Prior Criticism in Bayesian Meta-Analysis

Leonhard Held

University of Zurich[UZH]

GMDS 2023

September 19, 2023

Estimation vs. Criticism

J. R. Statist. Soc. A, (1980),
143, *Part* 4, *pp.* 383–430

Sampling and Bayes' Inference in Scientific Modelling and Robustness

By George E. P. Box

University of Wisconsin–Madison

[Read before the Royal Statistical Society at a meeting organized by the South Wales Local Group on Thursday, May 15th, 1980, the President Sir Claus Moser in the Chair]

Box (1980) distinguishes

– Estimation based on posterior distribution $f(\theta \mid \text{data, assumptions})$

Estimation vs. Criticism

J. R. Statist. Soc. A, (1980),
143, *Part* 4, *pp.* 383–430

Sampling and Bayes' Inference in Scientific Modelling and Robustness

By George E. P. Box

University of Wisconsin–Madison

[Read before the Royal Statistical Society at a meeting organized by the South Wales Local Group on Thursday, May 15th, 1980, the President Sir Claus Moser in the Chair]

Box (1980) distinguishes

- Estimation based on posterior distribution $f(\theta \mid \text{data, assumptions})$
- Criticism based on prior-predictive distribution $f(\text{data} \mid \text{assumptions})$

Estimation vs. Criticism

J. R. Statist. Soc. A, (1980),
143, *Part* 4, *pp.* 383–430

Sampling and Bayes' Inference in Scientific Modelling and Robustness

By George E. P. Box

University of Wisconsin–Madison

[Read before the Royal Statistical Society at a meeting organized by the South Wales Local Group on Thursday, May 15th, 1980, the President Sir Claus Moser in the Chair]

Box (1980) distinguishes

- Estimation based on posterior distribution $f(\theta \mid \text{data, assumptions})$
- Criticism based on prior-predictive distribution $f(\text{data} \mid \text{assumptions})$
- The assumptions include both model and prior assumptions.

Box's Tail Probability

Box (1980) suggests to compute

$$p_{\text{Box}} = \Pr\{f(\text{data} \mid \text{assumptions}) < f(\text{observed data} \mid \text{assumptions})\}$$

Box's Tail Probability

Box (1980) suggests to compute

$$p_{\text{Box}} = \Pr\left\{ f(\text{data} \mid \text{assumptions}) < f(\text{observed data} \mid \text{assumptions}) \right\}$$

– Quantifies compatability of model/prior with observed data.

Box's Tail Probability

Box (1980) suggests to compute

$$p_{\text{Box}} = \Pr\{f(\text{data} \mid \text{assumptions}) < f(\text{observed data} \mid \text{assumptions})\}$$

- Quantifies compatability of model/prior with observed data.
- Small values of p_{Box} indicate that model or prior are discredited by the observed data.

The Q-Test

- Model: Normal-normal model
- Prior: $\tau^2 = 0$
- Data: Differences between effect estimates $\hat{\theta}_i$ with standard errors σ_i

$\rightarrow$ f(observed data | assumptions) is the Q-statistic

$$Q = \frac{\sum_{i<j} w_i w_j (\hat{\theta}_i - \hat{\theta}_j)^2}{\sum_{i=1}^{k} w_i}$$

where $w_i = 1/\sigma_i^2$ are "fixed-effect" weights.

$\rightarrow$ p_{Box} simplifies to the p-value obtained from the null distribution $Q \sim \chi^2_{k-1}$

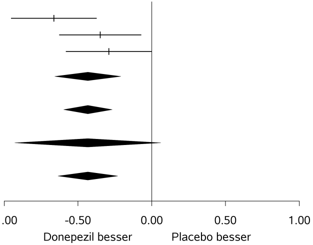
The Generalized Q-Test

- Model: Normal-normal model
- Null hypothesis: $\tau^2 = \tau_0^2$
- The generalized Q-statistic $Q(\tau_0^2)$ now uses "random-effects" weights $w_i = 1/(\sigma_i^2 + \tau_0^2)$.
- We still have $Q(\tau_0^2) \sim \chi^2_{k-1}$ if $\tau^2 = \tau_0^2$
- Solving $Q(\tau^2) = k - 1$ for τ^2 gives the Paule-Mandel estimate.

