The influence of dietary carbohydrate on perceived recovery status differs at the group and individual level – evidence of nonergodicity among endurance athletes

Authors: Jeffrey A. Rothschild[1*], Tom Stewart[1], Andrew E. Kilding[1], and Daniel J. Plews[1]

Affiliations:
[1] Sports Performance Research Institute New Zealand (SPRINZ), Auckland University of Technology, Auckland, New Zealand

*Corresponding author - Jeffrey Rothschild – Jeffrey.Rothschild@aut.ac.nz

Jeffrey Rothschild: ORCID: 0000-0003-0014-5878
Andrew Kilding: ORCID: 0000-0002-5334-8831

All authors have read and approved this version of the manuscript for pre-print.

Please cite as: Rothschild JA, Stewart T, Kilding AE, and Plews DJ. 2022. The influence of dietary carbohydrate on perceived recovery status differs at the group and individual level – evidence of nonergodicity among endurance athletes. SportRχiv.

Abstract

Purpose: Research findings are typically reported at the group level but applied to individuals. However, an emerging issue in sports science concerns nonergodicity — whereby group-level data cannot be generalized to individuals. The purpose of this study was to determine if the relationship between daily carbohydrate intake and perceived recovery status displays nonergodicity.

Methods: Fifty-five endurance athletes recorded daily measures of self-selected dietary intake, training, sleep, and subjective wellbeing for 12 weeks. We constructed linear models to measure the influence of daily carbohydrate intake on perceived recovery status while accounting for training load, sleep duration, sleep quality, and muscle soreness. Using linear model coefficients for carbohydrate intake we tested whether the distributions (mean and SD) differed at the group and individual levels (indicating nonergodicity). Additionally, a decision tree was created to explore factors that could provide an indication of an individual athlete's relationship between carbohydrate intake and perceived recovery status.

Results: Mean values were not different between group- and individual-level analyses, but SDs at the individual level were ~2.5 times larger than at the group level, indicating nonergodicity. Model coefficients for carbohydrate intake were negative for five participants, positive for four participants, and non-significant for 31 participants. The Kappa value measuring accuracy of the decision tree was 0.54, indicating moderate prediction accuracy.

Conclusion: For most individuals, carbohydrate intake did not influence recovery status. However, the influence of dietary carbohydrate intake on daily recovery differs at the group and individual level. Therefore, practical recommendations should be based on individual-level analysis.

Keywords: intraindividual variability, monitoring, carbohydrate, training load, decision tree, machine learning

List of Abbreviations:

ARIMA: Autoregressive Integrated Moving Average
CHO: Carbohydrate
CI: Confidence Intervals
PRS: Perceived Recovery Status
SD: Standard Deviation

1. Introduction

Sports nutrition guidelines recommend carbohydrate intake be modulated according to changes in exercise volume (Thomas et al., 2016), with the intention of optimizing training adaptation while ensuring adequate recovery. Under-fueling can cause low energy availability (Melin et al., 2019), while over-fueling can cause weight gain and potentially attenuate desired training adaptations (Bartlett et al., 2015). It is commonly reported that perceived ratings of wellness and recovery are sensitive to fluctuations in training load (Gastin et al., 2013; Thorpe et al., 2016), and sleep duration (Sawczuk et al., 2018). However, the influence of dietary intake on daily recovery during endurance training is less understood. During short-term periods of intensified endurance training, increasing energy and carbohydrate intake may attenuate symptoms of overreaching (Achten et al., 2004; Halson et al., 2004; Killer et al., 2017; Sousa et al., 2010), although it is unclear if this relationship between carbohydrate intake and daily recovery extends over longer time periods and/or across a range of training volumes in a practical setting.

Studies in sports science are typically conducted and reported at the group level yet applied at the individual level. It has been increasingly questioned whether group-level results can generalize to individuals (Chrzanowski-Smith et al., 2020; Fisher et al., 2018; Hill et al., 2021; Neumann et al., 2022), as group-level findings could conceal relevant inter-individual variability in response to a training stimulus or intervention (Glazier & Mehdizadeh, 2019). This could lead to sub-optimal training or nutrition prescriptions for an individual athlete. When the group-level variability of data does not resemble the individual-level variability, or when individual-level variability exhibits changing variance over time, the data is non-ergodic (Mangalam & Kelty-

Stephen, 2021; Molenaar & Campbell, 2009). Nonergodicity could lead studies to overestimate the accuracy of aggregated statistical estimates and in turn, the generalizability of conclusions between the group and individual. In light of this, nonergodicity has been suggested as a disservice to human subjects research (Fisher et al., 2018).

As previously described by Molenaar and Campbell (2009) and Neumann et al. (2022), a Cattell data box can be used to determine ergodicity of a given data set. This can be visualized as a 3-dimensional box with time, measured variables, and individual subjects as the dimensions. For group-level analysis of a variable of interest, a single time point is pooled across all subjects (e.g., all subjects on day 1), repeated for each additional time point, and summarized (e.g., mean, standard deviation (SD), confidence intervals (CI), etc.). For individual-level analysis the variable is analyzed across all time points separately for each subject and then summarized. If the structure of the group- and individual-level data differ (e.g., statistics of central tendencies, variations, and/or correlations of time series data), the process is considered non-ergodic and results obtained from standard analysis at the group level cannot be applied to the individual (Molenaar & Campbell, 2009).

Nonergodicity is relevant in the context of nutrition and training, as evidence-based practitioners and athletes often apply group-level research findings to the individual (Fullagar et al., 2019; Thomas et al., 2016). Therefore, the purpose of this study was to examine the relationship between daily carbohydrate intake and perceived recovery status and determine if group-level statistics can generalize to individual athletes. To do so, 55 endurance athletes recorded daily

measures of self-selected nutrition intake, exercise training, sleep habits, and subjective wellbeing for 12 weeks. We constructed linear models to measure the influence of daily carbohydrate intake on perceived recovery status the following morning while accounting for other factors such as training load, sleep, and muscle soreness. Using the model coefficient for carbohydrate intake we tested whether the distributions (mean and SD) differed at the group and individual levels. As an exploratory analysis, we also created a decision tree model to understand general traits of athletes that would predict a positive, negative, or non-significant model coefficient for carbohydrate intake. This could serve as the next step in understanding individual-level differences, and provide a direction for coaches and practitioners to make better decisions to support the individual athlete's needs.

