Statistical Analysis and Results

# 1. Statistical Analysis

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. 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, Trafimow, et al., 2019). For all analyses we opted to avoid dichotomising the existence of effects and therefore did not employ traditional null hypothesis significance testing on parameter estimates (Amrhein, Greenland, et al., 2019; McShane et al., 2019). Instead, we opted to take a model- (Sterba, 2009) and estimation-based approach (Cumming, 2014). 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. We focused primarily on qualitative examination of our results based on visualization of the data and models for fixed effects, and exploration of variances using random effects and visualising individual participant level model predictions.All analysis was performed in R (version 4.2.2, The R Foundation for Statistical Computing, 2022) and all data and code is presented in the supplementary materials (https://osf.io/kf9r3/).All data visualisations were made using ggplot2 (Wickham et al., 2022) and the patchwork package (Pedersen, 2022). The aim of our analysis was to explore how well the Zelemiq sensor device predicted blood lactate levels as measured from capillary samples by the Biosen C-Line, in addition to the agreement between lactate thresholds detected using either device. Both Zelemiq and blood lactate values were standardised within participants (i.e., clusters) by subtracting the participants mean and dividing by their standard deviation. Thus Zelemiq and blood lactate values were expressed in participants standard deviation units. A rolling mean was also calculated from the standardised Zelemiq raw data using a 10 sample window prior to the corresponding blood lactate values at that time point.

## 1.1 Prediction of blood lactate values from Zelemiq values

We firstly explored whether the Zelemiq values were able to predict the corresponding blood lactate values. The dependent variable in our model was therefore the standardised blood lactate levels, and the independent predictor variable was the standardised rolling average of the Zelemiq sensor data. Visual exploration of the data suggested a non-linear relationship between the two and so we opted to fit the Zelemiq values with a second-order polynomial function. A maximal (i.e., with random intercepts and slopes for both linear and quadratic terms) linear mixed effect model was estimated initially using Restricted Maximal Likelihood with the lme4 package (Bates et al., 2023) however this obtained a singular fit and even reduced models did not avoid this (other model checks for this structure are available in the supplementary materials - <https://osf.io/98ey6>). As such we refit the model using a Bayesian approach employing weakly regularising default priors to allow for estimation of the random effects using the brms package (Bürkner, 2017). The inclusion of random effects for the quadratic term however resulted in large values suggested chains had not converged and thus we removed this keeping on the linear term as a random slope. With the Zelemiq sensor data as a fixed second-order polynomial effect, and allowing random intercepts and random linear slopes by participant id, the model equation was as follows where and indicate they are the standardised variables:

All parameters in the model had values , trace plots demonstrated chain convergence, and the posterior predictive checks appeared appropriate (see [https://osf.io/e782c](https://osf.io/e782c%5D)). Given the novelty of the Zelemiq sensor device we did not have a clear intuition or informed opinion about what priors to set and so opted to use the default weakly regularising priors merely to allow for random effects estimation and “let the data speak”. Four Monte Carlo Markov Chains with 4000 warmup and 4000 sampling iterations were used in each model. Model predictions and 95% quantile intervals were visualised for both the global grand mean (i.e., the expectation of the posterior predictive distribution for the parameter ignoring cluster effects) in addition to the conditional effects (i.e., including cluster effects) and the posterior predictions, and also at the individual participant level using the marginaleffects package (Arel-Bundock et al., 2022). The standarised raw Zelemiq sensor data (i.e., unaveraged) was also visualised alongside the standardised blood lactate data.

## 1.2 Agreement of lactate thresholds detected using blood lactate values compared to Zelemiq values

We used the lactater package (Maturana, 2023) for the detection of blood lactate thresholds. We similarly used the standardised blood lactate and Zelemiq averaged values (i.e., and from the model described above) and as such did not explore threshold detection methods which relied on log transformations of the lactate response because of negative standardised values. For both blood lactate and Zelemiq values we determined thresholds for watts on the cycle ergometer using the Dmax, Exp-Dmax, ModDmax, and both LTP1 and LTP2 methods (Jamnick et al., 2018). We then examined the agreement between the watts at which each threshold was detected between the blood lactate and the Zelemiq values using the Bland-Altman 95% limits of agreement in addition to the concordance correlation coefficient () calculated using the SimpleAgree package (Caldwell, 2022). Frequentist confidence intervals for the bias and were calculated at the 95% level and for the upper and lower limits of agreement at the 90% level.

# 2. Results

## 2.1 Prediction of blood lactate values from Zelemiq values

Model parameters estimates and 95% quantile intervals are shown in [Table 1](#tbl-model) and [Figure 1](#fig-model) panel (A) shows the raw data and model predictions for each individual participant in addition to the participant level in panel (B) and model predictions in panel (C). The $\operatorname{Zelemiq}^\*\_{linear\space term}$ indicated that the rate of change in blood lactate at a standardised Zelemiq value of zero was 0.64 [95% quantile interval: 0.55, 0.72] and the positive $\operatorname{Zelemiq}^\*\_{quadratic\space term}$ was 0.17 [95% quantile interval: 0.11, 0.23] indicating a convex curve. Thus, blood lactate in standardised units increased exponentially with increasing Zelemiq standardised values. The random effects parameter estimates from the Bayesian model were negligible as expected given the singularity of the Frequentist model fit with Restricted Maximum Likelihood. The majority of variance in the model was attributable to the residual variation.

