Modeling the relationship between load and repetitions to failure in resistance training:

Mitter, Benedikt ; Zhang, Lei; Bauer, Pascal ; Baca, Arnold ; Tschan, Harald

DOI:
10.1080/17461391.2022.2089915

European Journal of Sport Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tejs20

Modelling the relationship between load and repetitions to failure in resistance training: A Bayesian analysis

Benedikt Mitter, Lei Zhang, Pascal Bauer, Arnold Baca & Harald Tschan

To cite this article: Benedikt Mitter, Lei Zhang, Pascal Bauer, Arnold Baca & Harald Tschan (2022): Modelling the relationship between load and repetitions to failure in resistance training: A Bayesian analysis, European Journal of Sport Science, DOI: 10.1080/17461391.2022.2089915

To link to this article: https://doi.org/10.1080/17461391.2022.2089915

EUROPEAN JOURNAL OF SPORT SCIENCE
https://doi.org/10.1080/17461391.2022.2089915

Modelling the relationship between load and repetitions to failure in resistance training: A Bayesian analysis

Benedikt Mitter [a], Lei Zhang [b], Pascal Bauer [a], Arnold Baca [a] and Harald Tschan [a]

[a]Centre for Sport Science and University Sports, University of Vienna, Vienna, Austria; [b]Social, Cognitive and Affective Neuroscience Unit, Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna, Vienna, Austria

ABSTRACT

To identify the relationship between load and the number of repetitions performed to momentary failure in the pin press exercise, the present study compared different statistical model types and structures using a Bayesian approach. Thirty resistance-trained men and women were tested on two separate occasions. During the first visit, participants underwent assessment of their one-repetition maximum (1-RM) in the pin press exercise. On the second visit, they performed sets to momentary failure at 90%, 80% and 70% of their 1-RM in a fixed order during a single session. The relationship between relative load and repetitions performed to failure was fitted using linear regression, exponential regression and the critical load model. Each model was fitted according to the Bayesian framework in two ways: using an across-subjects pooled data structure and using a multilevel structure. Models were compared based on the variance explained (R^2) and leave-one-out cross-validation information criterion (LOOIC). Multilevel models, which incorporate higher-level commonalities into individual relationships, demonstrated a substantially better fit (R^2: 0.97–0.98) and better predictive accuracy compared to generalised pooled-data models (R^2: 0.89–0.93). The multilevel 2-parameter exponential regression emerged as the best representation of data in terms of model fit, predictive accuracy and model simplicity. The relationship between load and repetitions performed to failure follows an individually expressed exponential trend in the pin press exercise. To accurately predict the load that is associated with a certain repetition maximum, the relationship should therefore be modelled on a subject-specific level.

KEYWORDS
Strength-endurance continuum; repetition maximum; maximum number of repetitions; prediction; critical load

1. Introduction

Modelling the relationship between exercise intensity and the maximum amount of realisable physical work has been an increasingly addressed objective in sports science (Bergstrom, Dinyer, Succi, Voskuil, & Housh, 2021). In resistance training, this relationship has previously been characterised by the term "strength-endurance continuum" (Campos et al., 2002) and can be described for a given exercise by modelling the external load as a function of the number of repetitions performed to momentary failure (RTF) (e.g. Reynolds, Gordon, & Robergs, 2006; Mayhew, Ball, Arnold, & Bowen, 1992). Precise knowledge about the interrelation between these two variables can be beneficial in various ways. First, it would enable the comprehensive determination of a person's exercise-specific physical fatigability in dependence of the external load. In contrast to previously documented methods of assessing strength endurance (Lawton, Cronin, & McGuigan, 2011), this approach may yield a descriptive indicator that is not limited to a single load, but rather describes fatigue resistance across a wide spectrum of loads (i.e. a "strength-endurance profile"). Second, it would enable the prediction of the maximum external load a person can move in a given exercise for any given number of repetitions. This involves the concept of estimating the one-repetition maximum (1-RM) based on the RTF that can be achieved at a submaximal load, which has been frequently investigated over the past decades (Reynolds et al., 2006; Mayhew et al., 1992; Braith, Graves, Leggett, & Pollock, 1993; Brzycki, 1993; Mayhew, Johnson, Lamonte, Lauber, & Kemmler, 2008; LeSuer, McCormick, Mayhew, Wasserstein, & Arnold, 1997; Brechue & Mayhew, 2009). Third, it would enable certain methods of resistance training prescription, like the standardisation of muscular exhaustion based on a predicted

CONTACT Benedikt Mitter benedikt.mitter@univie.ac.at Centre for Sport Science and University Sports, University of Vienna, Auf der Schmelz 6, Vienna 1150, Austria

Supplemental data for this article can be accessed online at https://doi.org/10.1080/17461391.2022.2089915

training maximum, a prescriptive approach which has been referred to as "relative intensity of set-repetition best" (Scott, Duthie, Thornton, & Dascombe, 2016; Suchomel, Nimphius, Bellon, Hornsby, & Stone, 2021). Previous studies that intended to model the relationship between external load and RTF mainly focused on applying across-subject regressions by means of either linear (Reynolds et al., 2006; Brzycki, 1993; Haff & Triplett, 2016; Adams & Beam, 2014) or exponential models (Reynolds et al., 2006; Mayhew et al., 1992; Sakamoto & Sinclair, 2006; Desgorces, Berthelot, Dietrich, & Testa, 2010). However, there is evidence suggesting that the strength-endurance relationship may succumb to considerable interindividual differences attributed to numerous factors, such as specificity to the tested exercise (Reynolds et al., 2006; Hoeger, Hopkins, Barette, & Hale, 1990), movement cadence (Sakamoto & Sinclair, 2006; LaChance & Hortobagyi, 1994) as well as the athletes' training experience (Braith et al., 1993; Brechue & Mayhew, 2009; Hoeger et al., 1990) and training background (Desgorces et al., 2010; Richens & Cleather, 2014). Considering these potential confounders, one could argue that any statistical modelling approach that generalises the relationship across different subjects without accounting for subject heterogeneity (i.e. "complete-pooling models") may result in a suboptimal representation of the strength-endurance relationship, therefore impeding both the fit and predictive accuracy of such models. Indeed, independent validation studies predominantly reported noticeable inaccuracy of 1-RM predictions based on complete-pooling models, especially when using lower relative loads (Reynolds et al., 2006; Mayhew et al., 2008; LeSuer et al., 1997; Brechue & Mayhew, 2009; Ware, Clemens, Mayhew, & Johnston, 1995). Several researchers have sought to overcome this issue and improve model validity by transposing the concept of critical power (Burnley & Jones, 2018) to dynamic resistance training (Bergstrom et al., 2021; Morton, Redstone, & Laing, 2014). The so-called *critical load* model, also referred to as critical lift or critical resistance model, introduced the idea of modelling strength endurance on an individual level (i.e. "no-pooling models") rather than a group-level, therefore treating the individual as the population of interest. While this concept may provide a valuable alternative for when data availability is limited, it should still be treated with caution in scientific research, since it implies a higher potential to overfit data (Gelman, 2006). A more promising solution may be expected by applying a multilevel model (i.e. "mixed model", also called "hierarchical model" or "partial pooling model") to the strength-endurance relationship, including both, group-level and subject-level parameters (Gelman, 2006). In particular, the use of Bayesian multilevel modelling seems promising, as simulation research has shown Bayesian parameter estimation to be more accurate than maximum likelihood estimation in small samples (Lee & Song, 2004). However, it has yet to be evaluated whether a multilevel modelling approach yields an advantage over the complete-pooling approach that has been primarily applied in research thus far. Furthermore, proposed models (linear regression, exponential regression, and critical load model) have yet to be compared among each other to determine which one provides the most appropriate representation of the strength-endurance relationship. The present study was designed to address these two issues using the example of the pin press exercise, which can be considered a variant of the bench press, in a resistance-trained population. Results will help to improve understanding of the relationship between load and RTF across a high-load range.

