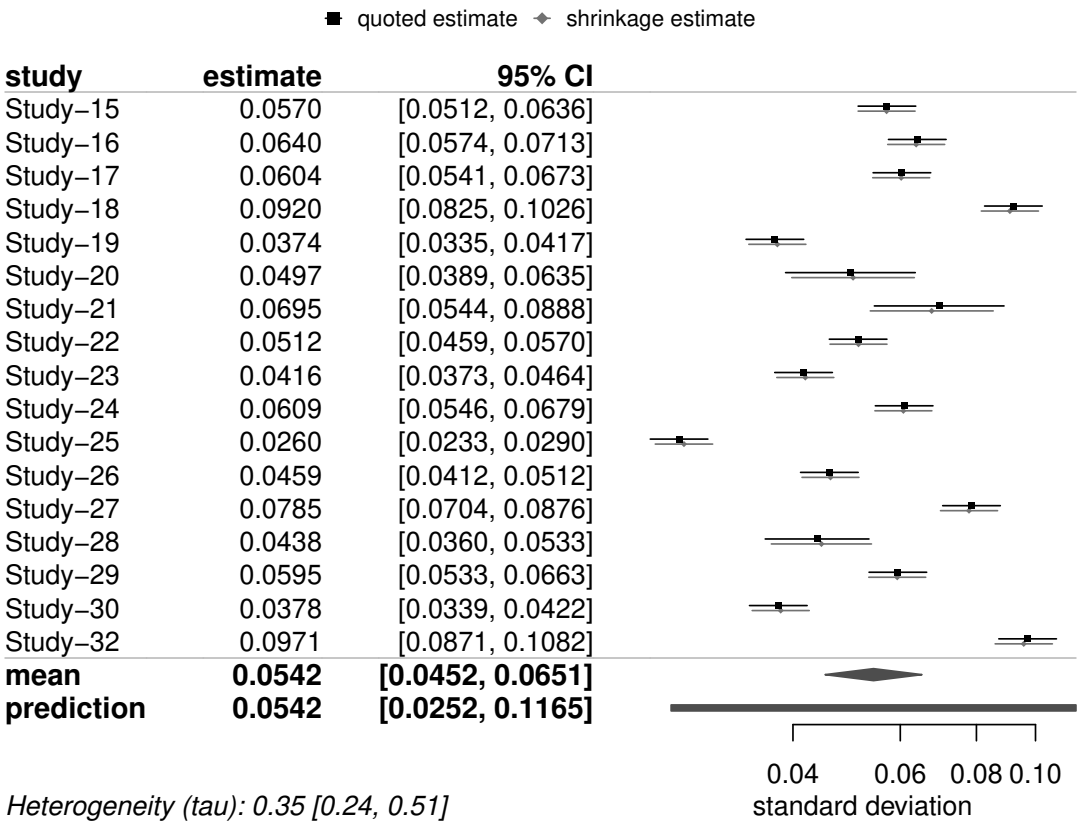
- actual measurement (average $y_{18} = 2.57$) is at 76% quantile,
  $\rightarrow$ consistent with historical means
- if all assumptions are met, percentage should follow a uniform distribution [8]

[8] T. Gneiting, F. Balabdaoui, A.E. Raftery. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society B*, **69**(2), 243–268, 2007.

Historical control data

Matching historical and concurrent data

- may perform analogous analysis based on historical + concurrent (logarithmic) standard deviations



- (observed standard deviation ($s_{18} = 0.06$) ends up at 61% quantile)

Historical control data

Jointly incorporating historical and concurrent data: MAC vs. MAP

- How can historical data augment the current study's control data? [9]

[9] A. Gurjanov, C. Vieira-Vieira, J. Vienenkoetter, L.A.I. Vaas, T. Steger-Hartmann. Replacing concurrent controls with virtual control groups in rat toxicity studies. *Regulatory Toxicology and Pharmacology* **148**:105592, 2024.

[10] 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

Historical control data

Jointly incorporating historical and concurrent data: MAC vs. MAP

- How can historical data augment the current study's control data? [9]
- based on the NNHM, two obvious options:
  - perform a joint meta analysis of all 18 studies, derive a shrinkage estimate for the 18th study (θ_{18}) ("**meta-analytic-combined (MAC) approach**")

[9] A. Gurjanov, C. Vieira-Vieira, J. Vienenkoetter, L.A.I. Vaas, T. Steger-Hartmann. Replacing concurrent controls with virtual control groups in rat toxicity studies. *Regulatory Toxicology and Pharmacology* **148**:105592, 2024.

