SportRxiv

PREPRINT - NOT PEER REVIEWED

The effects of supervision upon effort during resistance training: A Bayesian analysis of prior data and an experimental study of private strength clinic members

James Steele[1,2,3*], James Fisher[3,4], Tim Dettmann[1]

[1]Kieser Australia, Melbourne, Australia
[2]Steele Research Limited, Eastleigh, UK
[3]Department of Sport and Health, Solent University, Southampton, UK
[4]The Exercise Coach, Tennessee, USA

*Correspondence: james@steele-research.com

Cite as: James Steele et al. (2025). The effects of supervision upon effort during resistance training: A Bayesian analysis of prior data and an experimental study of private strength clinic members. *SportRxiv.* DOI: 10.51224/SRXIV.521

Supplementary Materials: https://osf.io/etb34/

Abstract

Supervision during resistance training (RT) may enhance strength gains by optimizing trainee effort. We investigated supervision's role in effort during RT in a unique setting with private strength clinics, where members train either unsupervised ("Core" membership) or supervised by a qualified exercise scientist ("Assisted" membership). Using both retrospective analysis of member training records and a prospective experimental study, we examined supervision's impact on exercise performance, measured as time under load (TUL), rating of perceived effort (RPE), and rating of perceived discomfort (RPD). Bayesian methods were applied, using empirically informed prior distributions from retrospective data to model the experimental study. The prior sample included ~1000 members training sessions from each membership type, while the experimental study involved 45 Core members performing both supervised and unsupervised sessions in randomized order, using their current training loads to momentary failure. Our findings suggest that, in real-world settings (in situ), exercise performance differed little between supervised and unsupervised training. However, in our experimental study, supervision improved TUL (Core = 125.12 [95%QI: 113.70, 131.90] sec; Assisted = 147.35 [95%QI: 134.29, 154.81] sec; contrast = -22.10 [95%QI: -26.60, -17.61] sec). In percentage points RPE was slightly higher with supervision in both prior real-world (Core = 53% [95%QI: 51%, 55%]; Assisted = 59% [95%QI: 57%, 61%]; contrast = -6% [95%QI: -8%, -4%]) and experimental settings (Core = 81% [95%QI: 75%, 86%]; Assisted = 87% [95%QI: 83%, 91%]; contrast = -6% [95%QI: -10%, -4%]), suggesting trainees push closer to failure under supervision. This was further supported by higher RPD during the experimental study (Core = 6.3 [95%QI: 5.1, 7.3]; Assisted = 7.5 [95%QI: 6.5, 8.3]; contrast = -1.2 [95%QI: -1.6, -0.9]). Overall, these results reinforce prior research on the benefits of supervision in RT, indicating that unsupervised trainees—especially in real-world conditions—likely train with suboptimal effort. Keywords: *supervision; effort; resistance training; bayesian*

All authors have read and approved this version of the manuscript.
The manuscript was last updated on February 18, 2025

Introduction

Recent work has highlighted the importance of supervision in optimising strength outcomes from resistance training (RT). In a recent systematic review and meta-analysis from Fisher et al. (2022) there was a moderate standardised mean effect (0.40 [95%CI: 0.06, 0.74]) of supervised vs unsupervised RT on strength outcomes synthesised from ten studies. Since then, a further study has reported that even in previously trained participants there may be benefit to strength outcomes from supervision (Coleman et al., 2023).

It has been speculated (Fisher et al., 2022) that an explanation for this effect may be due to the role that a trainer plays in prescribing load progression for the trainee, whereas when there is no clear load progression rule unsupervised trainees may be less likely to train with appropriate loads, and ultimately appropriate effort. Indeed, a recent meta-analysis examining the relative loads selected by trainees when given the ability to self-select highlights that they tend to choose loads which, whilst initially efficacious in novice trainees, become sub-optimal quickly as training experience progresses (~53% of one repetition maximum [1RM]) particularly when combined with the typical repetition ranges prescribed i.e., ~8 to 15 repetitions (Steele et al., 2022). Further, when given the opportunity to self-select load and the number of repetitions to complete there is evidence that trainees likely train with relatively low effort as determined from their proximity to momentary failure i.e., they select ~10 repetitions at ~53% 1RM (Nuzzo et al., 2024; Steele et al., 2022). Thus, whether self-selecting a load for a typically prescribed repetition range or self-selecting a load and self-selecting the number of repetitions to perform, most will naturally perform sets with an estimated proximity of ~10-20 repetitions shy of momentary failure.

As such it has also been suggested that a supervising trainer plays the role of providing motivation and enhancing trainee effort (Fisher et al., 2022). Indeed, recent survey studies highlight that trainees perceive supervision to have an important role in determining their motivation and resultant effort during RT which they also perceive to be important to achieve their training goals (Carlson et al., 2024; Fisher et al., 2023).

However, a recent meta-analysis of the dose-response relationship of proximity to momentary failure for strength and hypertrophy outcomes highlights that, while there is increased hypertrophy with closer proximity to failure, there is not a clear relationship for strength (Robinson et al., 2024). But a caveat is that there was limited data for proximities to failure >10 repetitions, and that the models were adjusted for load (average loads were typically ~75-85% of one repetition maximum). Thus, the results of this meta-analysis apply to proximity to failure *after* selecting load (i.e., intraset effort) and a higher relative load requires a greater effort all else being equal. When considering the average self-selected load of ~53% 1RM coupled with a typical ~10 repetitions per set equating to proximities to failure >10 repetitions (Steele et al., 2022) it seems plausible that unsupervised trainees may train with a less-than-optimal load, repetitions, and resultant effort to optimise strength outcomes. Ultimately, and in combination, supervision should provide trainees with feedback of their past performance, whether technical, effort- or program-based, and guidance toward their future performance (Fisher, under review).

We were afforded an opportunity to examine the role of supervision on effort during RT in a unique setting with several private strength clinics whose members train either unsupervised on a "Core" membership[1] or supervised by an exercise scientist on an "Assisted" membership both retrospectively and through a prospective experimental study. As such, using Bayesian methods, we examined samples of historical data from both types of member and in a sample of current Core members we investigated experimentally the impact of supervision from a qualified exercise scientist.

Methods

Experimental approach to the problem

The study was conducted at a selection of strength clinics operated by Kieser Australia recruiting from the existing pool of Core members at these locations. All Kieser Australia members are prescribed the same protocol. This consists of a single set of the resistance machine exercises prescribed on their current training programme card using a load that should permit them to reach momentary failure within a time-under-load (TUL) of 90-120 seconds (though an upper limit of 180 seconds TUL is enforced to avoid machines being occupied for too long on the clinic floor preventing other members from using them) using a ~ 12 seconds repetition duration (i.e., ~4:4 seconds concentric:eccentric actions with a 2 isometric second hold whilst still under load with tension on the involved musculature at the end of each concentric and eccentric muscle action). Core trainees are prompted to progress load by ~5% for the next session once a TUL of >120 seconds is achieved for a particular exercise before reaching momentary failure, and for Assisted trainees this is actioned by their supervising Exercise Scientist. Trainees are prompted to complete a session rating of perceived effort (RPE) at the end of their training session. An acute randomised cross-over design was employed to examine participants completing an RT session with and without direct supervision with one week apart (i.e., a Core vs Assisted session). Participants were instructed to complete a single set of the resistance machine exercises on their current training programme card using their current training loads to momentary failure. The TUL performed, and ratings of perceptions of effort (RPE) and discomfort (RPD) were recorded and compared between both conditions. This study was not pre-registered and as described below the sample size was justified based on logistical concerns and the analysis is considered exploratory.

In addition to the experimental study design noted above, we also took samples of Core and Assisted trainees from historical data records in Kieser Australia's database. We took separate random samples to obtain both TUL and session RPE data as not all members record their RPE. There were two reasons for examining these historical samples. Firstly, we used this data to generate empirical informative prior distributions to use in our Bayesian modelling of the experimental study data. Secondly, they allowed us to also explore the descriptive[2]

[1] Note, Core members do typically have a "Review" session every ~3 months with an exercise scientist to complete strength testing, and review their training programme cards including selection of exercises and weights.

