Resistance Training Optimization: A Study of Strength Gains Using Hevy App Data

Ruben Ahrens

June 28, 2024

Abstract

This study analyzes resistance training data. A recommender systems approach based on Singular Value Decomposition (SVD) is used to generate target weight advice for athletes and coaches accurately. This model is used to track athlete's strength over time. Subsequently, XGBoost has been employed to predict the strength gain or loss. From XGBoost's feature importances new training successful behaviours could be discovered. Because the XGBoost regressor has not been able to successfully predict the strength change, nothing could be learned from the feature importances. Although unsuccessful, this model has the potential to be useful for athletes.

1 Introduction

This study aims to analyze data on resistance training to help athletes achieve strength gains. Resistance training encompasses various disciplines [13], including Olympic-style weightlifting, bodybuilding, powerlifting [2], and strongman competitions. Each of these disciplines has distinct goals and methodologies, but are all among Centimeters Grams Seconds (CGS) disciplines on the sports data science cube. For example, bodybuilding focuses on achieving an aesthetic physique, with competitors judged based on their physical appearance. In contrast, powerlifting emphasizes lifting the heaviest weights possible in specific exercises, with performance judged on technique and weight lifted [21]. Although this study does not focus on any specific discipline, its findings are most directly applicable to powerlifting due to its emphasis on strength gain.

Resistance training workouts typically consist of several exercises, which can be categorized into multi-joint (MJ) and single-joint (SJ) exercises. Compound exercises involve multiple muscle groups working together to lift a weight, leading to significant fatigue and strength demands. SJ exercises target a single muscle group, while less effective for strength gain [17], when added to MJ movements more muscle hypertrophy has been observed [4]. Each exercise session includes multiple sets, with rest periods in between, and can be measured by the number of repetitions (reps), the weight lifted, one rep max (1RM), volume, and the Rating of Perceived Exertion (RPE). RPE is a subjective measure of how intense the exercise feels to the athlete. Research has shown that powerlifters can accurately report RPE [7]. 1RM is the maximum weight an exercise be performed, it can be calculated for any amount of reps. Volume is the weight multiplied by the reps, session volume can represent the load of a workout.

The data analyzed in this study was acquired from the *Hevy* social media platform, where athletes log their resistance training workouts. The Hevy platform allows for detailed tracking of workouts, including exercise type, equipment used, target muscles, and RPE for each set. Tracking workouts enables athletes to apply progressive overload, a key principle in strength training where weights and/or reps are incrementally increased to challenge the muscles continually. Previous research has shown that the social aspects of similar platforms, like Strava, can motivate athletes through peer interaction and competition [18]. However, there are also potential negative psychological effects due to the comparative nature of these apps.

The primary objective of this study is to identify the factors that most significantly contribute to strength gains in resistance training. This has been achieved using a recommender systems approach for calculating a representation of a user's strength over time. Subsequently, aggregate features are used to predict the change in this strength using a regression model. The features with the highest impact on this regression model are the features that should be

1.1 Related works

Few studies have applied data science or machine learning to resistance training data. One study modeled the relationship between different exercises [3]. It has been proven possible to automatically predict RPE based on heart rate and movement sensors [1]. Movement sensors have also been used to review athlete's technique [6]. Singular value decomposition [19], known from Simon Funk's Netflix recommendation algorithm, is used to get a 1RM score for users on exercises that they have not done.

1.2 Position on sports data science cube

Sports data science projects can be classified into a cube with application, discipline, and data type. The application for this project is improving physical performance. The discipline is CGS sports. as the objective of resistance training generally is to maximize training volume. The data type is journal data, logged by (amateur) athletes.

2 Methodology

In this section, the methods are applied to learn the indicators of strength gain. to do this, we first describe the data collection. Afterwards, we describe the data preprocessing. After data preprocessing, we will describe some data exploration. After data exploration, we will describe the method used to predict the most useful features to maximize strength gain.

2.1 Data Collection

The data was collected from the online social media platform *Hevy*, a platform dedicated to sharing weight-lifting workouts with your peers. For registered users, many accounts are publicly accessible. A list of (popular) *Hevy* users was obtained through the *Suggested Athletes* tile on the platform's homepage. There are in total 30 male and 30 female participants. User physiological data is not readily available, even though some athletes have their body composition in their profile bio. From these public accounts, the raw data was acquired using web scraping techniques, bypassing the API's restriction of only being able to acquire the data of your own account through the *Hevy* API or the platform's data export tool. The requests a web browser sends to load a user's data while scrolling their page have been used to simulate scrolling and download the files of each user's data.

