Athlete Load and Wellness Report: Extensive Report

Perry Battles

May 2025

1 Key Insights and Action Items

1.1 Key Insights and Action Items

- EDA demonstrated that average session heart rate and practice load likely express much of the same information. If collecting both is logistically intractable or difficult, not much is lost by sacrificing one. Because of the familiarity with heart rate data and zones, it may be more useful to present heart rate data to coaches and athletes.
- Sleep has by far the most outsized effect in mitigating athlete wellness by reducing fatigue and stress and increasing mood and motivation (see figures 3 through 18).
- Workload does not appear to affect sleep quality within the ranges in which it has been administered (figure 19), with the exception of a potential slight effect for practice load.
- Soreness was negatively associated with sleep to a much greater degree than with any other variable (see 21). Competition load may positively influence soreness.
- Athletes that performed more strength and conditioning in a given week experienced less stress (figure 15) and fatigue. However, these effects are likely very slight and possibly confounded by the positions of players engaging in the most strength and conditioning also performing less other training.

2 Introduction

The provided data outline day-by-day workload and wellness variables for each of 25 unique athletes on 334 different days in 2022. This report is aimed at the sports scientist, and discusses insights gained both from the exploration of this dataset and from subsequent modeling. Additionally, it outlines several action items derived from these insights designed to improve athlete performance and wellbeing by capitalizing upon the relationships discovered in the data.

3 Methods

3.1 Exploratory Data Analysis

Data were assessed for sparsity (e.g., number of missing values per feature), and distribution for the purposes of discovering patterns and guiding future modeling techniques. Additionally, multicollinearity was assessed to ascertain which variables could be included in the same set of predictors during modeling without undermining our ability to draw conclusions on the effect of one variable on the outcome of interest compared to another.

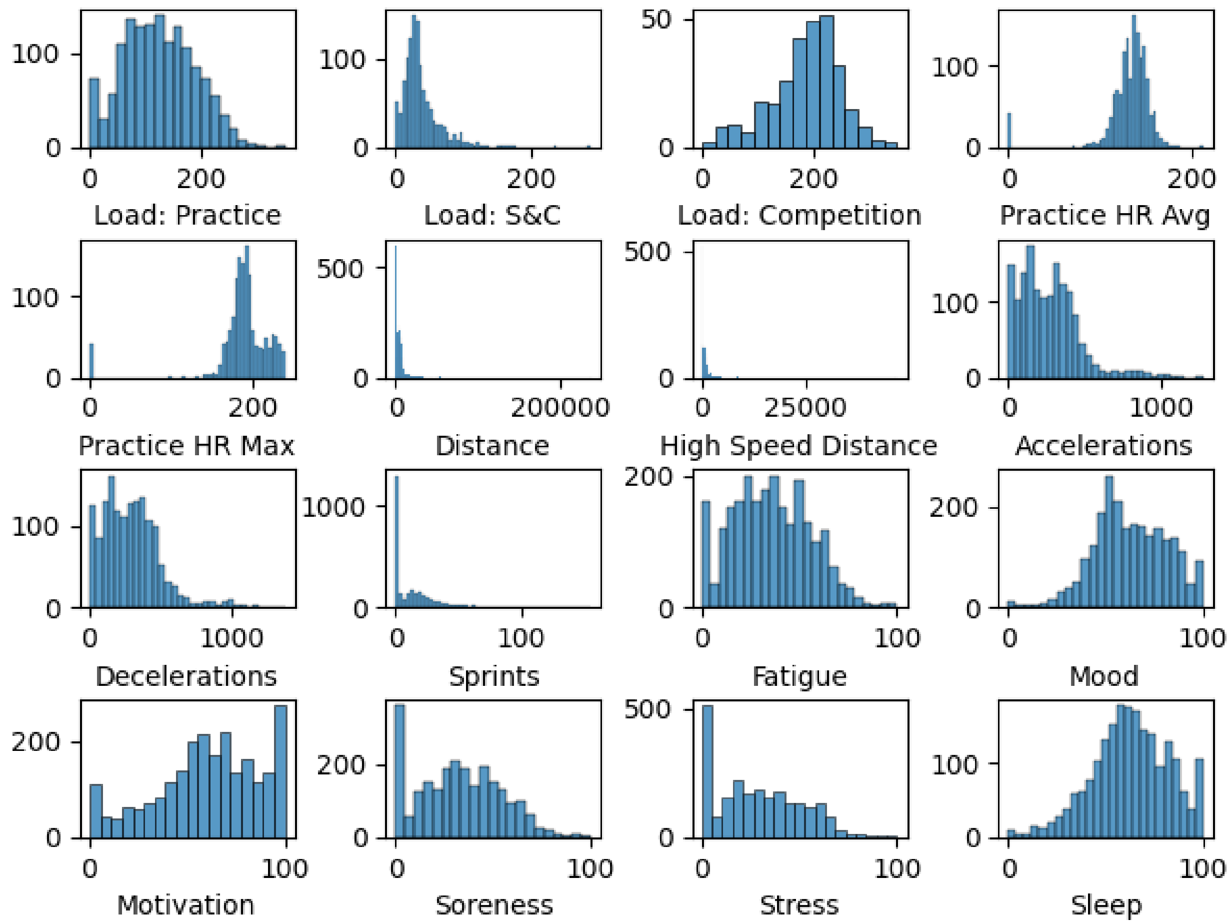

Figure 1: Histograms of all numeric variables in the provided data.

3.2 Model Fitting

Prior experience with workload quantification indicated that trends often emerge at the weekly and monthly levels which are often absent during at the day-to-day level. Hence, athlete data was binned into weekly increments, with some variables (e.g., sleep) being summed for each day in the week and others (e.g., motivation) being averaged. Once these values had been collected, change in each value from the prior time window was also calculated so that subsequent modeling could determine if certain variables might affect the absolute value of one outcome but not the change therein or vice versa.

Prior to model fitting, the data were split into training and testing sets using a 70/30 train/test split. The data were standardized using a robust scaler, as this does not rely upon the same statistical assumptions as a standard scaler. This is valuable given that not all of the outcome variables in the dataset are normally distributed (see figure 1). Robust scaling frees us from this assumption.

A grid search was fit to each model, such that several different estimators were employed and the best one selected for modeling each outcome on the basis of five-fold cross-validation.

In the first part of our analysis (see "Predicting Wellness Values using Workload and Sleep", below), we treated fatigue, motivation, mood, and stress as outcome variables; workload and sleep variables were treated as predictors. We used practice, strength and conditioning, and competition loads as predictors alongside sleep. As outcomes, we assessed both the wellness variable and the change in that wellness variable to determine if there were effects on one or the other.

In the second part of our analysis ("Predicting Sleep Disturbance using Workload Data", below), we moved sleep out of its predictor status and treated it as an outcome. In this situation, we wanted to assess how training might affect sleep quality. Here, practice, strength and conditioning, and competition load were used as predictors.

In the final part of our analysis ("Predicting Soreness using Workload, GPS, and Sleep", below), we attempted to predict soreness using a larger subset of the data, involving practice, strength and conditioning, and competition load, distance, high speed distance, sprints, and sleep as predictors.

4 Results

4.1 Exploratory Data Analysis

Investigations into the dataset indicated that the position groups are meaningfully different in terms of their duties and workload. For example, figure 4.1 indicates clear. The number of pairwise comparisons that would be performed to test whether each group was meaningfully different from the others would likely have created a fairly drastic Bonferroni correction and was thus not included here.

Several variables were found to be highly correlated with one another such that they could not both be included in a model as predictors at once without undermining the ability to make inferences about the effect of any one of those variables on the outcome. These were as follows:

- Practice load and average practice heart rate.
- Distance and accelerations.
- Distance and decelerations.
- Accelerations and decelerations.

From this information, it was inferred that practice load and average heart rate probably capture much of the same information, which makes sense physiologically. Furthermore, simply because of the nature of GPS data, the correlations between distance and the velocity change values made sense. As a result, we did include practice load and heart rate in the same set of predictors during modeling, nor did we include any two of distance, accelerations, and decelerations. Because of its overall intuitive quality and descriptiveness, we used distance.

Exploratory data analysis also revealed that none of the wellness variables appeared to be normally distributed 1. Since the fatigue, soreness, and stress variables appeared to follow a gamma distribution, a gamma generalized linear model with the log link function was included among the models that were fit in the grid search for each outcome variable. The log link function was chosen due to matrix multiplication issues that occurred in `NumPy` and R's `glm` package, which were not able to fit the data with some of the options they provided. This is likely due to some very large values in the dataset that created very large products, which was ameliorated by taking the log of the values via that particular linking function.

4.2 Predicting Wellness Values using Workload and Sleep

As shown in figure 3, sleep is by far the most impactful feature in terms of affecting an athlete's perceived fatigue level. Strength and conditioning workload is also impactful, as neither of the two variables have a confidence interval overlapping with zero. Practice and competition loads have a much more modest effect overall. The accompanying model fit can be seen in figure 4.