[10] 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

Historical control data

Jointly incorporating historical and concurrent data: MAC vs. MAP

- How can historical data augment the current study's control data? [9]
- based on the NNHM, two obvious options:
  - perform a joint meta analysis of all 18 studies,
    derive a shrinkage estimate for the 18th study (θ_{18})
    ("**meta-analytic-combined (MAC) approach**")
  - perform a meta analysis of 17 historical studies,
    use the predictive distribution as a prior for the 18th study (θ_{18})
    ("**meta-analytic-predictive (MAP) approach**")

[9] A. Gurjanov, C. Vieira-Vieira, J. Vienenkoetter, L.A.I. Vaas, T. Steger-Hartmann. Replacing concurrent controls with virtual control groups in rat toxicity studies. *Regulatory Toxicology and Pharmacology* **148**:105592, 2024.

[10] 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

Historical control data

Jointly incorporating historical and concurrent data: MAC vs. MAP

- How can historical data augment the current study's control data? [9]
- based on the NNHM, two obvious options:
  - perform a joint meta analysis of all 18 studies,
    derive a shrinkage estimate for the 18th study (θ_{18})
    ("**meta-analytic-combined (MAC) approach**")
  - perform a meta analysis of 17 historical studies,
    use the predictive distribution as a prior for the 18th study (θ_{18})
    ("**meta-analytic-predictive (MAP) approach**")
- it turns out, both approaches are **equivalent** [10]
- analogous approach may be used in the context of clinical trials
  (using historical data from control groups in similar experiments)

---

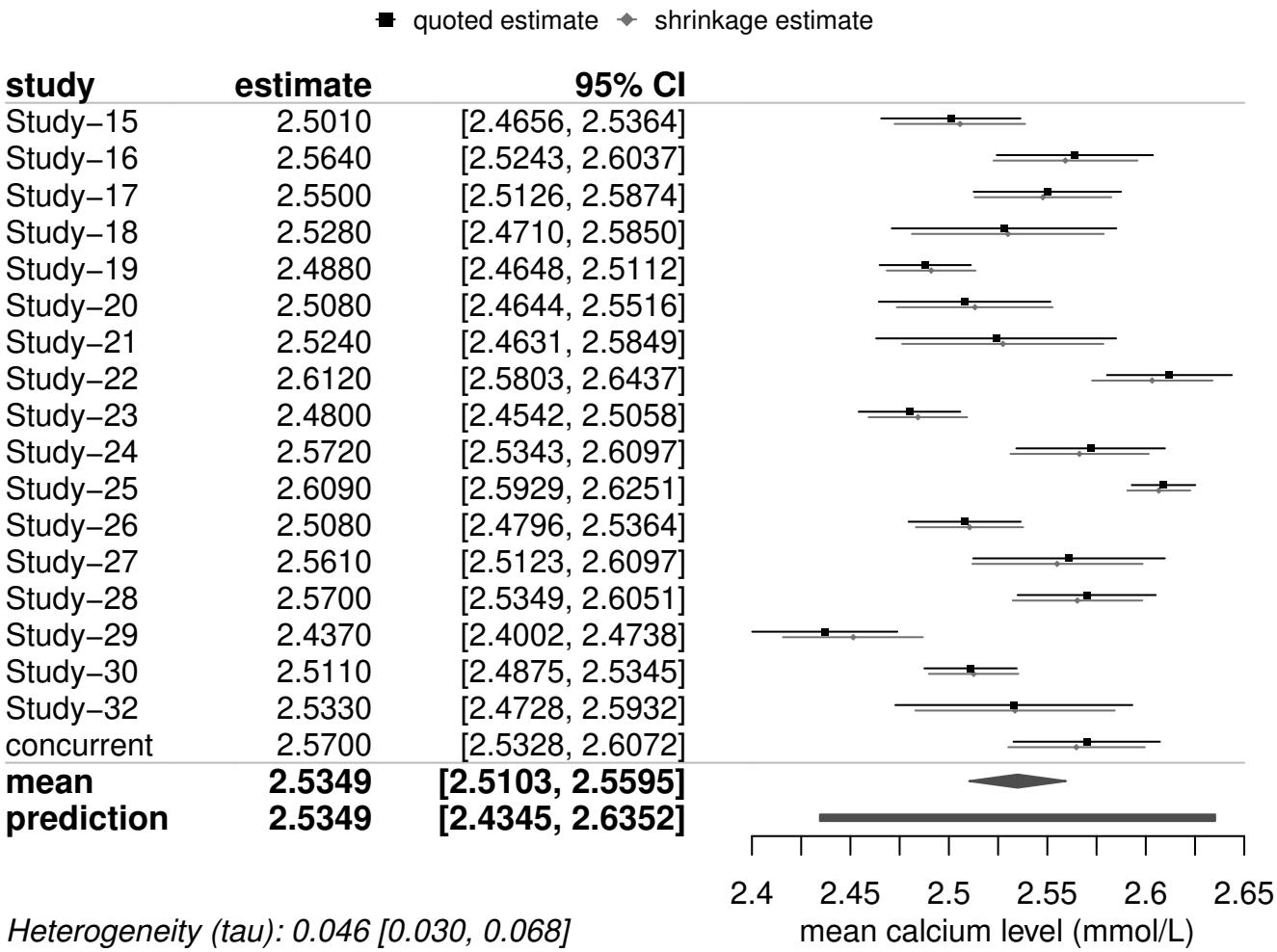
[9] A. Gurjanov, C. Vieira-Vieira, J. Vienenkoetter, L.A.I. Vaas, T. Steger-Hartmann. Replacing concurrent controls with virtual control groups in rat toxicity studies. *Regulatory Toxicology and Pharmacology* **148**:105592, 2024.

