SUSTAINABILITY & SPORTS SCIENCE JOURNAL

ORIGINAL ARTICLE

Evaluating WHOOP wearable metrics as predictors of Division I collegiate volleyball performance

Asha D. Phillips . *Department of Kinesiology. Lipscomb University. Nashville, TN, United States of America.*
Zachary F. Droll. *Department of Kinesiology. Lipscomb University. Nashville, TN, United States of America.*
Andrew J. Mauldin. *Department of Kinesiology. Lipscomb University. Nashville, TN, United States of America.*
Matthew D. Ruiz. *Department of Kinesiology. Lipscomb University. Nashville, TN, United States of America.*

ABSTRACT

Wearable devices are increasingly popular among athletes, yet their metrics' predictive value for sports performance remains unclear. Purpose: Evaluate whether WHOOP-derived metrics predict objective and subjective volleyball performance in Division I collegiate athletes. Methods: Fourteen University of Tennessee volleyball players (age 20 ± 1.44 years; playing experience 8.5 ± 3.01 years; WHOOP usage 2.08 ± 1.16 years) participated during off-season. WHOOP metrics included strain, recovery, sleep performance, and sleep debt. Performance outcomes included attacking efficiency, passing efficiency, and weekly Perceived Performance Team Sports Questionnaire (PPTSQ) scores. Linear regression and repeated-measures correlation analyses were conducted (α = .05). Results: Higher strain associated with reduced attacking efficiency (p = .0017, r^2 = .029), and better sleep performance associated with higher perceived performance (p = .0373, r^2 = .042). Despite statistical significance, associations showed weak predictive strength, accounting for minimal performance variability. Other metrics showed no significant relationships. Conclusion: WHOOP metrics showed minimal predictive value for volleyball performance in this off-season sample. However, associations involving strain and sleep suggest larger samples, and longer monitoring periods are needed to determine wearable data reliability for performance prediction.

Keywords: Health science, Athletic performance, Attacking efficiency, Strain, Training load.

Cite this article as:
Phillips, A. D., Droll, Z. F., Mauldin, A. J., & Ruiz, M. D. (2026). Evaluating WHOOP wearable metrics as predictors of Division I collegiate volleyball performance. *Sustainability and Sports Science Journal*, *4*(3), 182-191. https://doi.org/10.55860/HTMT4995

Corresponding author. *Department of Kinesiology. Lipscomb University. Nashville, TN, United States of America.*
E-mail: aphill10@go.olemiss.edu
Submitted for publication January 31, 2026.
Accepted for publication March 04, 2026.
Published May 07, 2026.
Sustainability and Sports Science Journal. ISSN 2990-2975.

Identifier: https://doi.org/10.55860/HTMT4995

INTRODUCTION

Wearable technology, including WHOOP and similar devices, is now commonly used to monitor athlete recovery and training load in applied sport settings. These devices provide continuous, non-invasive physiological data, however evidence supporting their ability to predict sport-specific performance remains limited. Despite this gap, adoption across professional organizations including the National Football League (NFL), the Professional Golfers' Association of America (PGA), and the Ladies Professional Golf Association (LPGA), reflects growing confidence in WHOOP for athlete monitoring. This industry wide confidence sparked the origins of this study to address if wearables could be used for more purposes than initially marketed.

Different technologies are used across the wearable market for 24/7 tracking, but WHOOP relies on photoplethysmography (PPG) to estimate heart rate (HR), heart rate variability (HRV), and sleep metrics. PPG sensors differ from electrocardiography (ECG) in its ability to collect continuous physiological data (Rehman et al., 2024). To test the reliability of PPG sensors, Bellenger et al. (2021) demonstrated that WHOOP 2.0 closely matched ECG for HR and HRV during sleep, and Rehman et al. (2024) found WHOOP 4.0 to provide dependable cardiovascular monitoring across 24-hour periods, with increased error during activity consistent with PPG limitations. These validation efforts are important because performance predictions depend not only on data availability, but also on measurement reliability. Bellenger et al. (2021) established that WHOOP 2.0 closely matched ECG for HR and HRV during sleep, and Rehman et al. (2024) found WHOOP 4.0 to provide dependable cardiovascular monitoring across 24-hour periods.

