Journal of Mathematical Psychology 88 (2019) 27–31

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

Commentary

Accounting for estimation uncertainty and shrinkage in Bayesian within-subject intervals: A comment on Nathoo, Kilshaw, and Masson (2018)[☆]

Daniel W. Heck [*,1]

University of Mannheim, Germany

HIGHLIGHTS

- Within-subject intervals facilitate visualizations in repeated-measures designs.
- Nathoo, Kilshaw, & Masson (2018) developed a Bayesian within-subject interval.
- The interval neglects estimation uncertainty and shrinkage of random effects.
- As a remedy, a fully Bayesian, two-step approach is proposed.

ARTICLE INFO

Article history:
Received 5 September 2018
Received in revised form 6 November 2018

Keywords:
Repeated-measures designs
Linear mixed model
Credible interval
Highest-density interval
Within-subject Bayesian inference

ABSTRACT

To facilitate the interpretation of systematic mean differences in within-subject designs, Nathoo, Kilshaw, and Masson (2018) proposed a Bayesian within-subject highest-density interval (HDI). However, their approach rests on independent maximum-likelihood estimates for the random effects which do not take estimation uncertainty and shrinkage into account. I propose an extension of Nathoo et al.'s method using a fully Bayesian, two-step approach. First, posterior samples are drawn for the linear mixed model. Second, the within-subject HDI is computed repeatedly based on the posterior samples, thereby accounting for estimation uncertainty and shrinkage. After marginalizing over the posterior distribution, the two-step approach results in a Bayesian within-subject HDI with a width similar to that of the classical within-subject confidence interval proposed by Loftus and Masson (1994).

1. Introduction

To illustrate the relative size of measurement errors versus systematic effects in repeated-measures designs, it has become standard practice to plot mean patterns with error bars based on within-subject confidence intervals (Loftus & Masson, 1994). Essentially, this approach removes the between-subject variance, which is usually not of interest in within-subject designs. Recently, Nathoo, Kilshaw, and Masson (2018) extended this work and proposed a Bayesian approach to within-subject interval estimation. Assuming Jeffreys noninformative prior, Nathoo et al. (2018) analytically derived the posterior distribution of the mean parameter μ_j conditional on the random-effect estimates. The resulting Bayesian highest-density interval (HDI) offers the advantage of having a more intuitive interpretation than classical confidence intervals (Hoekstra, Morey, Rouder, & Wagenmakers, 2014). Furthermore, the within-subject HDI by Nathoo et al. (2018) is always more narrow than the frequentist within-subject confidence interval by Loftus and Masson (1994).

However, the Bayesian HDI proposed by Nathoo et al. (2018) relies on an approach termed "conditional Bayesian inference" (p. 2). In contrast to a fully Bayesian analysis, the HDI is derived conditional on independent maximum-likelihood estimates for the random effects. By doing so, the random-effect estimates are treated as fixed values, and in turn, "the uncertainty in estimating the subject-specific random effects is not propagated to the width of the HDI" (Nathoo et al., 2018 p. 3). Hence, one of the main advantages of a fully Bayesian analysis is lost, namely, that statistical inference accounts for estimation error of all parameters simultaneously (Gelman et al., 2013). While acknowledging this

---

[☆] Data and R code of all analyses are available at the Open Science Framework: https://osf.io/mrud9/.

[*] Correspondence to: Department of Psychology, School of Social Sciences, University of Mannheim, B6 30-32 Room 313, D-68131 Mannheim, Germany.

E-mail address: heck@uni-mannheim.de.

[1] I thank Hansjörg Plieninger for comments on an earlier version of this manuscript. The author was supported by the research training group Statistical Modeling in Psychology (GRK 2277), funded by the German Research Foundation (DFG).

https://doi.org/10.1016/j.jmp.2018.11.002
0022-2496/

possible drawback, Nathoo et al. (2018) justified their approach as follows: “It is the conditioning on estimated random effects that removes the uncertainty that is not of interest, and this conditioning precludes fully Bayesian inference.”