[10] 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

Historical control data

Jointly incorporating historical and concurrent data: MAC approach

- MAC: perform meta analysis of 17 historical + 1 concurrent estimate, check “concurrent” shrinkage estimate

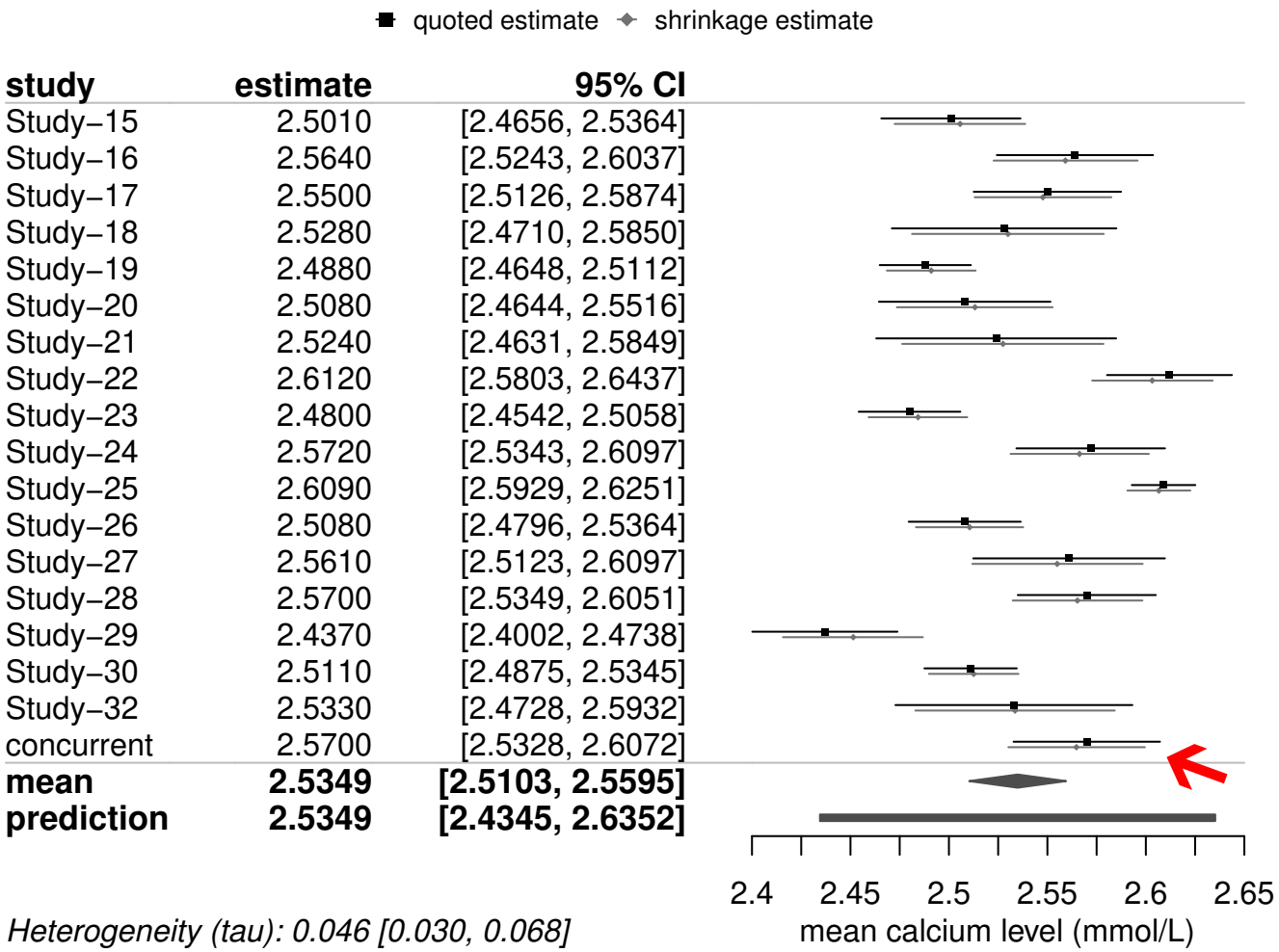