[2] Note that we do not claim this difference to be an estimated causal effect in this data given that we have not adjusted for confounders; for example, that different types of people might self-select into either Core or Assisted memberships and that this might confound comparisons of either TUL or session RPE.

difference between Core and Assisted training *in situ*, but also that we could compare the prior distribution for the random samples of members *in situ* with the posterior distribution of members completing their sessions in the experimental study to understand the extent to which study effects might also influence outcomes.

The experimental study component of this project was approved by the Southampton Solent University Health Exercise and Sport Science Ethics committee (id: STEELEAUG2016). All participants provided informed consent to participate in the study. Data in the prior sample was used in de-identified form and with member consent for data to be used for research purposes.

Participants

Experimental

The study was advertised at five Kieser Australia strength clinics and we sought to recruit members who had > 6 months prior training experience at the strength clinics, were healthy (no clinical conditions on their member record as recorded by a physiotherapist or exercise physiologist), both males and females (age +18 years), and without any current condition for which RT would be contraindicated. They were recruited from the existing member pool of Core members. We recruited a sample size based on considerations of what was logistically feasible given the availability of the exercise scientist staff at each clinic to conduct data collection and perform the supervised Assisted sessions and wanting to minimise the burden on day-to-day operation of the clinics. As such it was decided that a target of 50 participants across clinics would be acceptable. We ended up with n = 45 participants recruited across the five clinics.

Prior Sample

We queried the Kieser Australia database to generate random samples from historical data. We opted to generate reasonably large samples to ensure precise prior estimates, though did not use the entirety of the historical data (~50000 members, >10 years of data) so as to reduce computation time for Bayesian modelling. For TUL we limited data to Core and Assisted members training sessions that either were not led by an exercise scientist or were respectively[3], took the first training session after at least 6 months of prior training at Kieser Australia had been completed by each member, randomly sampled 1000 Core and 1000 Assisted members and then filtered to the resistance machines used in by members in the experimental study so that we had a selection of members across varied clinic locations and completing sessions with a selection of resistance machine exercises and had TUL data for each exercise. Thus we ended up with a sample of n = 1751 (Core = 849, Assisted = 902). For session RPE we performed a similar database query and randomly sampled 1000 Core and 1000 Assisted members who had reported session RPE values.

[3]Whilst members can have Core or Assisted memberships the former also have sessions that are supervised (i.e., their "Review" sessions noted in footnote[1]) and the latter can have unsupervised sessions (i.e., they can attend the clinic outside of their scheduled sessions with the exercise scientist and train unsupervised). Thus we limited our sample to only normal training sessions which had been either unsupervised or supervised for both Core and Assisted members respectively.

Protocols

All RT was performed using resistance machines (Kieser Training AG, Zurich, Switzerland) including the A1 (hip extension), A2 (torso flexion), A3 (hip abduction), B1 (leg extension), B6 (leg press), B7 (seated leg curl), C3 (torso arm i.e., pulldown), C7 (seated row), D5 (arm cross i.e., pectoral fly), D6 (chest press), D7 (seated dip), F2/F2.1 (abdominal flexion), and K2 (supported supinated grip pullup). Note, not all participants performed exercise using each machine, but had at least some of these machines in their current training cards. For the experimental study participants were instructed to attend two training sessions at least one-week apart where they completed either their current Core session (i.e., unsupervised) or an Assisted session (i.e., supervised). During the Core session the participant completed their training session as prescribed with the only difference for the Assisted session being that an exercise scientist supervised them providing instruction and motivation; however, they did not interfere in any physical way during the set (e.g. to spot or assist in completing a failing repetition). Each participant had an existing training programme card. In both conditions they utilised the current training loads for each machine that were recorded in the last session of their training programme card prior to participation in the study. Participants were instructed to ensure that they continue performing the exercise to momentary failure independently of the TUL achieved for the purposes of the study i.e., if they realised during the set that their prior load selection was evidently too low for the prescribed TUL range they should continue to momentary failure regardless off the TUL acheived. Momentary failure was defined as per Steele et al. (2017) i.e., the point at which, despite their greatest effort, participants are unable to continue concentrically contracting and moving the resistance, and this was clearly communicated to the participants. During each session for each exercise the trainees/trainers recorded their TUL achieved as they would do for their usual sessions using timers situated around the clinic in view during training specifically for this purpose. In addition, they recorded their RPE and RPD in that order immediately upon completing the exercise using previously validated scales for differentiating these perceptions (Steele, Fisher, McKinnon, et al., 2017); scripts and scales are available here: https://osf.io/ufvy8/.

The data generated from the sample of historical members was generated according to the standard protocol described above that all Kieser Australia members are prescribed. TUL was also recorded similarly using timers available about the clinic, or for some sessions using a more recently developed mobile phone application (Kieser Konnect, Kieser Australia) which is placed on a stand on the resistance machines and used to track TUL for each exercise in their training card which is programmed to the application. Session RPE was recorded using either the mobile application for Core members, or via the application used by staff during Assisted sessions and used a modified version of the Borg 6-20 scale where the scale is displayed as scrollable and the verbal anchors are “6 - None”, “7- Very, very light”, “13 – Somewhat hard”, “19 - Very, very hard”, and “20 – Maximal Exertion”.

Statistical Analysis

All code utilised for data preparation and analyses are available in either the Open Science Framework page for this project https://osf.io/etb34/ or the corresponding GitHub repository https://github.com/jamessteeleii/supervision_tul. We cite all software and packages used in the analysis pipeline using the `grateful` package (Rodriguez-Sanchez et al., 2023) which can be seen here: https://osf.io/ew79g.

As noted, the project was not pre-registered but involved exploratory analysis of the experimental and prior datasets. All analyses have been conducted within a Bayesian framework and all posterior estimates and their precision, along with conclusions based upon them, are interpreted continuously and probabilistically, considering priors, data quality, and all within the context of each outcome and the assumptions of the model employed as the estimator (Kruschke & Liddell, 2018). All models were run with 2000 warmup and 2000 sampling iterations and four Monte Carlo Markov Chains. Trace plots were produced along with $\hat{R}$ values to examine whether chains had converged, and posterior predictive checks for each model were also examined to understand the model implied distributions (see https://osf.io/u7g8c. We fit two sets of models for the TUL and RPE outcomes, one on the prior sample of data using weakly regularising priors in order to generate posterior distributions to inform the priors used in the other model using the experimental data, and a single model on the experimental data for the RPD outcome using weakly regularising priors. These models are described below. For all models we calculated average marginal effects as the global grand means for both predictions under each condition as well as contrasts between conditions and visualise both prior and posterior distributions along with median and 95% quantile interval (QI) estimates from these.

Time Under Load Analysis

Upon inspecting the raw distribution of the prior sample of data for TUL we noticed that there was a spike at 120 seconds concomitant with the top of the target TUL range prescribed to members under both Core and Assisted conditions (see Figure 1). Anecdotal reports from Kieser Australia staff had also prior to exploring the data suggested to us that, despite the prescription to train to momentary failure, many Core members were instead selecting loads and training only to the upper 120 second TUL range threshold irrespective of proximity to failure. We did not however expect this to be the same for the prior sample of Assisted members. As such, it suggested to us that there might be two processes underlying the observed data: a tendency to target a specific TUL threshold and to stop the exercise at that point, or to otherwise continue the exercise to momentary failure (or to stop the exercise based on some other process such as tolerable proximity to failure). Thus, we opted to use a hurdle type model comprising a Bernoulli distribution for the probability of having a TUL of 120 seconds, and a student t distribution for all other values of TUL[4]. We expected this may also be the case in our experimental dataset given we recruited Core members and so utilised the same type of model for this. As such this allowed us to examine both the expectations of

[4]A custom family for the hurdle-student t model was produced for use with the `brms` R package, adapted from the hurdle-normal distribution developed by Heiss (n.d.).

the global grand mean for TUL, in addition to the probabilities of having a TUL of 120 seconds.