Example: Donepezil vs. Placebo

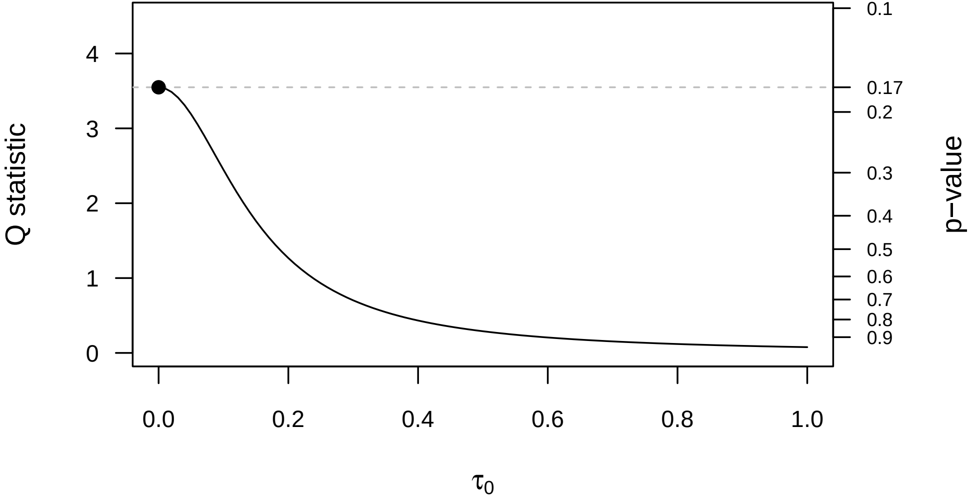
Donepezil vs. Placebo
DAD, CMCS, PDS

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]



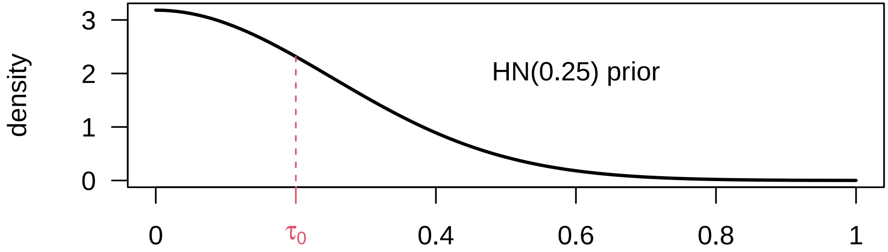
Heterogenität: Q=3.70, df=2, p=0.157, I^2=45.9%
Gesamteffekt (REM - DerSimonian-Laird): Z-Score=-3.78, p<0.001, Tau=0.135

Example: Donepezil vs. Placebo



Checking the Heterogeneity Prior

– Suppose we now have a half-normal prior $f(\tau)$ with $\mathrm{E}(\tau) = \tau_0$