2. Methods

2.1 Study design

Self-selected nutrition intake, exercise training, sleep habits, and subjective wellbeing of endurance athletes were monitored daily over a 12-week period. Throughout the study period, participants were free to perform any type of exercise and consume any type of diet. Results presented herein are from a wider study of endurance training and recovery. Data related to carbohydrate periodization (Jeffrey A Rothschild et al., 2022) and machine learning predictions (J.A. Rothschild et al., 2022) have been reported elsewhere. The study was open to male and females aged 18 or older who train at least seven hours per week, were using a smartphone app

to track their dietary intake at least five days per week, captured HRV daily, and tracked sleep using a wearable device. All study protocols and materials were approved by the Auckland University of Technology Ethics Committee (22/7), and all participants provided informed consent prior to starting the study.

2.2 Participants

Fifty-five endurance athletes (61.8% male, aged 42.6 ± 9.1 years, training 11.6 ± 3.9 hours per week) took part in the study. The primary sports represented were triathlon (67.3%), running (20.0%), cycling (10.9%), and rowing (1.8%). The self-reported competitive level included professional (2.6%), elite non-professional (qualify and compete at the international level as an age-group athlete, 34.6%), high-level amateur (qualify and compete at National Championship-level events as an age-group athlete, 25.6%), and amateur (enter races but don't expect to win, or train but do not compete, 37.2%) athletes.

2.3 Assessment of self-reported exercise

All exercise was recorded in Training Peaks software (TrainingPeaks, Louisville, CO, USA). Each session was noted for modality (e.g., bike, run, swim), duration, and session rating of perceived exertion (sRPE (Foster et al., 2021)) using the Borg CR100® scale, which offers additional precision compared with the CR10 scale (Clemente et al., 2019). Participants were instructed to rate their

perceived effort for the whole training session within 1-h of exercise, although sRPE scores are temporally robust from minutes to days following a bout of exercise (Foster et al., 2021).

2.4 Assessment of self-reported dietary intake

Details of dietary assessment have been described elsewhere (Jeffrey A Rothschild et al., 2022). Briefly, participants were instructed to maintain their typical dietary habits and record all calorie-containing food and drink consumed for the duration of the 12-week study, using the MyFitnessPal application (www.myfitnesspal.com) (Evenepoel et al., 2020). Incomplete days of tracking (2.2 ± 4.6% of days per participant) were removed from the data, and analysis of the calorie intake trend over time was performed for each participant as an additional check of reporting compliance (Jeffrey A Rothschild et al., 2022). Four participants were excluded from the analysis due to the detection of a downward trend in dietary reporting that could not be explained by changes in training load or body weight.

2.5 Assessment of sleep and subjective wellbeing

Nightly sleep duration was recorded using wearable devices, which included Oura ring, Whoop strap, Applewatch, Fitbit, and Garmin models as previously described (J.A. Rothschild et al., 2022). These consumer-grade devices offer adequate accuracy in detecting sleep-wake times, but not sleep staging (Chee et al., 2021; Chinoy et al., 2021; Miller et al., 2020; Roberts et al., 2020; Zaffaroni et al., 2019). Each morning participants answered four questions related to

subjective wellbeing based on the recommendations of Hooper and Mackinnon (1995). The perceived recovery status (PRS) scale (Laurent et al., 2011) was used to measure overall recovery with athletes manually typing a number into Training Peaks software. The 100-point version of the scale was used, which has been shown to provide more accurate measures of recovery than the 10-point scale (Clemente et al., 2019). In addition, ratings of life stress (1–7), sleep quality (1–7), and muscle soreness (1–10) were also recorded into the software each morning. Participants were familiarized with all scales prior to starting the study.

2.6 Data preparation

Training load was calculated for each workout as the product of sRPE and duration of exercise in minutes (Haddad et al., 2017), divided by 10 to account for the 100-point scale, and summed into daily totals. External load metrics such as heart rate, power, or pace were not collected because many athletes undertake activities that can't be quantified on a common scale such as strength training, yoga, or swimming without a HR monitor, and also because the sRPE is considered to be a valid and reliable method for calculating training load across modalities (Haddad et al., 2017). Seven-day rolling measures for training monotony (a measure of day-to-day variability in the weekly training load, calculated as average daily load divided by the standard deviation) and training strain (product of total weekly training load and training monotony) were calculated (Haddad et al., 2017). A sleep index score was calculated as the product of sleep duration and sleep quality (Sawczuk et al., 2021). Dietary macronutrient intake was converted to a relative intake (g per kg body mass) to allow for appropriate comparison between athletes.

Participants were excluded from the analysis if they were training on average less than 6 h per week (n = 8) or did not log at least 85% of the required data points (n = 3). Participants who did not complete the full 12 weeks due to illness, injury, or drop-out but completed at least 6 weeks of tracking were included in the analysis (n = 11). Among participants included in the analysis (n = 40), 2.5 ± 1.7 % of data points were missing. Missing values were imputed at the individual level using multiple linear regression and nearest neighbor algorithms for diet and training measures and using median values for other variables (Kuhn & Johnson, 2013).

2.7 Analysis

Following the recommendations of previous studies (Fisher et al., 2018; Neumann et al., 2022), we extracted a subset of data that was symmetrical (i.e., an equal number of participants and observations per participant) to equalize statistical power for analysis at the group and individual levels. Because we had 40 participants in the final analysis, 40 consecutive days were chosen beginning with day 8 to allow for an accurate calculation of training strain (which reflects the previous 7 days of training). Repeated measures correlation (Bakdash & Marusich, 2017) was used at the group level to examine the bivariate relationship between the morning (AM) PRS score and prior day carbohydrate intake. Pearson or Spearman correlations, depending on normality of the data as determined by the Shapiro-Wilk test, were used to examine the bivariate relationship between the AM PRS score and prior day carbohydrate intake for each individual.