Table 1. Subject characteristics.

	Male ($n = 19$)	Female ($n = 11$)
Age (y)	27.4 ± 3.7 [21.2–33.6]	26.9 ± 5.2 [20.2–35.9]
Experience in BP (y)	7.6 ± 3.0 [3.0–15.0]	3.5 ± 2.6 [1.0–10.0]
Height (cm)	180.9 ± 5.4 [171.5–191.6]	163.1 ± 5.1 [154.3–171.0]
Body mass (kg)	85.4 ± 7.2 [69.2–96.9]	63.4 ± 4.4 [55.2–69.7]
1-RM (kg)	112.2 ± 13.6 [85.0–142.5]	61.4 ± 10.0 [50.0–80.0]
Relative 1-RM (kg? kg^{-1})	1.32 ± 0.12 [1.15–1.50]	0.98 ± 0.19 [0.77–1.31]
RTF at 90%1-RM (n)	4.3 ± 0.9 [3.0–6.0]	4.3 ± 1.3 [2.0–6.0]
RTF at 80%1-RM (n)	7.6 ± 1.3 [5.0–10.0]	8.4 ± 1.6 [6.0–11.0]
RTF at 70%1-RM (n)	12.1 ± 2.4 [7.0–16.0]	13.1 ± 2.1 [9.0–15.0]

Note: Data are presented as mean ± SD [min – max]. BP: bench press; 1-RM: one-repetition maximum in the pin press; RTF, repetitions performed to momentary failure in the pin press.

2. Materials and methods

2.1. Participants

Nineteen men and eleven women with previous experience in resistance training voluntarily participated in the investigation. Descriptive characteristics of participants are summarised in Table 1. Inclusion criteria were: (a) being free of illness and injury, (b) being between 18 and 40 years of age, (c) having at least one year of regular training experience in the bench press exercise and (d) achieving a minimal relative 1-RM in the pin press of 1x body mass (men) or 0.75x body mass (women). Subjects were informed about benefits and potential risks related to their participation, completed a modified Physical Activity Readiness Questionnaire and signed an informed consent form prior to undergoing any test. All procedures were implemented in accordance with the ethical guidelines of the

Declaration of Helsinki and approved by a local ethical review committee (no. 00461).

2.2. Experimental design

Participants attended the laboratory on two days, separated by approximately 48 h. On day 1, subjects were assessed for body mass and height using a scale (Seca Model 877; SECA GmbH & Co. KG., Hamburg, Germany) and stadiometer (Seca Model 217; SECA GmbH&Co. KG., Hamburg, Germany). Afterwards, they were familiarised with the execution of the free-weight pin press exercise and followed a progressive loading test to determine their individual 1-RM. On day 2, participants completed sets to momentary failure at submaximal loads in descending order. Subjects were instructed to refrain from strenuous exercise and alcohol 24 h before tests and not to consume caffeine 6 h prior to testing. The exercise was performed in a Competition Combo Rack approved by the International Powerlifting Federation using a 20-kg barbell and calibrated weight plates (Eleiko, Halmstad, Sweden).

To provide participants a safe testing environment for performing sets to momentary failure and to reduce potential variability in RTF resulting from an inconsistent use of the stretch-shortening cycle, the pin press was executed according to the following movement specifications: in each repetition, subjects were required to lower the barbell onto two safety pins adjusted to a height that would allow for a distance between the barbell's lowest position and the participant's chest of up to 3 cm. Upon having the barbell come to rest on the safety pins, a researcher would provide the command "Press!", ordering the subject to perform the concentric phase of the movement at maximum intended velocity until reaching full extension of their elbows. When multiple repetitions were executed within a set performed to momentary failure, participants were further instructed to autonomously minimise the time holding the barbell with extended elbows in between repetitions in order to reach the point of momentary failure as quickly as possible. Throughout each set, they had to maintain their feet's position on the floor and keep their hip, shoulders and head in contact with the bench. A linear position transducer (GymAware Power Tool, Kinetic Performance Technologies, Canberra, Australia) was used to record mean concentric barbell velocity, to provide testers with feedback during the 1-RM assessment and help selecting appropriate load increments. The accuracy of the device has been scientifically validated before (Mitter et al., 2021) and its use for the assessment of mean velocity has been reported to provide good test-retest reliability (Dorrell, Moore, Smith, & Gee, 2019; Orange et al., 2020).