Besides not accounting for estimation uncertainty of the random-effect estimates, Nathoo et al. (2018) relied on descriptive person-means to estimate the random effects. However, these independent maximum-likelihood estimates do not provide shrinkage, meaning that the random-effect estimates are not pulled towards the overall group mean (Efron & Morris, 1977; Pinheiro & Bates, 2000). Nathoo et al. (2018) mentioned this drawback in the discussion and justified the reliance on the simple plug-in estimator by the “desire for simple closed forms that can be easily used by practitioners” (p. 7). While closed-form solutions are in general desirable and might increase the applicability of a method, more complex computational approaches can nowadays easily be used on any desktop computer, especially if the relevant code is shared and openly available. Leaving aside computational considerations, it is well known that shrinkage offers many benefits in linear mixed-effects models. Most importantly, it increases the average prediction accuracy by decreasing the variance of the random-effect estimates compared to independent maximum-likelihood estimates (James & Stein, 1961; Stein, 1956). Since the conditional Bayesian inference approach by Nathoo et al. (2018) relies on independent random-effect estimates without shrinkage, the between-subject variance (i.e., the random-effect variance) will generally be overestimated. Given that the proposed method aims at removing exactly this between-subject variance, the correction will generally overshoot and result in a too narrow within-subject HDI. As a consequence, the proposed Bayesian within-subject interval might lead to overconfidence in substantive conclusions.

As a remedy, I propose a fully Bayesian two-step approach that (a) takes into account the estimation uncertainty of random effects and (b) relies on shrinkage for the random-effect estimates. At its core, the proposed method can be seen as an application of Nathoo et al. (2018)’s closed-form solutions for the within-subject HDI within a fully Bayesian framework. However, instead of conditioning on independent maximum-likelihood estimates for the random effects, the within-subject HDI is computed repeatedly using posterior samples from the underlying linear mixed model. As shown in the next section, this two-step approach takes estimation uncertainty and shrinkage of the random effects into account.

2. Using posterior samples to propagate estimation uncertainty

Before outlining the proposed method, I provide a short description of the linear mixed model and the Bayesian within-subject HDI (for details, see Nathoo et al., 2018). In a within-subject design, responses Y_{ij} of a person $i = 1, \ldots, N$ in condition $j = 1, \ldots, C$ are modeled as:

$$Y_{ij} = \mu_j + b_i + \epsilon_{ij}, \quad \epsilon_{ij} \overset{iid}{\sim} N(0, \sigma_\epsilon^2), \tag{1}$$

where μ_j is the true mean in the jth condition, b_i the random effect of the ith subject, and σ_ϵ^2 the error variance of the residuals ϵ_{ij}. Assuming Jeffreys noninformative prior, Nathoo et al. (2018) derived the posterior distribution of the mean conditional on the random-effect estimates:

$$\left(\mu_j \mid \boldsymbol{Y}, \hat{b}_1, \ldots, \hat{b}_N\right) \sim t_{C(N-1)}\left(\hat{\mu}_j, \frac{SS_{S\times C}}{N(N-1)C}\right), \tag{2}$$

where $\hat{\mu}_j = \frac{1}{N}\sum_i Y_{ij}$ are the estimated condition means. Moreover, the independent maximum-likelihood estimates for the random effects $\hat{b}_i = M_{i\cdot} - M$ are based on the mean $M_{i\cdot} = \frac{1}{C}\sum_j Y_{ij}$ for the ith person and the overall mean $M = \frac{1}{NC}\sum_i\sum_j Y_{ij}$. Finally, the interaction sum-of-squares $SS_{S\times C}$ is computed as:

$$\begin{aligned} SS_{S\times C} &= \sum_j \sum_i \left(Y_{ij} - (M_{i\cdot} - M) - M_{\cdot j}\right)^2 \\ &= \sum_j \sum_i (Y_{ij} - \hat{b}_i - \hat{\mu}_j)^2. \end{aligned} \tag{3}$$