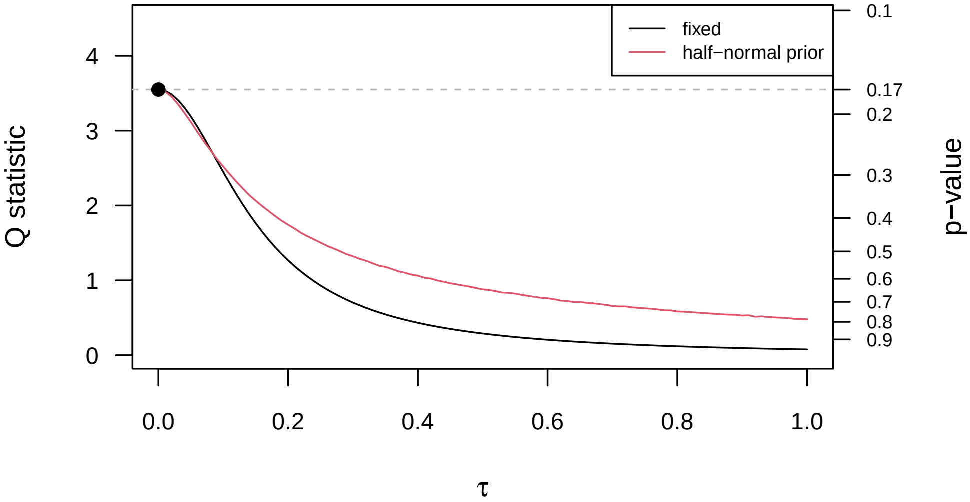
$\rightarrow$ p_{Box} is now based on

$$\widetilde{Q} = \int Q(\tau^2) f(\tau^2) d\tau^2$$

– The χ^2-distribution still holds because τ^2 is a pivot for $Q(\tau^2)$

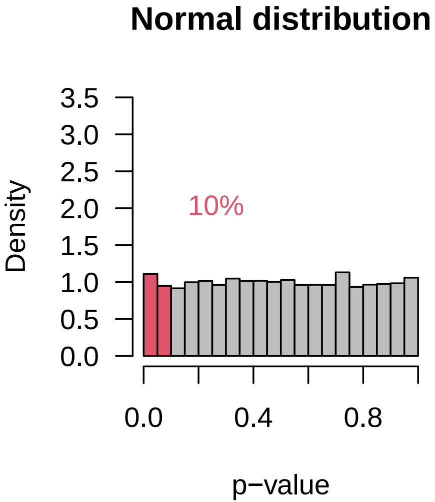
$\rightarrow$ p_{Box} can be easily calculated through Monte Carlo simulation.