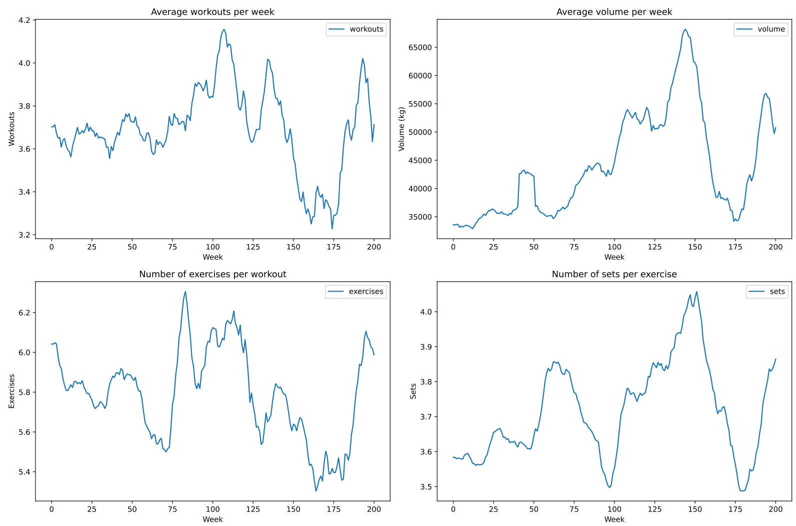

Figure 1: Statistics of workout details over time. Smoothed using convolution

2.2 Data preprocessing

By preprocessing the data, we move from a JSON database where each workout is an instance to a temporal tabular database. The features are: Username, Workout description, exercise title, set index, weight (kg), reps, rpe Only sets with an RPE of 8 or higher were included to filter out warmup sets and sets that have not been pushed close to failure. Since most athletes neglect to log RPE, we will not use this feature.

Start times were normalized to the days since each user's first recorded workout.

2.2.1 1RM Calculation

One rep max (1RM) is the maximum weight that an athlete can lift for one repetition for any given exercise. The estimated 1RM for each set was calculated using the Epley formula:

$$\text{1RM} = \text{weight} \times (1 + 0.0333 \times \text{reps}) \tag{1}$$

The Epley formula is among the most accurate formulae for calculating 1RM [20].

2.3 Data exploration

2.3.1 Training consistency

An important part of developing muscle strength and hypertrophy is staying consistent. Aiming to gain insight into the consistency of athlete's training schedules, we plot four different measures over a time period of 200 weeks. This time period is limited by only using data from weeks where at least five users have trained.

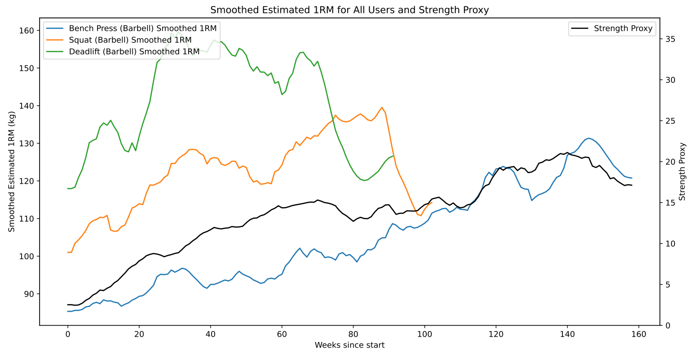

Figure 2: Smoothed Estimated 1RM for Bench Press, Squat, and Deadlift for All Users. The upper limit for the right y-axis has been decided by the final strength of the strongest person in the dataset. On average users have gotten stronger.

From figure 1, there seem to be two pairs of measures that are correlated. The average workouts per week drop and rise when the number of exercises per workout does. When users perform more sets per exercise, they also lift more volume in the relevant week.