Based on the modified conditional posterior distribution in Eq. (2), Nathoo et al. (2018) derived the critical boundaries for the within-subject HDI:

$$\hat{\mu}_j \pm \sqrt{\frac{SS_{S\times C}}{N(N-1)C}}\, t_{C(N-1),\alpha/2}. \tag{4}$$

The proposed extension of this Bayesian within-subject interval relies on a method by Ly et al. (2017) who used a two-step approach to correct the credibility interval of a correlation coefficient when some variables are contaminated by estimation error (for a user-friendly implementation, see Heck, Arnold, & Arnold, 2018). To apply the two-step approach to the Bayesian within-subject HDI, the linear mixed model in Eq. (1) is first fitted in a fully Bayesian framework. For this purpose, I follow Nathoo et al. (2018) in assuming Jeffreys noninformative prior for the condition means and the residual variance:

$$p(\mu_1, \ldots, \mu_C, \sigma_\epsilon^2) \propto \frac{1}{\sigma_\epsilon^2} \tag{5}$$

Moreover, additional assumptions are required to actually fit the model. First, following common practice in mixed-effects modeling (Gelman et al., 2013; Pinheiro & Bates, 2000), it is assumed that the random effects are normally distributed with variance τ^2:

$$b_i \overset{iid}{\sim} N(0, \tau^2). \tag{6}$$

Second, a prior distribution for the random-effect variance τ^2 is required. Here, we follow the recommendation of Rouder, Morey, Speckman, and Province (2012, p. 363) and assume an inverse-$\chi^2(1)$ distribution as a prior for the variance of the *standardized* random-effect terms b_i/σ_ϵ.[2] This has the advantage that the resulting within-subject HDI is based on the same prior distribution for τ^2 as that commonly used to compute default Bayes factors for ANOVA (Rouder et al., 2012).

Based on user-friendly software such as JAGS (Plummer, 2003) or Stan (Stan Development Team, 2018), Markov chain Monte Carlo (MCMC) methods allow us to draw $r = 1, \ldots, R$ random samples of the parameters from the posterior distribution. The resulting distribution of posterior samples $(\boldsymbol{\mu}^{(r)}, \boldsymbol{b}^{(r)}, \sigma_\epsilon^{(r)}, \tau^{(r)})$ contains information about the estimation uncertainty of all parameters (Gelman et al., 2013). Moreover, in a fully Bayesian analysis, the random-effect posterior samples $\boldsymbol{b}^{(r)}$ will show shrinkage, meaning that they are usually closer to zero than independent maximum-likelihood estimates (Ly et al., 2017). The amount of shrinkage depends on the specific data set, namely, on the number of observations within and between participants, as well as on the prior distribution for τ^2 (Marsman, Maris, Bechger, & Glas, 2016; Rouder, Morey, Verhagen, Swagman, & Wagenmakers, 2017).

Next, we apply the closed-form Equations (2), (3), and (4) by Nathoo et al. (2018) using the posterior samples $b_i^{(r)}$ and $\mu_j^{(r)}$ instead of the independent point estimates $\hat{b}_i$ and $\hat{\mu}_j$. Since the procedure is applied repeatedly for each of the R posterior samples,

---

[2] By assuming a prior for the variance of the *standardized* random effects, the prior for the variance τ^2 of the *unstandardized* random effects is scaled by the residual variance σ_ϵ^2 as follows: $(\tau^2/\sigma_\epsilon^2 \mid \sigma_\epsilon^2) \sim \text{Inverse-}\chi^2(1)$. Rouder et al. (2012) provide a detailed discussion of this prior and also propose an alternative, weakly informative prior distribution for the condition means $\boldsymbol{\mu}$.