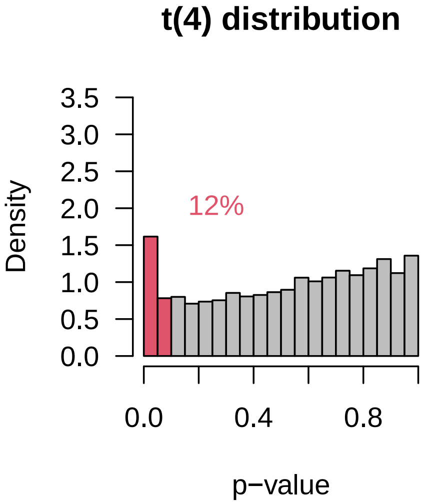
Example: Donepezil vs. Placebo



Type-I Error Assessment

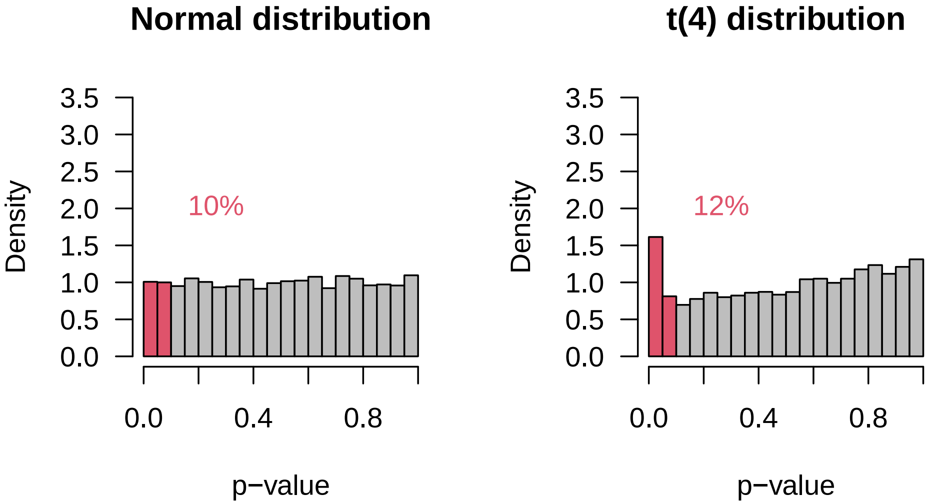
$k = 3$, $\tau_0 \sim f(\tau)$





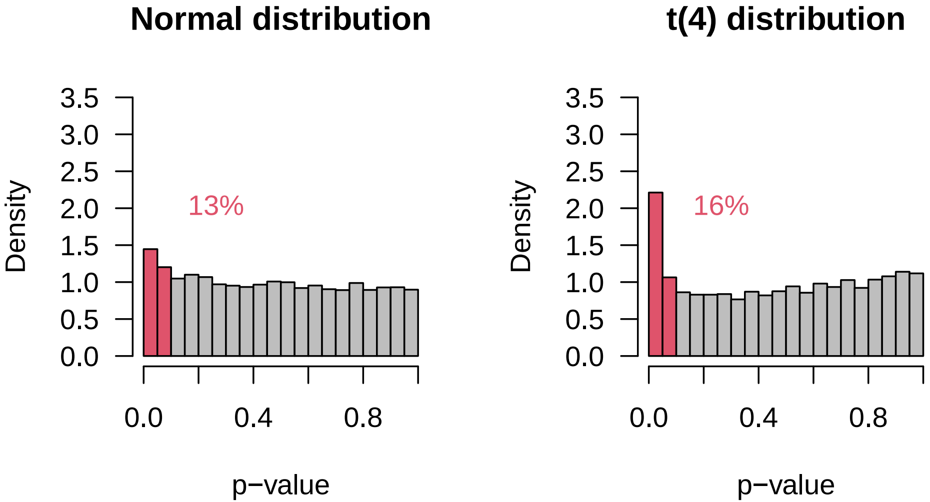
Power Assessment

$k = 3$, **fixed** $\tau_0 = 0.2 = \mathrm{E}(\tau)$



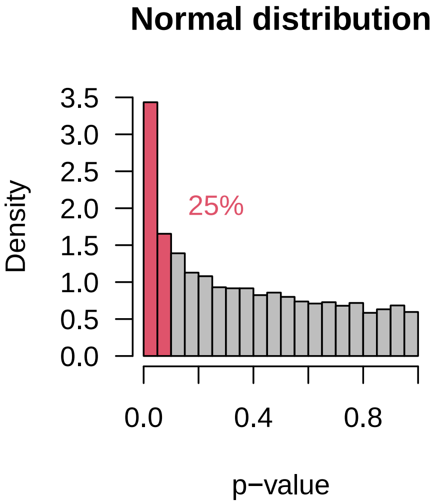
Power Assessment

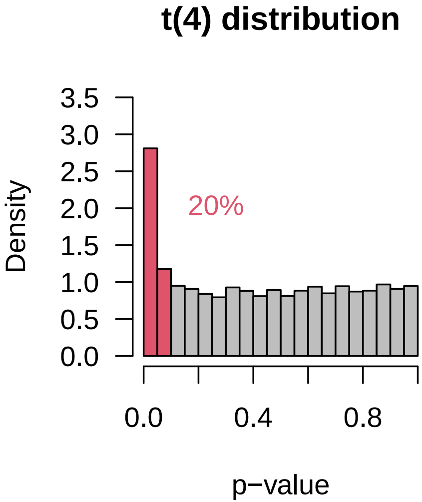
$k = 3$, **fixed** $\tau_0 = 0.4 = 2\,\mathrm{E}(\tau)$



Power Assessment

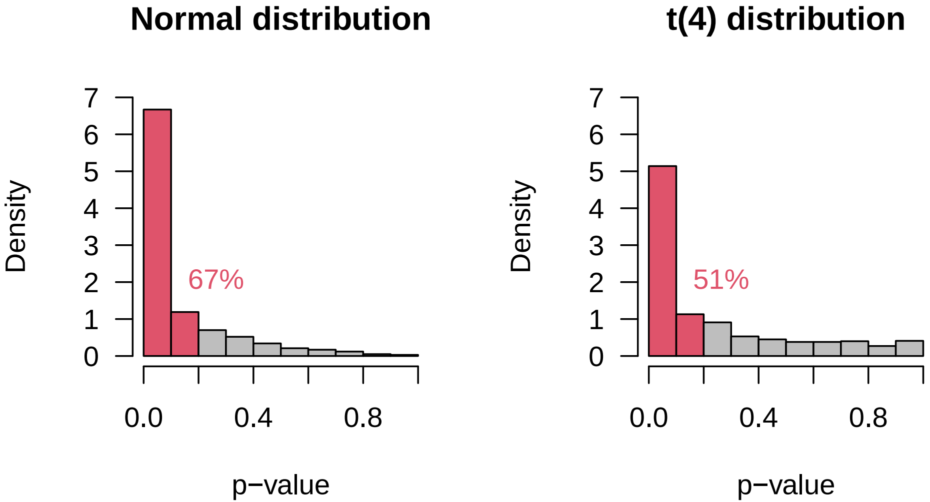
$k = 3$, fixed $\tau_0 = 0.8 = 4\,\mathrm{E}(\tau)$





Aggregated Power Assessment for 10 Studies

$k = 3$, **fixed** $\tau_0 = 0.8 = 4\,\mathrm{E}(\tau)$



Summary

- Q-Test can be generalized to check heterogeneity prior and other model assumptions.
- Has low power for meta-analyses with very small k.
- Power can be increased by summation of Q-statistic across meta-analyses.

Summary and Discussion

– Q-Test can be generalized to check heterogeneity prior and other model assumptions.
– Has low power for meta-analyses with very small k.
– Power can be increased by summation of Q-statistic across meta-analyses.

– Heterogeneity prior is not the only critical assumption
– Normality assumption may also be wrong due to publication bias etc.

References

Box, G. E. P. (1980). Sampling and Bayes' inference in scientific modelling and robustness (with discussion). Journal of the Royal Statistical Society, Series A, 143:383–430.