Previous studies of ergodicity have focused on comparisons of univariate distributions and bivariate correlations (Fisher et al., 2018; Neumann et al., 2022). However, the relationship between diet and recovery is likely also dependent on other factors relating to training and sleep. To account for this, linear regression models were constructed with AM PRS score specified as the dependent variable, and carbohydrate intake (g/kg) from the previous day, training load from the previous day, training strain (encompassing the previous 7 days), muscle soreness, and sleep index specified as independent variables. These variables were chosen because they had the highest importance scores in our predictive modeling study (J.A. Rothschild et al., 2022). The model coefficient for carbohydrate intake was the primary variable of interest. For group level analysis, models were made for all 40 athletes together on day 1 and repeated for each of the 40 days with the results summarized across days (as mean, SD, and 95% CIs). For individual-level analysis, a separate model was created for each athlete, and the results were then summarized. However, data at the individual level are a time series, which refers to a sequence of data points at equally spaced points in time and ordered chronologically (Schaffer et al., 2021). Time series data cannot be analyzed with common techniques such as linear modeling if the day-to-day observations are correlated with observations at previous time points (i.e., auto-correlated) and are not independent of each other, as key assumptions of linear regression are violated (James et al., 2021). Autoregressive Integrated Moving Average (ARIMA) models are commonly used in time series analysis to account for these issues (Schaffer et al., 2021). Therefore, for individual-level analyses we constructed ARIMA models using the Hyndman-Khandakar algorithm for automatic ARIMA modelling (Hyndman & Khandakar, 2008). To determine statistical significance,

95% CIs for means and SDs of the unstandardized regression coefficients were calculated, and values were considered significantly different if the 95% CIs did not overlap. Ergodicity can be confirmed if the mean and SD at the group and individual levels were not significantly different (Molenaar & Campbell, 2009). R-squared (R^2) was calculated as an overall measure of model accuracy.

To explore characteristics which might inform the individual responses to carbohydrate intake, a decision tree model was created to predict the classification of statistical significance for the model coefficient of prior day carbohydrate intake from the individual ARIMA models (non-significant, significantly positive, or significantly negative). The coefficients were organized into these three categories with practical application in mind. That is, coaches or nutritionists might benefit more from knowing if/how an individual responds to carbohydrate in this context, rather than getting a predicted model coefficient for the individual athlete.

Variables used in the decision tree model were age, training age, competitive level, primary sport, sex, BMI, percentage of training days performing fasted-state training, and average values of daily kcal intake (kcal/kg), daily carbohydrate, fat, and protein intake (g/kg), carbohydrate monotony (mean daily intake/SD), weekly training volume (h), training monotony, and training strain. All available data points were used for the decision tree models (n = 3,285, 82.1 ± 9.6 days per participant), rather than the 40-d subset used to compare group vs. individual responses to

obtain the most accurate picture of each individual's characteristics. Modeling was performed in R using the Tidymodels ecosystem (Kuhn & Wickham, 2020). Hyperparameters were tuned using 100 bootstrap resamples and model accuracy was established using 500 bootstrap resamples. Class imbalances were handled by up-sampling prior to tuning. Cohen's Kappa was used as the primary accuracy measure due to the imbalanced, multi-class nature of the outcome variable. Kappa accounts for the accuracy that would be generated simply by chance, producing values between −1 and 1. We interpret these values using the guidelines of Landis and Koch (1977), with values of 0–0.20 considered slight, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 substantial, and 0.81–1 as almost perfect. In addition, we report positive predictive value and negative predictive value (Altman & Bland, 1994). All analyses were carried out with R version 4.0.3 (The R foundation for Statistical Computing, Vienna, Austria). Descriptive statistics are provided as mean ± SD.

3. Results

During the 40-d period selected for the primary analysis, average participant training volume was 11.7 ± 3.3 h per week. Mean daily dietary intake was 38.8 ± 8.7 kcal/kg, 3.9 ± 1.5 g/kg carbohydrate, 1.9 ± 0.4 g/kg protein, and 1.7 ± 0.6 g/kg fat, and average sleep duration was 7.5 ± 0.7 hours per night. Bivariate repeated-measures correlation at the group level revealed a significant negative relationship between AM PRS and carbohydrate ingestion the prior day (r = -

0.09, 95% CI -0.14 to -0.04, $p < 0.001$), but this relationship varied considerably among individuals (Figure 1).

After accounting for prior day training load, 7-day training strain, muscle soreness, and sleep index via linear modeling, model coefficients for carbohydrate intake were negative for five participants (13%), positive for four participants (10%), and non-significant for 31 participants (78%, Figure 2). Mean values for model coefficients were similar between the group and individual (evidenced by overlapping CIs), whereas SDs were different, (i.e., non-overlapping CIs) indicating nonergodicity (Figure 3). Nonergodicity was also observed in the overall model accuracy. Mean R-squared values of the linear models were 0.32 (95% CI 0.29, 0.36), and 0.40 (95% CI 0.34, 0.46), for the group and individual models, respectively, and SD values were 0.11 (95% CI 0.09, 0.14) and 0.18 (95% CI 0.15, 0.23) for the group and individual models, respectively.

A decision tree was created to explore potential factors that could provide coaches or practitioners with an indication of an athlete's relationship between carbohydrate intake and perceived recovery status (Figure 4). The Kappa value was 0.54, indicating a moderate level of agreement. Positive predictive value was 0.50, and negative predictive value was 0.86. A confusion matrix of actual and predicted classes is shown in Figure 5.