we obtain not one, but R Bayesian within-subject HDIs. Moreover, instead of computing the residuals in Eq. (3) for all random-effect terms $b_i^{(r)}$, it is easier and conceptually more natural to estimate the interaction sum-of-squares $SS_{S\times C}$ in the within-subject HDI in Eq. (4) directly. For this purpose, it is important to recognize that the interaction sum-of-squares provides an estimate for the residual-error variance σ_ϵ^2 of the linear mixed model:

$$\hat{\sigma}_\epsilon^2 = \frac{SS_{S\times C}}{(N-1)C} \qquad (7)$$

Hence, instead of computing residuals in Eq. (3) using the samples $b_i^{(r)}$, it is more straightforward to use the posterior samples $\sigma_\epsilon^{(r)}$ directly to compute within-subject HDIs:

$$\mu_j^{(r)} \pm \frac{\sigma_\epsilon^{(r)}}{\sqrt{N}}\, t_{C(N-1),\alpha/2}. \qquad (8)$$

By summarizing the resulting distribution of HDIs (e.g., by taking the mean), we essentially marginalize over the posterior distribution of the random effects $b_i^{(r)}$ and the residual variance $\sigma_\epsilon^{(r)}$, thus taking estimation uncertainty and shrinkage into account (Ly et al., 2017). A user-friendly implementation of this two-step approach in R is available at the Open Science Framework (https://osf.io/mrud9/). Note that the approach rests on the assumption that the residual variance σ_ϵ^2 is constant across experimental conditions (cf. Eq. (1)). To accommodate violations of this sphericity assumption, the proposed two-step method can be generalized by applying Theorem 3 of Nathoo et al. (2018) based on MCMC samples from a heteroscedastic mixed-effects model with different residual variances σ_j for each experimental condition.

To illustrate the two-step approach, I applied the proposed method to the example data set by Loftus and Masson (1994) with $N = 10$ participants measured across $C = 3$ conditions, which was also used by Nathoo et al. (2018, their Table 1). Fig. 1 shows that the within-subject HDI by Nathoo et al. (2018) provides a single modified posterior distribution of the mean μ_j conditional on the independent maximum-likelihood estimates of the random effects (dashed black line, cf. Eq. (2)). In contrast, the fully Bayesian, two-step approach results in a complete distribution of such densities conditional on the posterior samples of the random effects (thin red lines). To summarize this distribution of within-subject HDIs, we marginalize over the posterior distribution of densities by computing the mean (solid black line). Similarly as for the probability density functions, we can compute a lower and an upper critical boundary for each of the posterior samples $r = 1, \ldots, R$ using Eq. (8) and then marginalize over the resulting estimates. The bottom of Fig. 1 shows that this two-step approach results in a relatively wide 95% within-subject HDI compared to the 95% within-subject HDI conditional on the maximum-likelihood estimates (half-width: ±0.53 versus ±0.42, respectively). Essentially, the larger width of the proposed within-subject interval reflects the estimation uncertainty and shrinkage of the random effects. Interestingly, the width of the posterior-averaged HDI was very similar to that of the classical within-subject confidence interval (half-width: ±0.52; Loftus & Masson, 1994).

The width of the posterior-averaged within-subject HDI was also very similar when choosing different prior distributions for the random-effect variance τ^2. The width of the within-subject interval was estimated to be ±0.52 when choosing (a) an improper uniform prior on the standard deviation τ of the random effects (which results in a proper posterior distribution for $C \geq 3$; Gelman, 2006) or (b) a half-Cauchy distribution on τ (Gelman, 2006). Overall, this sensitivity analysis shows that different prior distributions for the random-effect variance τ^2 result in almost identical within-subject HDIs, thus corroborating the conclusion that the conditional approach by Nathoo et al. (2018) results in a too narrow interval compared to a fully Bayesian approach.