2.3. One-repetition maximum assessment (day 1)

Participants followed a standardised general warm-up including 5 min of stationary cycling (Kettler X1, Trisport, Huenenberg, Switzerland) at a cadence of about 80 rpm and a constant power output of 1 W per kg body mass. Subsequently, they completed 2 min of unloaded dynamic mobilisation exercises comprising circumduction of the shoulders, flexion and extension of the elbows and circumduction of the wrists, followed by 10 repetitions of axial external rotation of the humerus against light elastic resistance. In the next step, subjects were required to estimate their 1-RM in the pin press, considering the previously described specifications for movement execution. A progressive loading scheme was applied to slowly approach the true 1-RM, using loads equivalent to 25%, 50%, 75%, 85% and 95% of the estimated 1-RM. The number of repetitions performed at each load and passive rest in between sets were standardised according to an established autoregulatory procedure (Mitter et al., 2021; Sánchez-Medina, Perez, & Gonzalez-Badillo, 2010) that bases set configurations on the achieved barbell velocity of the preceding set, which has been considered a good predictor of the actually applied relative load (Weakley et al., 2021). The rationale for employing this autoregulatory approach was to rudimentarily account for the possibility of subjects under- or overestimating their 1-RM and, consequently, being assigned an inadequate combination of actual warm-up loads, repetition numbers and rest periods that might result when assigning fixed values to inaccurate subjective estimates. Participants initially performed three repetitions with a 3 min break in between sets. Volume was adapted to two repetitions accompanied by a 4 min break, once mean velocity dropped below 1.0 $m \cdot s^{-1}$, and further reduced to a single repetition accompanied by 5 min of rest, once mean velocity fell below 0.65 $m \cdot s^{-1}$. Rest intervals were chosen corresponding to those reported for highly reliable 1-RM test protocols (Grgic, Lazinica, Schoenfeld, & Pedisic, 2020). After completing 95% of their estimated 1-RM, load increments were selected individually based on the participant's subjective feedback and achieved mean barbell velocity. Larger individual load increments of 2.5–10 kg were selected as long as the achieved mean concentric barbell velocity of the preceding attempt was above 0.2 $m \cdot s^{-1}$, which corresponds to recently reported norm values (mean + one standard deviation) of the velocity achieved at the 1-RM in the bench press (Weakley et al., 2021). Small load increments of 2.5 kg were selected once mean concentric barbell velocity fell below 0.2 $m \cdot s^{-1}$. The test was terminated once a subject could no longer press an assigned load across the full range of motion, suggesting that the 1-RM had been reached. On average, subjects required 2.5 ± 1.4 attempts to determine their 1-RM and reached a velocity at 1-RM of 0.13 ± 0.04 $m \cdot s^{-1}$.

2.4. Repetitions to failure assessment (day 2)

Participants were tested for the RTF in the pin press at loads corresponding to 90%, 80% and 70% of the previously determined 1-RM. Each RTF test was initiated by the same general warm-up applied for the 1-RM assessment on day 1. Subsequently, subjects completed three specific warm-up sets in the pin press, comprising three repetitions at 25%, three repetitions at 50% and two repetitions at 75% 1-RM. 3 min of rest were provided in between warm-up sets and 5 min of rest prior to each test to momentary failure. A test to momentary failure was terminated once the participant attempted to complete the concentric phase of a current repetition, but was unable to do so (Steele, Fisher, Giessing, & Gentil, 2017). To increase efficiency of data acquisition and, thus, limit participant drop out, the test protocol for the RTF assessment was designed for implementation within a single visit. Therefore, two methodological specifications were applied in order to minimise negative effects of accumulating fatigue on the completed RTF. First, the sequence of tested loads was fixed in a declining manner (i.e. set 1: 90%, set 2: 80%, set 3: 70% 1-RM), as research suggests that fatigue is more prevalent after sets performed to failure at lighter loads compared to heavier loads (Sánchez-Medina & González-Badillo, 2011). Second, subjects were granted a prolonged period of rest in between sets to failure (de Salles et al., 2009). For this purpose, each set to momentary failure was immediately followed by 5 min of passive rest. After that, subjects completed the general and specific warm-up described above in order maintain positive warm-up effects. This yielded an approximate 22 min in between sets to failure, while applying the same preparatory measures before each test.

2.5. Statistical modelling

The following four model types were used to quantify the relationship between load (expressed as a percentage of the 1-RM) as the dependent variable, and RTF as the independent variable:

$$Lin: load \sim Normal\ (\boldsymbol{a} + \boldsymbol{b} \cdot RTF,\ \sigma^2) \tag{1}$$

$$Ex2: load \sim Normal\ (\boldsymbol{a} \cdot e^{(\boldsymbol{b} \cdot RTF)},\ \sigma^2) \tag{2}$$

$$Ex3: load \sim Normal\ (\boldsymbol{c} + \boldsymbol{a} \cdot e^{(\boldsymbol{b} \cdot RTF)},\ \sigma^2) \tag{3}$$

$$Crit: load \sim Normal(\boldsymbol{L}'/(RTF - \boldsymbol{k}) + \boldsymbol{CL},\ \sigma^2) \tag{4}$$

The linear model (*Lin*, Equation (1)) describes the relationship as a simple 2-parameter linear regression, which has been assumed to be a convenient approximation at a high-load range (Reynolds et al., 2006; Brzycki, 1993; Haff & Triplett, 2016; Adams & Beam, 2014). The model contains an additive intercept parameter ***a*** and a slope coefficient ***b***. The exponential regression model (*Ex2*, Equation (2), and *Ex3*, Equation (3)) describes a curvilinear relationship between variables. Previous research predominantly advocated exponential models in the form of Equation (3), featuring a multiplicative parameter ***a***, an exponential curvature parameter ***b*** and an additive parameter ***c*** (Reynolds et al., 2006; Mayhew et al., 1992; Sakamoto & Sinclair, 2006; Desgorces et al., 2010). However, we also included Equation (2) as a simplified version of Equation (3) that omits the additive parameter ***c*** (8). Ultimately, Equation (2) can be rearranged to a simple linear regression model by applying a natural log transformation to the dependent variable, making the model easily computable. Finally, the critical load model (*Crit*, Equation (4)) entails a hyperbolic relationship between the variables (Bergstrom et al., 2021; Morton et al., 2014). The model comprises a curvature parameter ***L'***, a vertical asymptote parameter ***k*** and a horizontal asymptote parameter ***CL***. An illustrative description of model functions is provided in Figure 1.