For both the prior sample and experimental study data the following model was employed which included a fixed (i.e., population level) effect for the session type, Core or Assisted, where Assisted was coded as the intercept. The model also included random (i.e., cluster or group level) effects as intercepts for the location (i.e., the clinic the session was performed at), the member id, and the resistance machine used (note, due to the cross-over nature of the experimental design the random effects for location, member, and machine were modelled as nested in this dataset). These fixed and random effects were modelled as predictors for both the hurdle and student t components of the model. TUL was centred at zero for modelling (i.e., all values of TUL_i has 120 subtracted from them prior to modelling). Formally the model(s) for i^{th} TUL across condition c within the j^{th} location, k^{th} member, l^{th} machine were as follows:

$$
\begin{aligned}
&\text{TUL}_{ic} \sim \text{Hurdle log–student } t(\pi_{ic_{j[i],k[i],l[i]}}, \nu_y, \mu_{ic_{j[i],k[i],l[i]}}, \sigma_y) \ \ldots \text{ or alternatively,} \\
&\text{TUL}_{ic} \sim \begin{cases} 0 & \text{with probability } \pi_{ic_{j[i],k[i],l[i]}} \\ \text{Student } t(\mu_{ic_{j[i],k[i],l[i]}}, \nu_y, \sigma_y) & \text{with probability } 1 - \pi_{ic_{j[i],k[i],l[i]}} \end{cases} \\
&\textbf{Models for distribution parameters} \\
&\text{logit}(\pi_{ic}) = (\gamma_0 + \gamma_{0_{j[i],k[i],l[i]}}) + \gamma_1 \text{Condition}_{ic} && \text{120 seconds/not-120 seconds process} \\
&\mu_{ic} = (\beta_0 + b_{0_{j[i],k[i],l[i]}}) + \beta_1 \text{Condition}_{ic} && \text{Location parameter in student } t \text{ process} \\
&\gamma_{0_j} \sim \mathcal{N}(0, \sigma_{\gamma_{0_j}}), \text{ for location j} = 1, \ \ldots, \text{J} && \text{Variation in intercepts for location in hurdle model} \\
&\gamma_{0_k} \sim \mathcal{N}(0, \sigma_{\gamma_{0_k}}), \text{ for member k} = 1, \ \ldots, \text{K} && \text{Variation in intercepts for member in hurdle model} \\
&\gamma_{0_l}, \sim \mathcal{N}(0, \sigma_{\gamma_{0_l}}), \text{ for machine l} = 1, \ \ldots, \text{L} && \text{Variation in intercepts for machine in hurdle model} \\
&b_{0_j} \sim \mathcal{N}(0, \sigma_{b_{0_j}}), \text{ for location j} = 1, \ \ldots, \text{J} && \text{Variation in intercepts for location in student } t \text{ model model} \\
&b_{0_k} \sim \mathcal{N}(0, \sigma_{b_{0_k}}), \text{ for member k} = 1, \ \ldots, \text{K} && \text{Variation in intercepts for member in student } t \text{ model model} \\
&b_{0_l} \sim \mathcal{N}(0, \sigma_{b_{0_l}}), \text{ for machine l} = 1, \ \ldots, \text{L} && \text{Variation in intercepts for machine in student } t \text{ model model}
\end{aligned}
\tag{1}
$$

Weakly regularising priors were employed for prior sample dataset. These were the default priors in the `brms` package for all intercept terms which are scaled to the expected response value when all predictors are at their means and use a student t distribution with degrees of freedom $\nu = 3$, and variance terms similarly scaled with $\mu = 0$. The only change to this was to set reasonable upper and lower bounds for the prior condition effects to limit prediction of impossible TUL values. The priors were as follows:

$$
\begin{aligned}
&\textbf{Priors (prior sample data model)} \\
&\gamma_0 \sim \text{Logistic}(2, 0.1) && \text{Prior for intercept in hurdle model i.e., Assisted condition} \\
&\beta_0 \sim \text{Student } t(3, 3, 13.3) && \text{Prior for intercept in student } t \text{ model i.e., Assisted condition} \\
&\gamma_1, \beta_1 \sim \text{Student } t(3, 0, 15, \text{lb} = -60, \text{lb} = 60) && \text{Prior for Condition effects in both models i.e., Core minus Assisted condition} \\
&\sigma_y, \sigma_{\gamma_{0_j}}, \sigma_{\gamma_{0_k}}, \sigma_{\gamma_{0_l}}, \sigma_{b_{0_j}}, \sigma_{b_{0_k}}, \sigma_{b_{0_l}} \sim \text{Student } t(3, 0, 13.3, \text{lb} = 0) && \text{Prior for all variability parameters} \\
&\nu_y \sim \text{Gamma}(2, 1) && \text{Prior for student } t \text{ degrees of freedom}
\end{aligned}
\tag{2}
$$

We then took posterior parameter estimates from the prior sample model and used these to generate informative priors for the experimental data models which were as follows:

$$
\begin{aligned}
&\textbf{Priors (experimental sample data model)} \\
\gamma_0 &\sim \text{Student } t(3.28, -3.03, 0.12) && \text{Prior for intercept in hurdle model i.e., Assisted condition} \\
\beta_0 &\sim \text{Student } t(3.28, 2.79, 1.30) && \text{Prior for intercept in student } t \text{ model i.e., Assisted condition} \\
\gamma_1 &\sim \text{Student } t(3.28, -0.19, 0.11) && \text{Prior for Condition effects in hurdle model i.e., Core minus Assisted condition} \\
\beta_1 &\sim \text{Student } t(3.28, 0.07, 0.58) && \text{Prior for Condition effects in student } t \text{ model i.e., Core minus Assisted condition} \\
\sigma_y &\sim \text{Student } t(3, 10.59, 0.19, \text{lb} = 0) && \text{Prior for residual variability parameter} \\
\sigma_{\gamma_{0_j}} &\sim \text{Student } t(3, 0.31, 0.08, \text{lb} = 0) && \text{Prior for location variability parameter in hurdle model} \\
\sigma_{\gamma_{0_k}} &\sim \text{Student } t(3, 1.16, 0.07, \text{lb} = 0) && \text{Prior for member variability parameter in hurdle model} \\
\sigma_{\gamma_{0_l}} &\sim \text{Student } t(3, 0.15, 0.08, \text{lb} = 0) && \text{Prior for machine variability parameter in hurdle model} \\
\sigma_{b_{0_j}} &\sim \text{Student } t(3, 3.43, 0.62, \text{lb} = 0) && \text{Prior for location variability parameter in student } t \text{ model} \\
\sigma_{b_{0_k}} &\sim \text{Student } t(3, 10.2, 0.23, \text{lb} = 0) && \text{Prior for member variability parameter in student } t \text{ model} \\
\sigma_{b_{0_l}} &\sim \text{Student } t(3, 3.72, 0.93, \text{lb} = 0) && \text{Prior for machine variability parameter in student } t \text{ model} \\
\nu_y &\sim \text{Gamma}(2, 1) && \text{Prior for student } t \text{ degrees of freedom}
\end{aligned}
\tag{3}
$$

Rating of Perceived Effort and Discomfort Analysis

Despite the use of two different scales and protocols for data collection of perception of effort in our samples i.e., a session RPE using the 6-20 Borg scale for the prior sample and immediately post each exercise RPE using a 0-10 scale, both scales were anchored at their limits as in essence no effort (6 or 0) or maximal effort (20 or 10). As such, and as the intention was to utilise the prior sample data to assist in producing empirical prior distributions for our experimental data models, we opted to rescale both outcomes to lie on the $(0, 1)$ interval such that they reflected the percent effort that was reported as perceived (Steele, 2020). We then employed an ordered beta regression model (Kubinec, 2022) which employs a cutpoint process similar to ordered logistic models in order to model both the continuous responses on the $(0, 1)$ interval reflecting some effort though not maximal, and the degenerate response on the bounds $[0, 1]$ reflecting both no and maximal effort respectively. We approached the RPD similarly as this scale was also anchored between no discomfort and the maximum imaginable.

