ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN

SCIENZA E CULTURA DEL BENESSERE E DEGLI STILI DI VITA

Ciclo 36

**Settore Concorsuale:** 06/N2 - SCIENZE DELL'ESERCIZIO FISICO E DELLO SPORT

**Settore Scientifico Disciplinare:** M-EDF/02 - METODI E DIDATTICHE DELLE ATTIVITA' SPORTIVE

# SUBJECTIVE MONITORING OF PROFESSIONAL SOCCER PLAYERS: VALIDATION AND APPLICATION USING BIG DATA ANALYTICS

**Presentata da:** Carlo Simonelli

**Coordinatore Dottorato**

Laura Bragonzoni

**Supervisore**

Samuele Maria Marcora

**Co-Supervisore**

Lorenzo Mondini

Esame finale anno 2024

# Table of Contents

# Abstract

The first aim of the research program reported in this thesis was to develop predictive models of daily and match-related fatigue and investigate the factors associated with subjective recovery in professional soccer players by analysing the commonly used wellness measures. The second aim was to explore the validity and reliability of five single-item scales widely used in research and practice to measure the subjective status of professional soccer players.

In the first study we found, using big data analytics, that daily and match-day fatigue can be predicted with reasonable accuracy in six professional soccer teams monitored throughout the entire season (53,294 observations). This study also shows that psychological factors like stress and mood are important predictors of fatigue (mental fatigue). In the second study of four professional soccer teams (36,381 observations), we found that subjective recovery is primarily associated with fatigue and muscle soreness, and that these variables mediate 55% of the relationship between training load and subjective recovery. Albeit correlative, our findings also suggest that reducing mental fatigue and muscle soreness may help subjective recovery and performance of professional soccer players. In the third study involving 186 Italian soccer players, we investigated the validity and reliability of single-item measures of subjective Fatigue, Sleep Quality, Muscle Soreness, Stress and Mood. Although correlated to their criterion measures, these scales do not show a convergent validity to the criterion measures themselves. In conclusion, these practical and inexpensive single-item scales commonly used to monitor soccer players daily, do not appear to be a valid assessment of the variable they purport to measure.

# Structure of the thesis

This thesis is written in paper style. The thesis consists of a General Introduction, three Paper Chapters, and a General Discussion. The General Introduction (Chapter 1) includes a narrative review based on the theoretical framework of soccer and the aims of the thesis. Each Paper Chapter (Chapters 2, 3, and 4) consists of a stand-alone manuscript reporting studies presented at national and international congresses and being submitted to relevant international journals for publication. For consistency, I wrote all manuscripts in AMA style. I am the first author of all these manuscripts and the contribution of the co-authors of each original manuscript is detailed in the acknowledgments. A General Discussion (Chapter 5) concludes the thesis. The main purpose of this section is to integrate the findings across the Paper Chapters, critically discuss them and provide directions for future research. Additional documents and information not included in these chapters are collected in the appendixes.

# Acknowledgments

This path would not have been possible without the support of many people, who have contributed to the making of this journey.

My supervisor, Samuele Marcora. Thanks for making me second guess my every thought and understanding how real science and academia should be. It has been my privilege to know you as not only my supervisor, but as a friend.

My buddy and colleague Alessio Rossi. Thanks for the endless hours online discussing and working together.

My fellow PhD candidate, Federico Nigro. Thanks for your input and support. I'm grateful for this journey with you.

Borja, Dalton and all the other components of "Marcora's Lab", food for the mind.

All the athletes and coaches I've been working with during these years.

My parents, Elia and Marinella. My sister, Chiara. My roots.

My wife Salima, my sons Suley and Elia. You inspired me and gave me strength to reach every stage of this journey.

Elia, thank you for your smiles, hugs and kisses. Daddy loves you very much.

Suley, I hope that I can provide you with the same inspiration to achieve your life goals.

And finally, Salima, I will be forever thankful for the support you give me every day. I feel it in me. An important part of this journey is yours.

As this chapter is about to end, I look forward to our next exciting life adventures together.

**If you have always done it that way,**

**it is probably wrong.**

**Charles Kettering**

# CHAPTER 1: General Introduction

Soccer is one of the top participation and spectator sport in the world. Thus, due to its increasing popularity, as well as the amount of financial interest in the game, soccer is one of the most extensively researched intermittent team sports[1]. Indeed, there are plenty of subject areas that have benefitted from scientific knowledge gained from soccer including the natural and physical sciences, medicine and social sciences[1].

## The conceptual framework

As research on soccer, especially soccer training, is impressive, conceptual frameworks can be used to synthesize evidence, help understand the phenomenon studied, inform future research, and serve as a reference guide in practice. In the scientific process, conceptual frameworks allow hypotheses to be specified more precisely, even when the main theory's auxiliary or main assumptions are modified when predictions fail[2] .

In addition, the specificity of workload personalization requires that the entire monitoring process be targeted and completed at the same time, methodologically rigorous and multifaceted. The complex nature of soccer highlights some common problems that may also be found in other popular team sports. Impressive amount of data too often used to combine variables without a specific theoretical knowledge generates a complex analysis based on undefined constructs. Creating a successful monitoring strategy must be developed through all the current research processes and innovations in sport science.

Several theoretical frameworks have been developed and are nowadays employed in the field, sometimes representing a guide in understanding and guessing new resilient practical solutions.

Recently, an interesting model of physical training was published by Jeffries and colleagues. It appears to be the most exhaustive so far because the authors introduced significant new and expanded concepts. Training prescription of the coaches generates and external and an internal training load always mediated by individual and contextual factors. Subsequently in the training effects, it is highlighted how both the acute and chronic training effects of training should be furtherly subclassified in positive and negative, generating four distinct and related components of the construct. In addition, the result of this adaptation once again mediated with individual and contextual factors generates the sport performance outcomes. Improvement, as well as decrement or no change, is the full range of possible descriptions of those outcomes.

Investigating and classifying soccer training monitoring methodologies published, could facilitate the exchange of knowledge about the ideas adopted and developed with soccer players, and their scientific validity to practitioners making a conscious choice possible. Moreover, providing an exploitable interpretation that allows to underline gaps and redundancies present in the international scientific literature on the topic of soccer.

This conceptual framework for physical training is intended to illustrate the relationship between stimulus (internal training load), training effects and their measures and sport performance[2]. For the development of this conceptual framework applied to soccer, the most cited 150 papers were analyzed.

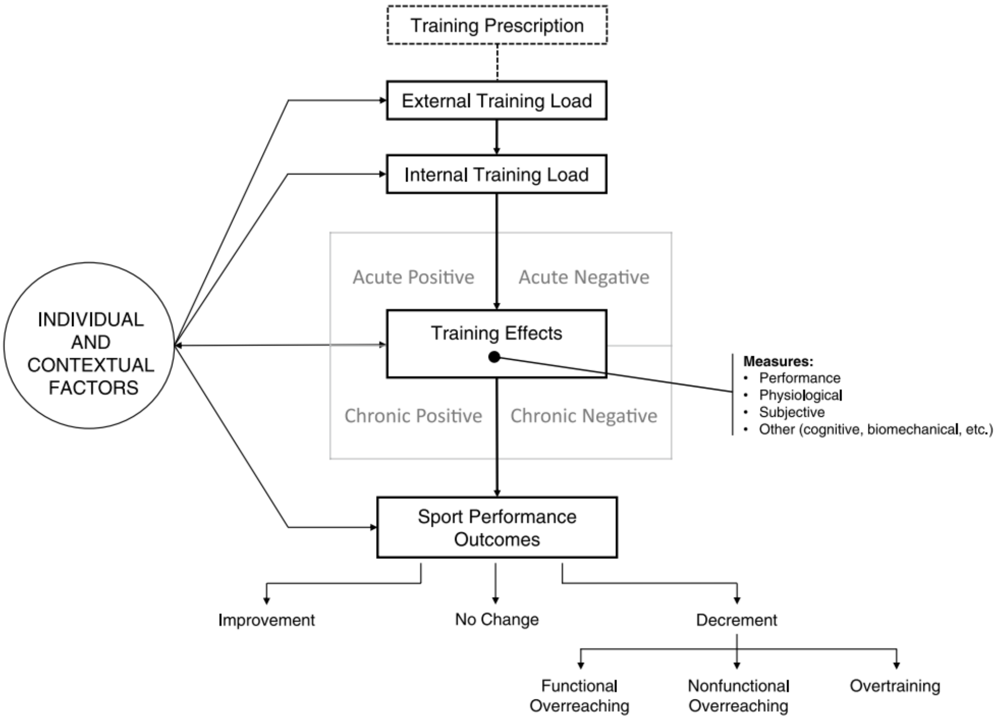



**Figure 1.** Conceptual framework of physical training as proposed by Jeffries et al.[2]

## Exercise prescriptions

In the conceptual framework, training prescriptions can be described as: "short (single training session) to long (multiannual periodization) plans defining the nature and organization of the exercises/training sessions supposed to target factors causally (directly or indirectly) related to sport performance. The training prescription is influenced and adapted based on performance models, contextual and individual factors, training effects, previous training load experienced by the athletes, and coach experience"[2]. Hence, exercise prescription represents the development of the training ideas to be turned into practice.