Heterogeneity (tau): 0.046 [0.030, 0.068]

Historical control data

Jointly incorporating historical and concurrent data: MAC approach

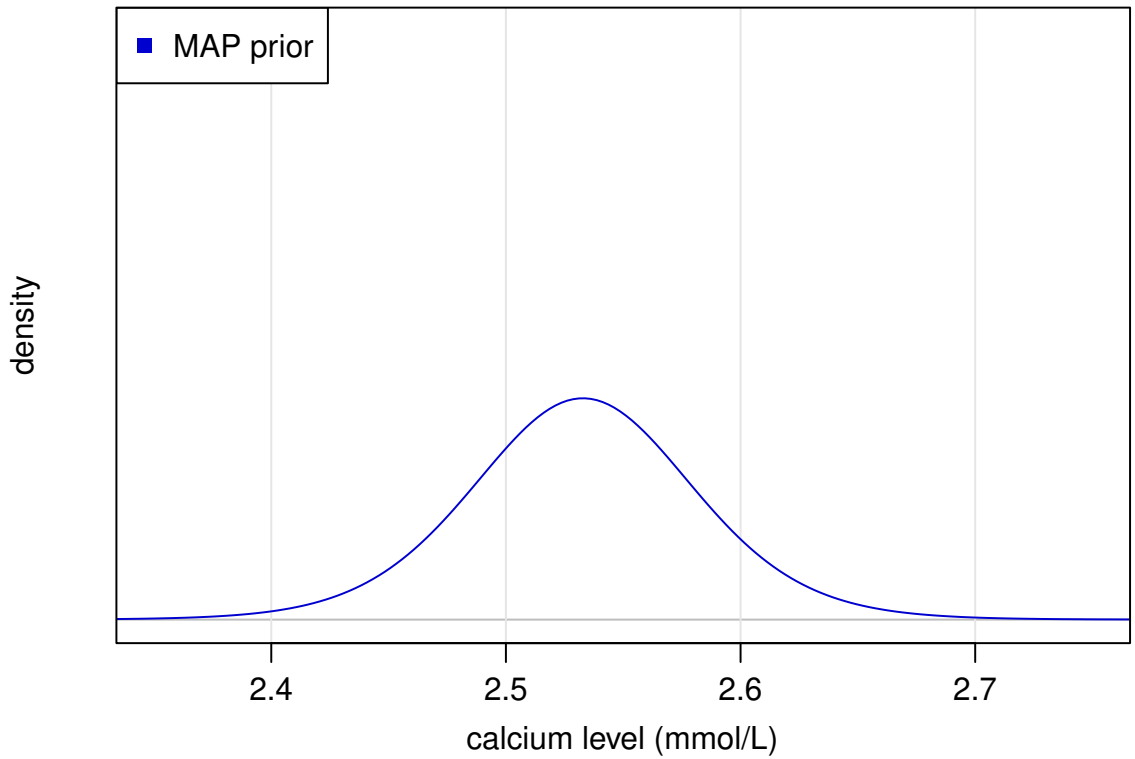
- MAC: perform meta analysis of 17 historical + 1 concurrent estimate, check “concurrent” shrinkage estimate



