Statistical Analysis and Results

# 1. Statistical Analysis

Statistical analysis of the data extracted was performed in R, (v 4.3.3; R Core Team, https://www.r-project.org/) and RStudio (v 2023.06.1; Posit, https://posit.co/). All code utilised for data preparation and analyses are available in either the Open Science Framework page for this project <https://osf.io/5fjq3/> or the corresponding GitHub repository <https://github.com/jamessteeleii/older_adults_RT_machines_MA>. The present analysis was not pre-registered as we had no a priori hypotheses and thus, given the pilot nature of this study, was considered exploratory and aimed at parameter estimation (Cumming, 2014) within a Bayesian meta-analytic framework (Kruschke & Liddell, 2018). For all analyses model parameter estimates and their precision, along with conclusions based upon them, were interpreted continuously and probabilistically, considering data quality, plausibility of effect, and previous literature, all within the context of each model. The renv package (Ushey et al., 2023) was used for package version reproducibility and a function based analysis pipeline using the targets package (Landau et al., 2023) was employed (the analysis pipeline can be viewed by downloading the R Project and running the function targets::tar\_visnetwork()). Standardised effect sizes were all calculated using the metafor package (Viechtbauer, 2023). The main package brms (Bürkner et al., 2023) was used in fitting all the Bayesian meta-analysis models. Prior and posterior draws were taken using tidybayes (Kay & Mastny, 2023) and marginaleffects (Arel-Bundock et al., 2023) packages. All visualisations were created using ggplot2 (Wickham et al., 2023), tidybayes, and the patchwork (Pedersen, 2023) packages. Where data to be extracted from included studies was reported in plots only we used the juicr package to extract this data (Lajeunesse, 2021) and the reproducible reports for this can be found in the online supplementary materials.

As noted, we adopted a Bayesian approach to the present meta-analysis. Specifically, we adopted an arm-based multiple treatment comparison (i.e., network) type model given that for the studies identified some, but not all, included a non-training control arm in addition to the machine-based RT arm (Hong et al., 2016). In typical contrast-based meta-analyses data is limited to the effect sizes for paired contrasts between arm and thus studies that include both arm (i.e., relative effects between non-training control vs machine based RT conditions); however, in arm-based analyses the data are the absolute effects within each arm and information is borrowed across studies to enable both within condition absolute, and between condition relative contrasts to be estimated. As in the present analysis we are only comparing two conditions we do not examine ranking methods as are typical in multiple treatment comparison models, but instead we focus on reporting the between condition relative contrast for non-training control vs machine-based RT. We fit two models: one for all function outcomes reported and one for all strength outcomes reported. Pre- to post-intervention period standardised effect sizes were calculated using Becker’s (Becker, 1988) for each outcome within each arm within each study. As such, data were hierarchical across three levels (i.e., effects within arms within studies) and so we included random intercepts using implicit nested coding across these levels. Given the arm-based model, we included a fixed categorical predictor using dummy coding indicating which condition a given arm within the study belonged to (i.e., non-training control or machine-based RT where the former was the intercept) and also allowed for this to be a random effect to enable the partial pooling of information across studies where there were not direct between condition relative contrasts present. We did not have any prior intuition or data available for the specific intervention in this population that was not included in the likelihood for the model anyway and so we adopted uninformed default weakly regularising priors for all parameters. We fit each model using four Monte Carlo Markov Chains each with 2000 warmup and 6000 sampling iterations. Trace plots were produced along with values to examine whether chains had converged, and posterior predictive checks for each model were also examined to understand the model implied distributions. These all showed good convergence with all values close to 1 and posterior predictive checks seemed appropriate distributions for the observed data (all diagnostic plots can be seen in the supplementary materials: <https://osf.io/z9u7s>). From each model we obtained draws from the posterior distributions for the conditional absolute estimates for each condition by study, the global grand mean absolute estimates for each condition, and the between condition relative contrast for non-training control vs machine-based RT in order to present probability density functions visually, and also to calculate mean and 95% quantile intervals (i.e., ‘credible’ or ‘compatibility’ intervals) for each estimate. These gave us the most probable value of the parameter in addition to the range from 2.5% to 97.5% percentiles.