## External training load

The term external load implicitly refers to the physical work undertaken by an athlete during the training session and prescribed in the training plan. Notably, this does not refer to 'work' in the physics sense (force x distance) but more so in a generic manner. Accordingly, the term external load accommodates quantification and prescription in a variety of manners, enabling the use of a diverse range of measures and metrics. Some common measures of external load include GPS derived units (speed, accelerations, etc.) and level of resistance[3–5]. In the specific context of soccer the use of GPS monitoring systems allowed to track the distance covered at a specific speed, accelerations, decelerations, accelerometer load, impacts, etc.[5–10] in small sided games or in different demands with the ball. Bujalance-Moreno et al. described various situations regarding small-sided games (SSG)[9]. Specifically, Casamichana et al.[11] and Dellal et al.[12–14], comparing different kinds of SSG, observed that the high intensity profile during friendly matches was higher than in SSGs, nevertheless, the global indicators of workload (work-rest ratio, player workload, and exertion index) and the distance covered per minute were higher for SSGs than for friendly matches. Regarding game formats, Aguiar et al.[15] found that the distance covered in the smallest format (2 vs. 2) was lower than in all other formats (3-,4- and 5-a-side), moreover, this format presented the lowest number of sprints. Casamichana et al.[16], Cihan[17] and Aasgaard and Kilding[18] investigated the effects of defensive strategies on external training load. Specifically, total distance and distance covered in high-intensity running zone significantly increased adopting tactical rules such as man-marking and double-man pressure in SSG protocols. Moreover, Castellano et al.[19] found that three indicators of external load (total distance covered, player load, and the work-rest ratio) decreased when goals/goalkeepers were included, but the number of accelerations was higher in games involving goals/goalkeepers. Arslan et al.[20] found that the distance covered at the high-intensity running was higher in SSGs with active rest compared to the same with passive rest. Furthermore, when the type of coach feedback was compared, no differences were found for the time-motion characteristics (Brandes & Elvers[21]). Furthermore, Giménez et al.[22] analysing the number of ball contacts, observed that during SSGs players performed their highest intensity of exercise (acceleration of >4 $m/s^2$) playing with one touch only.

## Internal Training load

Internal Training Load typically refers to the body responses experienced by an athlete during an exercise or a training session. Both physiological (e.g. heart rate, EMG, salivary, blood, and muscle samples) and psychological (e.g. rating of perceived exertion - RPE), are representative measures of internal responses during exercise. The invasive nature of some physiological measures of internal TL and related difficulties for a daily use can make this way of monitoring unpractical in a real-world scenario. Differently, the Rating of Perceived Exertion (RPE) is one of the most used time efficient tools to estimate the internal TL due to the fact that it is reliable, non-invasive and low-cost. Moreover, previous studies demonstrated its association with physiological variables such as heart rate, oxygen consumption, respiratory rate, ventilation, and blood lactate concentrations[23–26]. Impellizzeri et al.[27] compared the Borg RPE scale[24] as adapted by Foster[26] to Bannister's training impulse, describing the RPE method as simpler, cheaper, and less invasive. Furthermore, a combination of an external measure of the duration (volume) of specific exercise or session with an internal measure of intensity (e.g. the RPE of the session (sRPE)), has been also used to estimate the subjective players' Training Load rating (TL)[25] (TL = sRPE × training time). Moreover, in many different team sports such as soccer where the unpredictable nature of the game makes the TL management complex and competition occurs at least on a weekly basis, this specifical subjective approach is a simple, quick and inexpensive[28–30]. Hence, as players are required to peak with limited recovery between matches, monitoring internal TL is critical for practitioners for assessing the psychophysiological responses of their athletes to the external training workload stimulus performed during the scheduled training programme.

## Training Effects

Training effects can be described as the outcomes occurring after a single or multiple training sessions. In soccer, as in most team sports, all stimuli have acute and chronic responses during physical, technical and tactical soccer training. The ability to manage long-term adjustments throughout the competitive season is an important key for improving soccer performance. Specifically, if the adaptation to these *stimuli* lasts more than one training session but less than a week can be considered acute, contrarywise if the recovery takes more than a week or a microcycle, the resulting effect can be considered chronic.

Training effects can be further divided into positive effects and negative effects. Determining the positive or negative effects of training is extremely important to identify signs of expected outcomes prior to treatment. Formulating a process to measure the training effects for the validation process itself of metrics and training variables is paramount.

### Daily monitoring

In professional soccer, as well as in other professional athletes, balancing between fatigue resulting from match or training and recovery to be able to play or train again the day after in crucial. An imprecise relation between exercise prescription, subsequent fatigue and recovery, can lead to an accumulation of fatigue that can lead to overreaching or overtraining, whereas an excessive reduction in training leading to detraining. Both statuses are detrimental for performance[30]. Hence, many non-invasive monitoring tools serving as valid and reliable indicators of the fatigue and recovery status in athletes have been developed in the recent years. In a real-world scenario in professional soccer, those tools should be simple, quick, inexpensive, easy to administer sensitive to a specific training effect determined by acute or chronic adaptation of daily training load[30].

Furthermore, in team sports such as soccer, any monitoring assessment should be administered frequently during the long and congested competitive period, from summer camp on. During the competitive season, soccer players compete at least on a quasi-weekly basis, in some periods on two to three occasions across a 7-day period. This specific consequence highlights the fact that, to ensure the players are in the best physical and psychological condition, tools to evaluate the subjective status must be administered daily to track changes in internal training load or in acute / chronic training effect.

**Subjective wellness scales**

Subjective wellness questionnaire comprising four or more single-item scales, have been used extensively to assess the daily overall training effects of athletes during training and competition[28,31–35]. Saw et al.[36–38] reported how subjective measures could have greater sensitivity to fatigue, acute and chronic training load, in comparison to objective measures.

Recently, the use of non-invasive and psychometrically validated tools and scales, such as the Profile of Mood States[39], Daily Analyses of Life Demands for Athletes[23], Total Quality Recovery[40], and Recovery-Stress Questionnaire for Athletes[41] became a very common method to assess subjective status of athletes. Unfortunately, all the scales cited have weekly or monthly timeframes, that can be useful for individual sports in preparing a long-term competition or event, but not very useful in professional team sport in which competitions can happen even 3 times in a week. Moreover, these scales present a low applicability on a daily basis in a professional team-sport scenario, due to the number of items and their subsequent time-consuming characteristics,

Hence, customized daily wellness questionnaires became very common in team-sport scenarios for their quickness, easiness of use, potential usefulness and face validity, even though they have not been validated.

Specifically, AFL research has also shown custom subjective questionnaires to be sensitive to daily, within-weekly and seasonal changes in training load[31,42]. Moreover, in a pre-season camp, players daily training load was found to be significantly correlated with fatigue, sleep quality, stress, mood and muscle soreness[42]. During the course of the season, sensitivity of subjective ratings of 9 items (i.e. fatigue, general muscle, hamstring, quadriceps, pain/stiffness, power, sleep quality, stress, well-being) were found to be sensitive to weekly training manipulations, to periods of unloading during the season and to individual player characteristics[31].