2.3.2 1RM Development

To explore the development of the strength of athletes over time. We plotted the 1RM of all athletes over time since they started tracking their workouts on the app. For the plot, we focus on the three fundamental exercises of powerlifting: the barbell squat, the barbell deadlift, and the barbell deadlift. Additionally, we add a strength proxy that aims to model the general strength of each user. The strength proxy is a weighted sum of estimated 1RM, training frequency, and RPE:

Edge-padding convolution was applied to the series to smooth the data and reveal trends without the influence of short-term fluctuations.

The plot in Figure 2 shows the smoothed 1RM for all users over time for the Bench Press, Squat, and Deadlift exercises and the strength proxy.

We filtered the number of users to perform exercises so we don't take the data of only one user. for each week we need at least five users to have performed the exercises.

Additionally, we use the strength change as described in section 2.4 to sum up the accumulated strength. The y-axis limit will be set at the highest achieved strength out of all users but the line will be the average.

From figure 2, we can see that the athletes are unlikely to be beginners when starting to use the app as the reported 1RMs at day 0 have values of intermediate athletes.

The bench press has the slowest but steadiest growth. The squat and deadlift rise more quickly. But is not sustained by most users. Doing these exercises weekly might come with more fatigue accumulation as compared with the bench press.

From the strength proxy, we can see that on average, users have improved their strength steadily. Considering most users were not beginners when they started logging their workouts, this is promising for intermediate athletes.

Model	Parameter	Search Space
SVD	n_factors	{50, 100, **150**}
	lr_all	{0.002, **0.005**, 0.01}
	reg_all	{**0.02**, 0.05, 0.1}
CoClustering	n_cltr_u	{3, 5, **10**}
	n_cltr_i	{**3**, 5, 10}
	n_epochs	{10, 20, **30**}

Table 1: Hyperparameter Search Space and Best Parameters for SVD and CoClustering

Metric	**RMSE (kg)**	MAE (**kg**)	**FCP**
Baseline	0.522	0.341	0.821
SVD	**0.318**	**0.182**	**0.889**
SlopeOne	0.442	0.294	0.835
CoClustering	0.951	0.761	0.813

Table 2: Comparison of RMSE values with and without kg conversion

2.4 Recommender System

A recommender system was used to help athletes more quickly adapt to new exercises, reducing their acclimation period of starting a new exercise and preventing the athlete from training too lightly for the first sessions when doing a new exercise. The recommender system uses the maximal 1RM of each exercise for all users in the last six months, this time window is expected to capture a more truthful representation of a user's current strength than their all-time training history.

The *Surprise* library [9] has been used to evaluate the algorithms: Singular Value Decomposition (SVD) [11] CoClustering [5]. SlopeOne [14]. And a baseline [10]. All of the aforementioned algorithms work with a sparse u: user, i: item matrix: M. Each matrix index has a rating r_{ui}, in this study the "rating" is the 1RM for user u on exercise i. The performance of the recommender system was evaluated using Root Mean Squared Error (RMSE) [8], FCP (Fraction of Concordant Pairs) [12], and Mean Absolute Error (MSE) [8]. SVD and CoClustering have hyperparameters that can be optimized. SlopeOne and the baseline do not have any hyperparameters.

A 5-fold cross-validation train-test split is used, alongside a grid search to tune the hyperparameters if available. The search spaces and best parameters can be seen in table 1, and the metrics of these configurations on the test set can be seen in table 2.

The RMSE and MAE values from Table 2 demonstrate that the average prediction error for the 1RM values is approximately 0.318 kg and 0.178 kg, respectively. These relatively low error margins indicate that the model performs well in predicting the estimated 1RM values with a high degree of accuracy.

Subsequently, the model with the optimal parameters was trained on the full dataset to further enhance prediction accuracy. This refined model can be leveraged within the *Hevy* app to provide precise weight suggestions for target repetitions. Utilizing the Epley formula 1, users can calculate their target rep max (xRM) from the predicted 1RM, thereby tailoring their training more effectively.

2.5 Learning how to get strong

This section describes how the recommender system can be used to learn fundamental features for gaining strength. First, the features will be described alongside the target. Following, the model will be described and finally, the feature importance is shown.