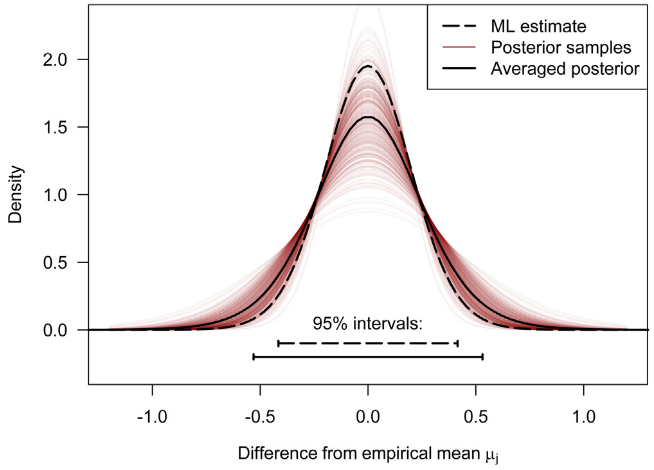

Fig. 1. Posterior distribution of the mean μ_j conditional on different estimates for the random effects. The analysis is based on the hypothetical data set by Loftus and Masson (1994) which includes $N = 10$ participants each providing $C = 3$ responses with condition means $M_{\cdot 1} = 11.0$, $M_{\cdot 2} = 13.0$, and $M_{\cdot 3} = 14.2$ and the residual sum-of-squares $SS_{S\times C} = 11.1$ (cf. Nathoo et al., 2018 their Table 1).

3. Simulation study

The differences between the within-subject confidence interval by Loftus and Masson (1994), the HDI by Nathoo et al. (2018), and the proposed posterior-averaged HDI can be studied more systematically in a Monte Carlo simulation. For this purpose, responses Y_{ij} were generated according to the model equations in Eqs. (1) and (6) for varying sample sizes of $N = 10, 20, \ldots, 100$. As data-generating values, the residual variance was fixed at $\sigma_\epsilon^2 = 1$, while the standard deviation of the random effects varied between $\tau = 0.25$, $\tau = 1$, and $\tau = 4$. Moreover, the number of experimental conditions was manipulated between $C = 2$, $C = 3$, and $C = 4$. For each of 1000 replications, posterior samples were drawn using two MCMC chains with 2000 iterations each to compute the posterior-averaged within-subjects HDI in Eq. (8).

Fig. 2 shows the mean width of the three different within-subject intervals across 1000 replications as a function of the number of participants N. Across rows, the panels differ by the number of within-subject conditions C, and across columns, the panels assume different random-effect variances τ^2. In line with the analytical proof by Nathoo et al. (2018), the Bayesian within-subject HDI was always more narrow than the corresponding classical confidence interval by Loftus and Masson (1994), an effect that diminished with the number of conditions C. Most importantly, the HDI by Nathoo et al. (2018) resulted in overconfidence when compared to the proposed posterior-averaged HDI which takes estimation uncertainty and shrinkage of the random effects into account. Moreover, the within-subject interval of the proposed two-step approach was almost identical to the within-subject confidence interval by Loftus and Masson (1994) when the amount of heterogeneity was substantial (i.e., if $\tau \geq \sigma_\epsilon$). In contrast, the posterior-averaged HDI was below the classical confidence interval but still above the conditional HDI when the random-effect parameter τ was smaller than the residual variance (i.e., when $\tau = \sigma_\epsilon/4$).

Overall, the simulation shows that the narrower width of the within-subject HDI – according to Nathoo et al. (2018), one of the key advantages of the conditional Bayesian HDI – may result in overconfidence compared to a fully Bayesian approach which takes estimation uncertainty and shrinkage of the random effects into account.