For both the prior sample and experimental study data the following model was employed which included a fixed (i.e., population level) effect for the session type, Core or Assisted, where Assisted was coded as the intercept. The model also included random (i.e., cluster or group level) effects as intercepts for the location (i.e., the clinic the session was performed at), the member id, and the resistance machine used for the experimental data model only (note, due to the session RPE reported in the prior sample data and that we sampled only one session we do not have member or machine random effects, and due to the cross-over nature of the experimental design the random effects for location, member, and machine were modelled as nested in this dataset). These fixed and random effects were modelled as predictors for both the mean and precision parameters of the ordered beta regression model. As noted, RPE in both datasets was rescaled to lie on the $(0, 1)$ interval. Formally the model(s) for i^{th} RPE across condition c within the j^{th} location, k^{th} member, l^{th} machine were as follows (note that the model for the prior dataset omits the random effect for machine):

$$
\begin{aligned}
\text{RPE}_{ic} &\sim \left\{ \begin{array}{cc} 0 & \text{with probability } \alpha_{ic_{j[i],k[i],l[i]}} \\ \in (0,1) & \text{with probability } \delta_{ic_{j[i],k[i],l[i]}} \\ 1 & \text{with probability } \gamma_{ic_{j[i],k[i],l[i]}} \end{array} \right\} \\
\text{logit}(\alpha_{ic}) &= 1 - (X'\beta - k_1) && \text{Probability of obtaining a 0} \\
\text{logit}(\delta_{ic}) &= [(X'\beta - k_1) - (X'\beta - k_2)]\text{Beta}(X'\beta, X'\beta_\phi) && \text{Probability of obtaining a value between 0 and 1} \\
\text{logit}(\gamma_{ic}) &= (X'\beta - k_2) && \text{Probability of obtaining a 1} \\
X'\beta &= (\beta_0 + b0_{j[i],k[i],l[i]}) + \beta_1 \text{Condition}_{ic} && \text{Vector of predictors} \\
b0_j &\sim \mathcal{N}(0, \sigma_{b0_j}), \text{ for location j} = 1, \ldots, \text{J} && \text{Variation in intercepts for location in student } t \text{ model model} \\
b0_k &\sim \mathcal{N}(0, \sigma_{b0_k}), \text{ for member k} = 1, \ldots, \text{K} && \text{Variation in intercepts for member in student } t \text{ model model} \\
b0_l &\sim \mathcal{N}(0, \sigma_{b0_l}), \text{ for machine l} = 1, \ldots, \text{L} && \text{Variation in intercepts for machine in student } t \text{ model model} \\
X'\beta_\phi &= (\beta_{\phi 0} + b\phi 0_{j[i],k[i],l[i]}) + \beta_{\phi 1} \text{Condition}_{ic} && \text{Vector of predictors} \\
b\phi 0_j &\sim \mathcal{N}(0, \sigma_{b\phi 0_j}), \text{ for location j} = 1, \ldots, \text{J} && \text{Variation in intercepts for location in student } t \text{ model model} \\
b\phi 0_k &\sim \mathcal{N}(0, \sigma_{b\phi 0_k}), \text{ for member k} = 1, \ldots, \text{K} && \text{Variation in intercepts for member in student } t \text{ model model} \\
b\phi 0_l &\sim \mathcal{N}(0, \sigma_{b\phi 0_l}), \text{ for machine l} = 1, \ldots, \text{L} && \text{Variation in intercepts for machine in student } t \text{ model model}
\end{aligned}
\tag{4}
$$

Weakly regularising priors were employed for prior sample dataset. These were the default priors in the `ordbetareg` package for all intercept terms which are scaled to the expected response value when all predictors are at their means and use a student t distribution with degrees of freedom $\nu = 3$, and variance terms similarly scaled with $\mu = 0$. We left the flat prior on the fixed effect for condition given we were primarily interested in letting the prior dataset speak for itself in generating a posterior distribution to use as prior in the experimental data model. As noted, the prior sample model only included the random effects for location. The priors were as follows:

$$
\begin{aligned}
&\textbf{Priors (prior sample data model)} \\
\beta_0 &\sim \text{Student } t(3, 0.6, 2.5) && \text{Prior for intercept of mean parameter i.e., Assisted condition} \\
\beta_{\phi 0} &\sim \text{Student } t(3, 0, 2.5) && \text{Prior for intercept of precision parameter i.e., Assisted condition} \\
\beta_1, \beta_{\phi 1} &\sim \text{Uniform}(\text{lb} = -\infty, u = \infty) && \text{Prior for Condition effects in both location and precision parameters i.e., Core minus Assisted condition} \\
\sigma_{b0_j}, \sigma_{b\phi 0_j} &\sim \text{Student } t(3, 0, 2.5, \text{lb} = 0) && \text{Prior for location variability parameters} \\
k_1, k_2 &\sim \text{Induced dirichlet}(1, 1, 1) && \text{Prior for cutpoint probabilities}
\end{aligned}
\tag{5}
$$

We then took posterior parameter estimates from the prior sample model and used these to generate informative priors for the experimental data model which were as follows:

Priors (experimental sample data model)

$$
\begin{aligned}
\beta_0 &\sim \text{Student } t(3, 0.35, 0.05) && \text{Prior for intercept of mean parameter i.e., Assisted condition} \\
\beta_{\phi 0} &\sim \text{Student } t(3, 2.01, 0.07) && \text{Prior for intercept of precision parameter i.e., Assisted condition} \\
\beta_1 &\sim \text{Student } t(3, -0.24, 0.04) && \text{Prior for Condition effects in mean parameter i.e., Core minus Assisted condition} \\
\beta_{\phi 1} &\sim \text{Student } t(3, -0.48, 0.06) && \text{Prior for Condition effects in precision parameter i.e., Core minus Assisted condition} \\
\sigma_{b0_j} &\sim \text{Student } t(3, 0.19, 0.04, \text{lb} = 0) && \text{Prior for location variability parameters} \\
\sigma_{b\phi 0_j} &\sim \text{Student } t(3, 0.26, 0.05, \text{lb} = 0) && \text{Prior for all other variability parameters} \\
\sigma_{b0_k}, \sigma_{b0_l}, \sigma_{b\phi 0_k}, \sigma_{b\phi 0_l} &\sim \text{Student } t(3, 0, 2.5, \text{lb} = 0) && \text{Prior for all variability parameters} \\
k_1, k_2 &\sim \text{Induced dirichlet}(1, 1, 1) && \text{Prior for cutpoint probabilities}
\end{aligned}
\tag{6}
$$

We did not have prior sample data for RPD and so for this outcome we used the same model as above for the RPE outcomes but with wholly default weakly regularising priors.

Results

In the experimental dataset we ended up with 135 observations for each outcome across for each of the two conditions. Our prior sample for TUL after filtering to the same sample of machines came from 1751 members (Core = 849, Assisted = 902) from 27 locations encompassing 11541 observations. The distributions of raw data from both the prior sample and experimental datasets can be seen in Figure 1.

Time Under Load

In our prior sample model there was little difference between Core and Assisted sessions in terms of both mean TUL and the probability of stopping at 120 seconds. The estimated TUL for Core sessions was 122.76 [95%QI: 120.25, 125.2] seconds, for Assisted sessions was 122.65 [95%QI: 120.17, 125.12] seconds, and the between condition contrast (Core minus Assisted) was 0.08 [95%QI: -1.02, 1.19] seconds. The probability of stopping at 120 seconds for Core sessions was 4% [95%QI: 3%, 5%], for Assisted sessions was 5% [95%QI: 4%, 6%], and the between condition contrast (Core minus Assisted) was -1% [95%QI: -2%, 0%].

Our experimental sample model showed far greater difference between Core and Assisted sessions in terms of both mean TUL and the probability of stopping at 120 seconds, in addition to the posterior estimates differing considerably from the prior sample estimates in terms of between condition contrasts for TUL and probability of stopping at 120 seconds, Assisted condition TUL, and both conditions probability of stopping at 120 seconds. The estimated TUL for Core sessions was 125.12 [95%QI: 113.7, 131.9] seconds, for Assisted sessions was 147.35 [95%QI: 134.29, 154.81] seconds, and the between condition contrast (Core minus Assisted) was -22.1 [95%QI: -26.59, -17.61] seconds. The probability of stopping at 120 seconds for Core sessions was 15% [95%QI: 10%, 24%], for Assisted sessions was 1% [95%QI: 1%, 3%], and the between condition contrast (Core minus Assisted) was 14% [95%QI: 8%, 23%].