Table 1: Model parameter estimates for both fixed and random effects.

| Model Term | Estimate | Lower 95\% QI | Upper 95\% QI |
| --- | --- | --- | --- |
| **Fixed Effects** | | | |
| $Intercept$ | -0.29 | -0.40 | -0.18 |
| $\operatorname{Zelemiq}^\*\_{linear\space term}$ | 0.64 | 0.55 | 0.72 |
| $\operatorname{Zelemiq}^\*\_{quadratic\space term}$ | 0.17 | 0.11 | 0.23 |
| **Random Effects** | | | |
| $\sigma\_{Intercept}$ | 0.05 | 0.00 | 0.16 |
| $\sigma\_{\operatorname{Zelemiq}^\*\_{linear\space term}}$ | 0.04 | 0.00 | 0.14 |
| $\rho\_{Intercept:\operatorname{Zelemiq}^\*\_{linear\space term}}$ | 0.04 | -0.94 | 0.96 |
| $\sigma\_{Residual}$ | 0.28 | 0.23 | 0.34 |
| Note: |  |  |  |
| QI = quantile interval |  |  |  |
| Note, estimates are in standard deviation units at the individual participant level. |  |  |  |

|  |
| --- |
| Figure 1: Panel (A) shows the standardised values for raw Zelemiq sensor output and Biosen C-Line blood lactate during the course of the incremental test where time has been normalised to 0-100% of the test duration and a locally estimated scatter smooth (LOESS) curve for each has been performed; panel (B) shows the individual participant level predicted values (thick lines) with 95% quantile intervals (ribbons) from the fitted model; panel (C) shows the global grand mean estimate (thick line) from the model with 95% quantile interval (dark grey ribbon) in addition to intervals for the conditional effects including the random effects of participants (lighter grey ribbon) and the posterior predictions (lightest grey ribbon). Individual points in each panel are the individual observed values of data; and panel (D) shows the scatterplots of lactate thresholds detected from either Zelemiq or blood lactate values using Dmax, Exp-Dmax, LTP1, LTP2, and ModDmax. |

## 2.2 Agreement of lactate thresholds detected using blood lactate values compared to Zelemiq values

Scatterplots of lactate thresholds calculated from either blood lactate or from the Zelemiq values using either Dmax, Exp-Dmax, LTP1, LTP2, or ModDmax can be see in [Figure 1](#fig-model) panel (D). In comparing the two devices there appeared to be the greatest agreement in threshold detected when using the ModDmax method with a bias of -0.95 [95% confidence interval:-13.85,11.95] Watts and lower and upper 95% limits of agreement of bias of -31.2 [95% confidence interval:-45.6,-16.8] and bias of 29.3 [95% confidence interval:14.9,43.7] Watts respectively. For the ModDmax method the $\_{CCC}$0.93 [95% confidence interval:0.77,0.98].

# 3. References

Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance. *Nature*, *567*(7748), 305–307. <https://doi.org/10.1038/d41586-019-00857-9>

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>

Arel-Bundock, V., Diniz, M. A., & Greifer, N. (2022). *Marginaleffects: Marginal Effects, Marginal Means, Predictions, and Contrasts*.

Bates, D., Maechler, M., Bolker [aut, B., cre, Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., Green, P., Fox, J., Bauer, A., & simulate.formula), P. N. K. (shared. copyright on. (2023). *Lme4: Linear Mixed-Effects Models using ’Eigen’ and S4*.

Bürkner, P.-C. (2017). Brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*, 1–28. <https://doi.org/10.18637/jss.v080.i01>

Caldwell, A. (2022). *SimplyAgree: Flexible and Robust Agreement and Reliability Analyses*.

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

Jamnick, N. A., Botella, J., Pyne, D. B., & Bishop, D. J. (2018). Manipulating graded exercise test variables affects the validity of the lactate threshold and VO2peak. *PLOS ONE*, *13*(7), e0199794. <https://doi.org/10.1371/journal.pone.0199794>

Maturana, F. M. (2023). *Lactater: Tools for Analyzing Lactate Thresholds*.

McShane, B. B., Gal, D., Gelman, A., Robert, C., & Tackett, J. L. (2019). Abandon Statistical Significance. *The American Statistician*, *73*(sup1), 235–245. <https://doi.org/10.1080/00031305.2018.1527253>

Pedersen, T. L. (2022). *Patchwork: The Composer of Plots*.

Sterba, S. K. (2009). Alternative Model-Based and Design-Based Frameworks for Inference From Samples to Populations: From Polarization to Integration. *Multivariate Behavioral Research*, *44*(6), 711–740. <https://doi.org/10.1080/00273170903333574>

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D., & RStudio. (2022). *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*.