2.5.1 Feature engineering

To model the strength gain effect of a week's training (week w, we selected several features. Since the strength gain is calculated based on the subsequent week's training ($w+1$), the chosen features include:

- **Sessions per week:** The number of workout sessions completed in a week.
- **Exercises per workout:** The average number of different exercises performed per workout session.
- **Sets per workout:** The average number of sets performed per workout session.
- **Reps per set:** The average number of repetitions performed per set.
- **%1RM per set:** The average percentage of the one-repetition maximum lifted per set, compared to the maximum 1RM recorded for the user up until that point (the cumulative maximum).
- **Fraction of Single Joint (SJ) exercises:** The proportion of isolation exercises.
- **Fraction of Multi-Joint (MJ) exercises:** The proportion of compound exercises.
- **Fraction of each muscle group:** The proportion of exercises targeting specific muscle groups.
- **Fraction of each equipment type:** The proportion of exercises utilizing specific types of equipment.

Muscle Groups

The muscle groups included in the analysis are as follows:

- Chest
- Shoulders
- Upper back
- Traps
- Quadriceps
- Lats
- Hamstrings
- Lower back
- Biceps
- Triceps
- Glutes
- Calves
- Abdominals
- Abductors
- Forearms
- Full body

Equipment Types

The types of equipment considered in the study include:

- Barbell
- Dumbbell
- Machine
- Resistance band
- None
- Other
- Plate
- Kettlebell
- Suspension

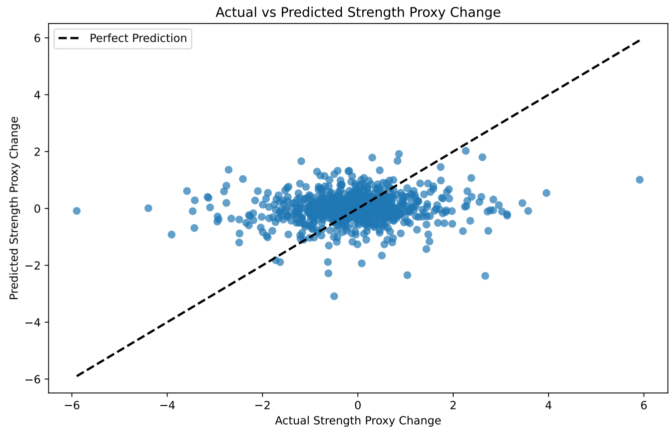

Figure 3: Actual vs Predicted Strength Proxy Change

Because the Muscle groups and equipment features are categorical, each muscle group and exercise type has been assigned its own feature. This feature is transformed by dividing the occurrences of each group/type by the total amount of sets for each user/week group. This ensures all features are numerical.

The target ground truth is the change in the average 1RM of each user of all exercises. comparing week w with week $w+1$

$$\Delta 1\text{RM}_u^w = \frac{1}{|I|}\sum_{i\in I} r_{ui}^{w+1} - \frac{1}{|I|}\sum_{i\in I} r_{ui}^{w} \tag{2}$$

Where u is a user, i is an exercise, I is the set of all exercises, and r_{ui} is the rating (1RM) of the user u for item i. The SVD model is not trained from the start however, the model described in section 2.4 is loaded and then the model is fine-tuned on week w and a copy of the model is fine-tuned on week $w+1$ data.

2.5.2 Regressor

XGBoost has been used to predict the target using the features. 100 estimators have been used. All features have been converted to numerical features. All features can therefore be transformed using a standard scaler that removes the mean and scales according to unit variance.

From figure 3 we can read that the model fails to learn the pattern in the data.

2.5.3 Feature Importance

The final step for generating advice for athletes is plotting the feature importance. After having trained the regression model, SHAP (SHapley Additive exPlanations) [15] was employed to extract the feature importances and impacts. As XGBoost is a tree-based model, we used SHAP's Tree Explainer [16]

Because of the high RMSE, the feature importances are not applicable for training an athlete. However, the plot serves as an illustration of how advice could look in a dashboard that athletes