In terms of change in fatigue compared to the previous week, sleep is again the most impactful coefficient, being associated with decreased perceptions of fatigue from one week to the next. One of the other coefficients have confidence intervals that do not overlap with zero 6.

Mood is also positively affected by sleep, with the other three coefficients having comparatively very modest effects, if any at all 8. The model fit for this analysis is also arguably the best of those presented here, making this a particularly strong conclusion (figure 8).

Sleep was the only variable whose coefficient confidence interval did not overlap with zero, suggesting a potential effect (see figure 9). However, the associated model fit was poor (see figure 10).

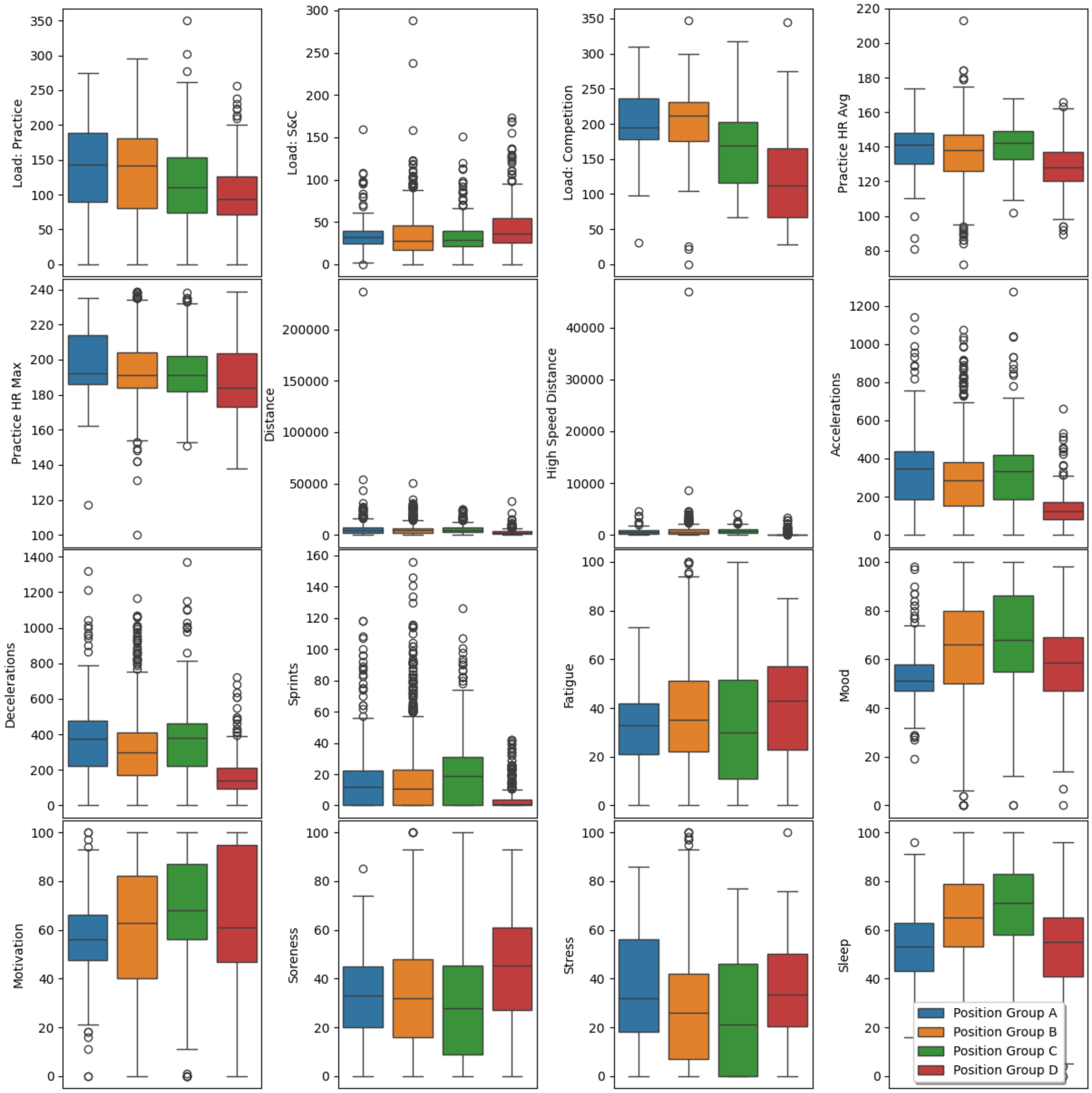

Figure 2: Box and whisker plots of each variable for the four position groups. Each plot displays distributions for position groups A, B, C, and D respectively, from left to right.

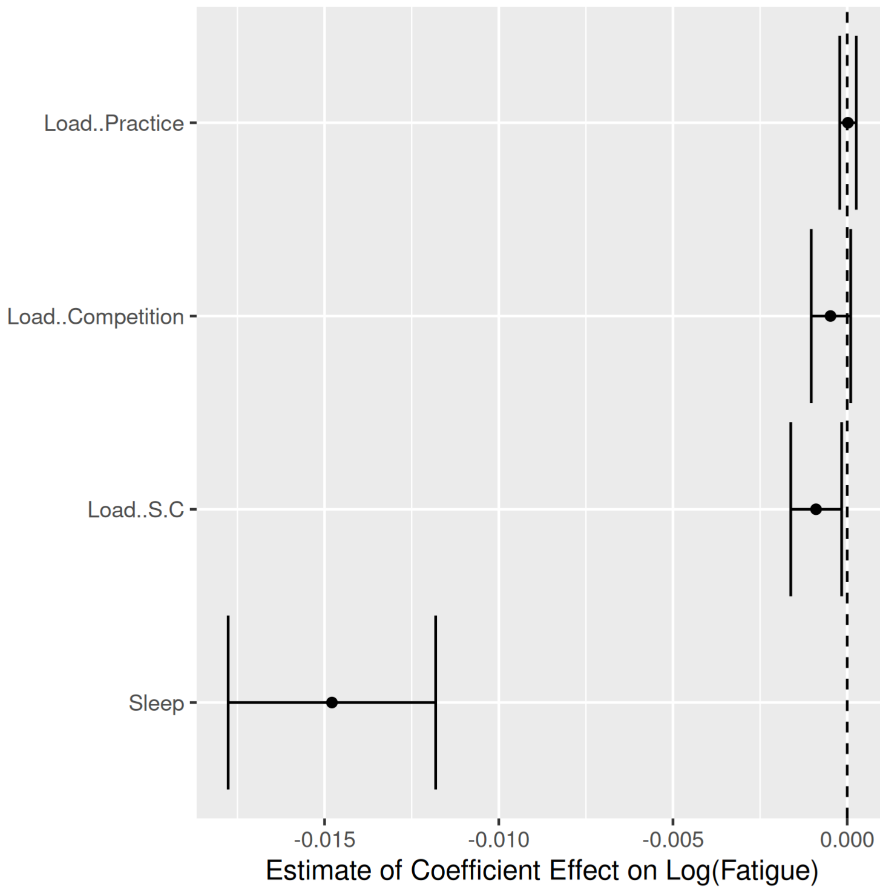

Figure 3: Coefficient magnitudes for gamma generalized linear regression of fatigue based on sleep and workload data.

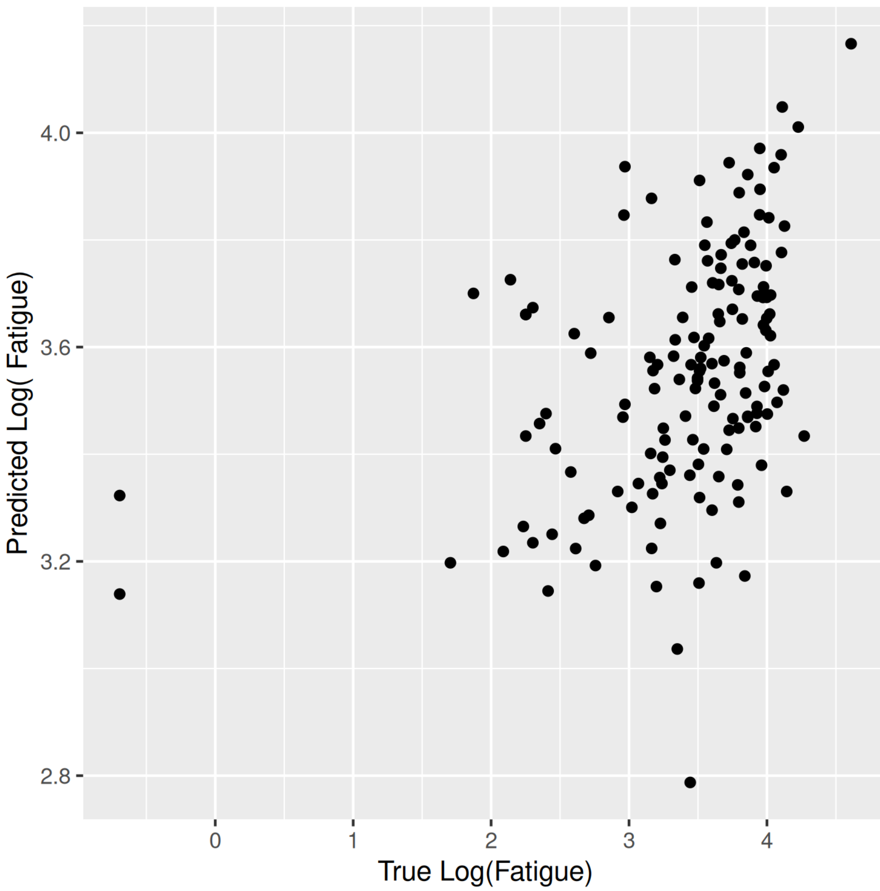

Figure 4: Scatterplot of predicted and true log-transformed outcome values for gamma generalized linear regression of fatigue based on sleep and workload data.