All marginal predictions (i.e., global grand means) for predictions and contrasts can be seen in Figure 2.

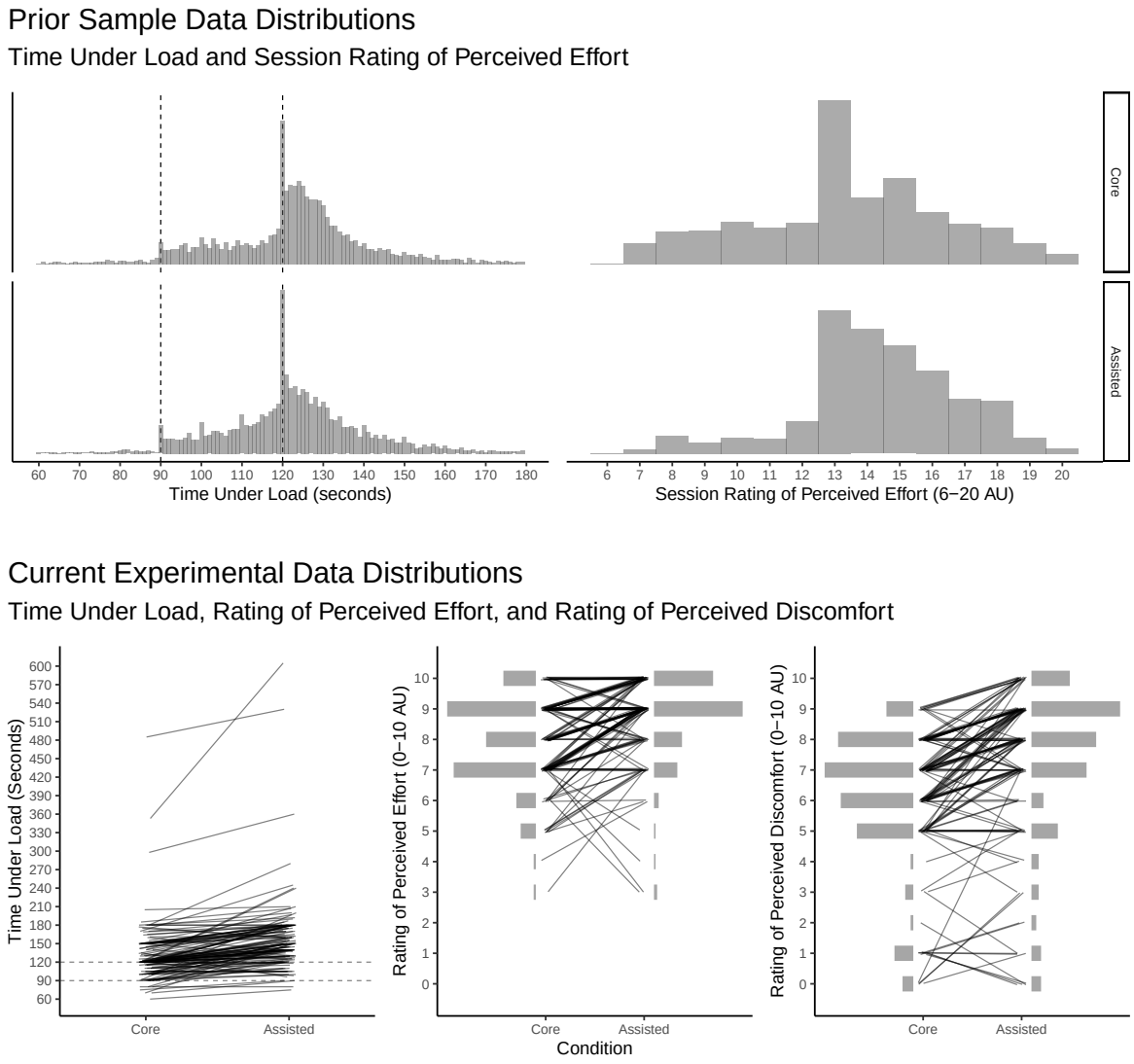

Figure 1: The top two panels show the distributions as histograms of both time under load and session rating of perceived effort in the prior sample of data, and the bottom three panels show the paired responses for time under load in addition to the paired responses and histograms for the rating of perceived effort and discomfort in the current experimental data (note, the prescribed target time under load range of 90-120 seconds is indicated by the vertical dashed lines in the top left panel and horizontal dashed lines in the bottom left panel).

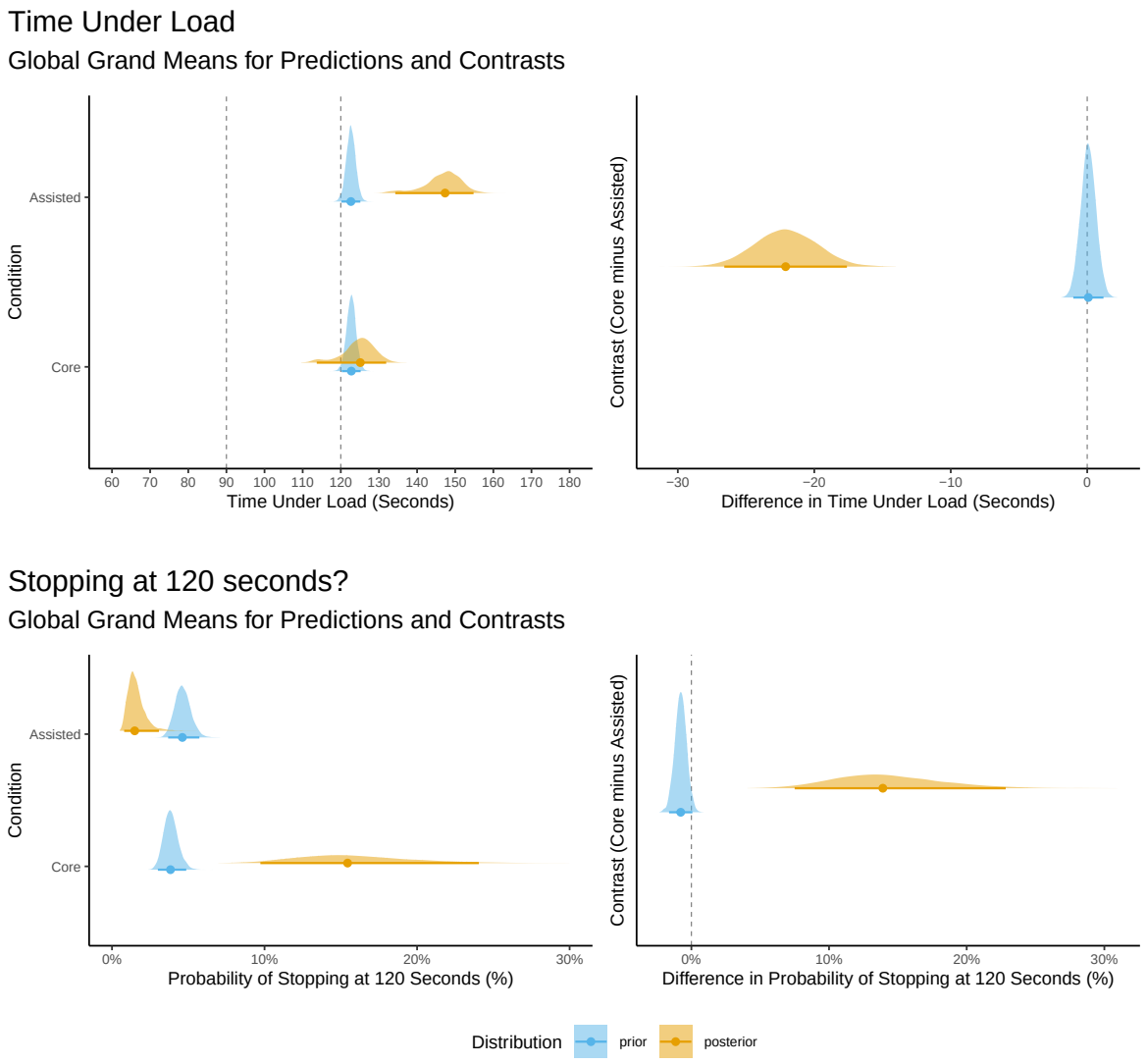

Figure 2: The top two panels show the prior (i.e., generated from the prior sample) and posterior (i.e., the updated distributions after observing the current experimental data) distributions with median (point) and 95% quantile intervals (error bar) of time under load (note, the prescribed target time under load range of 90-120 seconds is indicated by the vertical dashed lines in the top left panel) and the bottom panels show the probabilities of stopping at 120 seconds.