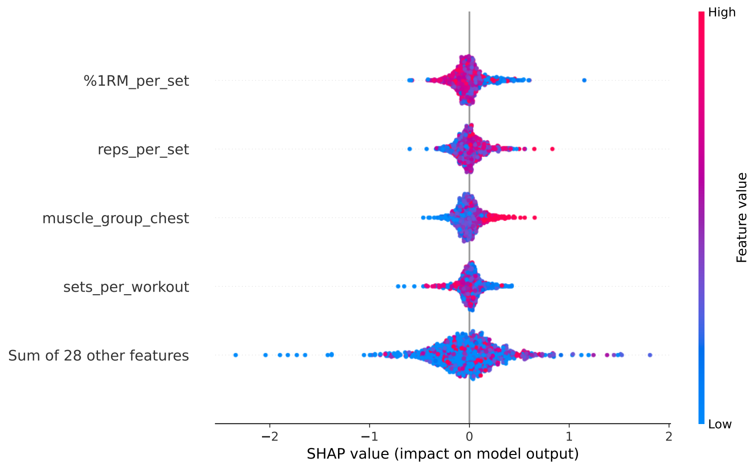

Figure 4: Features and their importance for predicting strength gain over a one week window. Ordered from important to unimportant from top to bottom.

and coaches could use. It would be very clear what the most important features for muscle strength gain are when the strength gain prediction model could be improved.

2.6 Application

From the methods proposed, we can get a useful application for athletes and personal trainers. The *Hevy* app could simply be updated to include weight suggestions for a desired rep range. A potential implementation of this is illustrated in figure 5.

3 Conclusion

The data exploration provides an overview of the behaviours and 1RM trends for Bench Press, Squat, and Deadlift exercises among users of the *Hevy* app. The recommender system, implemented using SVD, demonstrated promising accuracy in predicting 1RM values. This system can be used for a practical application, providing athletes with precise target weight advice for their training sessions in a third-party dashboard or the *Hevy* app.

However, the XGBoost regressor employed to discover important features for strength gains over a one-week window did not yield successful results. This suggests a need for more or different features. Training consistency, volume, and exercise specificity are features that could be explored more deeply to improve regression RMSE. If improved, personalized advice could be provided alongside the weight suggestions.

Future work should address the following areas to improve the accuracy of the model. The feature engineering should include more features like RPE, training volume, and days since the start to capture the nuances of training consistency and intensity. Exploring longer time windows (e.g., monthly) alongside or instead of weekly windows may provide a better understanding of the long-term effects of training patterns. RPE is a promising variable but lacks data. Future

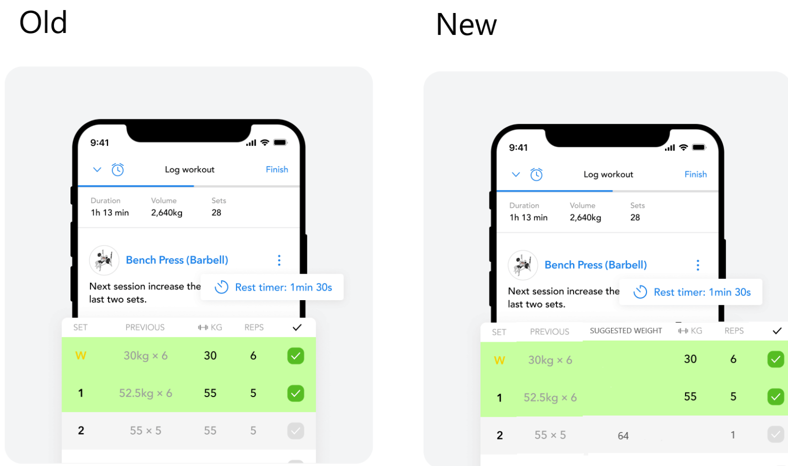

Figure 5: Mockup of Hevy weight suggestions based on predicted 1RM (64 kg) and filled-in target reps (1).

work could task participants to meticulously log all of their exercises to ensure a complete dataset. Additionally, more beginners should be included. The impact of adding female participants should also be explored to ensure a diverse dataset and to make sure the final application is accessible to everyone. Experimenting with different regression models and comparing their performance could yield better results. Investigating the prediction of absolute strength rather than changes in strength might identify effective training strategies used by the strongest athletes.

In conclusion, while the study has demonstrated a useful data science application for optimizing resistance training, further research is required to fully realize the benefits for athletes.

References

[1] Justin Amadeus Albert et al. "Using machine learning to predict perceived exertion during resistance training with wearable heart rate and movement sensors". In: *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE. 2021, pp. 801–808.

[2] Dan Austin and Bryan Mann. *Powerlifting: The complete guide to technique, training, and competition*. Human Kinetics, 2021.