In soccer players subjective ratings of fatigue, muscle soreness, sleep quality and stress have been used as post-match monitoring[43,44]. Significant large negative correlations were found between fatigue and s-RPE in professional players. Significant small to very large negative correlations were found for sleep quality, fatigue and muscle soreness with all internal and external variables[45].

Following the framework description, we therefore have four possible permutations of training effects: acute positive, acute negative, chronic positive and chronic negative training effects.

**Acute Positive effects**

Regarding acute positive effects, Dello Iacono and Seitz[46] showed the acute positive effects of two post-activation potentiation hip thrust-based protocols on subsequent sprint performance.  Guerra Jr. et al.[47] observed that acute plyometric and sled towing stimuli enhance jump performance in male soccer players. The protocol also indicated a significant difference on CMJ height across conditions, with ingestion of caffeine (60 minutes pre-test) that elicited a greater response. Morcillo et al.[48], assessed changes in metabolic and mechanical responses to a specific repeated sprint ability test. A nearly perfect correlation between CMJ height loss and the lactate concentration was found. Moderate correlation between speed loss and lactate concentration and a moderate correlation was also found between speed loss and post-RSA ammonia concentration finishing the RSA test[48].

**Acute Negative Effects**

Acute negative effects were observed comparing acute inflammatory responses, muscle damage and hormonal variations according to the eccentric training in soccer professional athletes with different genetic profiles of ACTN3 (XX, RX and RR)[49]. Recovery kinetics (performance, muscle damage, and neuromuscular fatigue) after speed-endurance training were also found to induce short-term neuromuscular fatigue lasting 24 to 72h[50].

Many published research on acute effect monitoring hormonal, psycho-physiological, anthropometric and contextual variables, were not included in the training effects as part of the sport performance outcomes[43,51–53].

### Chronic positive effects

Chronic effects of training can be described when changes are measurable one week after a treatment or microcycle. Wong et al.[54] in a concurrent muscolar strength and high intensity interval training intervention protocol discovered improvements in strength, jump height, 10-m and 30-m sprint times, distances covered in the Yo-Yo Intermittent Recovery Test and maximal aerobic speed[54]. Askling et al. showed that an addition of specific pre-season eccentric strength training for hamstrings, would be beneficial both for injury prevention and for performance enhancement in elite soccer players[55]. Dupont et al[56] highlighted an improvement in maximal aerobic speed and a decrement in the 40-m sprint time due to high intensity interval training. Rønnestad et al.[57] detected that one weekly strength maintenance session during the first 12 weeks of the season allowed professional soccer players to maintain the improved strength, sprint and jump performance achieved during a preceding 10-week preparatory period. Bujalance et al[9]. showed that Small Sided Games (SSGs) can lead to improvements in sprint speed, repeated sprint ability and change of direction, along with muscular and physiological adaptations.

### Chronic negative effects

A long-term negative effect of training is not usually induced voluntarily during a pre-season or a season, but can hypothetically happen in a contextual and methodological way of speaking. Casajus et al.[58] found no changes in the mean $VO_2max$ results in the first test (65.5 ml.kg-1.min-1) compared the second one (66.4 ml.kg-1.min-1). There were no significant differences in maximal heart rate and treadmill speed at $VO_2max$ as well.

## Sport Performance Outcomes

The results of the whole training process, from the balance between the positive and negative training effects, and the influence by contextual and individual factors (such as genetics, environment, psychological states, level of the opponent, etc) can be described as sport performance outcomes. The soccer-specific outcome can be measured in various ways using both absolute (e.g. distance at different speed, etc.) or relative and aggregate measures (e.g. winning/losing, etc.). Measures of tactical behaviour can be used as lower-level (causal) dimensions or proxies. Therefore, considering only variables sensitive to training, is important to understand what aspects of the training process could affect performance. Helgerud et al.[59], studied the effects of aerobic training on performance during soccer matches and soccer specific tests. Results showed that improvements in aerobic endurance in soccer players improved soccer performance by increasing the distance covered, enhancing work intensity, and increasing the number of sprints and involvements with the ball during a match.

## Individual and Contextual Factors

In many popular training models of literature, i.e. IR Banister and PerPot models, any other remaining element related to the primary variables of performance, is described as contextual or individual factor[60–62]. Jeffries et al.[2] described as contextual all the factors not part of the main process (physical training) such as environmental, social, cultural factors, etc. that can influence the training process or the training outcome (training effect and sport performance). These factors have an integrated relationship with all components of the conceptual framework, including bidirectionality with training effects. Contrarywise, characteristics of the individual athlete such as genetics, psychological traits and states, training background, etc. that can influence the training process or the training outcome should be described as individual factors[2]. These factors also have an integrated relationship with all components of the conceptual framework, including bidirectionality with training effects.

Rampinini et al.[10] found that total distance high intensity running and very high intensity running distance were influenced by the activity profile of opponent teams. They also showed that physical activity of the players is influenced by their playing position. Barrett et al.[53] analyzed the relationship between sRPE and playing position reporting that full backs had higher sRPE when compared to all other positions. They also analyzed the relationship between sRPE and opponent teams reporting higher sRPE-T (technical/cognitive exertion) for matches played against top teams compared to bottom and middle ranked teams. Mohr et al.[63] identified the relationship between quadriceps muscle temperature (Tm) and sprint performance evaluated during soccer matches. A major finding of the study was that both muscle and core temperature decreased markedly during the half-time period when players recovered passively. The lower body temperatures prior to the beginning of the second half were associated with a significant impairment in sprint performance. Conversely, when players performed a period of moderate-intensity exercise prior to the second half, body temperatures were maintained, and sprint performance did not deteriorate.

## Aims of the thesis

The first aim of the research program reported in this thesis was to develop predictive models of daily and match-related fatigue and investigate the factors associated with subjective recovery in Italian professional soccer players by analysing the commonly used Wellness scales. The second aim was to explore the validity and reliability of 5 single-item scales widely used in research and practice to measure the subjective status of professional soccer players.

Specifically, in Chapter 2, we present, using big data analytics and mediation analysis, the results of daily and match-day fatigue prediction in six professional soccer teams monitored throughout the entire season (53294 data collected). Specifically, our main aim was to clarify the predictors for daily and match-related fatigue in professional soccer players. Moreover, our second aim was to understand if any of the subjective variables mediates the relation between Training Load and subjective feeling of fatigue. A third and final aim was to assess the accuracy of the predictions.

In Chapter 3, we present, using big data analytics and mediation analysis, the correlates of the subjective judgement of recovery in six professional soccer teams monitored throughout the entire season (36381 data collected). Hence, our main aim was to understand the subjective correlates to the judgement of recovery in professional soccer players. Moreover, our second aim was to understand if any of the subjective variables mediates the relation between Training Load and subjective recovery. A third and final aim was to assess the accuracy of the predictions.

In Chapter 4, we wanted to assess the validity of the so-called Wellness single-item scales in Italian soccer players. Specifically, the first research question was to find whether these single-item measures of subjective Fatigue, Sleep Quality, Muscle Soreness, Stress and Mood are significantly correlated with their respective criterion measures (convergent validity). The second research question was to find if these single-item measures of subjective Fatigue, Sleep Quality, Muscle Soreness, Stress and Mood are consistent. The third research question was to confirm the existence of the wellness construct. The fourth and final research question was to understand if any of the single-item employed was redundant.