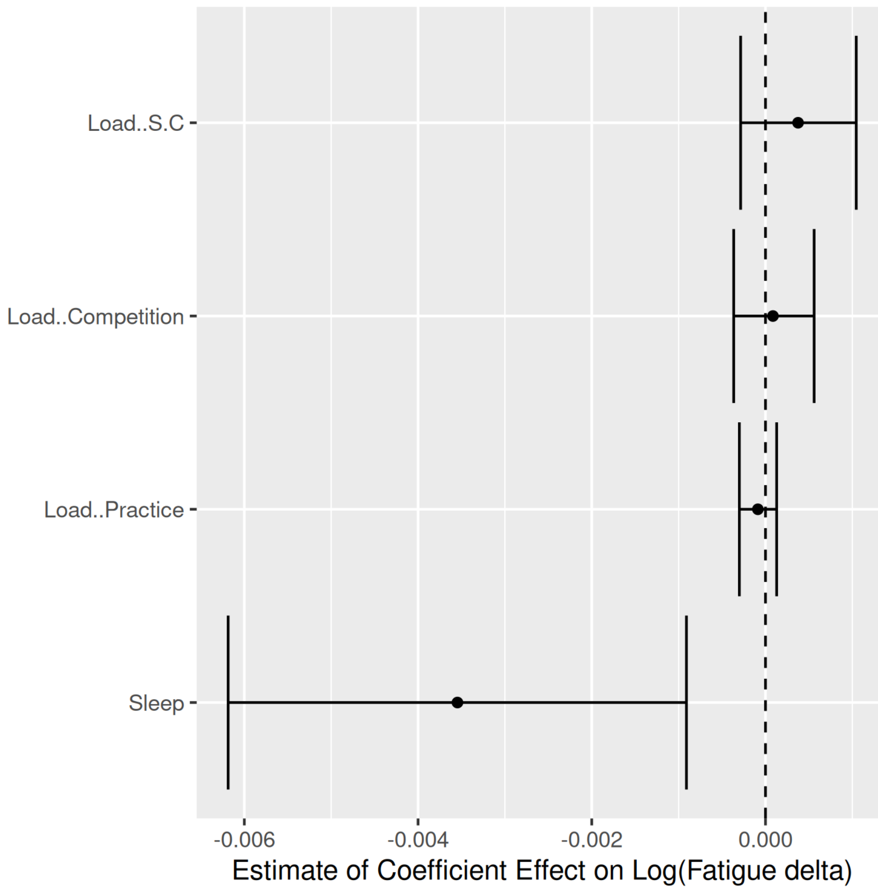

Figure 5: Coefficient magnitudes for gamma generalized linear regression of fatigue delta based on sleep and workload data.

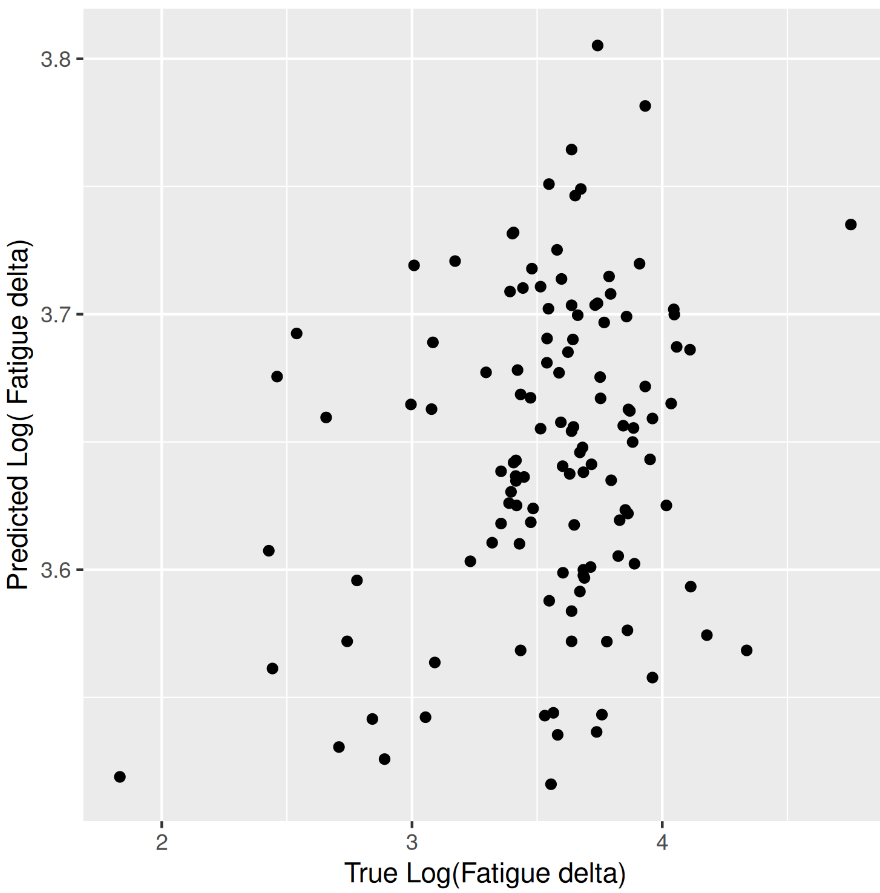

Figure 6: Scatterplot of predicted and true log-transformed outcome values for gamma generalized linear regression of fatigue delta based on sleep and workload data.

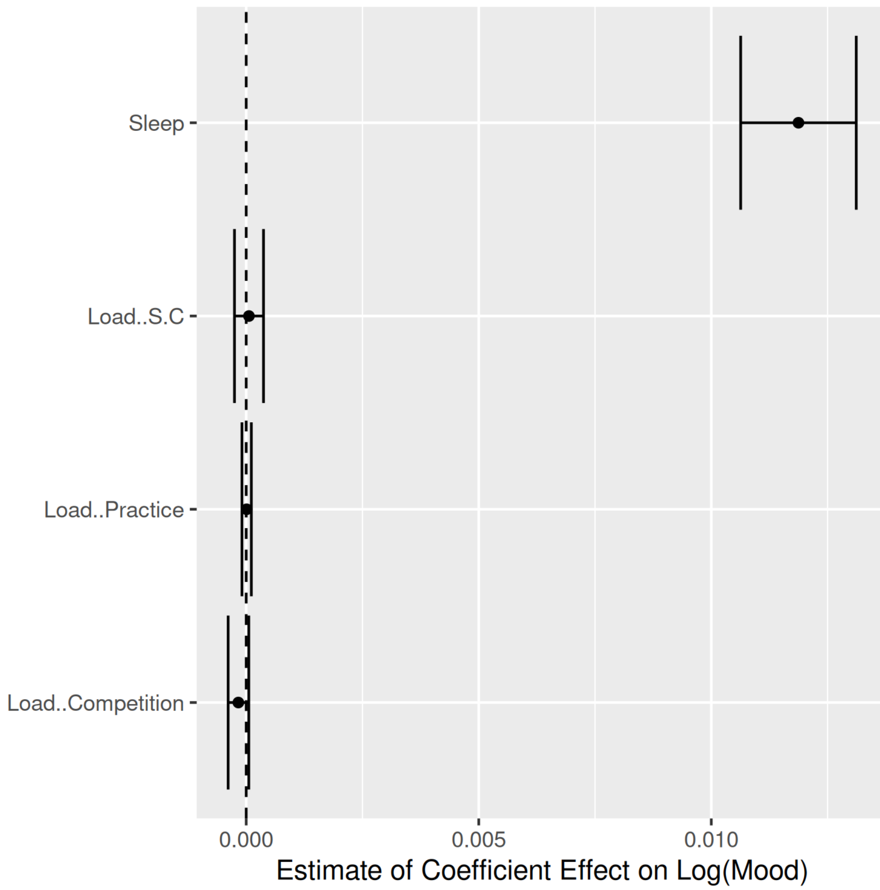

Figure 7: Coefficient magnitudes for gamma generalized linear regression of mood based on sleep and workload data.