Error rates increase during physical activity, which is consistent with known limitations of PPG. Overall, the study provided data that confirms WHOOP 4.0 is a reliable wearable for day and night HR and HRV measurements; however, it is important to note that activity types, day versus night collection, and sensor contact can add variability in these measurements. Sleep validation studies using WHOOP 3.0 also report acceptable accuracy for total sleep time and sleep–wake detection, though multi-stage sleep classification remains weaker (Miller et al., 2020; Berryhill et al., 2020).

While WHOOP provides valid physiological measurements, fewer studies have evaluated whether these data predict sport-specific performance. Some evidence links increased REM and total sleep duration with improved running performance (Allen et al., 2016). It is important to note, HRV findings remain inconsistent across studies, likely due to individual variability and demographic influences (Flatt et al., 2017; Makivic et al., 2013). Studies also indicate that, HRV shows relevance for psychological readiness, with lower HRV associated with reduced emotional self-regulation (Horvath et al., 2022).

Some studies have reported improved performance prediction when wearables metrics are combined with subjective wellness measures. However, these approaches often require extensive data collection, which can be time-consuming and may reduce athlete compliance in applied settings. (Taber et al., 2024; de Leeuw et al., 2022). However, identifying the most influential physiological and psychological predictors remains the highest priority. As a result, there is uncertainty whether wearable-derived metrics alone can provide meaningful insight into sport-specific performance. In turn, this led to the study design to assess the practicality and usefulness of using a wearable to track many variables that may have performance influences. Therefore, the purpose of this study was to examine whether WHOOP-derived metrics predict objective and subjective volleyball performance in Division 1 athletes. We hypothesized that WHOOP metrics would provide little predictive value for accurately predicting sport specific outcomes.

MATERIAL AND METHODS

Participants

In total, fourteen members of the University of Tennessee Spring 2025 volleyball team [age: 20 ± 1.44 years; playing experience: 8.5 ± 3.01 years; WHOOP (WHOOP Strap 4.0; WHOOP Inc., Boston, MA, USA) Usage history: 2.08 ± 1.16] years completed the study. All participants were regular WHOOP 4.0 users and provided individualized physiological data collection for eight weeks.

Table 1. Descriptive statistics of the sample n = 14.

Variable	Mean ± SD
Age (yrs)	20.21 ± 1.52
Playing experience (yrs)	8.5 ± 3.01
WHOOP experience (yrs)	2.07 ± 1.18
College experience (yrs)	2.79 ± 1.13

Measures

Physiological data

Each participant's WHOOP device collected continuous data, uploaded to a team cloud system, and subsequently exported to Excel for analysis. This study analyzed four WHOOP metrics: strain, recovery, sleep performance, and sleep debt.

Strain: Daily expenditure and workout intensity measured on a 0-21 scale using heart rate, accelerometer data, and cardiovascular/muscular load. Scores above 18 indicate significant physiological stress (WHOOP, 2023).

Recovery: Daily measurement of body readiness for performance, incorporating heart rate variability, resting heart rate, sleep performance, and respiratory rate.

Sleep performance: Ratio of actual sleep duration to individual sleep needs, calculated using sleep stages, daily strain, and sleep debt.

Sleep debt: Cumulative measurement of unmet sleep time in minutes, incorporating strain score, naps, and sleep duration.

Psychological data

Performance perception data were collected using the validated Perceived Performance in Team Sports Questionnaire (PPTSQ) (Gershgoren et al., 2021). Participants received a Google Forms survey each Friday and were required to submit it by Sunday night. This schedule aligned with the study’s goal of capturing consistent, week-to-week changes in perceived workload and wellness.

Performance data

Volleyball performance outcomes were collected weekly from a practice coder, attacking efficiency and passing efficiency were primary performance measures.

Attacking efficiency: Measured the effectiveness of attack attempts converted to scored points using the formula: (Kills - Errors) / Total Attacks. This measurement was chosen as the main statistic used to assess offensive success; higher attacking efficiency correlates with greater offensive success (Sitti et al., 2025).