Heterogeneity (tau): 0.046 [0.030, 0.068]

Historical control data

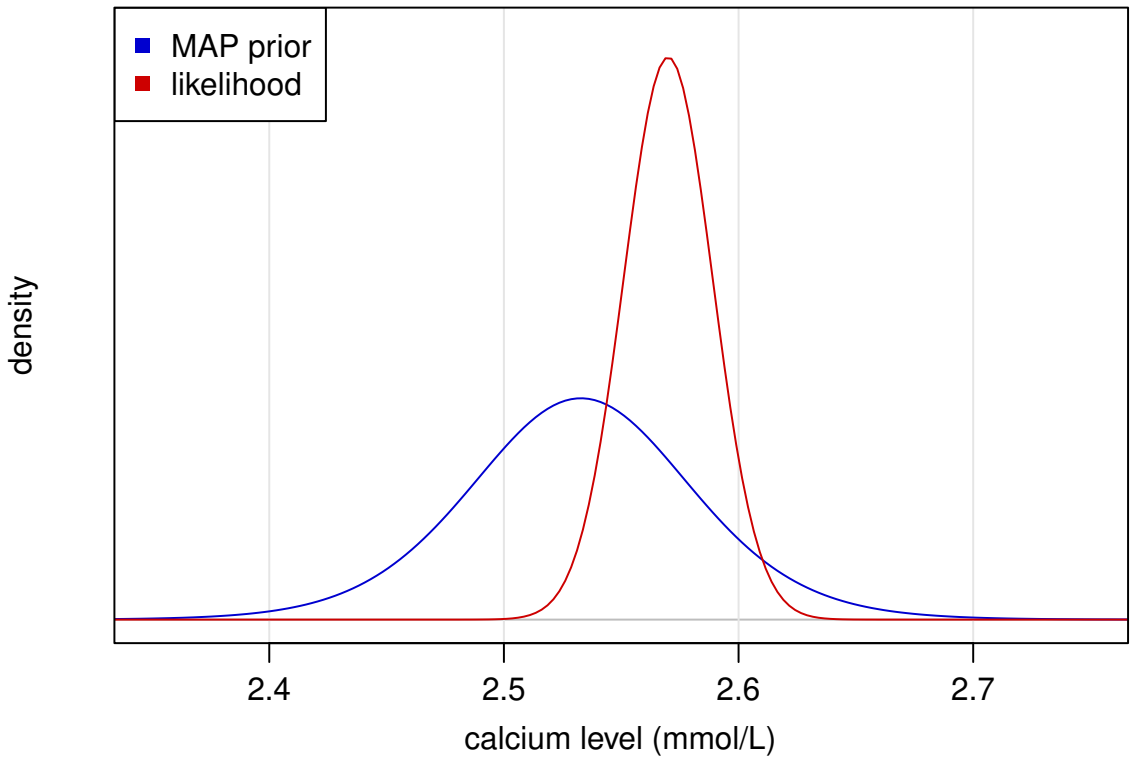
Jointly incorporating historical and concurrent data: MAP approach