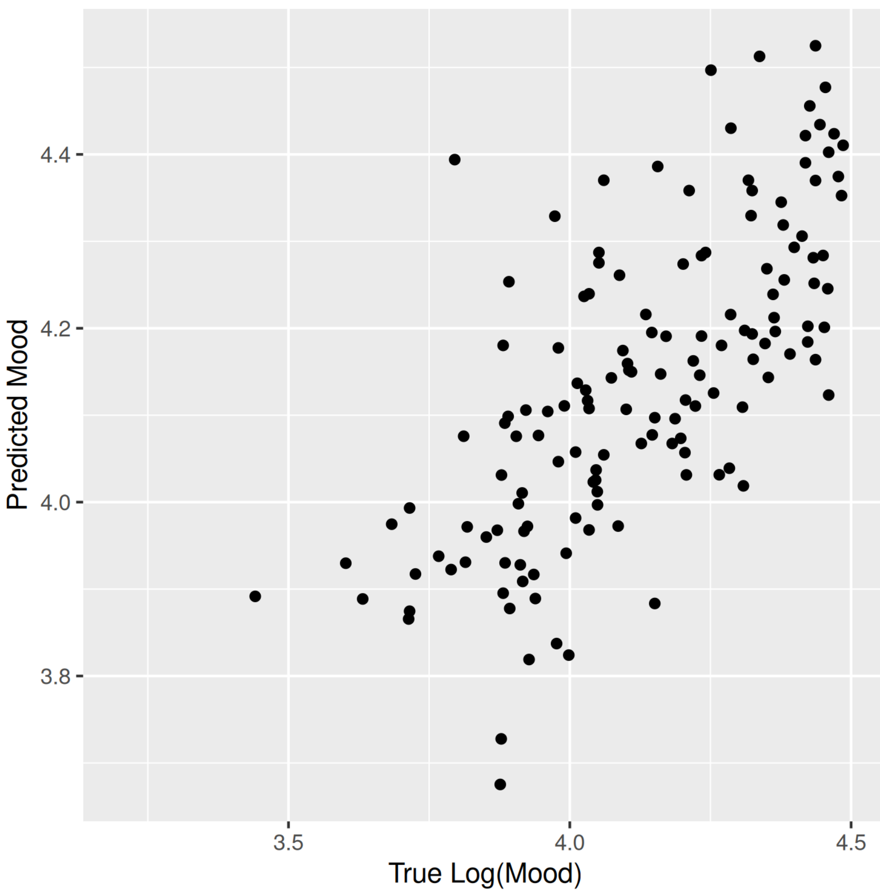

Figure 8: Scatterplot of predicted and true log-transformed outcome values for gamma generalized linear regression of mood based on sleep and workload data.

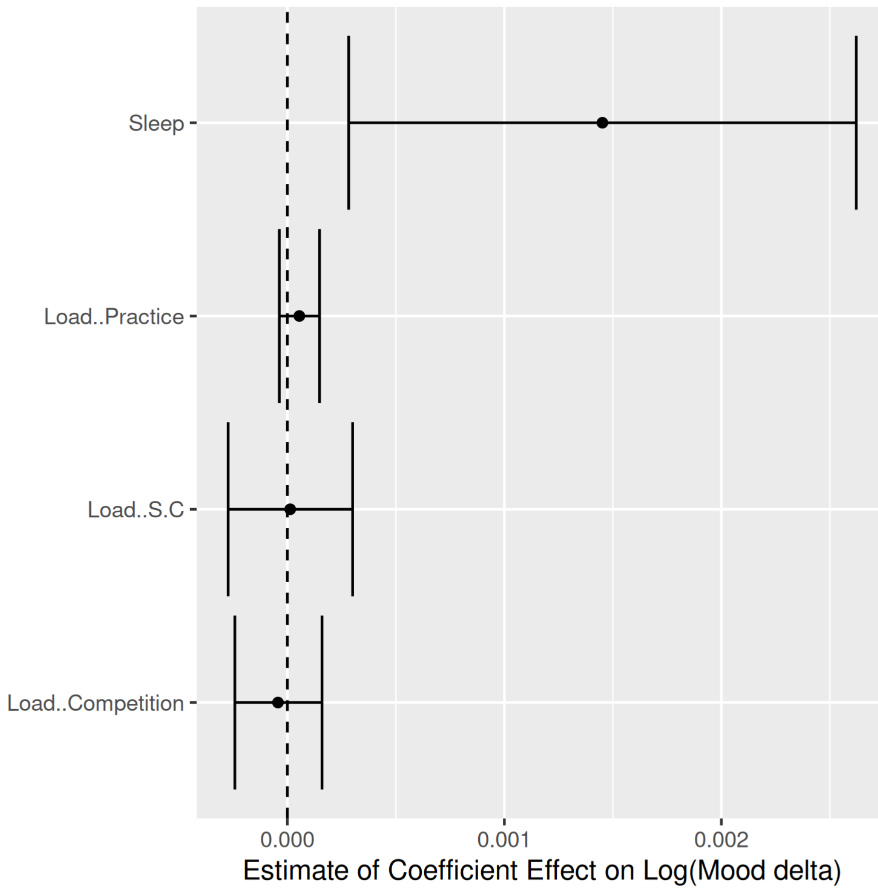

Figure 9: Coefficient magnitudes for gamma generalized linear regression of mood delta based on sleep and workload data.

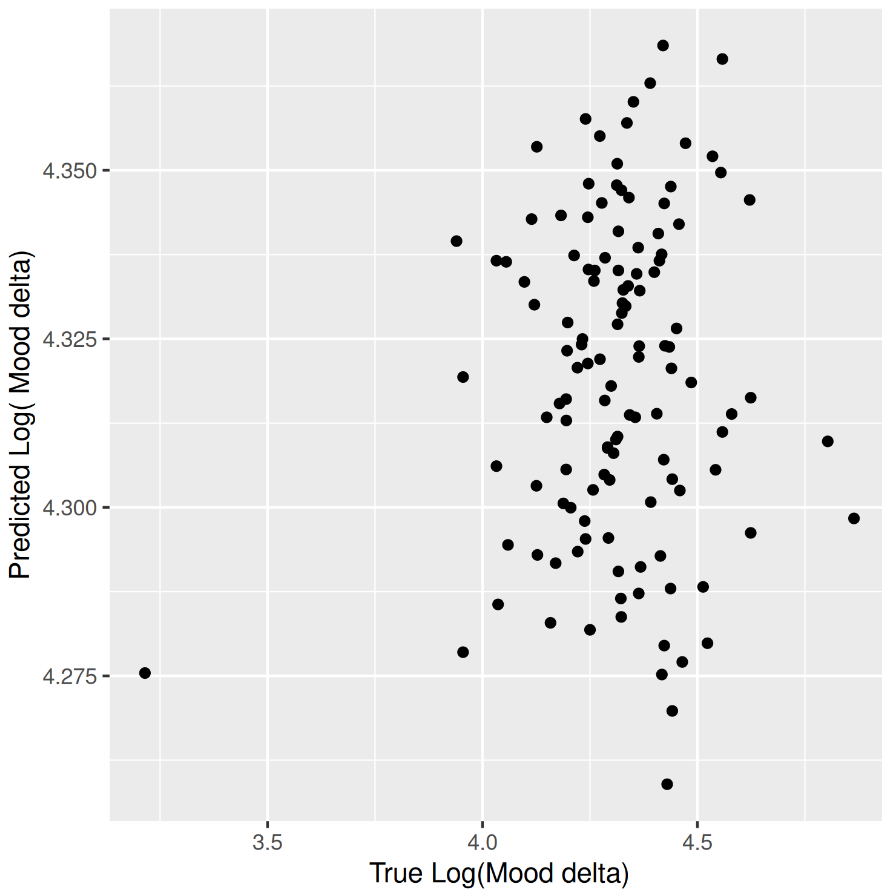

Figure 10: Scatterplot of predicted and true log-transformed outcome values for gamma generalized linear regression of mood delta based on sleep and workload data.