Passing efficiency: Assessed on a 3-point scale commonly used in collegiate settings, measuring passing quality based on offensive options created. This measurement was chosen as one of the most controlled settings in the sport, and many coaches use this metric to assess passing quality in competitions.

Figure 1. Representative of half a volleyball court, pass location determines a score of 0, 1, 2 or 3.

Procedures

Compliance records were determined by complete and accurate daily data obtained by WHOOP and weekly completion of PPTSQ. WHOOP compliance was 100%, and PPTSQ response rate was 93.75%. All participants provided written informed consent prior to participation. The study protocol was reviewed and approved by the Lipscomb University Institutional Review Board. All procedures were conducted in accordance with the Declaration of Helsinki and relevant institutional guidelines.

Statistical analysis

To address which WHOOP-derived metrics predict volleyball performance outcomes, analyses were performed assessing predictor strength versus volleyball performance in practice and scrimmage scenarios. Three analytical subsets were created: practice data, scrimmage data, and weekly practice averages versus weekly PPTSQ responses. Relationships examined included: strain versus passing/attacking, sleep performance versus passing/attacking, sleep debt versus passing/attacking, and recovery versus passing/attacking. For practice and scrimmage data, linear regression models (lm function in base R) and repeated measures correlation analyses (rmcorr package, Version 0.5.2; Bakdash & Marusich, 2017) were conducted. Linear regression assessed overall population-level relationships, while repeated measures correlation accounted for within-subject dependencies across multiple observations. PPTSQ responses were analyzed using linear regression models only, as these represented aggregated weekly data.

All analyses were performed using R (Version 4.4.1; R Core Team, 2024) in RStudio. All data sets were initially analyzed using linear regression models. Subsequently, repeated measures correlation analyses were conducted using the rmcorr package (Version 0.5.2; Bakdash & Marusich, 2017) to identify individual-level relationships between variables while accounting for within-subject dependencies. A 95% confidence interval was calculated for all estimates, and the alpha level was set at $\alpha = .05$.

RESULTS

Practice

Practice subset means: Strain = 16.68 ± 2.77, Recovery = 64.91 ± 21.3, Sleep Debt = 55.09 ± 40.5 minutes, Sleep Performance = 81.47 ± 15.3, Attacking Efficiency = 0.206 ± 0.286, Passing Efficiency = 1.803 ± 0.744. No statistically significant relationships were identified between passing efficiency and WHOOP predictors.

Linear regression analysis found higher strain was significantly associated with reduced attacking efficiency (p = .002), but the effect size was small (r^2 = .029; ~3% variance explained). Due to minimal explained variance, the results do not support that strain is strongly associated with the decrease in attacking efficiency and provides little practical predictive power. Repeated measures correlation analysis found no significance (p = .119).

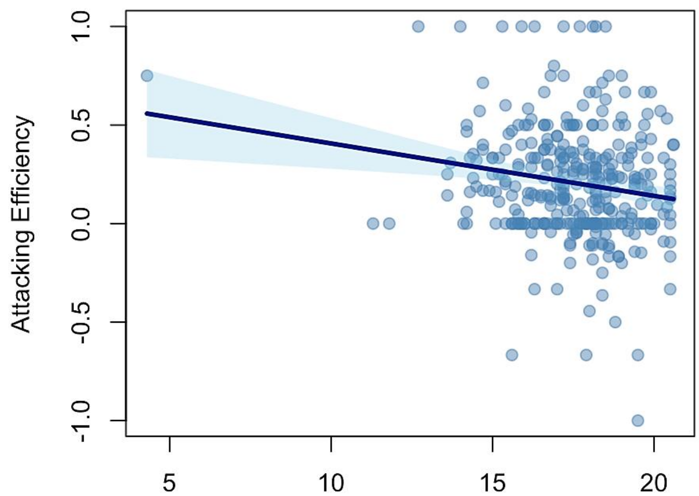

Figure 2. Linear regression – relationship between strain and attacking efficiency. Statical significance found (p = .0017, F(1,131) = 9.969, r^2 = .02924).