- MAP prior from meta-analysis of 17 historical control groups

Historical control data

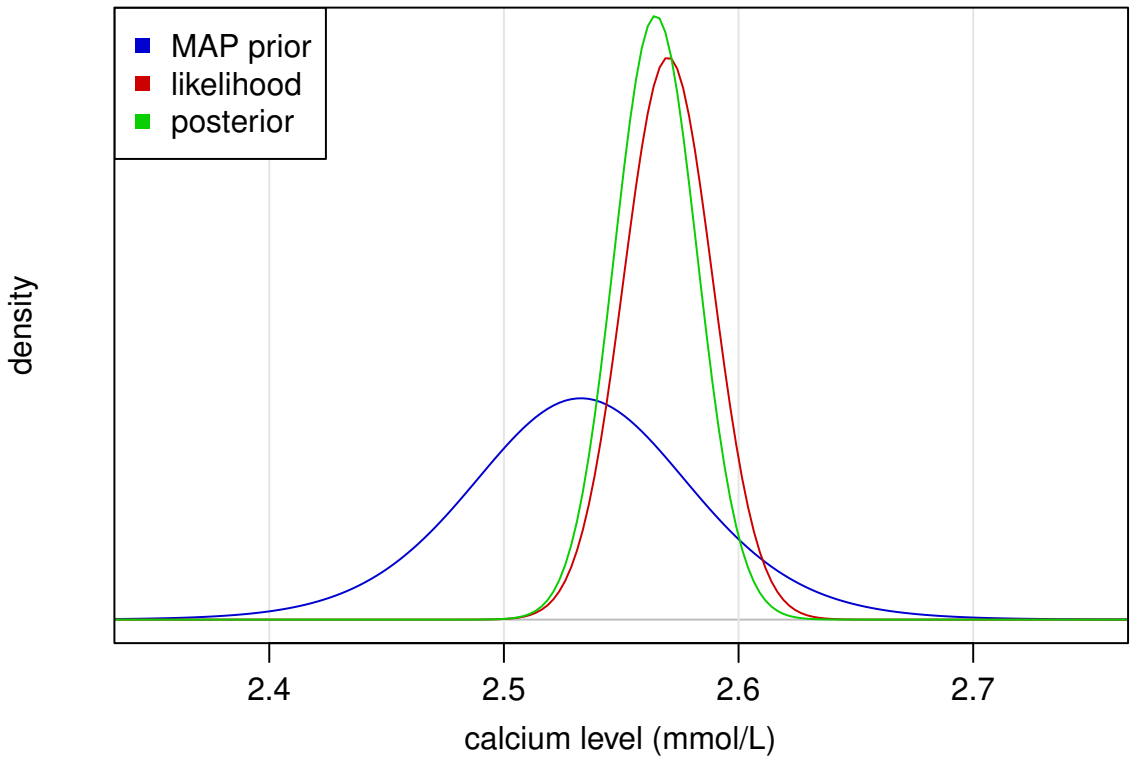
Jointly incorporating historical and concurrent data: MAP approach



- MAP prior from meta-analysis of 17 historical control groups
- “plain” likelihood from current (18th) control group (2.570 ± 0.0190)

Historical control data

Jointly incorporating historical and concurrent data: MAP approach



- MAP prior from meta-analysis of 17 historical control groups
- “plain” likelihood from current (18th) control group (2.570 ± 0.0190)
- shrinkage estimate for current (18th) control group (2.565 ± 0.0177)

Historical control data

Jointly incorporating historical and concurrent data: the shrinkage estimate

- moderate information gain here, since heterogeneity is relatively large ($\tau \approx \sqrt{n_i}\sigma_i \approx 0.05$: similar to within-study variability).
  $\rightarrow$ this leads to little borrowing / pooling.
- precision gain (interval width / standard errors): $\frac{0.0177}{0.0190} = 0.93$ corresponds to a $\approx 15\%$ increase in sample size [11]
- MAP prior's *effective sample size (ESS)* may formally be determined; [12] here also: ESS ≈ 1.4 (animals)

[11] C. Röver, T. Friede. Dynamically borrowing strength from another study through shrinkage estimation. *Statistical Methods in Medical Research*, **29**(1):293–308, 2020.

[12] B. Neuenschwander, S. Weber, H. Schmidli, A. O'Hagan. Predictively consistent prior effective sample sizes. *Biometrics*, **76**(2):578–587, 2020

Historical control data

Jointly incorporating historical and concurrent data: the effect estimates

- from here: may infer differences from *dose 0* to treatment group(s)

	"plain" control			MAP/MAC control		
comparison	Δ	s.e.	CI	Δ	s.e.	CI
dose 1	0.000	0.025	$[-0.048, 0.048]$	0.005	0.024	$[-0.041, 0.052]$
dose 2	0.070	0.023	[0.025, 0.115]	0.075	0.022	[0.033, 0.118]
dose 3	0.120	0.023	[0.075, 0.165]	0.125	0.022	[0.083, 0.168]

- as expected: slightly smaller standard errors also for contrasts

Historical control data

Jointly incorporating historical and concurrent data: the effect estimates

- *WAIT* — didn't we just **break randomization** ?