# References

1. Reilly T. Science and Soccer. Routledge; 2003.

2. Jeffries AC, Marcora SM, Coutts AJ, Wallace L, McCall A, Impellizzeri FM. Development of a Revised Conceptual Framework of Physical Training for Use in Research and Practice. Sports Med. 2022;52(4):709-724. doi:10.1007/s40279-021-01551-5

3. Impellizzeri FM, Marcora SM, Coutts AJ. Internal and External Training Load: 15 Years On. International Journal of Sports Physiology and Performance. 14(2):270-273. doi:10.1123/ijspp.2018-0935

4. Impellizzeri FM, Rampinini E, Marcora SM. Physiological assessment of aerobic training in soccer. Journal of Sports Sciences. 2005;23(6):583-592. doi:10.1080/02640410400021278

5. McLaren SJ, Macpherson TW, Coutts AJ, Hurst C, Spears IR, Weston M. The Relationships Between Internal and External Measures of Training Load and Intensity in Team Sports: A Meta-Analysis. Sports Med. 2018;48(3):641-658. doi:10.1007/s40279-017-0830-z

6. Mohr M, Krustrup P, Bangsbo J. Match performance of high-standard soccer players with special reference to development of fatigue. J Sports Sci. 2003;21(7):519-528. doi:10.1080/0264041031000071182

7. Mohr M, Krustrup P, Bangsbo J. Fatigue in soccer: A brief review. Journal of Sports Sciences. 2005;23(6):593-599. doi:10.1080/02640410400021286

8. Stolen T, Chamari K, Castagna C, Wisloff U. Physiology of Soccer. Sports medicine (Auckland, NZ). 2005;35:501-536. doi:10.2165/00007256-200535060-00004

9. Bujalance-Moreno P, Latorre-Román PÁ, García-Pinillos F. A systematic review on small-sided games in football players: Acute and chronic adaptations. J Sports Sci. 2019;37(8):921-949. doi:10.1080/02640414.2018.1535821

10. Rampinini E, Coutts A, Castagna C, Sassi R, Impellizzeri F. Variation in Top Level Soccer Match Performance. International journal of sports medicine. 2007;28:1018-1024. doi:10.1055/s-2007-965158

11. Casamichana D, Castellano J, Castagna C. Comparing the Physical Demands of Friendly Matches and Small-Sided Games in Semiprofessional Soccer Players. Journal of strength and conditioning research / National Strength & Conditioning Association. 2012;26:837-843. doi:10.1519/JSC.0b013e31822a61cf

12. A D, Drust B, Peñas C. Variation of Activity Demands in Small-Sided Soccer Games. International journal of sports medicine. 2012;33:370-375. doi:10.1055/s-0031-1295476

13. A D, Owen A, Wong DP, Krustrup P, van Exsel M, Mallo J. Technical and physical demands of small vs. large sided games in relation to playing position in elite soccer. Human movement science. 2012;31:957-969. doi:10.1016/j.humov.2011.08.013

14. A D, Varliette C, Owen A, Chirico E, Pialoux V. Small-Sided Games Versus Interval Training in Amateur Soccer Players: Effects on the Aerobic Capacity and the Ability to Perform Intermittent Exercises With Changes of Direction. Journal of strength and conditioning research / National Strength & Conditioning Association. 2011;26:2712-2720. doi:10.1519/JSC.0b013e31824294c4

15. Aguiar M, Gonçalves B, Botelho G, Lemmink KAPM, Sampaio J. Footballers' movement behaviour during 2-, 3-, 4- and 5-a-side small-sided games. Journal of Sports Sciences. 2015;33. doi:10.1080/02640414.2015.1022571

16. Casamichana D, Román-Quintana J, Castellano J, Calleja Gonzalez J. Influence of the Type of Marking and the Number of Players on Physiological and Physical Demands During Sided Games in Soccer. Journal of Human Kinetics. 2015;47:259-268. doi:10.1515/hukin-2015-0081

17. Cihan H. The effect of defensive strategies on the physiological responses and time-motion characteristics in small-sided games. Kinesiology. 2015;47:179-187.

18. Aasgaard M, Kilding A. Does Man Marking Influence Running Outputs and Intensity During Small-Sided Soccer Games? Journal of Strength and Conditioning Research. 2018;34:1. doi:10.1519/JSC.0000000000002668

19. Castellano J, Casamichana D, A D. Influence of Game Format and Number of Players on Heart Rate Responses and Physical Demands in Small-Sided Soccer Games. Journal of strength and conditioning research / National Strength & Conditioning Association. 2012;27. doi:10.1519/JSC.0b013e318267a5d1

20. Arslan E, Alemdaroğlu U, Köklü Y, Hazır T, Muniroglu S, Karakoc B. Effects of Passive and Active Rest on Physiological Responses and Time Motion Characteristics in Different Small Sided Soccer Games. Journal of Human Kinetics. 2017;60:123-132. doi:10.1515/hukin-2017-0095

21. Brandes M, Elvers S. Elite Youth Soccer Players' Physiological Responses, Time-Motion Characteristics, and Game Performance in 4 vs. 4 Small-Sided Games: The Influence of Coach

Feedback. Journal of Strength and Conditioning Research. 2017;31:2652-2658. doi:10.1519/JSC.0000000000001717

22. Gimenez J, Liu H, Lipińska P, Szwarc A, Rompa P, Ruano M. Physical responses of professional soccer players during 4 vs. 4 small-sided games with mini-goals according to rule changes. Biology of Sport. 2017;35:81. doi:10.5114/biolsport.2018.70754

23. Rushall BS. A tool for measuring stress tolerance in elite athletes. Journal of Applied Sport Psychology. 1990;2(1):51-66. doi:10.1080/10413209008406420

24. Borg G. Borg's Perceived Exertion And Pain Scales.; 1998.

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

26. Foster C, Florhaug J, Franklin J, et al. A New Approach to Monitoring Exercise Training. Journal of strength and conditioning research / National Strength & Conditioning Association. 2001;15:109-115. doi:10.1519/00124278-200102000-00019

27. Impellizzeri FM, Rampinini E, Coutts AJ, Sassi A, Marcora SM. Use of RPE-based training load in soccer. Med Sci Sports Exerc. 2004;36(6):1042-1047. doi:10.1249/01.mss.0000128199.23901.2f

28. Hooper SL, Mackinnon LT, Howard A, Gordon RD, Bachmann AW. Markers for monitoring overtraining and recovery. Med Sci Sports Exerc. 1995;27(1):106-112.

29. Coutts AJ, Reaburn P, Piva TJ, Rowsell GJ. Monitoring for overreaching in rugby league players. Eur J Appl Physiol. 2007;99(3):313-324. doi:10.1007/s00421-006-0345-z

30. R M, M D, C F, et al. Prevention, diagnosis, and treatment of the overtraining syndrome: joint consensus statement of the European College of Sport Science and the American College of Sports Medicine. Med Sci Sports Exerc. 2013;45(1):186-205. doi:10.1249/mss.0b013e318279a10a

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

32. McLean BD, Coutts AJ, Kelly V, McGuigan MR, Cormack SJ. Neuromuscular, endocrine, and perceptual fatigue responses during different length between-match microcycles in professional rugby league players. Int J Sports Physiol Perform. 2010;5(3):367-383. doi:10.1123/ijspp.5.3.367

33. Crowcroft S, McCleave E, Slattery K, Coutts AJ. Assessing the Measurement Sensitivity and Diagnostic Characteristics of Athlete-Monitoring Tools in National Swimmers. International Journal of Sports Physiology and Performance. 2017;12(s2):S2-95-S2-100. doi:10.1123/ijspp.2016-0406

34. Perri E, Simonelli C, Rossi A, Trecroci A, Alberti G, Iaia FM. Relationship Between Wellness Index and Internal Training Load in Soccer: Application of a Machine Learning Model. International Journal of Sports Physiology and Performance. 2021;16(5):695-703. doi:10.1123/ijspp.2020-0093

35. Lathlean TJH, Gastin PB, Newstead SV, Finch CF. A Prospective Cohort Study of Load and Wellness (Sleep, Fatigue, Soreness, Stress, and Mood) in Elite Junior Australian Football Players. International Journal of Sports Physiology and Performance. 2019;14(6):829-840. doi:10.1123/ijspp.2018-0372

36. Saw AE, Main LC, Gastin PB. Impact of Sport Context and Support on the Use of a Self-Report Measure for Athlete Monitoring. J Sports Sci Med. 2015;14(4):732-739.

37. Saw AE, Main LC, Gastin PB. Monitoring Athletes Through Self-Report: Factors Influencing Implementation. J Sports Sci Med. 2015;14(1):137-146.