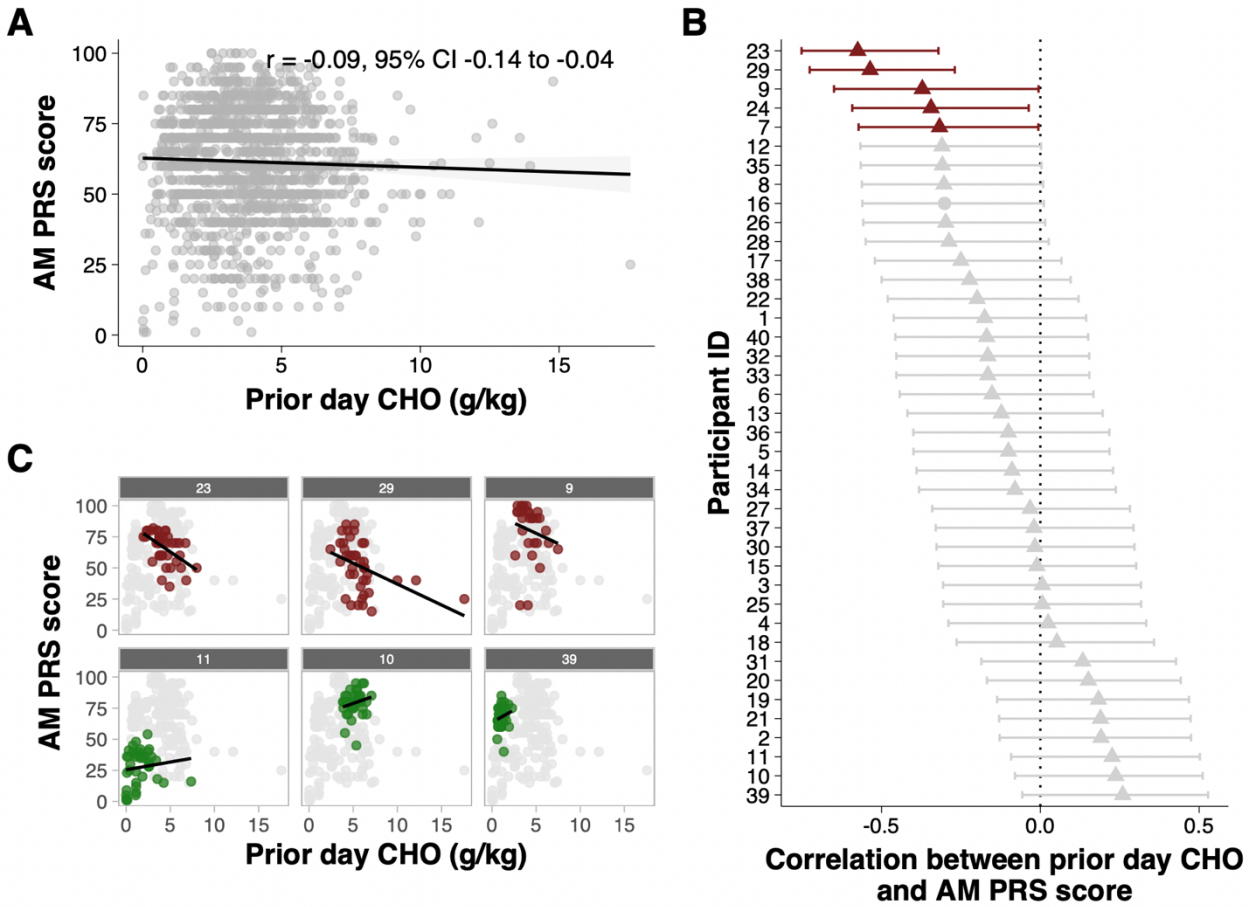


Figure 1. Bivariate correlations between AM perceived recovery status (PRS) and carbohydrate ingestion the prior day (g/kg). Values given in (A) used repeated-measures correlation analysis. Panel (B) shows Pearson (circle) or Spearman (triangle) correlation values and 95% confidence intervals for each participant, colored based on statistical significance ($p < 0.05$). Panel (C) shows example scatterplots for the participants with the three highest and three lowest correlation values. Numbers at the top of each panel in (C) relate to the participant ID shown in (B). The light grey points in (C) depict all points for the 6 participants shown in (C), with each individual's points shown in color (red for negative correlation values and green for positive).

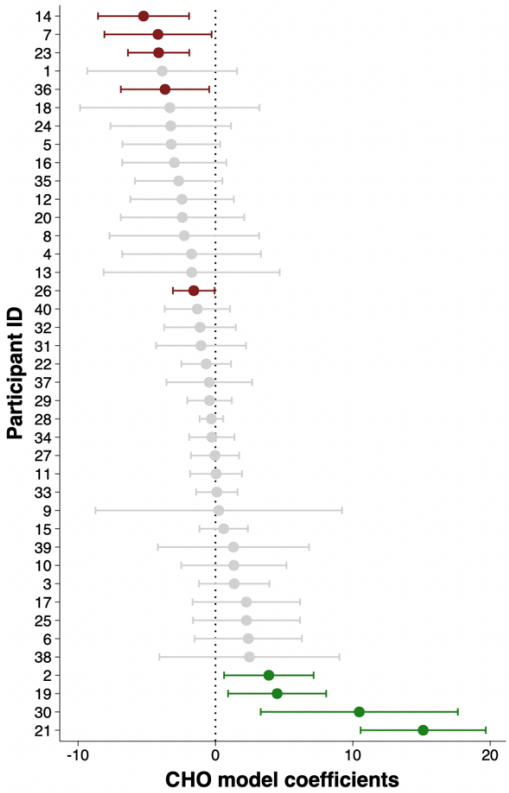
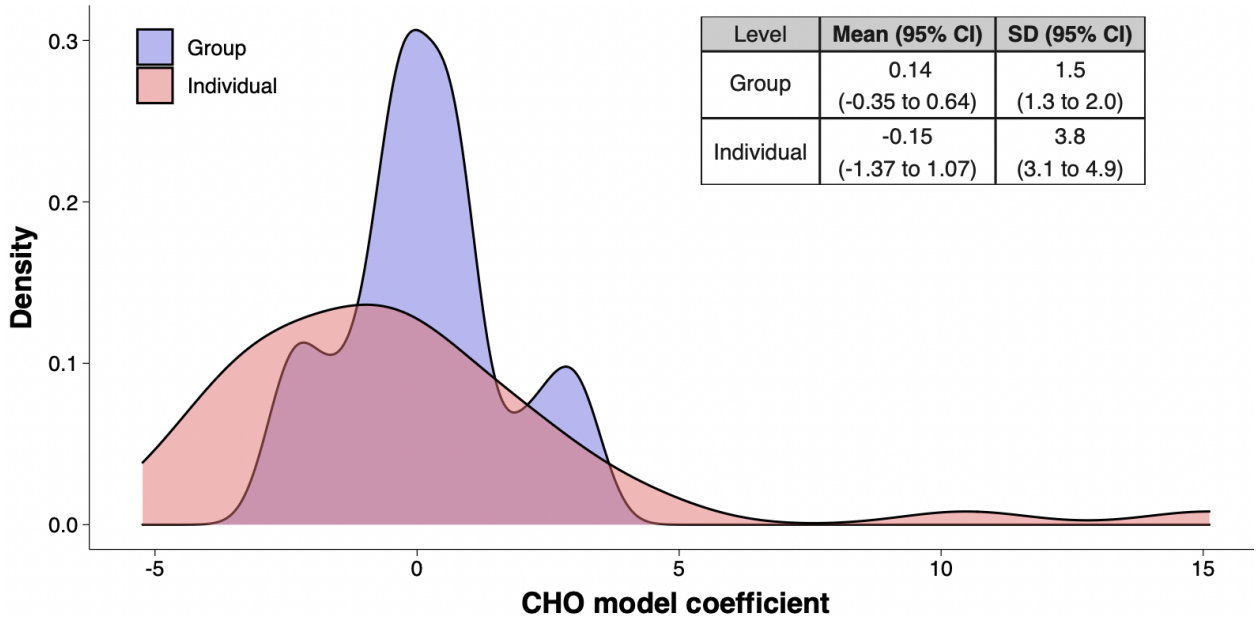