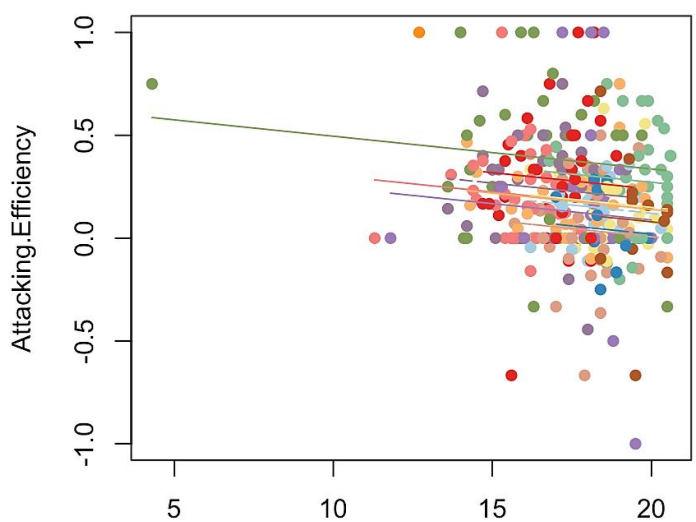

Figure 3. Repeated measures correlation – relationship between strain and attacking efficiency. No statistical significance found (p = .1199, r^2 = -.0871).

During practice sessions, neither linear regression nor repeated measures correlation analyses revealed significant relationships between WHOOP-derived metrics and performance outcomes. Particularly, attacking efficiency was not associated with recovery (LM: p = .847, rmcorr: p = .889), sleep debt (LM: p =

.219, rmcorr: p = .912), or sleep performance (LM: p = .204, rmcorr: p = .569). Likewise, passing efficiency showed no significant associations with recovery (LM: p = .617, rmcorr: p = .079), sleep debt (LM: p = .587, rmcorr: p = .080), sleep performance (LM: p = .5876, rmcorr: p = .053), or strain (LM: p = .9557, rmcorr: p = .053).

Scrimmage

Scrimmage subset means: Strain = 17.81 ± 2.11, Recovery = 61.71 ± 19.4, Sleep Debt = 42.78 minutes ± 40.4, Sleep Performance = 74.63 ± 13.2, Attacking Efficiency = 0.3406 ± 0.520, Passing Efficiency = 1.899 ± 0.576. No statistically significant relationships were found between attacking efficiency or passing efficiency and WHOOP predictors in scrimmage conditions.

Neither linear regression nor repeated measures correlation analyses identified significant relationships between WHOOP-derived metrics and performance outcomes. The analysis showed that, attacking efficiency was not associated with recovery (LM: p = .2188, rmcorr: p = .908), strain (LM: p = .7592, rmcorr: p = .646), sleep debt (LM: p = .219, rmcorr: p = .438), or sleep performance (LM: p = .2046, rmcorr: p = .568). Following the same pattern, passing efficiency showed no significant associations with recovery (LM: p = .306, rmcorr: p = .061), sleep debt (LM: p = .8612, rmcorr: p = .596), strain (LM: p = .360, rmcorr: p = .868), or sleep performance (LM: p = .392, rmcorr: p = .665).

PPTSQ v Averages

Weekly WHOOP predictor averages were measured against PPTSQ responses. The relationship between sleep performance and PPTSQ was statistically significant (p = .0373, r^2 = .0418), indicating that decreased sleep performance correlated with reduced perceived team performance. Conversely, repeated measures correlation analysis found no significance (p = .3209).

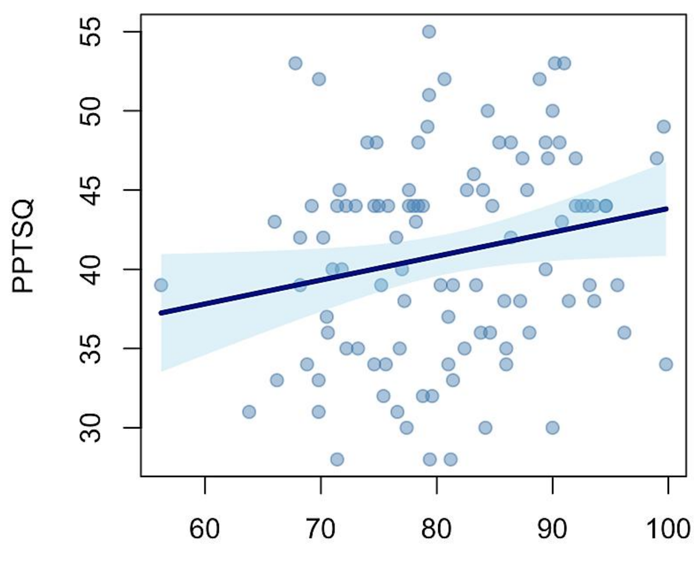

Figure 4. Linear regression – relationship between sleep performance and PPTSQ. Statistical significance found. (p = .0373, F(1,102 = 4.454, r^2 = .0418).