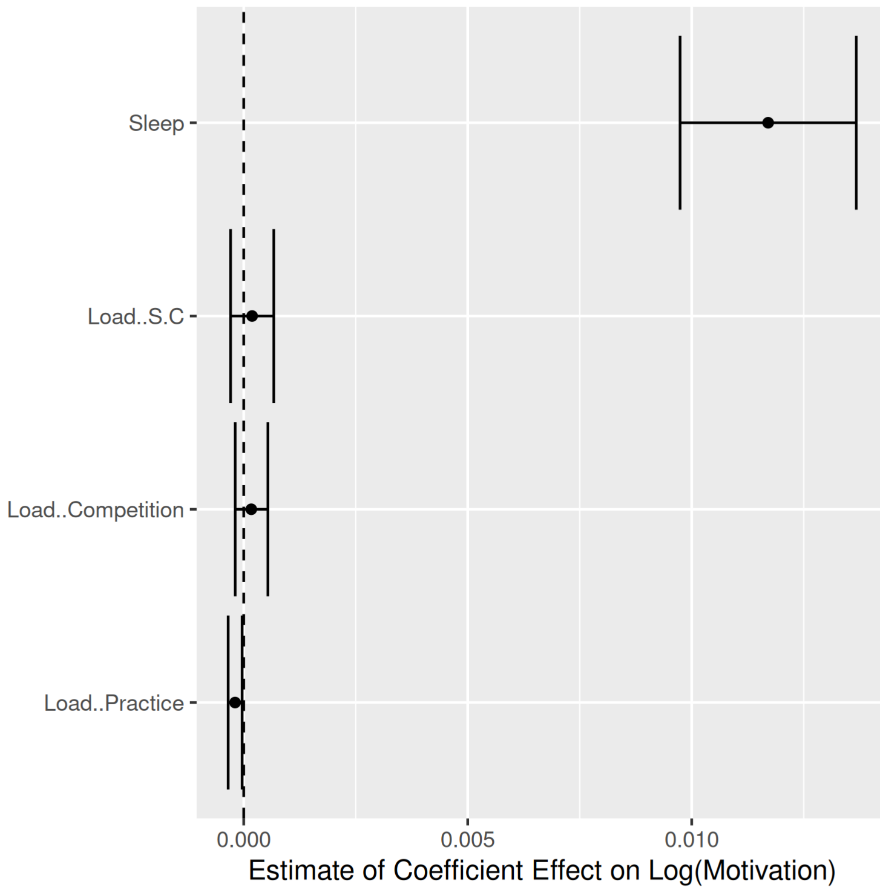

Figure 11: Coefficient magnitudes for gamma generalized linear regression of motivation based on sleep and workload data.

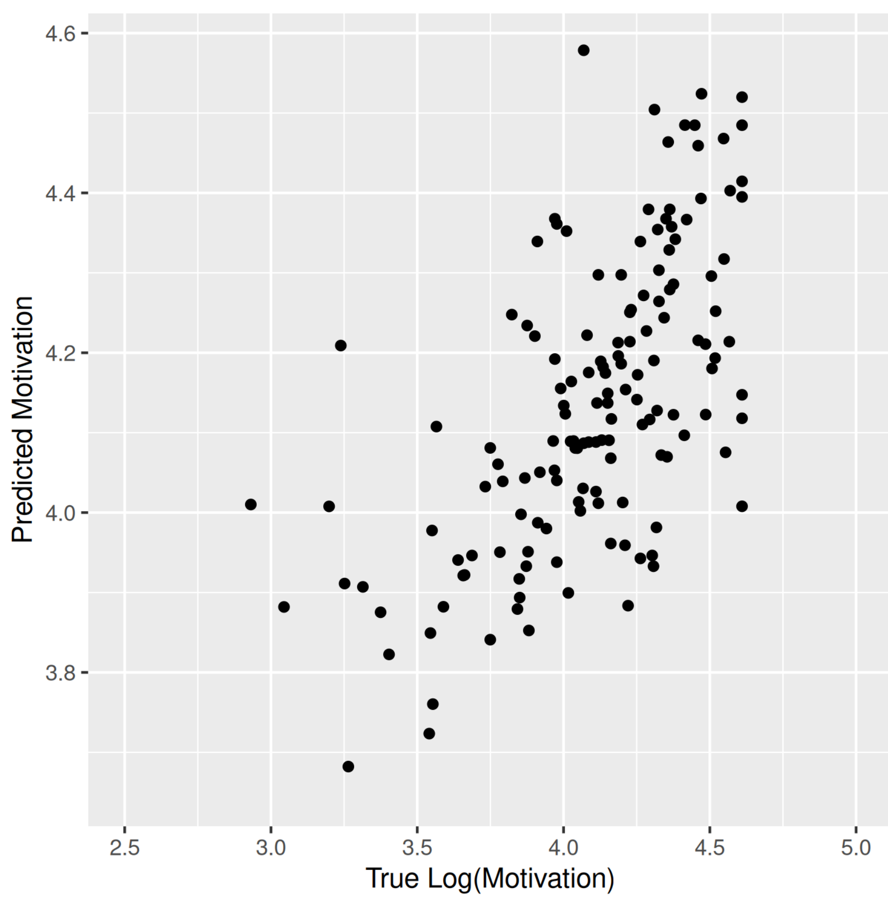

Figure 12: Scatterplot of predicted and true log-transformed outcome values for gamma generalized linear regression of motivation based on sleep and workload data.

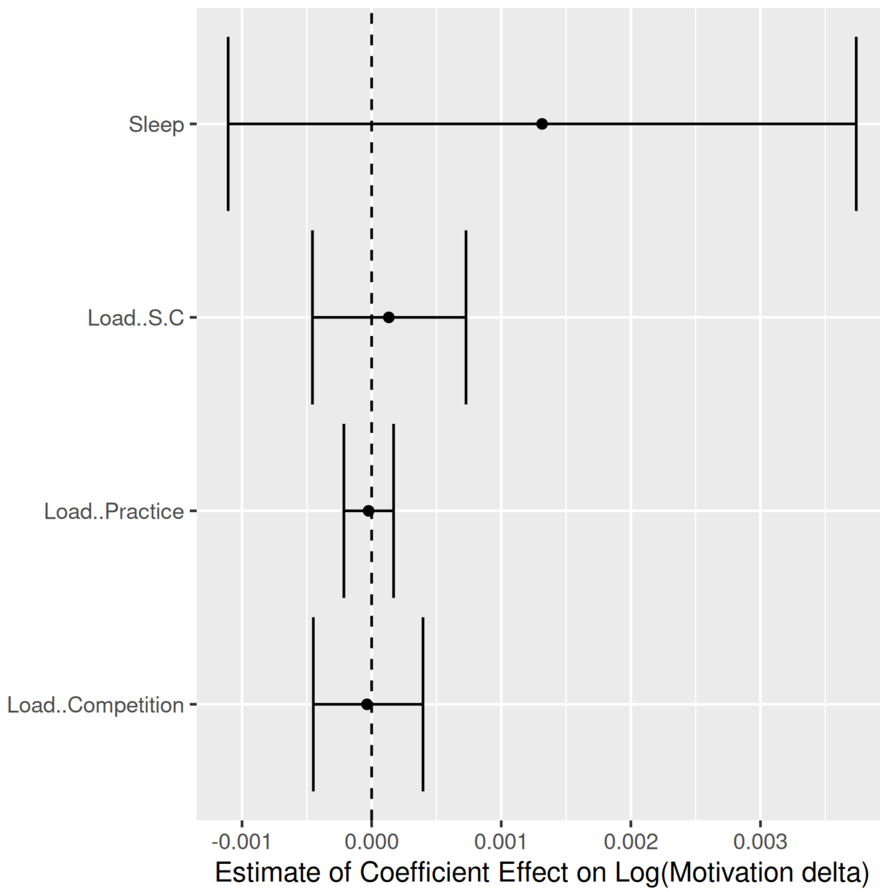

Figure 13: Coefficient magnitudes for gamma generalized linear regression of motivation delta based on sleep and workload data.

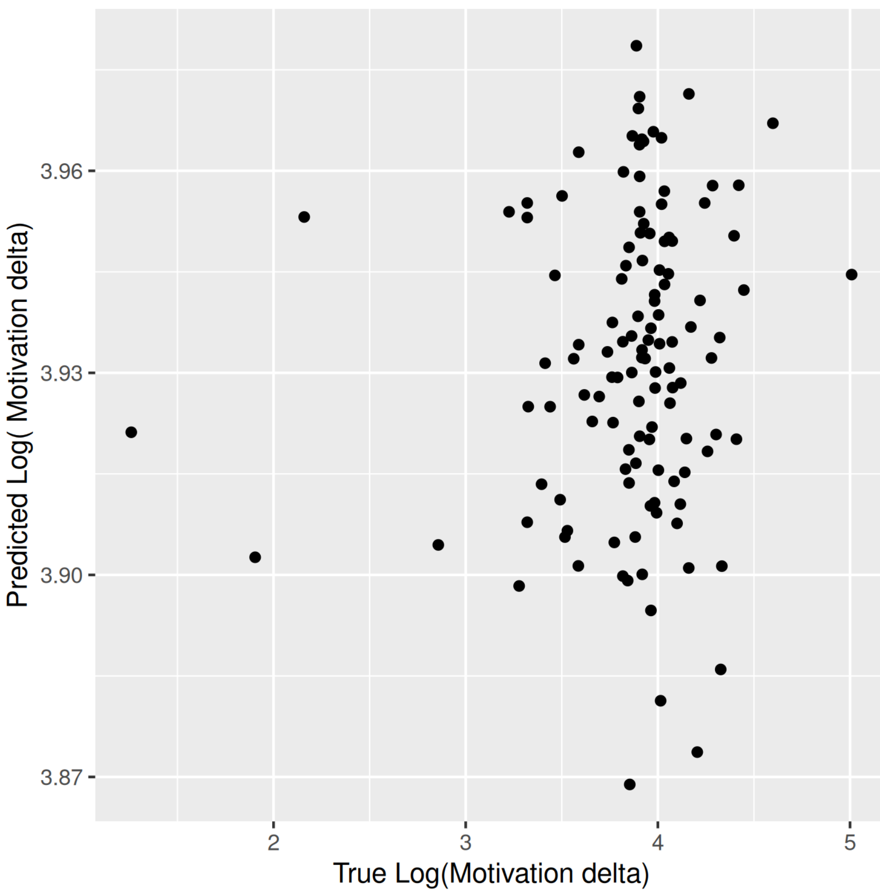

Figure 14: Scatterplot of predicted and true log-transformed outcome values for gamma generalized linear regression of motivation delta based on sleep and workload data.