2.6. Model fitting

Each model type (Equations (1) to (4)) was fitted according to two different model structures: first, a complete-pooling model (CPM) was calculated, including all data and containing only fixed effects, therefore not accounting for interindividual differences. Second, a multilevel model (MLM) was calculated, which, in addition to fixed effects, also added random effects for each subject. This implies that for the multilevel models, every subject-level parameter was fitted to the data of a single participant, assuming a higher-level group distribution of the respective parameter. For instance, it was assumed that a generic subject-level parameter "x" was drawn from a group-level normal distribution with the mean "μ_x" and variance "σ_x^2", namely, x ~ *Normal* (μ_x,

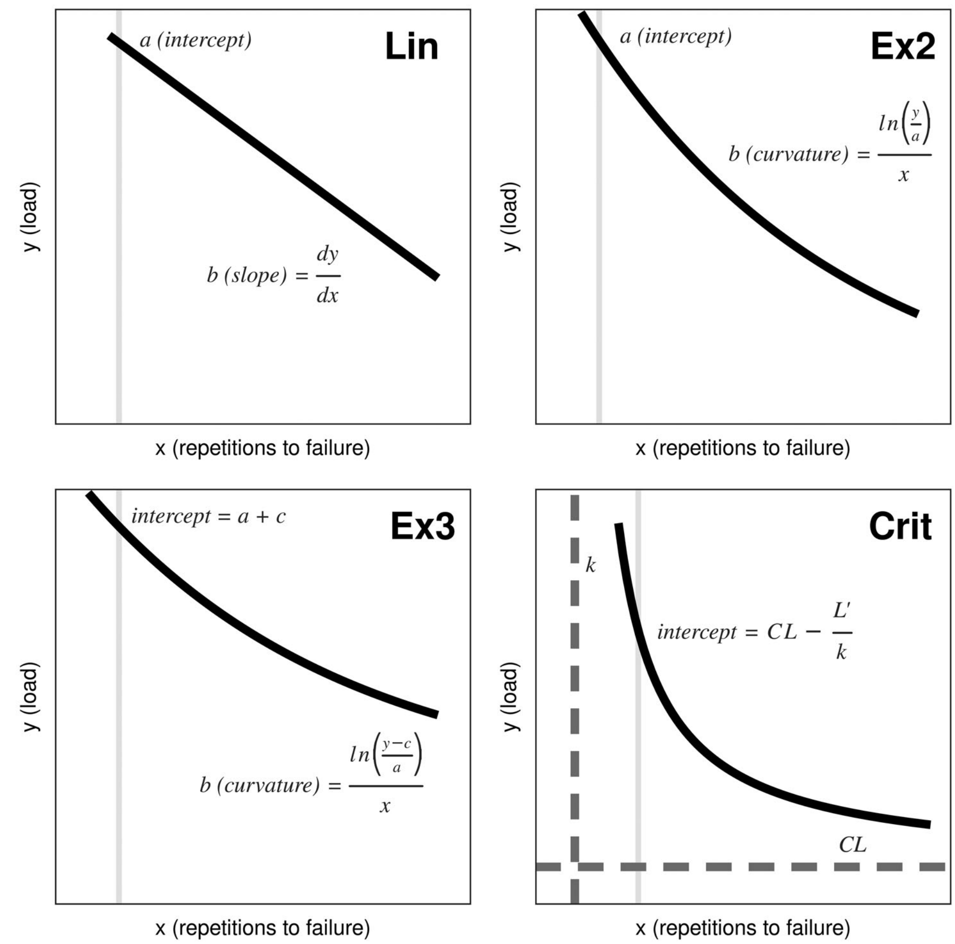

Figure 1. Illustrative examples of the investigated model types and description of model parameters. Solid black lines display model functions (extended slightly below 0); solid grey lines display the y-axis at x = 0; dashed grey lines display the vertical (k) and horizontal asymptote (CL) of the critical load model; intercepts mark the intersection of the model function and the y-axis at x = 0; Lin, linear regression model; Ex2, exponential 2-parameter regression model; Ex3, exponential 3-parameter regression model; Crit, critical load model.

σ_x^2). The possibility of correlated parameter structures within every multilevel model was accounted for by introducing a covariance matrix for the respective model's subject-level parameters. Therefore, eight different models were fitted that differed in model type (*Lin*, *Ex2*, *Ex3*, *Crit*) and structure (CPM, MLM).

Data analysis was conducted following a Bayesian approach, using the probabilistic programming language Stan (Carpenter et al., 2017), version 2.21.0, to estimate parameter distributions. Weakly informative priors were selected for variance parameters and the covariance matrix. Priors for the group-level parameters (fixed effects) of each model were defined by moment-matching a normal distribution to the posteriors of a preceding pilot study done on a separate sample of eight subjects. A prior sensitivity analysis was conducted to identify an appropriate scaling factor that would mitigate the influence of priors on posterior distributions, thus ensuring that pilot-derived priors were minimally informative. Further details on the pilot sample, prior selection and the sensitivity analysis are provided online (Supplemental digital material 1). Furthermore, sampling details and Stan codes are available online to enhance analytical reproducibility (Supplemental digital material 2).

2.7. Model evaluation

Models were compared in terms of model fit and model predictive accuracy. The model fit was analyzed by calculating a Bayesian R^2 distribution (Gelman, Goodrich, Gabry, & Vehtari, 2019) and interpreted according to the Maximum a Posteriori estimate (MAP) and the 90% Highest Density Interval (HDI) (Makowski, Ben-Shachar, & Lüdecke, 2019). Differences between R^2 posterior distributions were analyzed and interpreted according to their probability density overlap ($\cap R^2$) and deemed "substantial" for $\cap R^2 \leq 5\%$. Model predictive accuracy was evaluated by calculating the expected log predictive density and converting it into a measure of deviance labelled LOOIC (Vehtari, Gelman, & Gabry, 2017), whereas smaller values of LOOIC indicate higher predictive validity. Differences in LOOIC between models (ΔLOOIC) were complemented with an estimated standard error of difference (SE) (Vehtari et al., 2017) and considered to be substantial if they exceeded 4x the SE. In cases of model comparisons not indicating substantial differences in model fit or predictive accuracy, models were further evaluated according to their simplicity. Under respective circumstances, the logical principle of Occam's razor advocates that models with fewer parameters should be considered as more efficient. Posterior analysis was completed using R version 4.0.5 and the R packages *bayestestR* and *loo*.

3. Results

In all cases, the multilevel model resulted in a substantially better model fit compared to their complete-pooling counterpart ($\cap R^2$ < 0.1% for all comparisons). *Ex3* provided the highest R^2 among complete-pooling models, being substantially different from *Lin* ($\cap R^2$ < 0.1%), but not from *Ex2* ($\cap R^2 = 6.5\%$) and *Crit* ($\cap R^2 = 88.5\%$). The multilevel variant of *Crit* showed the best overall model fit, albeit not being substantially different from other multilevel models ($\cap R^2 = 13.6$–90.6%). Posterior distributions for R^2 are displayed in Figure 2.