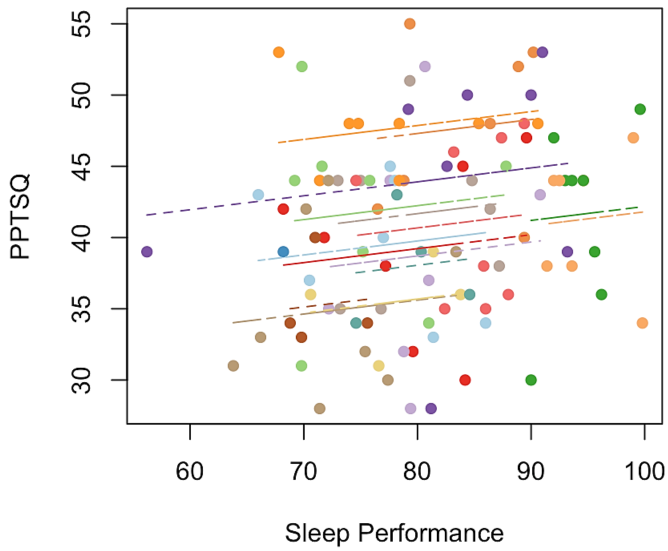

Figure 5. Repeated measures correlation - relationship between sleep performance and PPTSQ. No statistical significance found. (p = .3209).

No other significant relationships were found between recovery, sleep debt, or strain and PPTSQ scores (p = .3804, .1957, and .1144). These findings suggest limited predictive relationships between WHOOP metrics and volleyball performance, with only two statistically significant relationships identified.

DISCUSSION

This study examined the effectiveness of WHOOP for predicting volleyball performance in Division I athletes. The overall findings suggest that WHOOP variables demonstrate limited predictive validity for volleyball performance outcomes. However, two significant relationships warrant further investigation: strain versus attacking efficiency and sleep performance versus PPTSQ scores.

The significant negative correlation between strain and attacking efficiency suggests that elevated physiological stress may compromise attacking performance. Previous research supports the idea that systemic fatigue can decrease jump heights, which could be a possible physiological reason for the trend observed in the current investigation (de Leeuw et al., 2022). De Leeuw et al. (2022) indicated that strength training load management significantly influenced match performance in volleyball players when combined with perceived wellness measures, the results suggest that accumulated physiological stress affects performance outcomes. However, this study did not discover specific variables that would cause this relationship. WHOOP's strain algorithm quantifies cardiovascular and muscular exertion on a logarithmic scale (0-21), incorporating both heart rate-derived cardiovascular load and accelerometer-based muscular load (WHOOP, 2023). From a coaching perspective, chronic strain from intensive training and travel schedules may negatively impact competitive performance, therefore monitoring strain may help coaches flag early signs of fatigue or poor recovery. Zadeh et al. (2020) used a similar proactive approach in identifying injury risk through wearable monitoring, and that study reported success in preventing injuries that normally occur from overuse and overtraining. Even though success has been seen with this monitoring method, WHOOP's strain calculation incorporates multiple physiological variables, and individual components within this algorithm may demonstrate stronger predictive relationships than the overall score. WHOOP's algorithms may prevent examination of how individual physiological components contribute to composite metric scores, potentially masking stronger predictive relationships within individual metrics. Although the strain and attacking efficiency relationship was statistically significant, the small effect size only explains 3% of the observed variance, which limits practical applications for specific monitoring.