Figure 2. Individual participant model coefficients with 95% Confidence Intervals for the effect of prior day carbohydrate (CHO) intake (g/kg) on AM Perceived Recovery Status (PRS) score after accounting for prior day training load, 7-d training strain, muscle soreness, sleep index (product of sleep duration and sleep quality), using Autoregressive Integrated Moving Average (ARIMA) modeling. This can be interpreted as a change of AM PRS in the amount shown on the x-axis for every 1 g/kg increase in daily CHO intake, after holding everything else constant. Green indicates statistically significant positive values, red indicates statistically significant negative values, and grey indicates non-significant values ($p > 0.05$).



Level	Mean (95% CI)	SD (95% CI)
Group	0.14 (-0.35 to 0.64)	1.5 (1.3 to 2.0)
Individual	-0.15 (-1.37 to 1.07)	3.8 (3.1 to 4.9)

Figure 3. Density plots of the model coefficients for the effect of prior day carbohydrate (CHO) intake (g/kg) on AM Perceived Recovery Status (PRS) score after accounting for prior day training load, 7-d training strain, muscle soreness, and sleep index (product of sleep duration and sleep quality). Inset table shows mean, SD, and 95% confidence intervals for model coefficients for carbohydrate intake from group and individual level modeling.

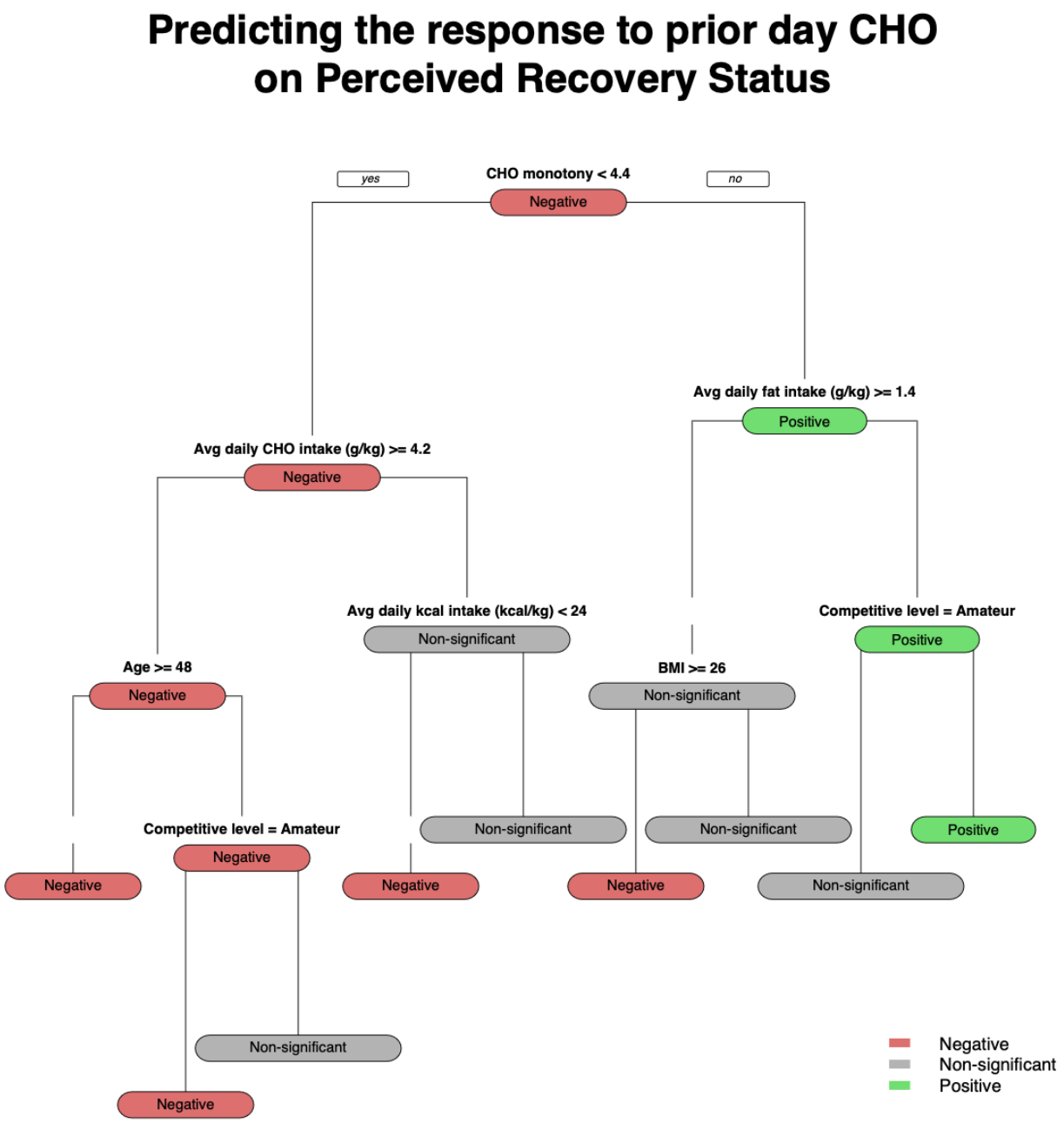


Figure 4. Decision tree predicting the response to prior day carbohydrate (CHO) intake on AM Perceived Recovery Status. Each node indicates the predicted class (negative, non-significant, or positive model coefficients). At each level, following the node to the left corresponds to yes, and following the node to the right corresponds to no.

Predicted class \ True class	Negative	Non-significant	Positive
Negative	5	1	4
Non-significant	0	27	0
Positive	0	3	0

Figure 5. Confusion matrix of actual and predicted classes for the decision tree model predicting the response to prior day carbohydrate ingestion on AM Perceived Recovery Status. Values shown in the dark green boxes indicate the number of correct predictions for each class.

4. Discussion

The aim of this study was to examine the relationship between daily carbohydrate intake and perceived recovery status and determine if group-level statistics can be generalized to individual athletes. The main outcomes are 1) daily carbohydrate intake does not influence perceived recovery status the following morning for most athletes, after accounting for other influential variables such as training load, muscle soreness, and sleep, 2) for those that are affected the influence can be positive or negative, 3) the data are non-ergodic, meaning group-level findings cannot be generalized to the individual, and 4) we build upon previous work using bivariate correlations to include linear model coefficients and offer a method for understanding the individual responses through a decision tree algorithm.

We observed a large discrepancy between inter- and intra-individual variation (i.e., nonergodicity), as SDs at the individual level were ~2.5 times larger than at the group level. This means there would be a difference when computing statistics by first averaging the data before the calculations versus first calculating the statistics for each individual before averaging these results (Hill et al., 2021). Furthermore, mean values may be misleading when determining the influence of carbohydrate intake on AM PRS. At the group level, a traditional interpretation would suggest carbohydrate has minimal influence on AM PRS after accounting for the other variables. Although this would be true for most athletes (~78% of our participants), model coefficients were positive for 10% and negative for 13% of our participants (Figure 2). This means the individual, rather than the group, should be placed at the level of analysis to avoid wrong conclusions (Hill et al., 2021).