38. Saw AE, Main LC, Gastin PB. Monitoring the athlete training response: subjective self-reported measures trump commonly used objective measures: a systematic review. Br J Sports Med. 2016;50(5):281-291. doi:10.1136/bjsports-2015-094758

39. Terry PC, Lane AM, Fogarty GJ. Construct validity of the Profile of Mood States-Adolescents for use with adults. Psychology of Sport and Exercise. 2003;4(2):125-139. doi:10.1016/S1469-0292(01)00035-8

40. Kenttä G, Hassmén P. Overtraining and Recovery: A Conceptual Model. Sports Med. 1998;26(1):1-16. doi:10.2165/00007256-199826010-00001

41. Kellmann M, Kallus KW. Recovery-Stress Questionnaire for Athletes: User Manual. Human Kinetics; 2001.

42. Buchheit M, Racinais S, Bilsborough JC, et al. Monitoring fitness, fatigue and running performance during a pre-season training camp in elite football players. Journal of Science and Medicine in Sport. 2013;16(6):550-555. doi:10.1016/j.jsams.2012.12.003

43. Ispirlidis I, Fatouros IG, Jamurtas AZ, et al. Time-course of Changes in Inflammatory and Performance Responses Following a Soccer Game. Clinical Journal of Sport Medicine. 2008;18(5):423. doi:10.1097/JSM.0b013e3181818e0b

44. Fatouros IG, Chatzinikolaou A, Douroudos II, et al. Time-course of changes in oxidative stress and antioxidant status responses following a soccer game. J Strength Cond Res. 2010;24(12):3278-3286. doi:10.1519/jsc.0b013e3181b60444

45. Oliveira RFS, Canário-Lemos R, Peixoto R, Vilaça-Alves J, Morgans R, Brito JP. The relationship between wellness and training and match load in professional male soccer players. PLOS ONE. 2023;18(7):e0289374. doi:10.1371/journal.pone.0289374

46. Dello Iacono A, Seitz LB. Hip thrust-based PAP effects on sprint performance of soccer players: heavy-loaded versus optimum-power development protocols. J Sports Sci. 2018;36(20):2375-2382. doi:10.1080/02640414.2018.1458400

47. Guerra Jr MA, Caldas LC, De Souza HL, et al. The acute effects of plyometric and sled towing stimuli with and without caffeine ingestion on vertical jump performance in professional soccer players. Journal of the International Society of Sports Nutrition. 2018;15(1):51. doi:10.1186/s12970-018-0258-3

48. Morcillo JA, Jiménez-Reyes P, Cuadrado-Peñafiel V, Lozano E, Ortega-Becerra M, Párraga J. Relationships Between Repeated Sprint Ability, Mechanical Parameters, and Blood Metabolites in Professional Soccer Players. The Journal of Strength & Conditioning Research. 2015;29(6):1673. doi:10.1519/JSC.0000000000000782

49. Pimenta EM, Coelho DB, Cruz IR, et al. The ACTN3 genotype in soccer players in response to acute eccentric training. Eur J Appl Physiol. 2012;112(4):1495-1503. doi:10.1007/s00421-011-2109-7

50. Recovery Kinetics After Speed-Endurance Training in Male Soccer Players in: International Journal of Sports Physiology and Performance Volume 15 Issue 3 (2020). Accessed October 28, 2023. https://journals.humankinetics.com/view/journals/ijspp/15/3/article-p395.xml

51. Silva J, Rumpf M, Maxime H, et al. Acute and Residual Soccer Match-Related Fatigue: A Systematic Review and Meta-analysis. Sports Medicine. 2018;48. doi:10.1007/s40279-017-0798-8

52. Ascensão A, Rebelo A, Oliveira E, Marques F, Pereira L, Magalhães J. Biochemical impact of a soccer match—Analysis of oxidative stress and muscle damage markers throughout recovery. Clinical biochemistry. 2008;41:841-851. doi:10.1016/j.clinbiochem.2008.04.008

53. Barrett S, McLaren S, Spears I, Ward P, Weston M. The Influence of Playing Position and Contextual Factors on Soccer Players' Match Differential Ratings of Perceived Exertion: A Preliminary Investigation. Sports. 2018;6. doi:10.3390/sports6010013

54. Wong DP, Chaouachi A, Chamari K, A D, Wisloff U. Effect of Preseason Concurrent Muscular Strength and High-Intensity Interval Training in Professional Soccer Players. Journal of strength and conditioning research / National Strength & Conditioning Association. 2009;24:653-660. doi:10.1519/JSC.0b013e3181aa36a2

55. Askling C, Karlsson J, Thorstensson A. Hamstring injury occurrence in elite soccer players after strength training with eccentric overload. Scandinavian journal of medicine & science in sports. 2003;13:244-250. doi:10.1034/j.1600-0838.2003.00312.x

56. Dupont G, Akakpo K, Berthoin S. The Effect of In-Season, High-Intensity Interval Training in Soccer Players. The Journal of Strength & Conditioning Research. 2004;18(3):584.

57. Rønnestad B, Nymark B, Raastad T. Effects of In-Season Strength Maintenance Training Frequency in Professional Soccer Players. Journal of strength and conditioning research / National Strength & Conditioning Association. 2011;25:2653-2660. doi:10.1519/JSC.0b013e31822dcd96

58. Casajus J. Seasonal variation in fitness variables in professional soccer player. The Journal of sports medicine and physical fitness. 2001;41:463-469.

59. Helgerud J, Engen LC, Wisløff U, Hoff J. Aerobic endurance training improves soccer performance.

60. Calvert TW, Banister EW, Savage MV, Bach T. A Systems Model of the Effects of Training on Physical Performance. IEEE Transactions on Systems, Man, and Cybernetics. 1976;SMC-6(2):94-102. doi:10.1109/TSMC.1976.5409179

61. Perl J. PerPot a meta-model and software tool for analysis and optimisation of load-performance-interaction. International Journal of Performance Analysis in Sport. 2004;4:61-73. doi:10.1080/24748668.2004.11868305

62. Morton R, Fitz-Clarke J, Banister E. Modeling human performance in running. Journal of applied physiology (Bethesda, Md : 1985). 1990;69:1171-1177. doi:10.1152/jappl.1990.69.3.1171

63. Mohr M, Krustrup P, Nybo L, Nielsen JJ, Bangsbo J. Muscle temperature and sprint performance during soccer matches – beneficial effect of re-warm-up at half-time. Scandinavian Journal of Medicine & Science in Sports. 2004;14(3):156-162. doi:10.1111/j.1600-0838.2004.00349.x

# CHAPTER 2: Prediction of daily and match-day subjective fatigue in professional soccer players: a machine learning approach

This paper was presented as an oral key presentation at XIII national SISMES congress in Milan, November 4-6, 2022 and as an oral presentation at the ECSS 2023 – European College of Sport Science in Paris, July 4-7.

# Prediction of daily and match-day subjective fatigue in professional soccer players: a machine learning approach

Carlo Simonelli[1]*, Alessio Rossi[2], Samuele M Marcora[1]

[1] Department for Life Quality Studies, University of Bologna, Italy;

[2] Independent researcher

* corresponding author: Carlo Simonelli (carlo.simonelli3@unibo.it)

## Abstract

**Background -** Predicting the state of fatigue in soccer players is useful to design training and optimize performance. Therefore, the aim of this study was to explore, using a framework of big data analytics, the most important predictors of daily and match-day fatigue in a group of professional soccer players using inexpensive and practical data monitoring tools.

**Methods –** Six professional soccer Italian professional third division (Serie C) teams took part in this study. Every morning, the players rated fatigue, sleep quality, muscle soreness, stress and mood. After each training session or match, the session Rating of Perceived Exertion (sRPE) was obtained and multiplied by duration to calculate the Training Load (TL). Finally, some contextual factors, (i.e. distance to previous and next match) were also recorded. A framework of machine learning models was trained and tested in order to assess their ability to predict the players' daily and match-day subjective fatigue.

**Results –** Machine learning models can accurately predict the players' fatigue (accuracy 79-84%) using practical and inexpensive training monitoring tools. Specifically, in the prediction of daily fatigue, the main related factor was the fatigue rating of the previous day. In the match-day fatigue prediction psychological factors of the previous day like stress and mood were the most influential factors.