In contrast, the positive correlation between sleep performance and perceived team performance suggests that improved sleep quality directly affects athletes' subjective assessments and team morale. This result is consistent with previous work highlighting how sleep quality relates to both psychological readiness and physical performance. For example, Allen et al. (2016) documented that greater REM sleep duration and total sleep duration correlated with improved running performance in Division I cross-country athletes. Berryhill et al. (2020), supported these findings by validating WHOOP's sleep measurements against polysomnography, demonstrating low bias in sleep duration measurements (bias: 13.8 minutes; precision: 17.8 minutes) and accurate measurement of REM and slow wave sleep (ICC: 0.74-0.85). Participants in this study recorded improved nighttime sleep quality while wearing WHOOP devices, suggesting potential benefits beyond passive monitoring.

WHOOP generates daily sleep performance scores through algorithms incorporating HR, HRV, sleep duration, sleep disturbances, and accumulated sleep debt. These scores provide a comprehensive assessment of sleep quality relevant to athletic recovery. From an applied perspective, coaches may recognize that disrupted sleep from intensive travel or training schedules may impair both individual and collective team performance. The relationship between sleep and psychological readiness represents an emerging area of research. Horvath et al. (2022) demonstrated that physiological markers like HRV correlate with emotional self-regulation capabilities, suggesting significant trends between physiological recovery and psychological performance. Even though this study found a statistically significant relationship between sleep and perceived performance, like the strain and attacking efficiency relationship, this association is only responsible for about 4% of the variance observed. With a negligible effect size, little to no practical guidelines can be determined when using WHOOP metrics as performance predictors from variables assessed in this study.

Although this study offers several practical interpretations, it is important to acknowledge its limitations. Firstly, the sample consisted of collegiate women's volleyball players from a single university which limits generalizability to other sports, competitive levels, or populations. On the same note, position-specific analyses further reduced sample sizes, particularly for passing efficiency assessments, which further limited statistical power. Secondly, off-season data collection may not reflect performance relationships during in-season play when training loads and stress significantly differ. Then when assessing the timeline, an eight-week duration may be insufficient to establish long-term predictive relationships or capture seasonal physiological adaptations. WHOOP's proprietary algorithms prevent examination of how individual physiological components contribute to composite scores, potentially masking stronger predictive relationships within specific metrics. Lastly, device accuracy may be affected by movement artifacts, compliance variations, and individual physiological differences. It is important to note, photoplethysmography-based measurements have inherent limitations during high-intensity or contact-based activities, which may have affected WHOOP composite score calculation reliability during practice and game performance.

Future research should focus on specific variables influencing strain-attacking performance and sleep-perceived performance relationships. Secondly, sample sizes should be expanded by including more athletes, sports, and competitive phases. Lastly, the inclusion and adaptation of sport-specific algorithms, may improve the practicality of performance prediction models.

CONCLUSION

This study provides limited evidence supporting WHOOP's effectiveness for predicting volleyball performance in Division I athletes. While findings are consistent with prior work demonstrating weak associations between passing efficiency and physiological markers, the limited relationship between strain and attacking efficiency differed from expectations based on previous research. Despite these limitations, the results offer relevant insight for applied sports performance. Especially in the context of monitoring athlete recovery and perceived readiness. Although, wearable technology offers accessible physiological data, bridging the gap between monitoring and reliable performance prediction remains a key challenge that warrants continued investigation.

AUTHORS CONTRIBUTIONS

All authors meet the criteria for authorship in accordance with established ethical guidelines. Asha Phillips conceived and designed the study, coordinated data collection, performed data management and preliminary analyses, interpreted results, drafted the manuscript, and incorporated revisions from all co-authors. Dr. Zachary Droll supervised all aspects of the study as thesis advisor and contributed to study conceptualization and design. Dr. Andrew Mauldin provided statistical expertise and assisted with interpretation of statistical findings. Dr. Matthew Ruiz contributed sports psychology expertise, provided guidance on implementation and interpretation of the Perceived Performance Team Sports Questionnaire (PPTSQ). All authors have read and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CONFLICT OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this manuscript.

AI USE DISCLOSURE

In accordance with current publishing ethics and transparency recommendations, artificial intelligence (AI) tools were used solely to assist with translation and language editing, with the aim of improving clarity and readability. No AI tools were used in the generation of scientific content, including the study design, data collection, analysis, interpretation of results, or the formulation of conclusions. The authors retain full responsibility for the content of the manuscript and confirm its originality, integrity, and accuracy.