Every multilevel model further provided a substantially higher predictive accuracy when compared to its complete-pooling counterpart. *Ex3* resulted in the lowest LOOIC among complete-pooling models, indicating a substantial difference from *Lin* (ΔLOOIC ± SE = 49.5 ± 9.5), but not from *Ex2* (ΔLOOIC ± SE = 26.6 ± 7.0) and *Crit* (ΔLOOIC ± SE = 1.9 ± 1.5). Across multilevel models, *Ex3* provided the highest predictive accuracy, although LOOIC was not substantially different from *Lin* (ΔLOOIC ± SE = 41.8 ± 14.4), *Ex2* (ΔLOOIC ± SE = 2.7 ± 7.8) and *Crit* (ΔLOOIC ± SE = 6.5 ± 2.2). Overall, the multilevel variant of *Ex2* emerged the most efficient model (Figure 3) due to its distinct similarity to the multilevel variants of *Ex3* and *Crit* in terms of model fit and predictive accuracy, while relying on fewer parameters. Furthermore, it yielded substantially better predictive accuracy (ΔLOOIC ± SE = 39.0 ± 7.4) compared to the multilevel variant of *Lin*. Statistics for model evaluation are summarised in Table 2.

Posterior distributions of group-level parameters (fixed effects) and subject-level parameters (random effects) calculated for the multilevel *Ex2* model are displayed in Figure 4. Group-level parameters were estimated at 102.76 (90% HDI = [102.24, 103.29]) for the intercept ***a*** and at −0.032 (90% HDI = [−0.034, −0.030]) for the curvature parameter ***b***. Subject-level parameters yielded homogeneous estimates for the intercept (between-subject coefficient of variation [90% HDI] = 0.1% [0.0, 0.9]), but considerable variance for the curvature parameter (between-subject coefficient of variation [90% HDI] = 19.7% [15.4, 25.9]).

4. Discussion

The objective of the present study was to investigate the relationship between external load and the RTF in the pin press exercise. In contrast to the greater part of published research on the topic, we did not confine our analysis to a single proposed model, but rather included several previously documented models to address two major issues: first, we aimed to determine whether a modelling approach that expresses individual relationships with higher-level commonalities (i.e. a multilevel model structure) offers substantial advantages in comparison to the traditional modelling approach that pools data without differentiation between subjects. Second, we compared four different models (Equations (1) to (4)) to identify which one provides the best approximation to the relationship in terms of model fit and predictive accuracy. Analysis was conducted using a sampling-based Bayesian method, which is considered helpful in situations with relatively small samples (Lee & Song, 2004). In addition, Bayesian methods allow the inclusion of prior information into the parameter estimation process, which may be beneficial to a priori rule out improbable values, given that adequate prior knowledge about parameters is available. Our findings yield further insight into latent structures of the strength-endurance continuum and provide practitioners with a novel and more accurate approach to calculate loads corresponding to a given repetition maximum.

4.1. Multilevel vs. complete-pooling models

To the best of our knowledge, this was the first investigation to compare pooled data modelling on the relationship between relative load and the RTF to a multilevel approach that specifies parameter expressions on an individual level. Complete-pooling model structures demonstrated both, a worse model fit and lower predictive accuracy compared to multilevel model structures, which may be attributed to noticeable variance of the RTF at lower relative loads (Figure 4). These results support the assumption that traditionally communicated models deploying only group-level parameters

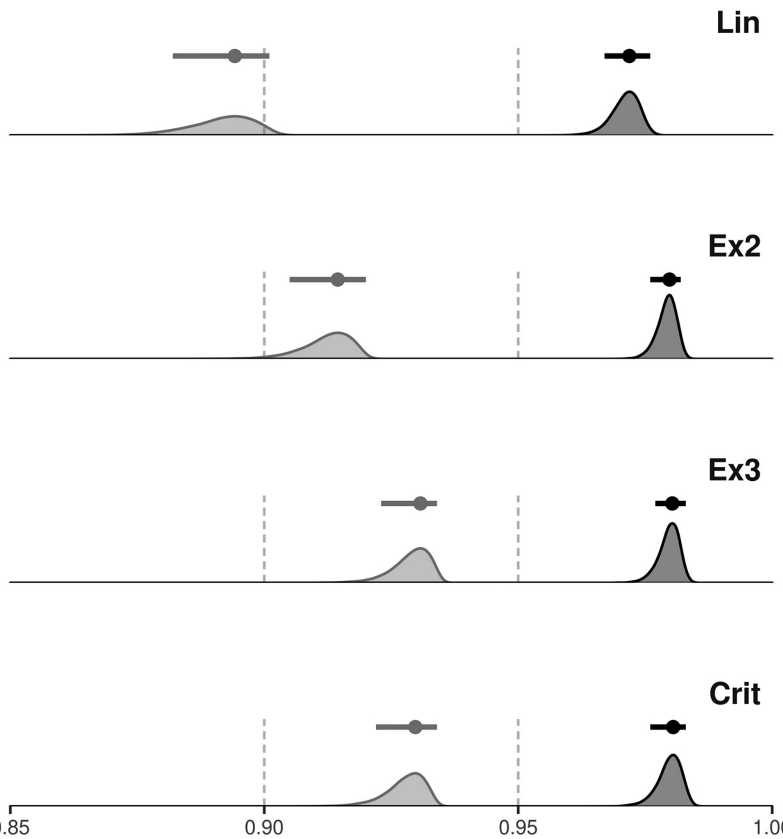

Figure 2. Comparison of model fit (R^2 posterior distributions). Dark grey distributions illustrate multilevel models; light grey distributions illustrate complete-pooling models; points represent maximum a posteriori (MAP) estimates; error bars display 90% highest density intervals (HDIs). Lin, linear regression model; Ex2, exponential 2-parameter regression model; Ex3, exponential 3-parameter regression model; Crit, critical load model.

(e.g. Adams & Beam, 2014; Brzycki, 1993; Desgorces et al., 2010; Mayhew, Ball, Arnold, & Bowen, 1992; Reynolds, Gordon, & Robergs, 2006; Sakamoto & Sinclair, 2006) do not sufficiently account for interindividual variation in the RTF that can be performed at a given relative load. Practitioners who apply respective models drawn from literature should be conscious of a potential estimation error, especially at lighter loads. Improved predictive accuracy can be expected by modelling the relationship between load and RTF on an individual level based on subject-specific data. However, application of this concept requires practitioners to assess the RTF at multiple different loads under comparable psycho-physiological conditions.

