Statistical Analysis and Results

# 1. Statistical Analysis

The present analysis was not pre-registered as we had no a priori hypotheses and, given the limited sample size due to resource constraints, was considered exploratory. Inferential statistics were treated as highly unstable local descriptions of the relations between model assumptions and data in order to acknowledge the inherent uncertainty in drawing generalised inferences from single and small samples (Amrhein et al., 2019). For all analyses we opted to take an estimation-based approach typical when examining validity and reliability. That is, we provide point estimates and the uncertainty in them for the statistical parameters reported. Two sets of models were employed exploring the JHs recorded from CMJ trials; one to examine the agreement between the IM method using the Force Decks (i.e., gold-standard) and each of the other methods, and one to explore the test-retest reliability for each method. Given we had nested data whereby each participant provided three trials for each method on two separate testing sessions we adopted a mixed effects limits of agreement approach (Parker et al., 2020). This allowed us to estimate mean bias for each method compared with the gold-standard for agreement utilising both first and second session data, and any test retest bias for reliability, in addition to to upper and lower limits of agreement providing 95% coverage probabilities for both agreement and test-retest reliability. Models were fit using the lme4 package and using Restricted Maximum Likelihood Estimation. For each of the models fit we used nonparametric case based bootstrapping resampling 10000 times at the individual participant level and refitting models in order to construct 95% quantile intervals for both the mean bias and limits of agreement estimates. Bootstrapping was performed using the lmeresampler package.

## 1.1 Agreement

For estimation of bias and limits of agreement regarding the agreement between the gold-standard and each device we utilised the following mixed effects model structure:

Where is the difference between measurements taken between the two devices (i.e., ), where the device indexed by is the gold-standard and device indexed by is the comparison device, for participant during session and for trial . Here is the overall mean of the between device differences (i.e., the mean bias), is the random effect for the participant, is the random effect for the session which is nested within participant, and is the error term. The 95% limits of agreement can then be calculated as:

with the square root of the total variance providing an estimate of the standard deviation of the differences for use in the conventional Bland-Altman limits of agreement calculation. A model was fit for each device in comparison o the gold-standard.

## 1.2 Reliability

For estimation of bias and limits of agreement regarding the test-retest reliability between each test session for each device we utilised the following mixed effects model structure:

Where is the difference between measurements taken between the two sessions for a given device (i.e., ), where the session indexed by is the first test session and the session indexed by is the second test session, for participant and for trial (note we use the superscript to distinguish this from the agreement model. Here is the overall mean of the between session differences (i.e., the mean bias), is the random effect for the participant, and is the error term. The 95% limits of agreement can then be calculated as:

with the square root of the total variance providing an estimate of the standard deviation of the differences for use in the conventional Bland-Altman limits of agreement calculation. A model was fit for each device in order to examine its test-retest reliability.

## 1.3 Open data, code, and materials

All data and code is presented in the supplementary materials (https://osf.io/e7qg5/). The renv package was used for package version reproducibility and a function based analysis pipeline using the targets package was employed (the analysis pipeline can be viewed by downloading the R Project and running the function targets::tar\_visnetwork()). We cite all packages used in the analysis pipeline below.

We used R version 4.3.3 (R Core Team, 2024) and the following R packages: glue v. 1.7.0 (Hester & Bryan, 2024), knitr v. 1.47 (Xie, 2014, 2015, 2024), lme4 v. 1.1.35.3 (Bates et al., 2015), lmeresampler v. 0.2.4 (Loy et al., 2024), patchwork v. 1.2.0 (Pedersen, 2024), quarto v. 1.4 (**quarto?**), renv v. 1.0.3 (Ushey & Wickham, 2023), rmarkdown v. 2.27 (Allaire et al., 2024; Xie et al., 2018, 2020), tarchetypes v. 0.9.0 (Landau, 2021a), targets v. 1.7.0 (Landau, 2021b), tidyverse v. 2.0.0 (Wickham et al., 2019).

# 2. Results

## 2.1 Agreement

The exact mean bias and limits of agreement along with their bootstrapped interval estimates can be seen for the agreement of each device in comparison to the gold-standard (impulse momentum method using the Force Decks) in [Figure 1](#fig-agree-plot). Both the MyJump App and the Optojump demonstrated very similar degrees of mean bias of around ~2 cm overestimation, and a similar width to their limits of agreement ranging ~11-12 cm about the mean bias. The Jump Mat however had a much larger mean bias of ~10 cm overestimation yet a similar range for their limits of agreement as the other devices.

|  |
| --- |
| Figure 1: Mean bias (thick horizontal line) and 95% limits of agreement (dotted horizontal lines) along with 95% quantile interval estimates for agreement of each device in comparison to the gold-standard (Force Decks impulse-momentum method). |

## 2.2 Reliability

The exact mean bias and limits of agreement along with their bootstrapped interval estimates can be seen for the test-retest reliability of each device between sessions in [Figure 1](#fig-agree-plot). All devices demonstrated minimal mean bias between sessions each typically less than 1 cm, and all demonstrated a similar width to their limits of agreement ranging ~10 cm about the mean bias.

|  |
| --- |
| Figure 2: Mean bias (thick horizontal line) and 95% limits of agreement (dotted horizontal lines) along with 95% quantile interval estimates for each the test-retest reliability of device between sessions. |

# 3. References

Allaire, J., Xie, Y., Dervieux, C., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W., & Iannone, R. (2024). *rmarkdown: Dynamic documents for R* [Manual].

Amrhein, V., Trafimow, D., & Greenland, S. (2019). Inferential Statistics as Descriptive Statistics: There Is No Replication Crisis if We Don’t Expect Replication. *The American Statistician*, *73*(sup1), 262–270. <https://doi.org/10.1080/00031305.2018.1543137>

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

Hester, J., & Bryan, J. (2024). *glue: Interpreted string literals* [Manual].

Landau, W. M. (2021a). *tarchetypes: Archetypes for targets* [Manual].

Landau, W. M. (2021b). The targets R package: A dynamic Make-like function-oriented pipeline toolkit for reproducibility and high-performance computing. *Journal of Open Source Software*, *6*(57), 2959.

Loy, A., Steele, S., & Korobova, J. (2024). *lmeresampler: Bootstrap methods for nested linear mixed-effects models* [Manual].

Parker, R. A., Scott, C., Inácio, V., & Stevens, N. T. (2020). Using multiple agreement methods for continuous repeated measures data: A tutorial for practitioners. *BMC Medical Research Methodology*, *20*(1), 154. <https://doi.org/10.1186/s12874-020-01022-x>

Pedersen, T. L. (2024). *patchwork: The composer of plots* [Manual].

R Core Team. (2024). *R: A language and environment for statistical computing* [Manual]. R Foundation for Statistical Computing.

Ushey, K., & Wickham, H. (2023). *renv: Project environments* [Manual].

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. <https://doi.org/10.21105/joss.01686>

Xie, Y. (2014). knitr: A comprehensive tool for reproducible research in R. In V. Stodden, F. Leisch, & R. D. Peng (Eds.), *Implementing reproducible computational research*. Chapman and Hall/CRC.

Xie, Y. (2015). *Dynamic documents with R and knitr* (2nd ed.). Chapman and Hall/CRC.

Xie, Y. (2024). *knitr: A general-purpose package for dynamic report generation in R* [Manual].

Xie, Y., Allaire, J. J., & Grolemund, G. (2018). *R markdown: The definitive guide*. Chapman and Hall/CRC.

Xie, Y., Dervieux, C., & Riederer, E. (2020). *R markdown cookbook*. Chapman and Hall/CRC.