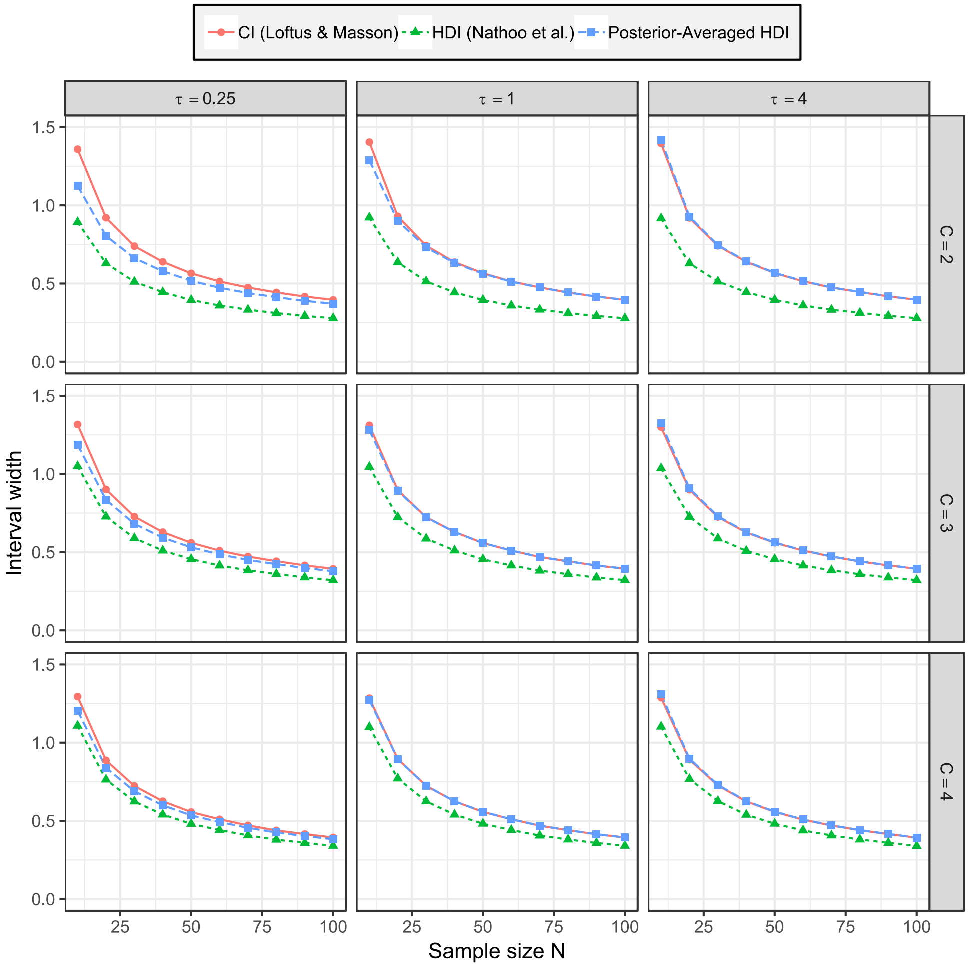

Fig. 2. Mean width of within-subject intervals as computed with three different methods for 1000 replications for each simulation condition. In all cases, the data-generating residual variance was $\sigma_\epsilon^2 = 1$. Note that the lines for the confidence interval (CI; Loftus & Masson, 1994) and the posterior-averaged HDI overlap for $\tau \geq 1$ and $C \geq 3$.

4. Discussion

The Bayesian within-subject interval derived by Nathoo et al. (2018) offers some conceptual advantages such as having a more intuitive interpretation than classical confidence intervals. However, the within-subject HDI is defined conditional on the random effects and thus makes the critical assumption that the random-effect parameters are known and identical to the maximum-likelihood estimates $\hat{b}_i$. Hence, the approach does not take into account estimation uncertainty and shrinkage of the random effects, which is a fundamental advantage of a fully Bayesian inference (Gelman et al., 2013). As a remedy, I proposed a two-step approach in which posterior samples from the linear mixed-effects model are plugged into the closed-form solutions by Nathoo et al. (2018). This has the conceptual advantage of fitting the random-effect parameters in a fully Bayesian framework instead of conditioning on a single set of point estimates for the random effects. By averaging over the posterior distribution of the random effects, the resulting within-subject HDI takes both estimation uncertainty and shrinkage into account (Ly et al., 2017). As expected, the proposed method results in wider within-subject HDIs compared to treating the random effects as being known (Nathoo et al., 2018).