Previous studies have used bivariate correlations to explore ergodicity (Fisher et al., 2018; Neumann et al., 2022). In this context, bivariate correlations could be misleading because athletes often increase carbohydrate intake on days with higher training loads (Jeffrey A Rothschild et al., 2022). Because of the multifactorial nature of day-to-day recovery, we created linear models to account for these additional factors while focusing the analysis on daily carbohydrate intake. Subjective muscle soreness and sleep index were included because they are two of the most important factors predicting AM PRS scores, as reported by us (x) and others (Gastin et al., 2013; Sawczuk et al., 2021; Thorpe et al., 2016). Training strain was included in the model to account for potential residual fatigue from the previous seven days of training. Training

strain (the product of training load and training monotony) is high when high training loads are combined with low variability of load, and low when athletes complete either low training loads or have regular variation in training (Haddad et al., 2017). Together, these variables account for a substantial amount of the variance in PRS scores and allow a more focused look at the influence of carbohydrate intake.

The beliefs and practices surrounding nutrition and training vary widely among athletes (Rothschild et al., 2020; Rothschild et al., 2021). Although it could be tempting to try and find unifying answers to some of the contrasting beliefs held by athletes (e.g., the positive or negative influence of fasted-state training or increasing carbohydrate intake), the current study underscores the idea that what's best for one athlete may not be best for another. It is also noteworthy that athletes in this study were undertaking self-selected training programs, and results cannot be generalized to short-term periods of intensified training, where increasing energy and/or carbohydrate intake has been shown to attenuate symptoms of overreaching (Achten et al., 2004; Halson et al., 2004; Killer et al., 2017; Sousa et al., 2010).

As a way of translating the interindividual variability from a statistical concept to practical application, a decision tree model was created. Variables such as age, sex, BMI, competitive level, training volume, and habitual dietary patterns were included to better understand what traits or qualities might be related to a certain response to carbohydrate intake. Although interpretation of the decision tree is challenged by the small number of athletes presenting significant model coefficients for carbohydrate intake and the inability of the model to accurately predict positive

coefficients (Figure 5), it can serve as a starting point for understanding how an athlete might be expected to respond to carbohydrate intake. The most important variables were carbohydrate monotony, followed by average daily carbohydrate and fat intake. Among athletes with low carbohydrate monotony scores (i.e., larger daily variations in carbohydrate intake), those with a higher average daily carbohydrate intake were likely to have a negative response to carbohydrate intake whereas those with a lower daily carbohydrate intake were more likely to have a non-significant effect of carbohydrate intake on AM PRS score (Figure 4). Athletes with a higher daily fat intake were also less likely to be influenced by changes in daily carbohydrate intake (Figure 4). The model displayed moderate accuracy (Kappa value of 0.54), although the ability of the model to learn from the data was challenged by the small and imbalanced data set. As shown in Figure 5, negative and non-significant outcomes were able to be predicted very well, but the model did not accurately predict any positive responders. Nevertheless, we feel this approach can be adopted by others who wish to better understand individual responses to a given intervention or stimulus.

There are several limitations to this study, primarily related to the use of self-report measures. Data integrity was checked based on the number of missing values, and by looking for trends in dietary reporting that could not be explained by changes in training load or body weight. However, it is possible that participants did not always enter data as accurately as possible. There is also the risk of bias in reporting if an athlete is aware that a coach or a researcher will be seeing their data, answering based on what they think is desirable. In addition, the limited number of

data points, particularly with unbalanced classes, made training and interpreting the decision tree model challenging.