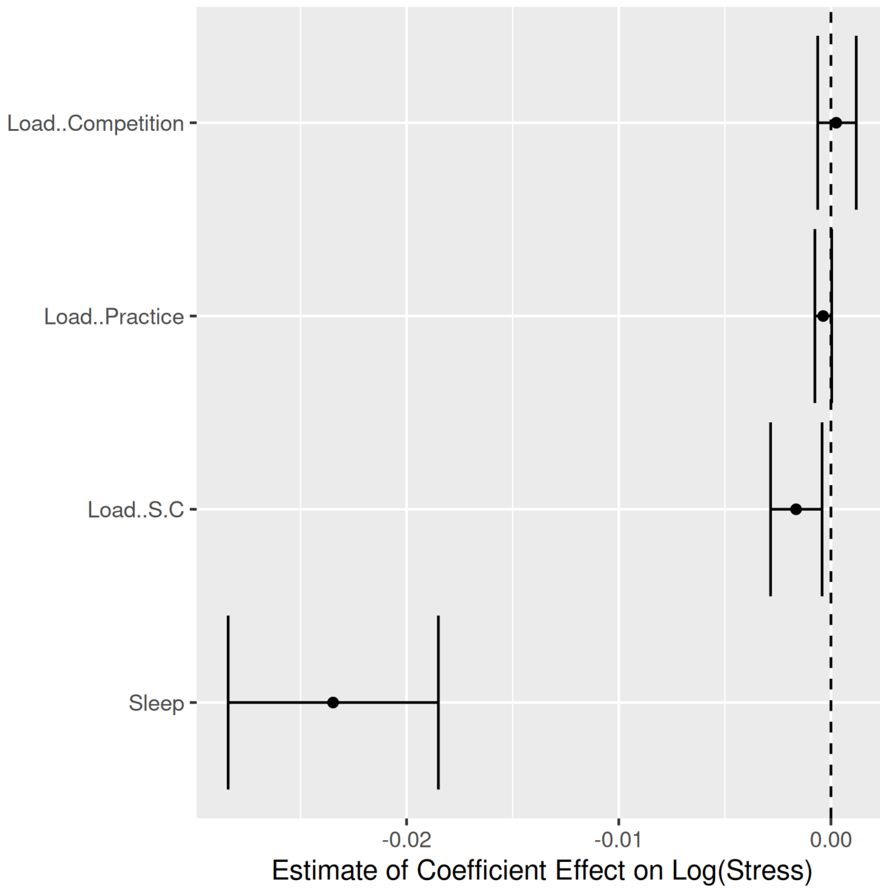

Figure 15: Coefficient magnitudes for gamma generalized linear regression of stress based on sleep and workload data.

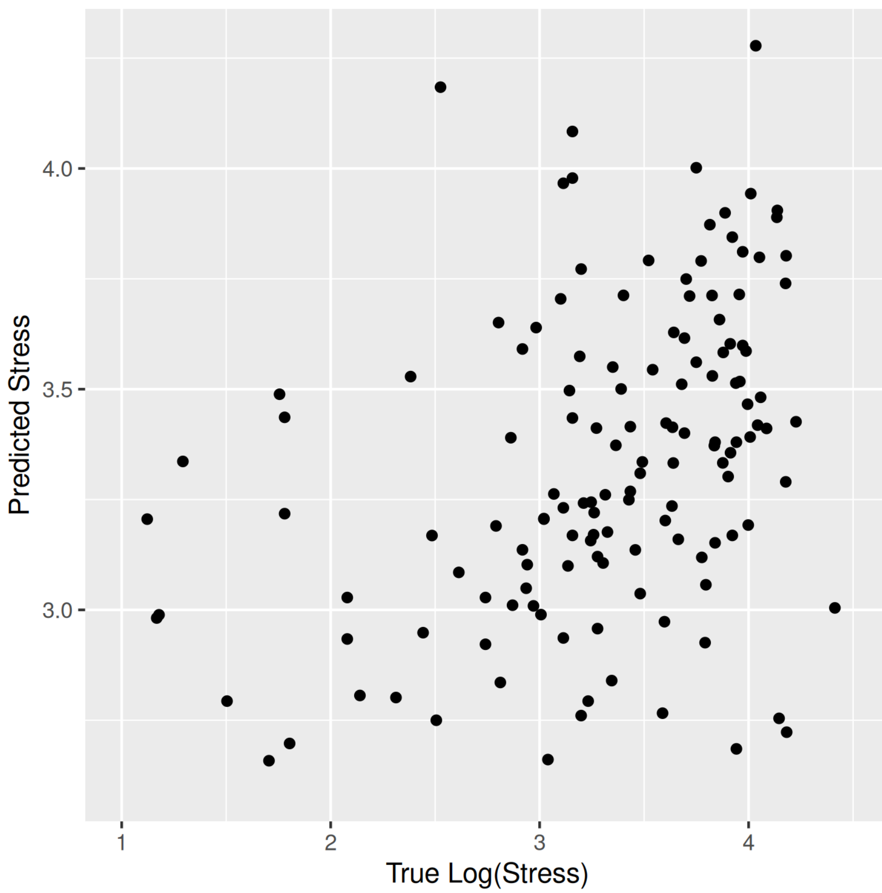

Figure 16: Scatterplot of predicted and true log-transformed outcome values for gamma generalized linear regression of stress based on sleep and workload data.

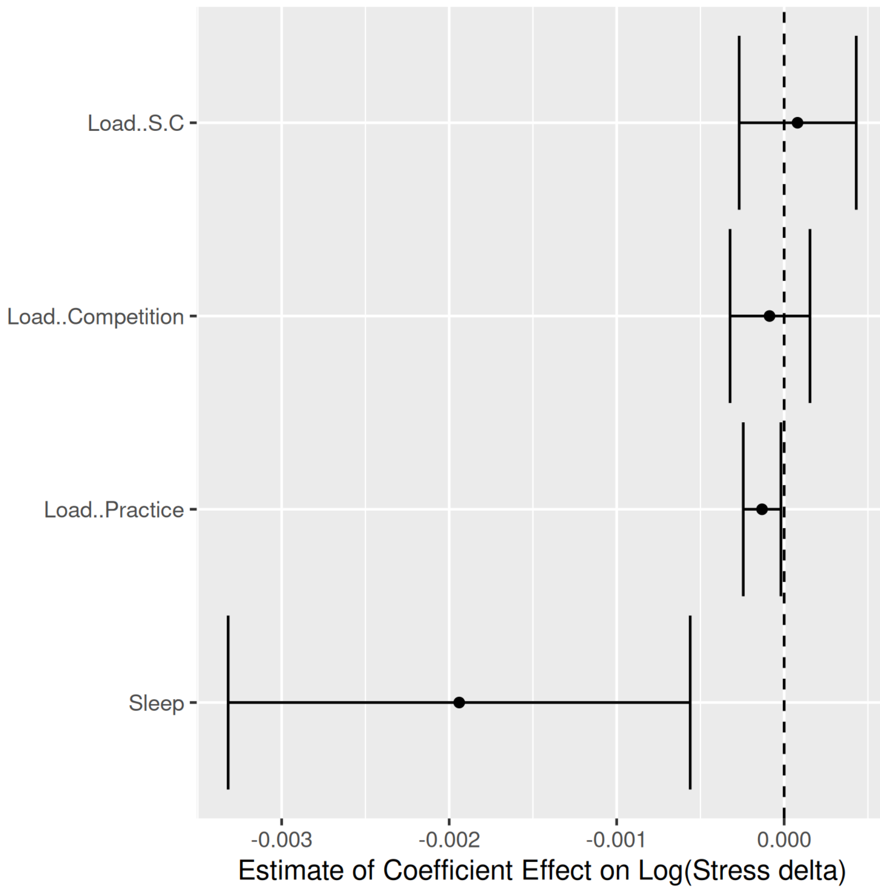

Figure 17: Coefficient magnitudes for gamma generalized linear regression of stress delta based on sleep and workload data.