More surprisingly, the width of the posterior-averaged within-subject HDI matched that of the classical within-subject interval by Loftus and Masson (1994) when the random-effect variance was at least as large as the residual variance (i.e., when $\tau^2 \geq \sigma_\epsilon^2$). However, despite the numerical similarity in some scenarios, one should refrain from computing the classical within-subject confidence interval and simply interpreting it as a Bayesian within-subject HDI. First, such an approach is conceptually not valid, and second, one cannot be sure whether the classical confidence interval is similar to the posterior-averaged HDI for a specific data set. Instead, I advocate to fit the mixed-effects model in a fully Bayesian framework and then apply the proposed two-step approach, which can easily be accomplished using the R scripts provided in the supplementary material (https://osf.io/mrud9/). Moreover, future work might provide an analytical solution for the posterior-averaged within-subject HDI that does not require MCMC samples from the posterior distribution.

The main issue of conditioning on the random-effect estimates is related to biased maximum likelihood estimates for the residual variance σ_ϵ^2 which have been proposed for classical within-subject confidence intervals. To simplify computations, Cousineau (2005) first "standardized" the within-subjects data set by subtracting the participant means from the actual observations Y_{ij} while adding the overall mean (i.e., $Y_{ij}^* = Y_{ij} - M_{i\cdot} + M$). Second, Cousineau (2005) computed confidence intervals for each experimental condition based on these standardized observations Y_{ij}^*. However, Morey (2008) showed that this standardization procedure results in a biased estimate of the residual variance and thus in too narrow within-subject intervals. The standardization approach by Cousineau (2005) bears some similarity with the within-subject HDI by Nathoo et al. (2018) which is defined conditional on the maximum likelihood estimates for the random effects $\hat{b}_i = M_{i\cdot} - M$. Essentially, the random effects are treated as if they were perfectly known in both cases, that is, when standardizing the observations in the classical approach (i.e., $Y^*_{ij} = Y_{ij} - \hat{b}_i$) and when conditioning on the estimates $\hat{b}_i$ in the conditional Bayesian framework in Eq. (2). This conceptual similarity also becomes clear when comparing the different estimates for the residual variance. Conditional on the random effects, Nathoo et al. (2018) derived the estimate $\hat{\sigma}^2_\epsilon = SS_{S\times C}/((N-1)C)$, which differs from the original, unbiased estimate in Loftus and Masson (1994) by the factor $C/(C-1)$—which is exactly the multiplicative correction factor derived by Morey (2008) for the standardization approach of Cousineau (2005). Overall, this comparison of classical and Bayesian within-subject intervals highlights the importance of treating the random effects as unknown parameters of the underlying statistical model.

The proposed two-step approach fits all parameters of the mixed-effects model in a fully Bayesian framework and thereby accounts for estimation uncertainty and shrinkage of the random effects. However, it is important to keep in mind that the proposed extension requires an additional assumption that is not present in the approach based on conditional Bayesian inference by Nathoo et al. (2018). Specifically, in the fully Bayesian analysis, it is assumed that the random effects are normally distributed. This additional assumption is very common in mixed-effects modeling (Bates, Mächler, Bolker, & Walker, 2015; Gelman et al., 2013; Pinheiro & Bates, 2000) and justified by the fact that the benefits of shrinkage usually outweigh the costs of making a specific distributional assumption (Efron & Morris, 1977). Moreover, even if the distribution of random effects is misspecified, it can be shown that the random-effect estimates converge to the true distribution in the population under certain conditions (Marsman et al., 2016). As a second limitation, the proposed two-step approach requires posterior samples from the mixed-effects model which renders its application computationally more costly compared to the closed-form solutions by Nathoo et al. (2018). However, given the availability of easy-to-use MCMC software (Plummer, 2003; Stan Development Team, 2018), the conceptual benefits of drawing conclusions within a fully Bayesian framework outweigh the computational costs.