[3] Vinh Huy Chau. "Powerlifting score prediction using a machine learning method". In: *Mathematical Biosciences and Engineering* 18.2 (2021), pp. 1040–1050.

[4] Paulo Gentil et al. "Effect of adding single-joint exercises to a multi-joint exercise resistance-training program on strength and hypertrophy in untrained subjects". In: *Applied Physiology, Nutrition, and Metabolism* 38.3 (2013), pp. 341–344.

[5] Thomas George and Srujana Merugu. "A scalable collaborative filtering framework based on co-clustering". In: *Fifth IEEE International Conference on Data Mining (ICDM'05)*. IEEE. 2005, 4–pp.

[6] Rylea Hart, Heather Smith, and Yanxin Zhang. "The development of an automated assessment system for resistance training movement". In: *Sports Biomechanics* (2024), pp. 1–33.

[7] Eric R Helms et al. "Self-rated accuracy of rating of perceived exertion-based load prescription in powerlifters". In: *The Journal of Strength & Conditioning Research* 31.10 (2017), pp. 2938–2943.

[8] Timothy O Hodson. "Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not". In: *Geoscientific Model Development Discussions* 2022 (2022), pp. 1–10.

[9] Nicolas Hug. "Surprise: A Python library for recommender systems". In: *Journal of Open Source Software* 5.52 (2020), p. 2174. DOI: 10.21105/joss.02174. URL: https://doi.org/10.21105/joss.02174.

[10] Yehuda Koren. "Factor in the neighbors: Scalable and accurate collaborative filtering". In: *ACM Transactions on Knowledge Discovery from Data (TKDD)* 4.1 (2010), pp. 1–24.

[11] Yehuda Koren, Robert Bell, and Chris Volinsky. "Matrix factorization techniques for recommender systems". In: *Computer* 42.8 (2009), pp. 30–37.

[12] Yehuda Koren and Joseph Sill. "Collaborative filtering on ordinal user feedback". In: *Twenty-third international joint conference on artificial intelligence*. Citeseer. 2013.

[13] Mark E Lavallee and Tucker Balam. "An overview of strength training injuries: acute and chronic". In: *Current sports medicine reports* 9.5 (2010), pp. 307–313.

[14] Daniel Lemire and Anna Maclachlan. "Slope One Predictors for Online Rating-Based Collaborative Filtering". In: *CoRR* abs/cs/0702144 (2007). arXiv: cs/0702144. URL: http://arxiv.org/abs/cs/0702144.

[15] Scott M Lundberg and Su-In Lee. "A Unified Approach to Interpreting Model Predictions". In: *Advances in Neural Information Processing Systems 30*. Ed. by I. Guyon et al. Curran Associates, Inc., 2017, pp. 4765–4774. URL: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf.

[16] Scott M. Lundberg et al. "From local explanations to global understanding with explainable AI for trees". In: *Nature Machine Intelligence* 2.1 (2020), pp. 2522–5839.

[17] Antonio Paoli et al. "Resistance training with single vs. multi-joint exercises at equal total load volume: effects on body composition, cardiorespiratory fitness, and muscle strength". In: *Frontiers in physiology* 8 (2017), p. 1105.

[18] Hayley C Russell, Charlie Potts, and Emma Nelson. ""If It's not on Strava it Didn't Happen": Perceived Psychosocial Implications of Strava use in Collegiate Club Runners". In: *Recreational Sports Journal* 47.1 (2023), pp. 15–25.

[19] Ruslan Salakhutdinov and Andriy Mnih. "Bayesian probabilistic matrix factorization using Markov chain Monte Carlo". In: *Proceedings of the 25th international conference on Machine learning*. 2008, pp. 880–887.

[20] Gianni F. Maddalozzo Terry M. Wood and Rod A. Harter. "Accuracy of Seven Equations for Predicting 1-RM Performance of Apparently Healthy, Sedentary Older Adults". In: *Measurement in Physical Education and Exercise Science* 6.2 (2002), pp. 67–94. DOI: 10.1207/S15327841MPEE0602_1. eprint: https://doi.org/10.1207/S15327841MPEE0602_1. URL: https://doi.org/10.1207/S15327841MPEE0602_1.

[21] S Kyle Travis et al. “Tapering and peaking maximal strength for powerlifting performance: a review”. In: *Sports* 8.9 (2020), p. 125.