5. Conclusion

Our findings suggest the influence of dietary carbohydrate intake on daily recovery differs at the group and individual level. Therefore, inferences may not be generalized from the group to the individual, and practical recommendations should be based on individual analysis. Furthermore, at the group level, the previous day's carbohydrate intake did not influence the perceived recovery status of athlete training ~12 h per week. This research also adds to the literature around ergodicity in sports science, an emerging concept that should be routinely considered as part of the statistical analysis process. Future research in athletes should focus on individual responses in order to prescribe the correct nutrition, and likely training, to maximize each athlete's performance and training.

Statements and Declarations

Ethics approval This study was performed in line with the principles of the Declaration of Helsinki. All study protocols and materials were approved by the Auckland University of Technology Ethics Committee (22/7), and all participants provided informed consent prior to starting the study.

Competing Interests No sources of funding were used in the preparation of this manuscript. The authors have no conflicts of interest related to this work.

Author contributions JAR conceived and designed the research, collected data, performed the analysis, and wrote the first draft of the manuscript. All authors edited and revised the manuscript and approved the final version of the manuscript.

Code availability

The R code used in this analysis is publicly available at https://github.com/Jeffrothschild/nonergodicity_code

References

Achten, J., Halson, S. L., Moseley, L., Rayson, M. P., Casey, A., & Jeukendrup, A. E. (2004). Higher dietary carbohydrate content during intensified running training results in better maintenance of performance and mood state. *J Appl Physiol (1985), 96*(4), 1331-1340. doi:10.1152/japplphysiol.00973.2003

Altman, D. G., & Bland, J. M. (1994). Diagnostic tests 2: Predictive values. *BMJ, 309*(6947), 102. doi:10.1136/bmj.309.6947.102

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated Measures Correlation. *Front Psychol, 8*, 456. doi:10.3389/fpsyg.2017.00456

Bartlett, J. D., Hawley, J. A., & Morton, J. P. (2015). Carbohydrate availability and exercise training adaptation: too much of a good thing? *Eur J Sport Sci, 15*(1), 3-12. doi:10.1080/17461391.2014.920926

Chee, N., Ghorbani, S., Golkashani, H. A., Leong, R. L. F., Ong, J. L., & Chee, M. W. L. (2021). Multi-Night Validation of a Sleep Tracking Ring in Adolescents Compared with a Research Actigraph and Polysomnography. *Nat Sci Sleep, 13*, 177-190. doi:10.2147/NSS.S286070

Chinoy, E. D., Cuellar, J. A., Huwa, K. E., Jameson, J. T., Watson, C. H., Bessman, S. C., . . . Markwald, R. R. (2021). Performance of seven consumer sleep-tracking devices compared with polysomnography. *Sleep, 44*(5). doi:10.1093/sleep/zsaa291

Chrzanowski-Smith, O. J., Piatrikova, E., Betts, J. A., Williams, S., & Gonzalez, J. T. (2020). Variability in exercise physiology: Can capturing intra-individual variation help better understand true inter-individual responses? *Eur J Sport Sci, 20*(4), 452-460. doi:10.1080/17461391.2019.1655100

Clemente, F. M., Rabbani, A., & Araujo, J. P. (2019). Ratings of perceived recovery and exertion in elite youth soccer players: Interchangeability of 10-point and 100-point scales. *Physiol Behav, 210*, 112641. doi:10.1016/j.physbeh.2019.112641

Evenepoel, C., Clevers, E., Deroover, L., Van Loo, W., Matthys, C., & Verbeke, K. (2020). Accuracy of Nutrient Calculations Using the Consumer-Focused Online App MyFitnessPal: Validation Study. *J Med Internet Res, 22*(10), e18237. doi:10.2196/18237

Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proc Natl Acad Sci U S A, 115*(27), E6106-E6115. doi:10.1073/pnas.1711978115

Foster, C., Boullosa, D., McGuigan, M., Fusco, A., Cortis, C., Arney, B. E., . . . Porcari, J. P. (2021). 25 Years of Session Rating of Perceived Exertion: Historical Perspective and Development. *Int J Sports Physiol Perform, 16*(5), 612-621. doi:10.1123/ijspp.2020-0599

Fullagar, H. H. K., Harper, L. D., Govus, A., McCunn, R., Eisenmann, J., & McCall, A. (2019). Practitioner Perceptions of Evidence-Based Practice in Elite Sport in the United States of America. *J Strength Cond Res, 33*(11), 2897-2904. doi:10.1519/JSC.0000000000003348

Gastin, P. B., Meyer, D., & Robinson, D. (2013). Perceptions of wellness to monitor adaptive responses to training and competition in elite Australian football. *J Strength Cond Res, 27*(9), 2518-2526. doi:10.1519/JSC.0b013e31827fd600

Glazier, P. S., & Mehdizadeh, S. (2019). Challenging Conventional Paradigms in Applied Sports Biomechanics Research. *Sports Med, 49*(2), 171-176. doi:10.1007/s40279-018-1030-1

Haddad, M., Stylianides, G., Djaoui, L., Dellal, A., & Chamari, K. (2017). Session-RPE Method for Training Load Monitoring: Validity, Ecological Usefulness, and Influencing Factors. *Front Neurosci, 11*, 612. doi:10.3389/fnins.2017.00612

Halson, S. L., Lancaster, G. I., Achten, J., Gleeson, M., & Jeukendrup, A. E. (2004). Effects of carbohydrate supplementation on performance and carbohydrate oxidation after intensified cycling training. *J Appl Physiol (1985), 97*(4), 1245-1253. doi:10.1152/japplphysiol.01368.2003

Hill, Y., Meijer, R. R., Van Yperen, N. W., Michelakis, G., Barisch, S., & Den Hartigh, R. J. (2021). Nonergodicity in protective factors of resilience in athletes. *Sport Exerc Perform Psychol, 10*(2), 217.