**Conclusion –** Sport scientists and coaches can use this framework of big data analytics to simulate the effects of different training programs in order to maximize players' readiness and reduce the potential drops in performance associated with daily and match-day fatigue.

**Keywords:** wellness; rating of perceived exertion; readiness; multidimensional approach.

# Introduction

Fatigue is a complex and multifaceted phenomenon normally induced by prolonged physical and/or cognitive tasks. It has two main dimensions: i) subjective fatigue, represented by feelings of high effort required to perform a task, tiredness and lack of energy, and ii) a reduction in objective measures of physical and/or cognitive performance[1].

In soccer players, fatigue is induced by both matches and training. During soccer matches, fatigue has been observed through a reduction in measures of physical performance: i) after short-term intense periods in both halves,[2,3] ii) in the initial phase of the second half[3,4], and iii) towards the end of the game[2,3,5,6]. In these studies, fatigue was quantified as decrements in leg muscle strength or in maximal speed, acceleration, and deceleration. In professional soccer players, the perception of fatigue or a specific training session is related to a performance reduction (maximum voluntary contractions, sprint performance decrements, external workload due to a combination of central and peripheral factors. Moreover, the effects of fatigue are also detrimental for technical abilities like passing and shooting[7] both subsequent to matches[8,9] as to training sessions[10,11]. Tactical behaviours also seem to be negatively affected by fatigue[12–14]. From a subjective point of view, increased ratings of fatigue have been observed during intensified training periods[15–21] or following matches[21–23].

These studies were important to quantify fatigue in soccer players and demonstrate its negative impact on performance. However, these employed data collected in a relatively small number of players over short periods of time (from a single match to few weeks of training). Such small data sets are not sufficient to develop predictive models of fatigue in soccer players. Understanding the factors that can predict the fatigue state of soccer players can be useful to anticipate the performance capacity of a specific day. With regards to match day, predicting fatigue can be used to ensure that players are not fatigued so that they can perform optimally. With regards to predicting fatigue before a training session, it can be useful to modulate the load of a specific session in order to optimize individual training as much as possible. To have the chance of an individual modulation for each training, nowadays the collection of daily measures has become very common in soccer. Specifically, in professional soccer, it is common practice to collect many measures such as: i) distance, speed and acceleration variables for external training load, ii) sRPE and heart rate for internal training load, iii) wellness and/or recovery questionnaires as subjective measures. This daily monitoring practice enables the collection of very large datasets through one or more seasons. Those collections can be useful to create machine learning predictive models for next-day fatigue or match day related fatigue to help practitioners modulate training in a day-by-day real world scenario.

In this study we applied this machine learning approach to thousands of data resulting from the daily monitoring of each player from six professional soccer teams during a full competitive season, from summer camp to the eventual playoffs. The aim of the study was to develop predicting models that can help identify the most important factors that can predict daily fatigue and fatigue specifically related to the day of the match (match-day fatigue). Fatigue here refers to the fatigue state experienced by the player in the morning before performing any training or match.

## Methods

### Subjects

Six Italian professional third division (Serie C) soccer teams were recruited for this study for a total of 171 players. Six complete championship seasons were recorded from summer camp on, for each team. The data recording was conducted between July 17th 2017 and May 11th 2023. None of the teams have coaches or tactical ideas in common. Descriptive statistics of the players are provided in Table 1. After a detailed description of the procedure and possible risks, players voluntarily decided to participate by signing an informed consent. The project was conducted according to the Declaration of Helsinki and was approved by the Bioethics Committee of the University of Bologna.

**Table 1** Players' descriptive statistics (mean ± standard deviation).

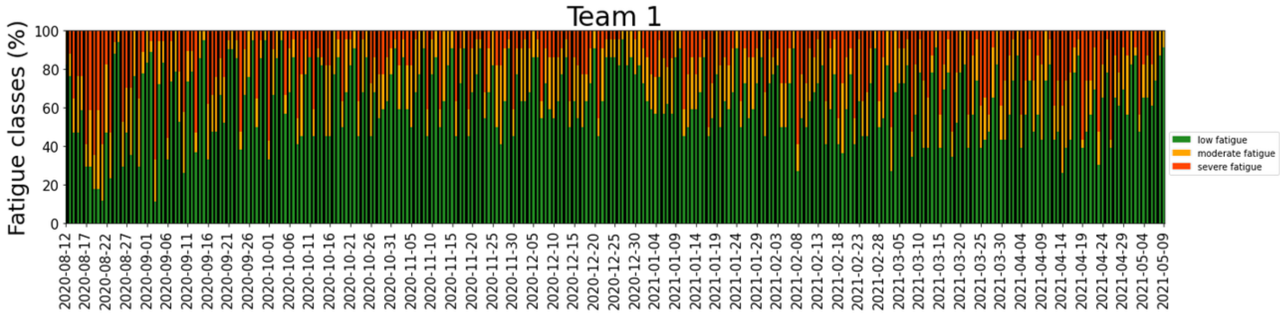
| Teams | Age (yrs) | Height (cm) | Weight (kg) |
|---|---|---|---|
| *Team 1* | 22.78±5.91 | 181±4 | 78.94±4.51 |
| *Team 2* | 21.14±3.44 | 182±5 | 79.12±7.56 |
| *Team 3* | 24.87±4.96 | 181±6 | 74.25±8.55 |
| *Team 4* | 21.91±3.95 | 180±4 | 73.17±5.87 |
| *Team 5* | 23.68±4.97 | 182±6 | 77.72±5.57 |
| *Team 6* | 21.91±3.95 | 180±5 | 75.37±7.12 |