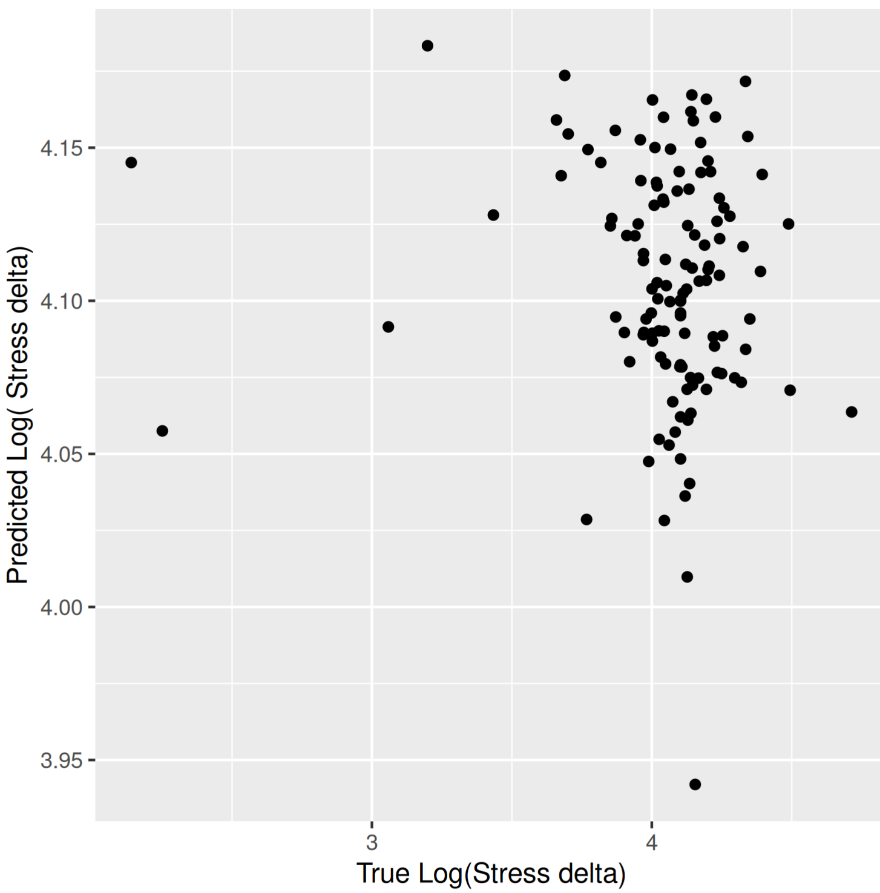

Figure 18: Scatterplot of predicted and true log-transformed outcome values for gamma generalized linear regression of stress delta based on sleep and workload data.

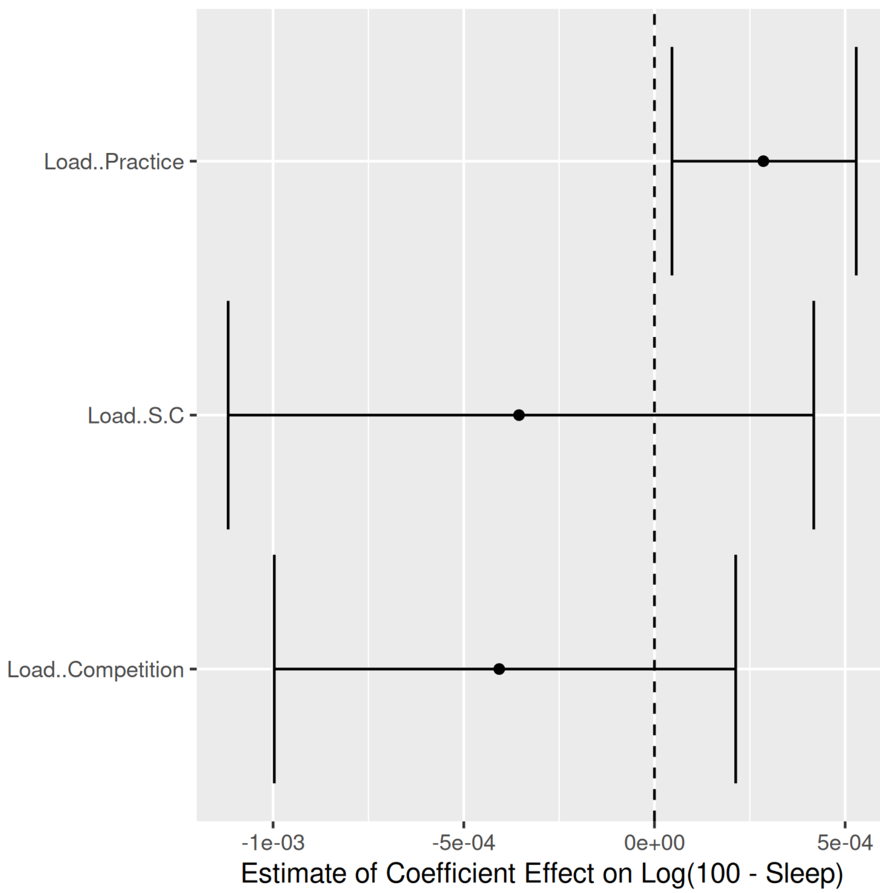

Figure 19: Coefficient magnitudes for gamma generalized linear regression of sleep based on workload data.

Sleep also positively affects motivation, being by far the most influential variable. The workload variables have a much lesser contribution, if any at all (see figure 11).

No coefficient had a confidence interval that did not overlap with zero (see figure 13). Additionally, the model fit was equivocal (see figure 14). Hence, there does not appear to be an effect of any workload variable or sleep on the change in motivation from one week to another.

Stress is negatively associated with sleep. Although no causal claims can be made on the basis of these analyses, and because stress is often anecdotally co-occurring with lower levels of sleep, it is difficult to determine whether one variable drives the other here. Interestingly, strength and conditioning load is also negatively associated with stress (see figure 15). However, the model fit here is rather poor both in absolute terms and compared to the other outcomes in question, making it difficult to justify any firm conclusions on the basis of the data (figure 16).

The model for change in stress from week to week had one of the least conclusive fits in the present analysis (see figure 16). Therefore, firm conclusions cannot be made on the basis of the model fit.

4.3 Predicting Sleep Disturbance using Workload Data

The quality of the model fit for sleep based on workload data is somewhat low due to the presence of many duplicate ground truth values for sleep disturbance. Because of this, too much cannot be confidently inferred on the basis of the results. Tentatively, the null hypothesis can be retained that workload does not meaningfully affect sleep quality, with the possible exception of practice load. Although there may be physiological reasons to assume that this is not true, the data do not substantiate any relationship between

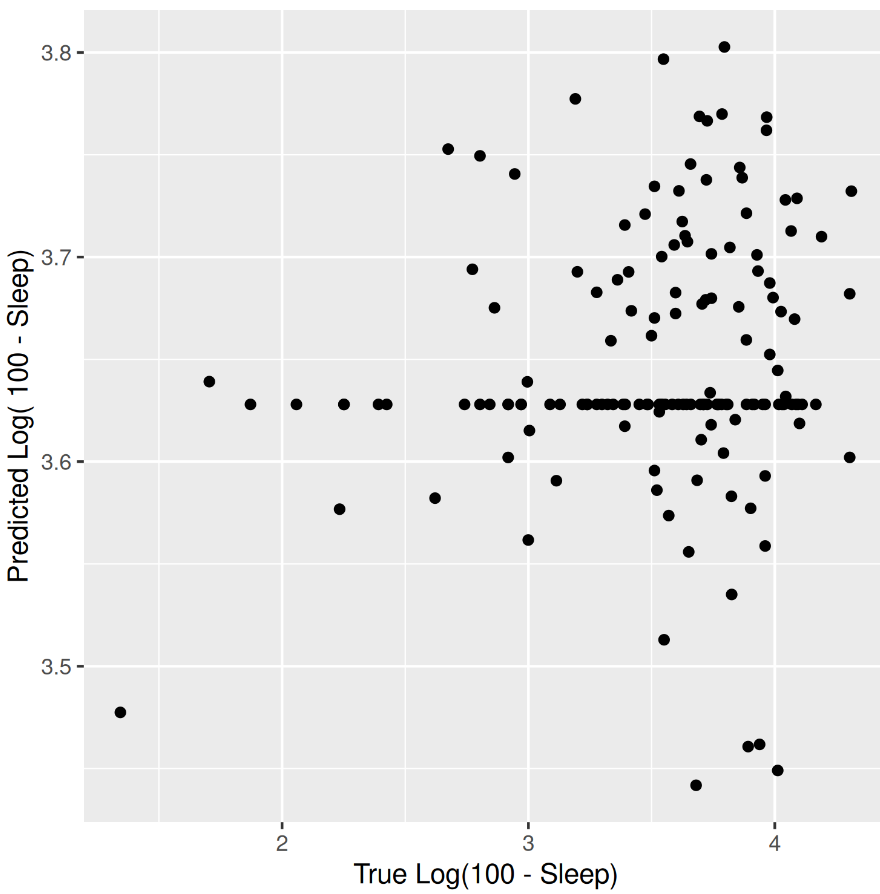

Figure 20: Scatterplot relating predicted sleep disturbance to true sleep disturbance.