4.2. Linear vs. exponential vs. critical load models

Based upon our findings, the strength-endurance continuum appears to follow a curvilinear trend at loads of 70% 1-RM and higher, which can be modelled effectively using an exponential regression or the critical load model. The results are in accordance with earlier publications comparing linear to 3-parameter exponential regression models, whereas authors reported a better across-subject fit for the nonlinear model, as indicated by the variance explained (R^2) and standard error of estimate (Reynolds et al., 2006; Desgorces et al., 2010). In the present study, the 3-parameter exponential model (Equation (3)), that has been previously proposed on numerous occasions (Reynolds et al., 2006; Mayhew et al., 1992; Sakamoto & Sinclair, 2006; Desgorces et al., 2010) showed a slightly better model fit and predictive accuracy than its 2-parameter alternative (Equation (2)) for the pooled-data fit, although the difference was not deemed substantial. In case of the multilevel fit, Equation (2) resulted in exceptionally similar estimates of R^2 and LOOIC. Despite our analysis not showing a statistical advantage of the 2-parameter exponential regression model, it exceeds both the 3-parameter exponential regression model and the critical load model in terms of simplicity, as indicated by the number of model parameters. Therefore, we endorse that applying the 2-parameter exponential regression model to subject-specific data yields the best representation of a person's strength-endurance relationship, without adding unnecessary complexity to the model.

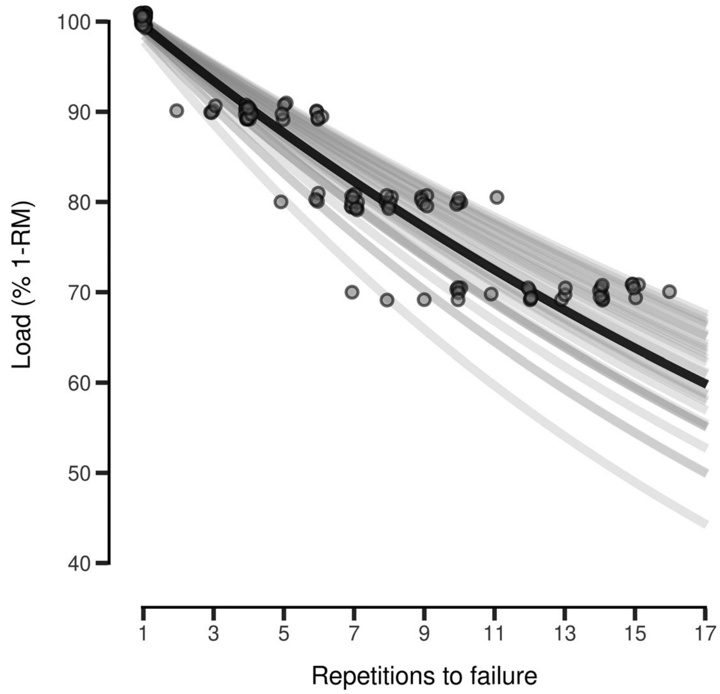

Figure 3. The strength-endurance relationship represented by the multilevel 2-parameter exponential regression model. Points represent subject data (jittered illustration); solid black line displays the group-level model; grey lines display subject-level models. 1-RM, one-repetition maximum.

4.3. Parameter analysis

Subject-level intercepts of the 2-parameter exponential regression model (parameter ***a*** in Equation (2)) only showed a small deviation from the group-level parameter, which may be attributed to load being normalised to the individual 1-RM. However, the curvature parameter (***b*** in Equation (2)) showed considerable variation between subjects, suggesting that it may constitute the main influence on the individual manifestation of the strength-endurance relationship. Therefore, future studies should consider employing a comprehensive analysis on potential confounders that may affect estimates of the curvature parameter. While an additional evaluation of model parameters was beyond the scope of the present study, an exploratory analysis of subject characteristics and their effect on subject-level parameters is provided online for interested readers (Supplemental digital material 3).

Table 2. Comparison of 8 models, ranked from best to worst model performance.

Rank	Model	n_P_g (n_P_s)	ΔLOOIC	SE	R^2 MAP [90% HDI]
1	Ex2 [MLM]	2 (60)	0.0	0.0	0.980 [0.976; 0.982]
2	Ex3 [MLM]	3 (90)	−2.7	8.0	0.980 [0.977; 0.983]
3	Crit [MLM]	3 (90)	3.8	7.3	0.981 [0.976; 0.983]
4	Lin [MLM]	2 (60)	39.0*	7.4	0.972 [0.967; 0.976]
5	Ex3 [CPM]	3 (0)	104.4*	17.7	0.931 [0.923; 0.934]
6	Crit [CPM]	3 (0)	106.3*	17.8	0.930 [0.922; 0.934]
7	Ex2 [CPM]	2 (0)	131.0*	18.0	0.915 [0.905; 0.920]
8	Lin [CPM]	2 (0)	153.9*	16.9	0.894 [0.882; 0.901]

Note: Models are ranked according to their fit, predictive accuracy and simplicity. n_P_g, number of group level parameters (fixed effects); n_P_s, number of subject level parameters (random effects); ΔLOOIC, difference in LOOIC compared to the most efficient model Ex2 [MLM] (lower values indicating better predictive performance, * indicating a substantial difference); SE, standard error of the difference in LOOIC; R^2, variance explained; MAP, Maximum a Posteriori estimate; HDI, Highest Density Interval. Ex2, exponential 2-parameter regression model; Ex3, exponential 3-parameter regression model; Crit, critical load model; Lin, linear regression model; MLM, multilevel structure; CPM, complete-pooling structure.

4.4. Limitations

Readers should consider that the present study investigated the strength-endurance relationship only in the specific case of the pin press using a highly controlled exercise technique without standardising movement cadence. Hence, the multilevel 2-parameter exponential regression may not necessarily provide the best approximation for other exercises that follow a different distribution of the RTF across loads, which questions the transferability of the present findings to a standard touch-and-go bench press. Additionally, our findings only cover for relative loads of 70% 1-RM and above, therefore neglecting model validity at lower loads. Finally, models were calculated based on data acquired during two visits, whereas tests to momentary failure were exclusively conducted during the second visit without randomising the order of tests. Therefore, the possibility of an order effect influencing the acquired

