



Contribution ID: 102

Type: not specified

A study of prior sensitivity in random effects models for decisions in interlaboratory comparisons: credible intervals versus Bayesian evidence

Random effects models are widely used in interlaboratory comparisons to estimate between-laboratory variability τ and assess degrees of equivalence of laboratories [1]. In this context, decisions are often based on 95% credible intervals [2], while Bayesian hypothesis testing provides an alternative probabilistic framework based on Bayes factors for assessing laboratory effects [3].

This work compares these two decision paradigms, focusing on the sensitivity of conclusions to the prior specification of the between-laboratory standard deviation. We investigate how credible interval based decisions and Bayes factor based hypothesis testing behave under different weakly informative priors on τ , including Half-Cauchy prior, inverse Chi-square and data-informed scale choices.

We further assess the robustness of credible interval and Bayes factor conclusions using posterior predictive simulations [4]. This predictive additional step to standard analysis allows us to examine whether decisions remain stable under replicated data scenarios, and to identify cases where apparent evidence may not be supported by predictive behavior.

Results from simulation studies and real interlaboratory data from CCQM-K53 [5] show that credible interval decisions are relatively stable with respect to prior choices, whereas Bayes factors exhibit strong sensitivity to the prior on τ . Posterior predictive checks provide additional insight to the robustness of the decisions. These findings emphasize the importance of combining inferential and predictive perspectives when making decisions in metrological applications such as laboratory equivalence assessments.

[1] Toman B, Possolo, A, Laboratory effects models for interlaboratory comparisons, *Accred Qual Assur* (2009) 14:553–563

[2] CCQM-KCWG/01, Guidelines for the CCQM KCWG on the review of CCQM CMCs for inclusion in the key comparison database, 2020.

[3] Wübbeler G, Bodnar O, Elster C. Bayesian hypothesis testing for key comparisons. *Metrologia*. 2016;1131-8.

[4] Raghu N Kacker RK Alistair Forbes, Sommer KD. Bayesian posterior predictive p-value of statistical consistency in interlaboratory evaluations. *Metrologia*. 2008:512-23.

[5] Lee J, Lee JB, Moon DM, Kim JS, van der Veen AMH, Besley L, et al. Final report on international key comparison CCQM-K53: Oxygen in nitrogen. *Metrologia*. 2010;47

Special/ Invited session

Classification

Both methodology and application

Keywords

Bayesian hypothesis testing, Bayes factors, posterior predictive distribution, degrees of equivalence, prior knowledge

Primary authors: Dr WÜBBELER, Gerd (PTB); DEMEYER, Séverine (Laboratoire National de Métrologie et d'Essais)

Presenter: DEMEYER, Séverine (Laboratoire National de Métrologie et d'Essais)

Track Classification: Measurement Uncertainty and Metrology