Pre-Analysis Plan: A robustness replication of Cox et al. (2023)*

Preliminary Authors: Björn S. Siepe†, Matthias Kloft,
Semih C. Aktepe, Daniel W. Heck

March 1st, 2024

Abstract

This document contains the pre-analysis plan for the robustness replication of Cox et al. (2023). It is a part of the Institute for Replication (I4R) project to replicate articles published in Nature Human Behaviour.

Brief summary of the original paper

Cox et al. (2023) performed a Bayesian meta-analysis of acoustic features of infant-directed speech (IDS) compared to adult-directed speech (ADS) using data from 88 unique studies. They investigated five parameters: fundamental frequency (f_0), f_0 variability, vowel space area, articulation rate, and vowel duration. They performed the same analyses for each of these parameters separately. Briefly, they first calculate three-level intercept-only random effect models with language, study, and measures as random effects for each parameter. They then use task, environment, age, and language as predictors of the effect size in a hierarchical Bayesian model. They then compared this full model with models without each of the predictor variables based on leave-one-out cross-validation. They then mostly interpret the effect sizes of the model with the highest stacking weight for each of the five parameters in Table 1 of their manuscript.

Given the estimated effect sizes on the five parameters, the authors conclude that IDS differs robustly from ADS, and that these findings still hold when accounting for potential publication bias (with one noted exception). They further note that age had an influence on f_0, articulation rate, and vowel duration, whereas f_0 variability and vowel space were stable during development.

Reproducibility

We will assess the computational reproducibility of the author's results by executing all code provided by the authors on the Open Science Framework.[1] We will check if the code runs without error. We will additionally check if we can re-create the figures and numerical results in the main manuscript. While re-running the code, we will check it for coding errors.

---

[1]Available at https://osf.io/hc7me/.

Robustness Replication

We will perform a robustness replication of the main results in the paper. Each of the following re-analyses applies to all five parameters studied. We note that we do not consider any of these alternatives to be necessarily superior to the analysis choices made by Cox et al. (2023). Due to the computational complexity of the planned statistical analyses, we will perform some of the robustness analyses on a subset of the reported analyses.

We define a replication as an effect that shows the same direction as in the original paper and whose 95% Credible Interval excludes 0, or, in the case of a null effect in the original paper, also includes 0. We will additionally interpret the potential differences in estimated effect sizes qualitatively.

1. Use model averaging to account for model uncertainty and publication bias.
The authors assessed publication bias using the `PublicationBias` R package (Mathur and VanderWeele 2020). This method enables researchers to estimate the minimum severity of publication bias that would be required to attenuate the Credible Interval of the effect size estimate to include zero. However, the authors did not directly estimate pooled effect sizes that are adjusted for publication bias. Additionally, they did not account for model uncertainty regarding the inclusion of random effects in their meta-analysis. As an alternative, we will use the R package `RoBMA` (Bartoš et al. 2023) to estimate an ensemble of meta-analytic models and then use Bayesian model averaging to combine the effect-size estimates of all models based on their posterior probability. In this way, we can obtain a model-averaged estimate for the effect size while considering various forms of publication bias. Specifically, we will adopt the moderator models selected by Cox et al. (2023) (as indicated in Table 1) and fit this model with the `RoBMA` default model ensemble. We will compare the resulting estimates for all parameters of interest to the estimates of the original paper. Should the model not converge due to computational reasons, we will try to incrementally reduce its complexity until we achieve satisfactory convergence.

2. Use different model fit assessments.

The authors used stacking weights to compare model fit between different meta-regression models. As an alternative, we will use posterior model probabilities obtained with bridge sampling in the `brms` package (Bürkner 2017) and the WAIC information criterion (also obtained with `brms`) to compare models. We will compare both the selection of the best-fitting model and the ordering of models to the corresponding results of the original paper. We define the same model order as a full replication and the same full best-fitting model (while others may differ in their rank) as a partial replication.

3. Use different sampling settings for most complex models

The authors used relatively unconventional settings for the Hamiltionian Monte Carlo (HMC) No U-Turn sampler (NUTS) in Stan (Stan Development Team 2023), as they set `adapt_delta` $= 0.999$ and the `max_treedepth` $= 20$. We will change these sampling settings to the Stan default (`adapt_delta` $= 0.8$, `max_treedepth` $= 10$) for the models presented in Table 1 of their manuscript. We will then inspect model convergence summaries (such as $\hat{R}$) to check for possible problems in model specification (e.g., lack of convergence or identifiability of certain parameters). If such potential issues are indicated, for example, with respect to the rather complex random-effects structure, we will incrementally decrease model complexity and compare the results against those reported in the paper.

References

Bartoš, F., Maier, M., Wagenmakers, E.-J., Doucouliagos, H. and Stanley, T. D.: 2023, Robust Bayesian meta-analysis: Model-averaging across complementary publication bias adjustment methods, *Research Synthesis Methods* **14**(1), 99–116.

Bürkner, P.-C.: 2017, brms: An R Package for Bayesian Multilevel Models Using Stan, *Journal of Statistical Software* **80**, 1–28.

Cox, C., Bergmann, C., Fowler, E., Keren-Portnoy, T., Roepstorff, A., Bryant, G. and Fusaroli, R.: 2023, A systematic review and Bayesian meta-analysis of the acoustic features of infant-directed speech, *Nature Human Behaviour* **7**(1), 114–133.

Mathur, M. B. and VanderWeele, T. J.: 2020, Sensitivity analysis for publication bias in meta-analyses, *Journal of the Royal Statistical Society. Series C, Applied Statistics* **69**(5), 1091–1119.

Stan Development Team: 2023, Stan Modeling Language Users Guide and Reference Manual.
URL: *https://mc-stan.org*