[13] 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

Historical control data

Jointly incorporating historical and concurrent data: the effect estimates

- *WAIT* — didn't we just **break randomization** ?
- yes

[13] 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

Historical control data

Jointly incorporating historical and concurrent data: the effect estimates

- *WAIT* — didn't we just **break randomization** ?
- yes, . . . but in a transparent / controlled way

[13] 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

Historical control data

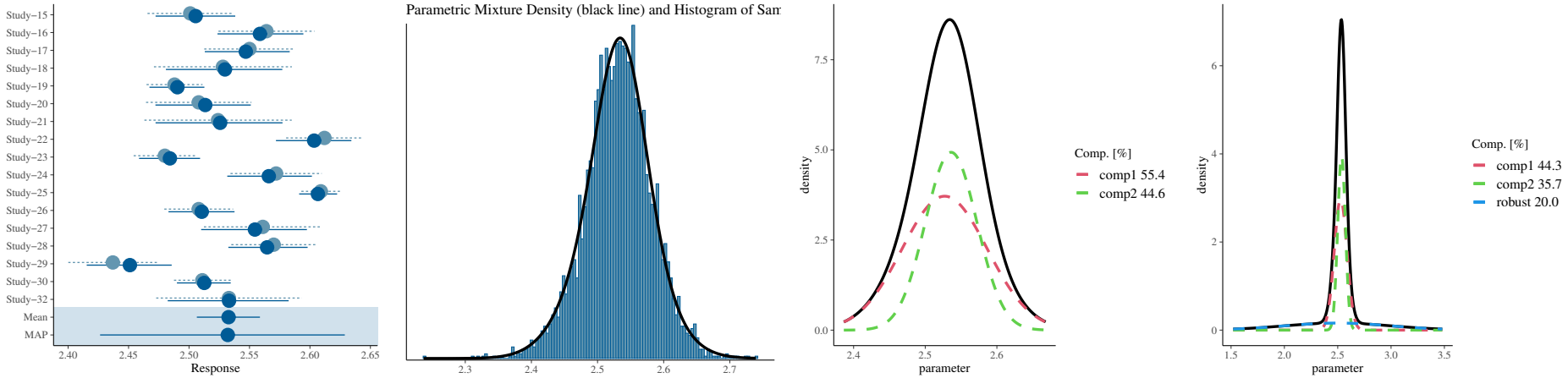
Jointly incorporating historical and concurrent data: the effect estimates

- *WAIT* — didn't we just **break randomization** ?
- yes, . . . but in a transparent / controlled way
- historical information expressed in (MAP) prior
- "robustification" possible [13]
  - deliberate specification of a "more conservative" prior
  - explicit anticipation of possible prior/data conflict
- resulting operating characteristics may be evaluated

[13] 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

Historical control data

Practical implementation: the `RBesT` package



- MAP priors from historical control data implemented in the **RBesT** R package [14, 15]
- includes flexible modeling via MCMC implementation (normal, binomial, Poisson, . . .)
- includes ESS computations, trial design evaluation,. . .

[14] `http://cran.r-project.org/package=RBesT`

[15] S. Weber, Y. Yi, J.W. Seaman, T. Kakizume, H. Schmidli. Applying meta-analytic-predictive priors with the R Bayesian evidence synthesis tools. *Journal of Statistical Software*, **100**(19):1–32, 2021

Discussion

- consistency checking
  - check of mean or standard deviation may be accommodated in meta-analysis
  - may want to extend to several parameters
    (e.g., mean and s.d. simultaneously — multivariate meta-analysis? cluster analysis?)
- recycling historical information
  - small gain in present example
    (due to large heterogeneity, $\tau \approx \sqrt{n_i}\sigma_i \approx 0.05$)
  - general tradeoff: number of studies vs. heterogeneity
    (might be better to consider fewer, more homogeneous studies;
    or: may need to look out for / focus on endpoints with small heterogeneity relative to between-individual variance)
  - "robustification" option: may anticipate possibility of (MAP-) prior/data conflict by specifying a heavier-tailed prior
  - for now, variances (standard errors) were assumed "fixed" / "known". Consideration of estimation uncertainty (and pooling?) also possible (may be sensible for relatively small sample sizes!)