Hooper, S. L., & Mackinnon, L. T. (1995). Monitoring overtraining in athletes. Recommendations. *Sports Med, 20*(5), 321-327. doi:10.2165/00007256-199520050-00003

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *J Stat Softw, 27*, 1-22.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning : with applications in R*. New York: Springer.

Killer, S. C., Svendsen, I. S., Jeukendrup, A. E., & Gleeson, M. (2017). Evidence of disturbed sleep and mood state in well-trained athletes during short-term intensified training with and without a high carbohydrate nutritional intervention. *J Sports Sci, 35*(14), 1402-1410. doi:10.1080/02640414.2015.1085589

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* (Vol. 26): Springer.

Kuhn, M., & Wickham, H. (2020). Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles. Retrieved from https://www.tidymodels.org

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/843571

Laurent, C. M., Green, J. M., Bishop, P. A., Sjokvist, J., Schumacker, R. E., Richardson, M. T., & Curtner-Smith, M. (2011). A practical approach to monitoring recovery: development of

a perceived recovery status scale. *J Strength Cond Res, 25*(3), 620-628. doi:10.1519/JSC.0b013e3181c69ec6

Mangalam, M., & Kelty-Stephen, D. G. (2021). Point estimates, Simpson's paradox, and nonergodicity in biological sciences. *Neurosci Biobehav Rev, 125*, 98-107. doi:10.1016/j.neubiorev.2021.02.017

Melin, A. K., Heikura, I. A., Tenforde, A., & Mountjoy, M. (2019). Energy Availability in Athletics: Health, Performance, and Physique. *Int J Sport Nutr Exerc Metab, 29*(2), 152-164. doi:10.1123/ijsnem.2018-0201

Miller, D. J., Lastella, M., Scanlan, A. T., Bellenger, C., Halson, S. L., Roach, G. D., & Sargent, C. (2020). A validation study of the WHOOP strap against polysomnography to assess sleep. *J Sports Sci, 38*(22), 2631-2636. doi:10.1080/02640414.2020.1797448

Molenaar, P. C., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current directions in psychological science, 18*(2), 112-117.

Neumann, N. D., Van Yperen, N. W., Brauers, J. J., Frencken, W., Brink, M. S., Lemmink, K., . . . Den Hartigh, R. J. R. (2022). Nonergodicity in Load and Recovery: Group Results Do Not Generalize to Individuals. *Int J Sports Physiol Perform, 17*(3), 391-399. doi:10.1123/ijspp.2021-0126

Roberts, D. M., Schade, M. M., Mathew, G. M., Gartenberg, D., & Buxton, O. M. (2020). Detecting sleep using heart rate and motion data from multisensor consumer-grade wearables, relative to wrist actigraphy and polysomnography. *Sleep, 43*(7). doi:10.1093/sleep/zsaa045

Rothschild, J., Kilding, A. E., & Plews, D. J. (2020). Prevalence and Determinants of Fasted Training in Endurance Athletes: A Survey Analysis. *Int J Sport Nutr Exerc Metab, 30*(5), 345-356. doi:10.1123/ijsnem.2020-0109

Rothschild, J. A., Kilding, A. E., & Plews, D. J. (2021). Pre-Exercise Nutrition Habits and Beliefs of Endurance Athletes Vary by Sex, Competitive Level, and Diet. *J Am Coll Nutr, 40*(6), 517-528. doi:10.1080/07315724.2020.1795950

Rothschild, J. A., Morton, J. P., Stewart, T., Kilding, A. E., & Plews, D. J. (2022). The quantification of daily carbohydrate periodization among endurance athletes during 12 weeks of self-selected training: presentation of a novel Carbohydrate Periodization Index. *medRxiv*.

Rothschild, J. A., Stewart, T., Kilding, A. E., & Plews, D. J. (2022). Predicting daily recovery during long-term endurance training using machine learning analysis. *SportRχiv*.

Sawczuk, T., Jones, B., Scantlebury, S., & Till, K. (2018). The influence of training load, exposure to match play and sleep duration on daily wellbeing measures in youth athletes. *J Sports Sci, 36*(21), 2431-2437. doi:10.1080/02640414.2018.1461337

Sawczuk, T., Jones, B., Scantlebury, S., & Till, K. (2021). Influence of Perceptions of Sleep on Well-Being in Youth Athletes. *J Strength Cond Res, 35*(4), 1066-1073. doi:10.1519/JSC.0000000000002857

Schaffer, A. L., Dobbins, T. A., & Pearson, S. A. (2021). Interrupted time series analysis using autoregressive integrated moving average (ARIMA) models: a guide for evaluating large-scale health interventions. *BMC Med Res Methodol, 21*(1), 58. doi:10.1186/s12874-021-01235-8

Sousa, M., Madsen, K., Simões, H., Pereira, R., Negrão, C., Mendonça, R., . . . Silva, M. (2010). Effects of carbohydrate supplementation on competitive runners undergoing overload

training followed by a session of intermittent exercise. *Eur J Appl Physiol, 109*(3), 507-516. doi:10.1007/s00421-010-1388-8

Thomas, D. T., Erdman, K. A., & Burke, L. M. (2016). American College of Sports Medicine Joint Position Statement. Nutrition and Athletic Performance. *Medicine and Science in Sports and Exercise, 48*(3), 543-568. doi:10.1249/MSS.0000000000000852

Thorpe, R. T., Strudwick, A. J., Buchheit, M., Atkinson, G., Drust, B., & Gregson, W. (2016). Tracking Morning Fatigue Status Across In-Season Training Weeks in Elite Soccer Players. *Int J Sports Physiol Perform, 11*(7), 947-952. doi:10.1123/ijspp.2015-0490

Zaffaroni, A., Coffey, S., Dodd, S., Kilroy, H., Lyon, G., O'Rourke, D., . . . Penzel, T. (2019). Sleep Staging Monitoring Based on Sonar Smartphone Technology. *Annu Int Conf IEEE Eng Med Biol Soc, 2019*, 2230-2233. doi:10.1109/EMBC.2019.8857033