# 2. Results

## 2.1 Functional Outcomes

The model examining functional outcomes included 15 studies containing 33 separate arms (8 non-training control and 15 machine-based RT) reporting 54 within arm effects. The global grand mean estimate for the between condition relative contrast (i.e., machine-based RT minus non-training control) was 0.63 [95% credible interval: 0.23,1.04], though there was considerable heterogeneity in the magnitude of effects between studies ( = 0.66 [95% credible interval: 0.37,1.07]). An ordered forest plot of conditional study level estimates for absolute within condition effects, the global grand mean estimates for absolute within condition effects, and the global grand mean estimate for the between condition relative contrast including interval estimates and posterior probability distributions are shown in [Figure 1](#fig-function-model)

|  |
| --- |
| Figure 1: An ordered forest plot for functional outcomes of conditional study level estimates for absolute within condition effects including individual effect sizes as points (top panel), the global grand mean estimates for absolute within condition effects (bottom left panel), and the global grand mean estimate for the between condition relative contrast (top right panel) including interval estimates and posterior probability distributions |

## 2.2 Strength Outcomes

The model examining strength outcomes included 11 studies containing 24 separate arms (6 non-training control and 11 machine-based RT) reporting 60 within arm effects. The global grand mean estimate for the between condition relative contrast (i.e., machine-based RT minus non-training control) was 0.61 [95% credible interval: 0.21,1.01], though there was considerable heterogeneity in the magnitude of effects between studies ( = 0.57 [95% credible interval: 0.32,1]). An ordered forest plot of conditional study level estimates for absolute within condition effects, the global grand mean estimates for absolute within condition effects, and the global grand mean estimate for the between condition relative contrast including interval estimates and posterior probability distributions are shown in [Figure 2](#fig-strength-model)

|  |
| --- |
| Figure 2: An ordered forest plot for strength outcomes of conditional study level estimates for absolute within condition effects including individual effect sizes as points (top panel), the global grand mean estimates for absolute within condition effects (bottom left panel), and the global grand mean estimate for the between condition relative contrast (top right panel) including interval estimates and posterior probability distributions |

# 3. References

Arel-Bundock, V., cre, cph, Diniz, M. A., Greifer, N., & Bacher, E. (2023). *Marginaleffects: Predictions, Comparisons, Slopes, Marginal Means, and Hypothesis Tests*.

Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology*, *41*(2), 257–278. <https://doi.org/10.1111/j.2044-8317.1988.tb00901.x>

Bürkner, P.-C., Gabry, J., Weber, S., Johnson, A., Modrak, M., Badr, H. S., Weber, F., Ben-Shachar, M. S., Rabel, H., Mills, S. C., & Wild, S. (2023). *Brms: Bayesian Regression Models using ’Stan’*.

Cumming, G. (2014). The New Statistics: Why and How. *Psychological Science*, *25*(1), 7–29. <https://doi.org/10.1177/0956797613504966>

Hong, H., Chu, H., Zhang, J., & Carlin, B. P. (2016). A Bayesian missing data framework for generalized multiple outcome mixed treatment comparisons. *Research Synthesis Methods*, *7*(1), 6–22. <https://doi.org/10.1002/jrsm.1153>

Kay, M., & Mastny, T. (2023). *Tidybayes: Tidy Data and ’Geoms’ for Bayesian Models*.

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, *25*(1), 178–206. <https://doi.org/10.3758/s13423-016-1221-4>

Lajeunesse, M. J. (2021). *Juicr: Automated and Manual Extraction of Numerical Data from Scientific Images*.

Landau, W. M., Warkentin, M. T., Edmondson, M., Oliver, S., Mahr, T., & Company, E. L. and. (2023). *Targets: Dynamic Function-Oriented ’Make’-Like Declarative Pipelines*.

Pedersen, T. L. (2023). *Patchwork: The Composer of Plots*.

Ushey, K., cre, Wickham, H., Software, P., & PBC. (2023). *Renv: Project Environments*.

Viechtbauer, W. (2023). *Metafor: Meta-Analysis Package for R*.

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D., Posit, & PBC. (2023). *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*.