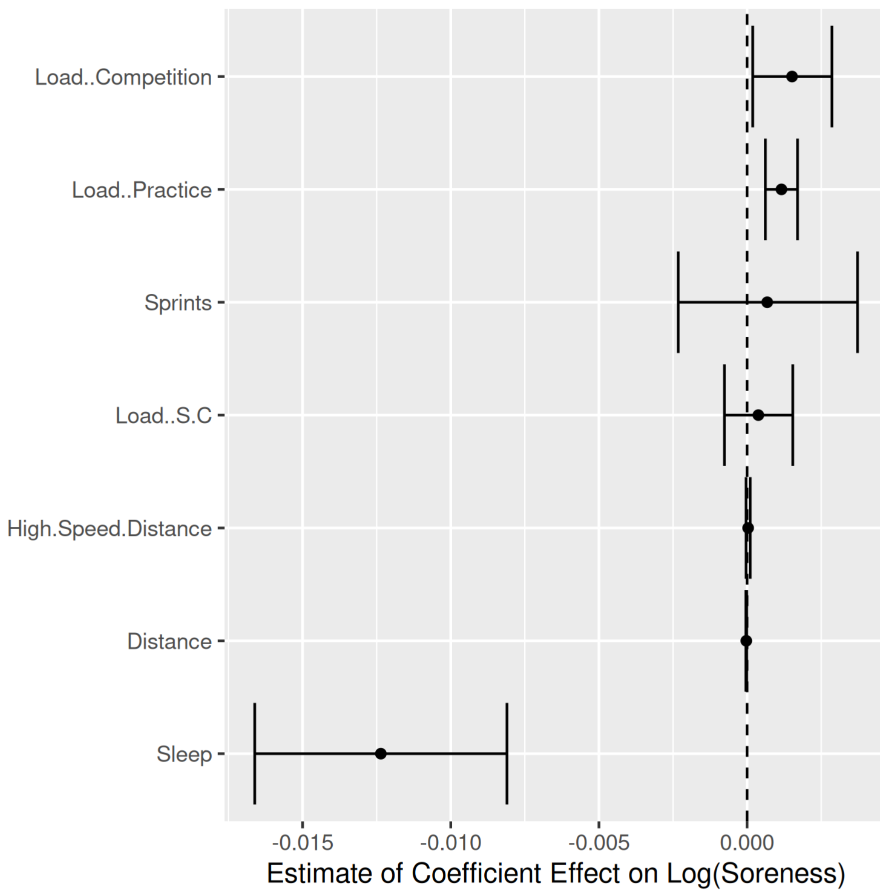

Figure 21: Coefficient magnitudes for gamma generalized linear regression of sleep based on workload data.

workload and sleep disturbance in this cohort.

4.4 Predicting Soreness using Workload, GPS, and Sleep

In predicting soreness on the basis of sleep, load, and GPS data, sleep was once again the most powerful indicator of how sore athletes would be in a given week. Competition load may also be positively associated with soreness, having a confidence interval that did not overlap with zero (see figure 21).

5 Conclusion

5.1 Key Insights and Action Items

- EDA demonstrated that average session heart rate and practice load likely express much of the same information. If collecting both is logistically intractable or difficult, not much is lost by sacrificing one. Because of the familiarity with heart rate data and zones, it may be more useful to present heart rate data to coaches and athletes.
- Sleep has by far the most outsized effect in mitigating athlete wellness by reducing fatigue and stress and increasing mood and motivation (see figures 3 through 18).
- Workload does not appear to affect sleep quality within the ranges in which it has been administered (figure 19), with the exception of a potential slight effect for practice load.

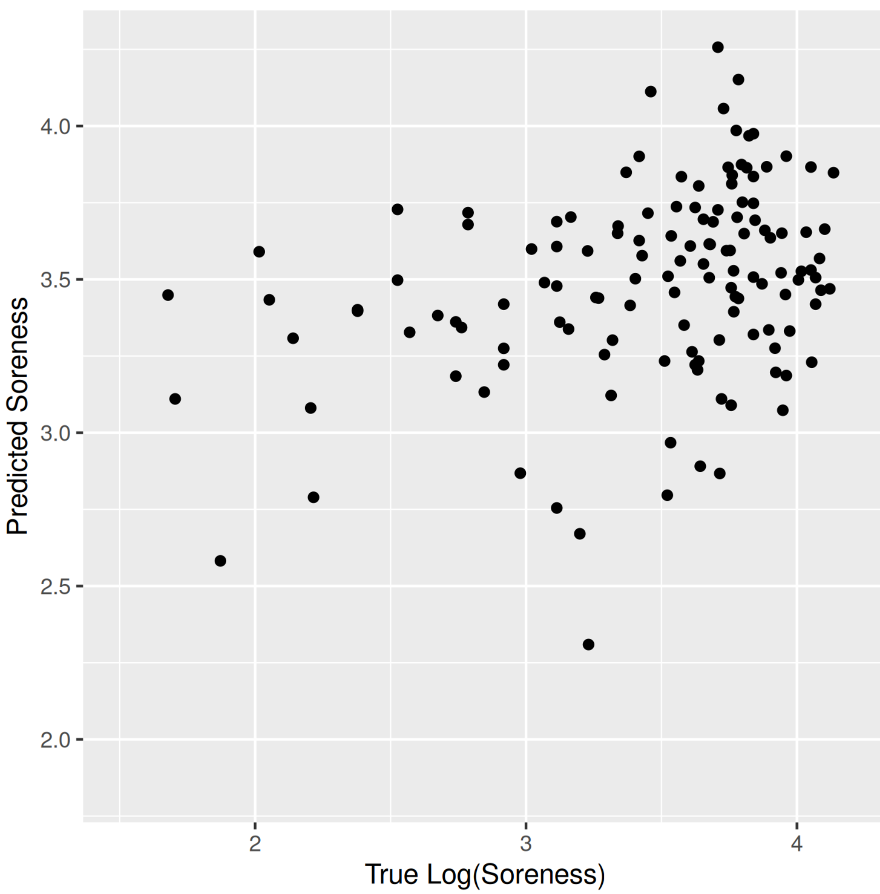

Figure 22: Scatterplot relating predicted sleep disturbance to true sleep disturbance.

- Soreness was negatively associated with sleep to a much greater degree than with any other variable (see 21). Competition load may positively influence soreness.
- Athletes that performed more strength and conditioning in a given week experienced less stress (figure 15) and fatigue. However, these effects are likely very slight and possibly confounded by the positions of players engaging in the most strength and conditioning also performing less other training.

5.2 Directions for Future Work

One question which this report has not addressed is the extent to which the variation in workload and wellness data is simply a function of athlete position group. Although there are some trends which apply across the athlete groups (e.g., the impact of sleep), and while there are certainly athletes within a particular group that might not be best served by a model of their workload and wellness that relies on their position alone, there are probably some trends that could provide helpful coaching heuristics. This would be a fruitful direction for future work.

Very relatedly, it would be interesting to see if there are underlying latent positions that are different from those presented here. Principal component analysis or stochastic neighbor embedding (see [1]) could help discover if there are more underlying position than those described by the position groups and look for potential informative patterns there. Otremba [2] did something very similar with pitch categories in SmartPitch. Manifold learning could provide some directions for more tailored recommendations for players in these groups.

The generalized linear model with a gamma family was chosen because of the distribution of certain outcome variables. This is valuable from an inferential standpoint, as models with assumptions that are most appropriate to the data are most suited for inferential claims about the individual coefficients. However, it may be that a Bayesian ridge regression, random forest regressor, or even basic multilayer perceptron would outperform the generalized linear model in terms of prediction accuracy. Whether or not the tradeoff in interpretability is worthwhile for these models would depend somewhat on the magnitude of the increase in prediction accuracy. However, this tradeoff could also be ameliorated to some degree with the use of Shapley values [3].

References

[1] Geoffrey E Hinton and Sam Roweis. Stochastic neighbor embedding. *Advances in neural information processing systems*, 15, 2002.

[2] Stephen Eugene Otremba Jr. *SmartPitch: Applied machine learning for professional baseball pitching strategy.* PhD thesis, Massachusetts Institute of Technology, 2022.

[3] Erik Štrumbelj and Igor Kononenko. Explaining prediction models and individual predictions with feature contributions. *Knowledge and information systems*, 41:647–665, 2014.