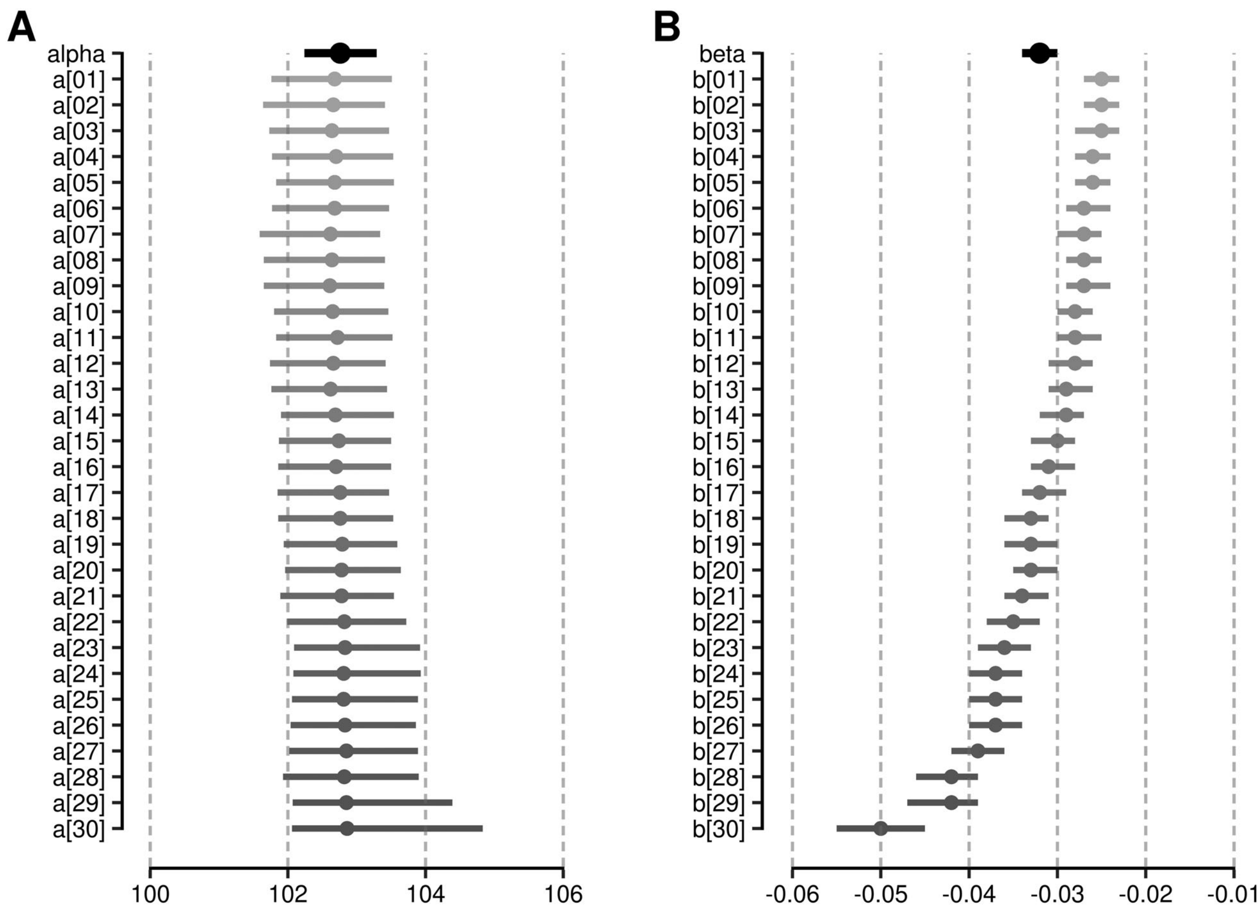

Figure 4. Posterior summary of the multilevel 2-parameter exponential regression model, including the intercept ***a*** (panel A) and curvature parameters ***b*** (panel B). Points represent maximum a posteriori (MAP) estimates; error bars display 90% highest density intervals [HDI]. Alpha, group-level parameter (fixed effect); a[i], individual-level parameter (random effect) of subject *i*.

data cannot be ruled out. While a similar approach to single-visit testing with a fixed order of trials has recently been proposed for the valid assessment of critical power (Triska et al., 2021), our data provide no conclusion whether subjects truly initiated each set to momentary failure under fully rested conditions. Future research should therefore target two important objectives: first, different methodological approaches of assessing RTF at multiple loads should be compared and it should be evaluated how they influence the estimated strength-endurance relationship on a subject-level (e.g. effects of single-visit vs. multiple-visit data acquisition). Second, the multilevel relationship between load and RTF should be investigated using a variety of exercises with less restrictive movement specifications, including the touch-and-go bench press.

5. Conclusion

The present study supplies evidence that the strength-endurance continuum, described by the relationship between relative load and the number of repetitions performed to failure, displays substantial interindividual variation. Practitioners and researchers can address this issue by modelling the relationship on an individual level, whereas the 2-parameter exponential regression evidently constitutes the most efficient model for this purpose in the pin press exercise.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported there is no funding associated with the work featured in this article.

ORCID

Benedikt Mitter http://orcid.org/0000-0001-6056-5853
Lei Zhang http://orcid.org/0000-0002-9586-595X
Pascal Bauer http://orcid.org/0000-0003-1867-2422
Arnold Baca http://orcid.org/0000-0002-1704-0290
Harald Tschan http://orcid.org/0000-0002-8848-1827

References

Adams, G. M., & Beam, W. C. (2014). *Exercise physiology laboratory manual*. New York, NY: McGraw-Hill. 338 p.

Bergstrom, H. C., Dinyer, T. K., Succi, P. J., Voskuil, C. C., & Housh, T. J. (2021). Applications of the critical power model to dynamic constant external resistance exercise: A brief review of the critical load test. *Sports (Basel)*, *9*(2), 15.

Braith, R. W., Graves, J. E., Leggett, S. H., & Pollock, M. L. (1993). Effect of training on the relationship between maximal and submaximal strength. *Medicine & Science in Sports & Exercise*, *25*(1), 132–138.

Brechue, W. F., & Mayhew, J. L. (2009). Upper-body work capacity and 1RM prediction are unaltered by increasing muscular strength in college football players. *Journal of Strength and Conditioning Research*, *23*(9), 2477–2486.

Brzycki, M. (1993). Strength testing – predicting a one-rep max from reps-to-fatigue. *Journal of Physical Education, Recreation & Dance*, *64*(1), 88–90.

Burnley, M., & Jones, A. M. (2018). Power–duration relationship: Physiology, fatigue, and the limits of human performance. *European Journal of Sport Science*, *18*(1), 1–12.

Campos, G. E. R., Luecke, T. J., Wendeln, H. K., Toma, K., Hagerman, F. C., Murray, T. F., Ragg, K., Ratamess, N., Kraemer, W., & Staron, R. S. (2002). Muscular adaptations in response to three different resistance-training regimens: Specificity of repetition maximum training zones. *European Journal of Applied Physiology*, *88*(1-2), 50–60.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1), 32.

de Salles, B. F., Simão, R., Miranda, F., Da Novaes, J. S., Lemos, A., & Willardson, J. M. (2009). Rest interval between sets in strength training. *Sports Medicine*, *39*(9), 765–777.