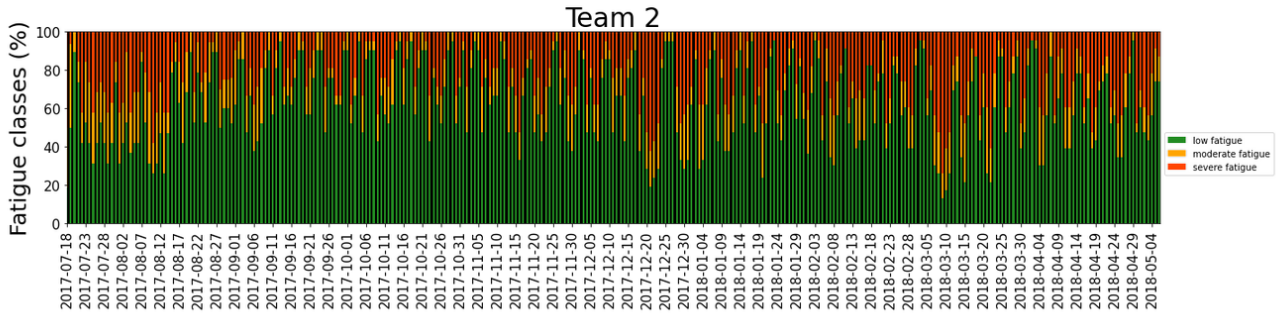
## Data collection

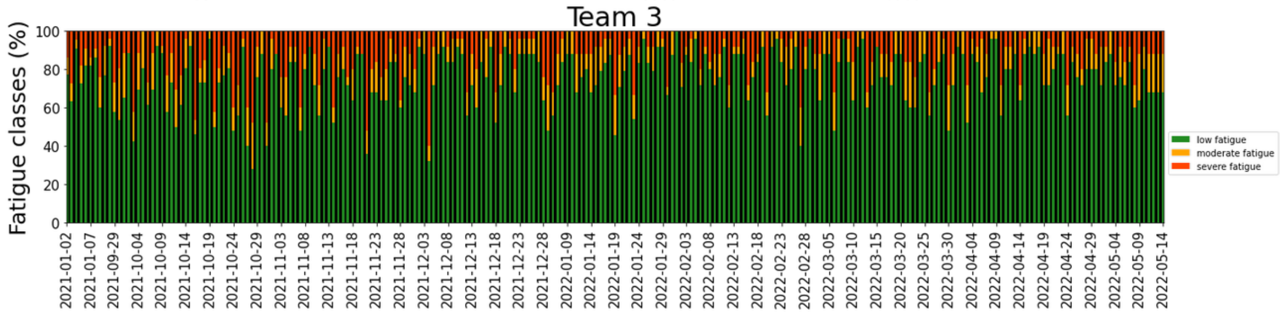
Every day, usually in private and at a consistent time in the morning, in any case before each training session or match, the players a Wellness Questionnaire (WQ)[24]. Specifically, players provided subjective ratings of fatigue, sleep quality, muscle soreness, stress and mood. Each single-item scale was a 5-point Likert scale where 1 and 5 indicate the highest and lowest values of wellness for each item, respectively. Finally, about 30 minutes after the end of each training session or match, the players rated how hard the training session or match was (sRPE) by using the 10-point scale (CR-10 Borg' scale), where 0 refers to resting state and 10 to maximal effort[25]. The Training Load (TL) for each training session or match was computed as the product between sRPE and the duration in minutes (time). In each analysis, TL refers to the load of the training session or match performed the day before the subjective ratings of recovery, fatigue, sleep quality, muscle soreness, stress and mood were provided. Training load can be simply described as a label attributed to a higher-order construct overarching other interrelated sub-dimensions such as sRPE and time specifically[26]. The players were familiar with the rating system having completed the process over the preseason period and been instructed in its use by the head of performance at the club. In order to take into consideration the intraindividual difference in individual rating, all the scores (i.e., fatigue, muscle soreness, sleep quality, mood, and stress) and TL were normalized by players between 0 and 1 that refer to minimum and maximum score values, respectively. In order to create the multiclass dependent features for our machine learning problems, the Fatigue scores were split by players' data distribution in three main classes as showed in Table 2: Low (lower than 33rd percentile); Moderate (between 33rd and 66th percentiles); Severe (higher than 66th percentile). Moreover, the distribution of the Fatigue classes as the season went for all the teams are provided in Figure 1.

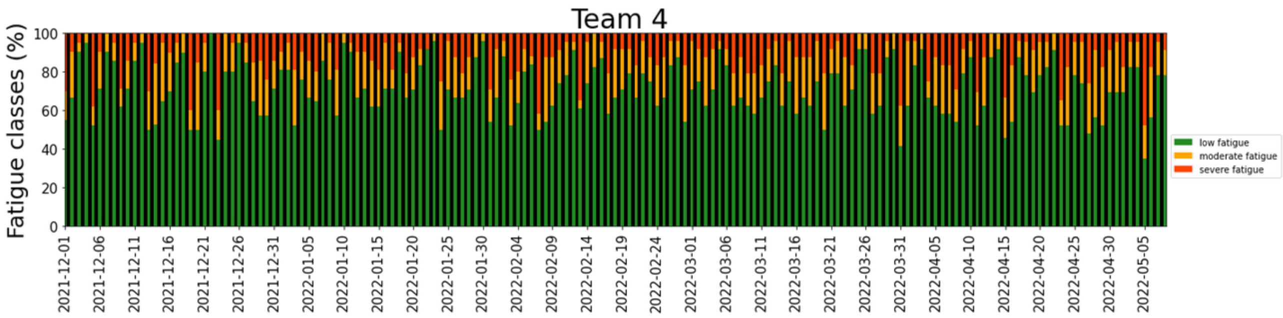
**Table 2.** Fatigue classes' descriptive statistics. SD, min and max refer to standard deviations, minimum values and maximal values, respectively, while 25%, 50% and 75% refer to the interquartile values.

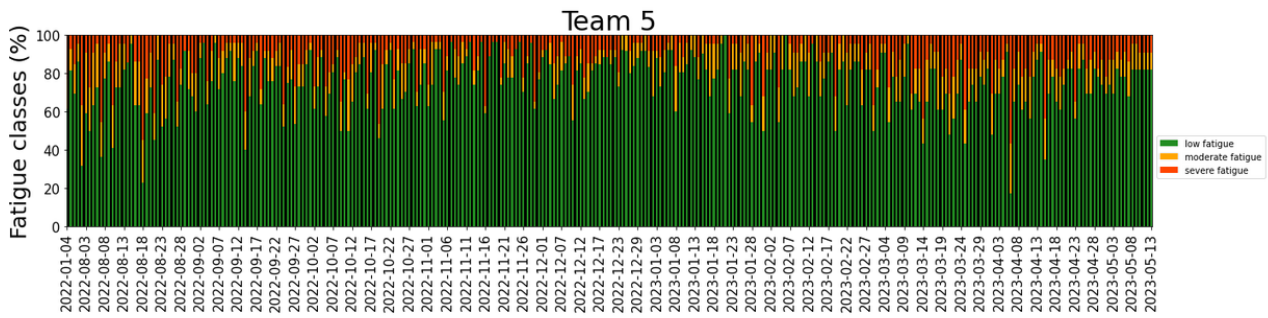
| Team | Class | Count | Mean | SD | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|---|
| *Team 1* | *Low* | 3619 | 1.76 | 0.43 | 1 | 2 | 2 | 2 | 2 |
| | *Moderate* | 1098 | 2.44 | 0.50 | 2 | 2 | 2 | 3 | 3 |
| | *Severe* | 1006 | 3.24 | 0.52 | 2 | 3 | 3 | 3 | 5 |
| *Team 2* | *Low* | 3777 | 1.61 | 0.49 | 1 | 1 | 2 | 2 | 2 |
| | *Moderate* | 882 | 2.56 | 0.50 | 2 | 2 | 3 | 3 | 3 |
| | *Severe* | 1441 | 3.23 | 0.45 | 3 | 3 | 3 | 3 | 5 |
| *Team 3* | *Low* | 3966 | 1.59 | 0.49 | 1 | 1 | 2 | 2 | 2 |
| | *Moderate* | 539 | 2.62 | 0.51 | 2 | 2 | 3 | 3 | 4 |
| | *Severe* | 724 | 3.23 | 0.45 | 3 | 3 | 3 | 3 | 5 |
| *Team 4* | *Low* | 2589 | 1.71 | 0.45 | 1 | 1 | 2 | 2 | 2 |
| | *Moderate* | 604 | 2.61 | 0.51 | 2 | 2 | 3 | 3 | 4 |
| | *Severe* | 433 | 3.33 | 0.63 | 2 | 3 | 3 | 4 | 5 |
| *Team 5* | *Low* | 2500 | 1.46 | 0.74 | 1 | 1 | 2 | 3 | 3 |
| | *Moderate* | 994 | 2.43 | 0.84 | 2 | 2 | 3 | 3 | 4 |
| | *Severe* | 1105 | 2.73 | 0.75 | 2 | 3 | 3 | 4 | 5 |
| *Team 6* | *Low* | 3276 | 1.63 | 0.48 | 1 | 1 | 2 | 2 | 2 |
| | *Moderate* | 797 | 2.28 | 0.45 | 2 | 2 | 3 | 3 | 3 |
| | *Severe* | 861 | 3.17 | 0.43 | 2 | 3 | 3 | 3 | 5 |