In sum, I have highlighted the importance of considering estimation uncertainty and shrinkage of random effects when computing within-subject HDIs in a Bayesian framework. By applying the closed-form solution of Nathoo et al. (2018) within the context of a fully Bayesian analysis, a within-subject interval is obtained that has an intuitive interpretation (Hoekstra et al., 2014) while taking estimation uncertainty and shrinkage of the parameters into account.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. http://dx.doi.org/10.18637/jss.v067.i01.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*, 42–45. http://dx.doi.org/10.20982/tqmp.01.1.p042.

Efron, B., & Morris, C. (1977). Stein's paradox in statistics. *Scientific American*, *236*, 119–127. Retrieved from https://www.jstor.org/stable/24954030.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, *1*, 515–534. Retrieved from https://projecteuclid.org/euclid.ba/1340371048.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). Boca Raton, FL: CRC.

Heck, D. W, Arnold, N. R., & Arnold, D. (2018). TreeBUGS: An R package for hierarchical multinomial-processing-tree modeling. *Behavior Research Methods*, *50*, 264–284. http://dx.doi.org/10.3758/s13428-017-0869-7.

Hoekstra, R., Morey, R. D., Rouder, J. N., & Wagenmakers, E.-J. (2014). Robust misinterpretation of confidence intervals. *Psychonomic Bulletin & Review*, *21*, 1157–1164. http://dx.doi.org/10.3758/s13423-013-0572-3.

James, W., & Stein, C. (1961). Estimation with quadratic loss. In *Proceedings of the fourth berkeley symposium on mathematical statistics and probability, Vol. 1* (pp. 361–379). Retrieved from https://projecteuclid.org/euclid.bsmsp/1200512173.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*, 476–490. http://dx.doi.org/10.3758/BF03210951.

Ly, A., Boehm, U., Heathcote, A., Turner, B. M., Forstmann, B., Marsman, M., et al. (2017). A flexible and efficient hierarchical Bayesian approach to the exploration of individual differences in cognitive-model-based neuroscience. In A. A. Moustafa (Ed.), *Computational models of brain and behavior* (pp. 467–479). Wiley-Blackwell. http://dx.doi.org/10.1002/9781119159193.ch34.

Marsman, M., Maris, G., Bechger, T., & Glas, C. (2016). What can we learn from Plausible Values? *Psychometrika*, *81*, 274–289. http://dx.doi.org/10.1007/s11336-016-9497-x.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*, 61–64. http://dx.doi.org/10.20982/tqmp.04.2.p061.

Nathoo, F. S., Kilshaw, R. E., & Masson, M. E. J. (2018). A better (Bayesian) interval estimate for within-subject designs. *Journal of Mathematical Psychology*, *86*, 1–9. http://dx.doi.org/10.1016/j.jmp.2018.07.005.

Pinheiro, J. C., & Bates, D. M. (2000). *Statistics and Computing*. *Mixed-effects Models in S and S-PLUS*. New York, NY: Springer-Verlag.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing, Vol. 124* (p. 125). Austria: Vienna.

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*, 356–374. http://dx.doi.org/10.1016/j.jmp.2012.08.001.

Rouder, J. N., Morey, R. D., Verhagen, J., Swagman, A. R., & Wagenmakers, E.-J. (2017). Bayesian analysis of factorial designs. *Psychological Methods*, *22*, 304–321. http://dx.doi.org/10.1037/met0000057.

Stan Development Team. (2018). Stan modeling language users guide and reference manual. Version 2.18.0. Retrieved from http://mc-stan.org/.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the third berkeley symposium on mathematical statistics and probability, Vol. 1* (pp. 197–206). Retrieved from https://projecteuclid.org/euclid.bsmsp/1200501656.