Rating of Perceived Effort and Discomfort

In our prior sample model there was a small difference between Core and Assisted sessions in RPE. The estimated RPE as a percentage for Core sessions was 53% [95%QI: 51%, 55%], for Assisted sessions 59% [95%QI: 57%, 61%], and the between condition contrast (Core minus Assisted) was -6% [95%QI: -8%, -4%]. The ϕ parameter for precision of the beta distribution was higher in the Assisted condition suggesting that RPE reported was more clustered around the estimated RPE values as compared to in the Core condition. The ϕ parameter for Core sessions was 1.53 [95%QI: 1.4, 1.67], for Assisted sessions 2.01 [95%QI: 1.87, 2.15], and the between condition contrast (Core minus Assisted) was -0.48 [95%QI: -0.6, -0.35].

Our experimental sample model showed a similar magnitude of difference between Core and Assisted sessions in RPE, though the posterior estimates for RPEs were considerably greater compared with the prior sample estimates. The estimated RPE as a percentage for Core sessions was 81% [95%QI: 75%, 86%], for Assisted sessions 87% [95%QI: 83%, 91%], and the between condition contrast (Core minus Assisted) was -6% [95%QI: -10%, -4%]. Similarly to the prior data, the ϕ parameter for precision of the beta distribution in our experimental sample was higher in the Assisted condition suggesting that RPE reported was more clustered around the estimated RPE values as compared to in the Core condition. The ϕ parameter for Core sessions was 2.89 [95%QI: 2.35, 3.36], for Assisted sessions 3.34 [95%QI: 2.78, 3.81], and the between condition contrast (Core minus Assisted) was -0.46 [95%QI: -0.61, -0.22].

The RPD results showed a similar pattern to the RPE albeit with a slightly larger magnitude of difference between conditions. The estimated RPD as a percentage for Core sessions was 6.3 [95%QI: 5.1, 7.3], for Assisted sessions 7.5 [95%QI: 6.5, 8.3], and the between condition contrast (Core minus Assisted) was -1.2 [95%QI: -1.6, -0.9]. The ϕ parameter for precision of the beta distribution was higher in the Core condition suggesting that RPD reported was more clustered around the estimated RPD values as compared to in the Assisted condition, though the contrast for this was less precise than in the RPE models. The ϕ parameter for Core sessions was 3.52 [95%QI: 2.56, 5.02], for Assisted sessions 2.91 [95%QI: 1.98, 3.55], and the between condition contrast (Core minus Assisted) was 0.59 [95%QI: -0.25, 2.36].

All marginal predictions (i.e., global grand means) for predictions and contrasts can be seen in Figure 3.

Discussion

Our study is a first to offer insights into the role of supervision (i.e., Core vs Assisted sessions) during RT both *in situ* in an ecologically valid real world setting in addition to during an experimental study. Under standardised training protocol prescription (i.e., single sets of RT using resistance machines to momentary failure using a fixed repetition duration using a load that should permit momentary failure in a target TUL range of 90-120 seconds) we examined exercise performance as the TUL, in addition to RPE and RPD. Our key findings were that under real world settings in our prior sample there did not appear to be a difference in TUL between Core and Assisted sessions nor the probability of whether members targeted the

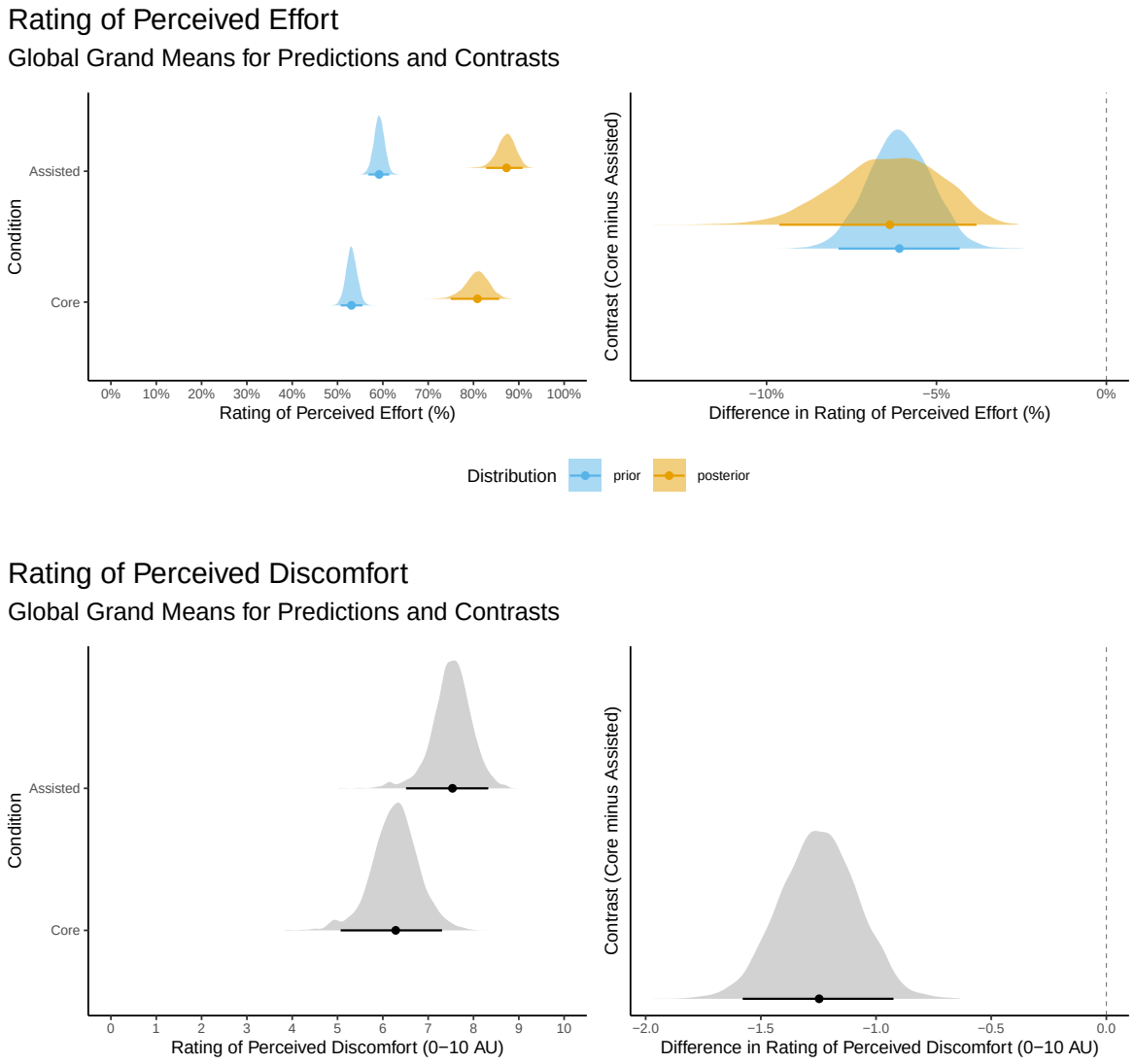

Figure 3: The top two panels show the prior (i.e., generated from the prior sample) and posterior (i.e., the updated distributions after observing the current experimental data) distributions with median (point) and 95% quantile intervals (error bar) of rating of perceived effort on the percentage scale, and the bottom panels show posterior disributions only for rating of perceived discomfort transformed back to the arbitrary unit scale (0,10).

upper bound of the prescribed TUL range (i.e., 120s), yet in our experimental study there was a clear benefit to TUL performance when under supervision and reduction in the probability of targeting the upper prescribed TUL range. There was a small difference in RPE reported with and without supervision under both real world and experimental conditions with members reporting higher RPE in Assisted training conditions, though RPE on the percentage points scale was notably higher during the experimental study compared with the prior sample for both Core and Assisted sessions. Lastly, RPD was also higher in the experimental study under supervised Assisted sessions compared to Core sessions. These results in general support prior work highlighting the importance of supervision during RT (Carlson et al., 2024; Fisher et al., 2022, 2023) and that under unsupervised, and perhaps more so during real world conditions, trainees likely train with suboptimal effort.