REFERENCES

Allen, C. (2016). Using physiological big data to predict cross country performance (Doctoral dissertation).
Bakdash, J. Z., & Marusich, L. R. (2017). Repeated measures correlation. Frontiers in psychology, 8, 456. https://doi.org/10.3389/fpsyg.2017.00456

Bellenger, C. R., Miller, D. J., Halson, S. L., Roach, G. D., & Sargent, C. (2021). Wrist-based photoplethysmography assessment of heart rate and heart rate variability: Validation of WHOOP. Sensors, 21(10), 3571. https://doi.org/10.3390/s21103571

Berryhill, S., Morton, C. J., Dean, A., Berryhill, A., Provencio-Dean, N., Patel, S. I., … Parthasarathy, S. (2020). Effect of wearables on sleep in healthy individuals: A randomized crossover trial and validation study. Journal of Clinical Sleep Medicine, 16(5), 775-783. https://doi.org/10.5664/jcsm.8340

de Leeuw, A. W., van Baar, R., Knobbe, A., & van der Zwaard, S. (2022). Modeling match performance in elite volleyball players: Importance of jump load and strength training characteristics. Sensors, 22(20), 7996. https://doi.org/10.3390/s22207996

Flatt, A. A., Esco, M. R., & Nakamura, F. Y. (2017). Individual heart rate variability responses to preseason training in high-level female soccer players. Journal of Strength and Conditioning Research, 31(2), 531-538. https://doi.org/10.1519/JSC.0000000000001529

Gershgoren, L., Blatt, A., Sela, T., & Tenenbaum, G. (2021). Does the league table lie? The development and validation of the perceived performance in team sports questionnaire (PPTSQ). Frontiers in Psychology, 11, 615018. https://doi.org/10.3389/fpsyg.2020.615018

Horvath, E., Kovacs, M. T., Toth, D., & Toth, L. (2022). A study of the relationship between anxiety, cognitive emotion regulation and heart rate variability in athletes. Journal of Physical Education and Sport, 22(2), 528-534. https://doi.org/10.7752/jpes.2022.02067

Makivić, B., Nikić Djordjević, M., & Willis, M. S. (2013). Heart rate variability (HRV) as a tool for diagnostic and monitoring performance in sport and physical activities. Journal of Exercise Physiology Online, 16(3), 103-131.

Miller, D. J., Sargent, C., & Roach, G. D. (2022). A validation of six wearable devices for estimating sleep, heart rate and heart rate variability in healthy adults. Sensors, 22(16), 6317. https://doi.org/10.3390/s22166317

Rehman, R. Z. U., Chatterjee, M., Manyakov, N. V., Daans, M., Jackson, A., O'Brisky, A., … Morris, M. (2024). Assessment of physiological signals from photoplethysmography sensors compared to an electrocardiogram sensor: A validation study in daily life. Sensors, 24(21), 6826. https://doi.org/10.3390/s24216826

Sitti, K., & Rangubhet, K. (2025). Volleyball spiking analysis for scoring success: performance indicators in the 2024 VNL women's teams. International Journal of Sports Science and Physical Education, 10(1), 8-17. https://doi.org/10.11648/j.ijsspe.20251001.12

Taber, C. B., Sharma, S., Raval, M. S., Senbel, S., Keefe, A., Shah, J., … Kaya, T. (2024). A holistic approach to performance prediction in collegiate athletics: Player, team, and conference perspectives. Scientific Reports, 14(1), 1162. https://doi.org/10.1038/s41598-024-51658-8

WHOOP. (2023, November 17). How does WHOOP strain work? Retrieved from [Accessed 2026, 12 March]: https://www.whoop.com/us/en/thelocker/how-does-whoop-strain-work-101/

Zadeh, A., Taylor, D., Bertsos, M., Tillman, T., Nosoudi, N., & Bruce, S. (2021). Predicting sports injuries with wearable technology and data analysis. Information Systems Frontiers, 23(5), 1023-1037. https://doi.org/10.1007/s10796-020-10018-3