Desgorces, F. D., Berthelot, G., Dietrich, G., & Testa, M. S. A. (2010). Local muscular endurance and prediction of 1 repetition maximum for bench in 4 athletic populations. *Journal of Strength and Conditioning Research*, *24*(2), 394–400.

Dorrell, H. F., Moore, J. M., Smith, M. F., & Gee, T. I. (2019). Validity and reliability of a linear positional transducer across commonly practised resistance training exercises. *Journal of Sports Sciences*, *37*(1), 67–73.

Gelman, A. (2006). Multilevel (hierarchical) modeling: What it can and cannot do. *Technometrics*, *48*(3), 432–435.

Gelman, A., Goodrich, B., Gabry, J., & Vehtari, A. (2019). R-squared for Bayesian regression models. *The American Statistician*, *73*(3), 307–309.

Grgic, J., Lazinica, B., Schoenfeld, B. J., & Pedisic, Z. (2020). Test-retest reliability of the one-repetition maximum (1RM) strength assessment: A systematic review. *Sports Medicine - Open*, *6*(1), 31.

Haff, G., & Triplett, N. T. (Eds.). (2016). *Essentials of strength training and conditioning*. Champaign, IL: Human Kinetics. 735 p.

Hoeger, W. W., Hopkins, D. R., Barette, S. L., & Hale, D. F. (1990). Relationship between repetitions and selected percentages of one repetition maximum: A comparison between untrained and trained males and females. *Journal of Strength and Conditioning Research*, *4*(2), 47–54.

LaChance, P. F., & Hortobagyi, T. (1994). Influence of cadence on muscular performance during push-up and pull-up exercise. *Journal of Strength and Conditioning Research*, *8*(2), 76–79.

Lawton, T. W., Cronin, J. B., & McGuigan, M. R. (2011). Strength testing and training of rowers: A review. *Sports Medicine*, *41*(5), 413–432.

Lee, S.-Y., & Song, X.-Y. (2004). Evaluation of the Bayesian and maximum likelihood approaches in analyzing structural equation models with small sample sizes. *Multivariate Behavioral Research*, *39*(4), 653–686.

LeSuer, D. A., McCormick, J. H., Mayhew, J. L., Wasserstein, R. L., & Arnold, M. D. (1997). The accuracy of prediction equations for estimating 1-RM Performance in the bench press, squat, and deadlift. *Journal of Strength and Conditioning Research*, *11*(4), 211–213.

Makowski, D., Ben-Shachar, M., & Lüdecke, D. (2019). Bayestestr: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software*, *4*(40), 1541.

Mayhew, J. L., Ball, T. E., Arnold, M. D., & Bowen, J. C. (1992). Relative muscular endurance performance as a predictor of bench press strength in college men and women. *Journal of Strength and Conditioning Research*, *6*(4), 200–206.

Mayhew, J. L., Johnson, B. D., Lamonte, M. J., Lauber, D., & Kemmler, W. (2008). Accuracy of prediction equations for determining one repetition maximum bench press in women before and after resistance training. *Journal of Strength and Conditioning Research*, *22*(5), 1570–1577.

Mitter, B., Hölbling, D., Bauer, P., Stöckl, M., Baca, A., & Tschan, H. (2021). Concurrent validity of field-based diagnostic technology monitoring movement velocity in powerlifting exercises. *Journal of Strength and Conditioning Research*, *35*(8), 2170–2178.

Morton, R. H., Redstone, M. D., & Laing, D. J. (2014). The critical power concept and bench press: Modeling 1RM and repetitions to failure. *International Journal of Exercise Science*, *7*(2), 152–160.

Orange, S. T., Metcalfe, J. W., Marshall, P., Vince, R. V., Madden, L. A., & Liefeith, A. (2020). Test-retest reliability of a commercial linear position transducer (GymAware PowerTool) to measure velocity and power in the back squat and bench press. *Journal of Strength and Conditioning Research*, *34*(3), 728–737.

Reynolds, J. M., Gordon, T. J., & Robergs, R. A. (2006). Prediction of one repetition maximum strength from multiple repetition maximum testing and anthropometry. *Journal of Strength and Conditioning Research*, *20*(3), 584–592.

Richens, B., & Cleather, D. J. (2014). The relationship between the number of repetitions performed at given intensities is different in endurance and strength trained athletes. *Biology of Sport*, *31*(2), 157–161.

Sakamoto, A., & Sinclair, P. J. (2006). Effect of movement velocity on the relationship between training load and the number of repetitions of bench press. *Journal of Strength and Conditioning Research*, *20*(3), 523–527.

Sánchez-Medina, L., & González-Badillo, J. J. (2011). Velocity loss as an indicator of neuromuscular fatigue during resistance training. *Medicine & Science in Sports & Exercise*, *43*(9), 1725–1734.

Sánchez-Medina, L., Perez, C. E., & Gonzalez-Badillo, J. J. (2010). Importance of the propulsive phase in strength assessment. *International Journal of Sports Medicine*, *31*(2), 123–129.

Scott, B. R., Duthie, G. M., Thornton, H. R., & Dascombe, B. J. (2016). Training monitoring for resistance exercise: Theory and applications. *Sports Medicine*, *46*(5), 687–698.

Steele, J., Fisher, J., Giessing, J., & Gentil, P. (2017). Clarity in reporting terminology and definitions of set endpoints in resistance training. *Muscle & Nerve*, *56*(3), 368–374.

Suchomel, T. J., Nimphius, S., Bellon, C. R., Hornsby, W. G., & Stone, M. H. (2021). Training for muscular strength: Methods for monitoring and adjusting training intensity. *Sports Medicine*, *51*(10), 2051–2066.

Triska, C., Hopker, J., Wessner, B., Reif, A., Tschan, H., & Karsten, B. (2021). A 30-Min rest protocol does not affect W′, Critical power, and systemic response. *Medicine & Science in Sports & Exercise*, *53*(2), 404–412.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432.

Ware, J. S., Clemens, C. T., Mayhew, J. L., & Johnston, T. J. (1995). Muscular endurance repetitions to predict bench press and squat strength in college football players. *Journal of Strength and Conditioning Research*, *9*(2), 99–103.

Weakley, J., Mann, B., Banyard, H. G., McLaren, S., Scott, T., & Garcia-Ramos, A. (2021). Velocity-based training: From theory to application. *Strength & Conditioning Journal*, *43*(2), 31–49.