We will first discuss the findings from the prior sample of Core and Assisted members. In this *in situ* dataset we find for both Core and Assisted members there is clear evidence of targeting a specific TUL based on the prescribed TUL range of 90-120 seconds. For Core members this is perhaps unsurprising given that, in repetition count based RT prescription, when given the opportunity to self-select repetition number people tend to target ~10 repetitions regardless of the load they self-select and that also most RT prescriptions tend to emphasise that particular target repetition number (Steele et al., 2022). Thus, that members are also shooting for targets under this TUL focused RT protocol suggests that this might be characteristic of most people training under real world unsupervised conditions. It is however surprising that this is also the case for Assisted members and that TUL differs very little between Core and Assisted sessions. However, the lack of difference in TUL itself is not necessarily indicative of suboptimal training effort, and we did observe slightly higher reported RPE during Assisted training sessions compared to Core sessions. Thus it might be that during Assisted sessions trainees might be more likely to use loads more appropriate for the prescribed TUL range and thus train with closer proximity to failure as compared to under Core sessions in addition to greater clustering of RPE around higher values (indicated by the ϕ parameter in the ordered beta regression). This may be due to load progression protocols being more closely followed by supervising exercise scientists as compared to those training unsupervised where, unless using initially heavy loads or in previously well trained persons, most unsupervised trainees do not progress load sufficiently (Fisher et al., 2022; Steele et al., 2022). In future work we plan to explore more thoroughly the causal effects of supervision in both load progression and also isometric strength outcomes collected between Core and Assisted members; but, an initial descriptive examination of load progression using linear-logarithmic growth models known to theoretically describe strength progression with time from RT well (Gschneidner et al., 2024; Latella et al., 2024; Steele et al., 2023) suggests that load progressions are typically similar between membership types (see https://osf.io/wcv6g)[5]. In fact, despite the greater RPE reported under Assisted sessions, in percentage points RPE under both Core and Assisted sessions was considerably lower in our prior sample (Core = 53% [95%QI: 51%, 55%], Assisted = 59% [95%QI: 57%, 61%]) compared with our experimental study (Core = 81% [95%QI:

[5]Similarly to the other prior historical samples used, we took a sample of 1000 Core and 1000 Assisted members fitting a linear-logarithmic model of time in weeks interacting with membership type (Core vs Assisted) with random intercepts for location, machine, and participant id, and random slopes for log time for each level. The model was fit using Restricted Maximum Likelihood estimation with the `lme4` package (Bates et al., 2015).

75%, 86%], Assisted = 87% [95%QI: 83%, 91%]) suggesting that members in these real world conditions may not be training with sufficient effort whether being supervised or not and many perhaps just "going through the motions" (Steele, Fisher, Skivington, et al., 2017).

In contrast to the prior sample, our experimental study clearly showed a benefit to TUL performance with supervision suggesting that it might enhance trainee effort. Further, given we recruited existing Core members and had them use their current training loads, it also reinforces that members may be typically selecting loads that are too light and also training with considerable distance from momentary failure. Though the prescribed protocol *in situ* in the Kieser Australia clinics is for trainees to perform each exercise to momentary failure, despite the average TUL under Core conditions being similar between prior sample and experimental sample (see Figure 2), we can see in Figure 1 that many were capable of achieving very long TUL before reaching momentary failure in both the Core and Assisted sessions. Interestingly whilst the probability of targeting the 120 second threshold was similar in the prior sample between session types, in the experimental study this was far greater in the Core session condition. We speculate that it might be a result of many members, whom were typically not achieving the 120 second TUL upper threshold in their prior sessions, interpreting our explicit instructions in the study to continue to momentary failure as being to ensure that they continue to this upper TUL range target. Whereas the wide range of TULs, and particularly some of the very high ones, might suggest that others more literally interpreted the instruction to continue to momentary failure. In addition to greater average RPE in both conditions, there was also greater clustering of RPE reported under the Assisted condition suggesting that members might have achieved closer proximity to failure whilst supervised. It was also the case that ratings of perceived discomfort were higher under supervision which, as previous evidence has suggested increases with greater proximity to failure (Refalo et al., 2022, 2023, 2025), also supports that members trained with greater effort whilst supervised.

Of interest is the considerable difference between the *in situ* prior historical sample and the experimental study. That under the explicit experimental study instructions to train to momentary failure members were likely to train with a greater effort under both unsupervised and supervised conditions, as evidenced by the TUL results, is reinforced by the greater RPE in percentage points compared to that reported in the prior sample. Indeed, this may reflect some degree of the so called "Hawthorne Effect" in both members and the exercise scientists supervising sessions wherein both were more likely to train/supervise training to a higher degree of effort and thus performance whilst under the auspice of the explicit experimental study instructions and observation[6]. Whilst this so called effect is not particularly well understood in its specifics there is general agreement that both participants and researchers react differently under observational settings such as research (McCambridge et al., 2014; Paradis & Sutkin, 2017) as seems to be observed here.

The limitations of the present results and conclusions should be noted. Firstly, the lack of difference in our prior sample of members regarding TUL does not necessarily indicate that even during Assisted training sessions members are training sub-optimally as this is a comparison

[6]At Kieser Australia the exercise scientists undertake an "exam" / "assessment" after 3-months of being employed and anecdotally we have noted from staff conducting these observations that those exercise scientists will change how they run a session when they are being watched as part of this process.

unadjusted for other factors possibly causal of the TUL during training (member characteristics, load used and load progression). However, the session RPE data does perhaps corroborate the inference that, whilst training with slightly greater effort during Assisted training, members *in situ* may still be training with suboptimal effort compared to what occurs under experimental study conditions when explicitly the instruction to train to momentary failure is reinforced. But, a further issue with this is that two separate tools/protocols for capturing RPE were used in the prior sample (session RPE with an adapted Borg 6-20 scale) vs the experimental study (scales taken from Steele et al. (2017)). Session RPE is historically captured using the Borg Category Ratio Scale (Foster, 1998) but for reasons unclear a choice was made when originally implementing this data capture to use the 6-20 scale. The adaptation to the 6-20 scale in particular involved changes to verbal anchors and it was noticed in the raw distribution of data (see Figure 1) that there was an abundance of values of 13 reported. Whether or not the tendency for lower RPE in percentage points in the prior sample is necessarily caused by reporting behaviours due to scale construction or due to members actually training with lower effort *in situ* is not clear. It is planned that for future member data collection Kieser Australia will move towards an exercise by exercise approach using adaptations of the Steele et al. (2017) scales. The similar degree of contrast between Core and Assisted conditions in both prior sample and experimental study data though does reinforce the role of supervision in influencing trainee effort, or at least reported perception of effort. Lastly, despite the impact of supervision on trainee effort, it is still relatively unclear the extent to which it influences adaptation to training. As noted, Robinson et al. (2024) suggest that higher effort through load and/or proximity to failure is important for outcomes and Fisher et al. (2022) reported a moderate standardised mean effect in favour of supervision for strength adaptation (0.40 [95%CI: 0.06, 0.74]) but estimates of these effects are still relatively imprecise. Kieser Australia also collect isometric strength test data from their members and so in future work we intend to explore this longitudinally and generate causal estimates regarding the role of supervision during RT among other variables from a large sample of members.

Conclusions

To our knowledge this study is the first examine the role of supervision during RT both *in situ* in an ecologically valid real world setting in addition to during an experimental study. Under real world settings there was little difference in exercise performance, yet in our experimental study there was a clear benefit to performance when under supervision. There was a small difference in RPE reported with and without supervision under both real world and experimental conditions suggesting that under supervision trainees train with greater proximity to failure, which was also supported by greater RPD under supervision. These results in general support prior work highlighting the importance of supervision during RT and that under unsupervised, and perhaps more so during real world conditions, trainees likely train with suboptimal effort.

Funding Information

No direct funding was received for this study. Operational costs for its delivery were absorbed by Kieser Australia and so provided *in-kind* value.

Contributions

All authors conceived the idea for the project and designed the methods. TD coordinated and oversaw the data collection. JS conducted the statistical analyses and produced data visualisations. JS and JF drafted the initial manuscript. All authors contributed to editing the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We would like to extend our sincere gratitude to all the members and exercise scientists at Kieser Australia without whom this project would not have been possible. We would also like to extend our thanks to Myles Moore for his feedback on the manuscript.

Data and Supplementary Material Accessibility

All data and code utilised for data preparation and analyses are available in either the Open Science Framework page for this project https://osf.io/etb34/ or the corresponding GitHub repository https://github.com/jamessteeleii/supervision_tul. Other supplementary analyses and plots are also available there.

References

Bates, D., Macherl, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using Lme4 [Journal {{Article}}]. *Journal of Statistical Software*, *67*(1), 1–48.

Carlson, L., Hauger, M., Vaughan-Wenner, G., & Fisher, J. P. (2024). Male and Female Perceptions of Supervision During Strength Training. *Sports*, *12*(11), 301. https://doi.org/10.3390/sports12110301

Coleman, M., Burke, R., Benavente, C., Piñero, A., Augustin, F., Maldonado, J., Fisher, J. P., Oberlin, D., Vigotsky, A. D., & Schoenfeld, B. J. (2023). Supervision during resistance training positively influences muscular adaptations in resistance-trained individuals. *Journal of Sports Sciences*, *41*(12), 1207–1217. https://doi.org/10.1080/02640414.2023.2261090

Fisher, J., Androulakis-Korakakis, P., Giessing, J., Helms, E., Schoenfeld, B., Smith, D., & Winett, R. (2023). Supervision during Resistance Training: A Comparison of Trainer and Trainee Perceptions: Perceptions of Supervision. *International Journal of Strength and Conditioning*, *3*(1). https://doi.org/10.47206/ijsc.v3i1.256

Fisher, J., Steele, J., Wolf, M., Androulakis Korakakis, P., Smith, D., & Giessing, J. (2022). The Role of Supervision in Resistance Training; an Exploratory Systematic Review and Meta-Analysis. *International Journal of Strength and Conditioning*, *2*(1). https://doi.org/10.47206/ijsc.v2i1.101

Foster, C. (1998). Monitoring training in athletes with reference to overtraining syndrome. *Medicine & Science in Sports & Exercise*, *30*(7), 1164.

Gschneidner, D., Carlson, L., Steele, J., & Fisher, J. (2024). *The effects of lengthened-partial range of motion resistance training of the limbs on arm and thigh muscle cross-sectional area: A multi-site cluster trial*. SportRxiv. https://doi.org/10.51224/SRXIV.485

Heiss, A. (n.d.). A guide to modeling outcomes that have lots of zeros with Bayesian hurdle lognormal and hurdle Gaussian regression models. In *Andrew Heiss*. https://doi.org/10.59350/ety2j-09566.

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective [Journal Article]. *Psychonomic Bulletin & Review*, *25*, 178–206.

Kubinec, R. (2022). Ordered Beta Regression: A Parsimonious, Well-Fitting Model for Continuous Data with Lower and Upper Bounds. *Political Analysis*, 1–18. https://doi.org/10.1017/pan.2022.20

Latella, C., van den Hoek, D., Wolf, M., Androulakis-Korakakis, P., Fisher, J. P., & Steele, J. (2024). Using Powerlifting Athletes to Determine Strength Adaptations Across Ages in Males and Females: A Longitudinal Growth Modelling Approach. *Sports Medicine (Auckland, N.Z.)*, *54*(3), 753–774. https://doi.org/10.1007/s40279-023-01962-6

McCambridge, J., Witton, J., & Elbourne, D. R. (2014). Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects. *Journal of Clinical Epidemiology*, *67*(3), 267–277. https://doi.org/10.1016/j.jclinepi.2013.08.015

Nuzzo, J. L., Pinto, M. D., Nosaka, K., & Steele, J. (2024). Maximal Number of Repetitions at Percentages of the One Repetition Maximum: A Meta-Regression and Moderator Analysis of Sex, Age, Training Status, and Exercise. *Sports Medicine*, *54*(2), 303–321. https://doi.org/10.1007/s40279-023-01937-7

Paradis, E., & Sutkin, G. (2017). Beyond a good story: From Hawthorne Effect to reactivity in health professions education research. *Medical Education*, *51*(1), 31–39. https://doi.org/10.1111/medu.13122

Refalo, M. C., Helms, E. R., Hamilton, D. L., & Fyfe, J. J. (2022). Towards an improved understanding of proximity-to-failure in resistance training and its influence on skeletal muscle hypertrophy, neuromuscular fatigue, muscle damage, and perceived discomfort: A scoping review. *Journal of Sports Sciences*, *40*(12), 1369–1391. https://doi.org/10.1080/02640414.2022.2080165

Refalo, M. C., Helms, E. R., Hamilton, D. L., & Fyfe, J. J. (2023). Influence of Resistance Training Proximity-to-Failure, Determined by Repetitions-in-Reserve, on Neuromuscular Fatigue in Resistance-Trained Males and Females. *Sports Medicine - Open*, *9*(1), 10. https://doi.org/10.1186/s40798-023-00554-y

Refalo, M. C., Helms, E. R., Hamilton, D. L., & Fyfe, J. J. (2025). The Effect of Proximity-To-Failure on Perceptual Responses to Resistance Training. *European Journal of Sport Science*,

25(3), e12266. https://doi.org/10.1002/ejsc.12266

Robinson, Z. P., Pelland, J. C., Remmert, J. F., Refalo, M. C., Jukic, I., Steele, J., & Zourdos, M. C. (2024). Exploring the Dose–Response Relationship Between Estimated Resistance Training Proximity to Failure, Strength Gain, and Muscle Hypertrophy: A Series of Meta-Regressions. *Sports Medicine*, *54*(9), 2209–2231. https://doi.org/10.1007/s40279-024-02069-2

Rodriguez-Sanchez, F., cre, cph, Jackson, C. P., Hutchins, S. D., & Clawson, J. M. (2023). *Grateful: Facilitate Citation of R Packages*.

Steele, J. (2020). *What is (perception of) effort? Objective and subjective effort during attempted task performance*. PsyArXiv. https://doi.org/10.31234/osf.io/kbyhm

Steele, J., Fisher, J. P., Giessing, J., Androulakis-Korakakis, P., Wolf, M., Kroeske, B., & Reuters, R. (2023). Long-Term Time-Course of Strength Adaptation to Minimal Dose Resistance Training Through Retrospective Longitudinal Growth Modeling. *Research Quarterly for Exercise and Sport*, *94*(4), 913–930. https://doi.org/10.1080/02701367.2022.2070592

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. https://doi.org/10.1002/mus.25557

Steele, J., Fisher, J., McKinnon, S., & McKinnon, P. (2017). Differentiation between perceived effort and discomfort during resistance training in older adults:Reliability of trainee ratings of effort and discomfort,and reliability and validity of trainer ratings of trainee effort. *Journal of Trainology*, *6*(1), 1–8. https://doi.org/10.17338/trainology.6.1_1

Steele, J., Fisher, J., Skivington, M., Dunn, C., Arnold, J., Tew, G., Batterham, A. M., Nunan, D., O'Driscoll, J. M., Mann, S., Beedie, C., Jobson, S., Smith, D., Vigotsky, A., Phillips, S., Estabrooks, P., & Winett, R. (2017). A higher effort-based paradigm in physical activity and exercise for public health: Making the case for a greater emphasis on resistance training. *BMC Public Health*, *17*(1), 300. https://doi.org/10.1186/s12889-017-4209-8

Steele, J., Malleron, T., Har-Nir, I., Androulakis-Korakakis, P., Wolf, M., Fisher, J. P., & Halperin, I. (2022). Are Trainees Lifting Heavy Enough? Self-Selected Loads in Resistance Exercise: A Scoping Review and Exploratory Meta-analysis. *Sports Medicine*, *52*(12), 2909–2923. https://doi.org/10.1007/s40279-022-01717-9