Team 1
Fatigue classes (%)
low fatigue
moderate fatigue
severe fatigue


Team 2
Fatigue classes (%)
low fatigue
moderate fatigue
severe fatigue


Team 3
Fatigue classes (%)
low fatigue
moderate fatigue
severe fatigue


Team 4
Fatigue classes (%)
low fatigue
moderate fatigue
severe fatigue


Team 5
Fatigue classes (%)
low fatigue
moderate fatigue
severe fatigue

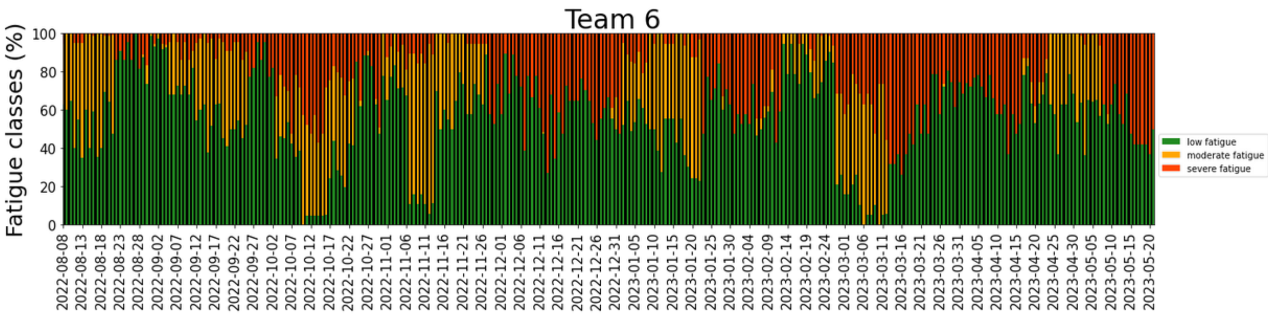



**Figure 1.** Distribution of Fatigue classes as the season went by for each team.

## Data preprocessing for daily analysis

For the daily analysis, data preprocessing is mandatory to perform the analysis correctly. Data preprocessing permits to aggregate the data time series to create indexes that provide more details about players' history. In particular, in addition to the daily values, two types of aggregations were computed for each independent feature (i.e., fatigue, sleep quality, muscle soreness, stress, mood ratings of the current day, sRPE and time of the previous day): i) exponential weighted moving average of past 7 days (Acute); ii) exponential weighted moving average of past 28 days (Chronic). The weight of both Acute and Chronic aggregation methods was computed with a specified decay in terms of span (2/(SPAN+2))[27].

## Data analysis

In this section we describe all the analysis conducted in this study. In particular, cross-correlation analysis (section 2.4.1.) was performed in order to detect time series association between TL and fatigue. The results of this analysis will permit us to assess the time lag between the two-time series. Based on the result of this analysis a framework of big data analytics was developed in order to predict the Fatigue of the next day and match (section 2.4.2.). Finally, in accordance with the cross-correlation and prediction analyses, a mediation analysis was conducted in order to assess if the Fatigue is directly affected by TL or is mediated by other self-reported wellness scores (section 2.4.3.). All the analyses were performed by using python 3.8 programming language.

## Cross-correlation analysis

In order to assess the time series relationship of TL to fatigue, a cross-correlation analysis was conducted. Cross-correlation is a measure of similarity of two series that permits to objectively determine how well they match up with each other and at what point the best match occurs. The correlation coefficient between the two times series could be ranged from -1.0 to +1.0. The closer the cross-correlation value is to 1, the more the two times series are similar.

## Machine learning approach

Two different analyses were performed, the first one (A) was developed by predicting next day fatigue by the preceding fatigue, sleep quality, muscle soreness, stress and mood perceptions and TL, while the second one (B) refers to the prediction of match-day fatigue, in particular:

- Daily fatigue (A) takes into consideration the history of the players in addition to daily features. In particular, the training history was computed as Acute (moving average of the previous 7 days) and Chronic workloads (moving average of the previous 28 days). To this aim, the Exponential Weighted Moving Average (EWMA) function with span decay on each normalized independent feature was used. Moreover, some contextual features that could be relevant to the perceived players' wellness status were also recorded: i) distance to previous match ($MD_{plus}$); ii) distance to next game ($MD_{minus}$). The main objective of this analysis is to understand fatigue tail and variations regardless of the day in the weekly schedule. Furthermore, all the chronic, acute, daily and contextual variables converge in the prediction of next day fatigue (fatigue d+1);
- Match-day fatigue (B) is constructed by focusing on match day fatigue. The predictive variables used for this prediction are based on the daily determinants on a weekly schedule (from d-1 to d-6 for all the variables, i.e. sleep, muscle soreness, fatigue, stress, mood, sRpe and time). Furthermore, all the variables analyzed for a weekly microcycle converge in the prediction of match-day fatigue. Only weeks with 1 match per week were used for the analysis.

A framework of big data analytics was developed in order to predict players' fatigue of the next day (daily fatigue) and on the match one (match-day fatigue). To these aims we train and test four machine learning models: i) Decision Tree classifier (DTC); ii) XGBoost classifier (XGB); iii) Random Forest Classifier (RFC); iv) Logistic regression (LR). Additionally, to assess the validity of the models trained, we compare the prediction results with a stratified dummy classifier model (Bs). Actually, Bs predicts the classes using a simple rule based on classes' distribution in the train set. This classifier is useful as a simple baseline to compare with the real classifiers. Even if machine learning models show high predictive accuracy, if the Bs performance is similar or higher than the trained models, these latter models are not able to accurately predict the output classes.

The models were validated by a train-test approach. This validation approach randomly splits the dataset into two main datasets, i.e. train and test sets. The train set (70% of the entire dataset) was used to train the machine learning models, while the test set (30% of the entire dataset) was used to evaluate the prediction ability of these models. The train and test split are performed by using a stratified approach, which permits to split the example in the dataset in the train and test sets in accordance with the distribution of WI values.

In the train set of both the validation approaches, a recursive feature elimination with 3-folds cross-validation (RFECV) was performed to select the most important features to fatigue prediction. This approach permits to reduce the feature dimension space increasing the interpretability of the models and their accuracy.

Precision, Recall and F1-score for each class and the accuracy were computed to detect the model's goodness. Precision (specificity) is the ratio of correctly predicted positive observations to the total predicted positive observations, while recall (sensitivity) is the ratio of correctly predicted positive observations to all observations in the actual class. Additionally, F1-score is the weighted mean of precision and recall. Finally, Accuracy is the ratio of correctly predicted observations to the total observations.

To globally and locally explain the decision-making process of the models, we compute SHapley Additive exPlanations (SHAP) values that allow us to explore the relationships between variables for predicted cases. In particular, SHAP assigns to each feature an importance value for a particular prediction (based on a linear function) permitting to evaluate the influence of each feature to final prediction. Moreover, the collective SHAP values can show how much each predictor contributes, either positively or negatively, to the target variable. Understanding why a model makes a certain prediction can be as crucial as the prediction's accuracy in many applications. Actually, inspecting the reasoning underlying the model's decisions can provide more profound insights into the differences in Fatigue classes.

## Mediation analysis

TL is, by definition, the only logical key that the practitioner can use to affect and modulate the fatigue of the next day for each player. In order to detect if and how the relationship between TL and Fatigue d+1 (Daily fatigue) or TL d-1 and Fatigue (match-day fatigue) is direct or mediated by the other perceived daily items (i.e. mood, sleep quality, muscle soreness, and stress) a mediation analysis was conducted. In order to take in consideration the assumption of related pairs, a repeated measures mediation analysis was performed. In particular, Marginal Regression Model using

Generalized Estimating Equations based on Autoregressive covariance structure was used to meet both time series and repeated measures assumptions.

# Results

## Cross-correlation

Section S1 in the supplementary material shows that TL is strongly related to Fatigue of the next day in all the soccer teams. This result indicates that the TL of the current day affects the fatigue ($r>0.88$) of the following days in all the soccer teams. For this reason, additionally to contextual features, the machine learning models are developed in order to predict Fatigue of the next day based on the TL of the current day, the past TL, the items of wellness index until the current day.

## Daily fatigue (A)

### Fatigue prediction

Table S1 in supplementary material provides the prediction goodness of the machine learning models on train-test scenarios. In particular for Fatigue prediction, XGB shows the highest predictive goodness for Fatigue in all the teams (Team 1: accuracy = 0.81; Team 2: accuracy = 0.80; Team 3: accuracy = 0.80; Team 4: accuracy = 0.79, Team 5: accuracy = 0.85, Team 6: accuracy = 0.87). All the machine learning models for Fatigue (Supplementary material Table S1) show higher predictive ability compared to the baseline ($B_s$) corroborating the fact that these models are able to detect patterns into the data that permits discrimination among Fatigue classes.

Figure 2 and Table 3 show the importance of each feature on the decision-making process for daily fatigue in all the six soccer teams.

The most important variable for daily fatigue prediction appears to be fatigue (30.4%). The other determinant variables for prediction are $MD_{plus}$ (15.1%), stress (11.7%), $MD_{minus}$ (11.3%), and mood (6.2%). Hence almost all the most important variables are related to the day previous to the prediction and to the contextual factors representing the distance of the training to the matches.

The general feature influence of every variable for all the teams was reported in Figure 3. This figure shows the relationship between individuals' features values and the probability of being part of a specific class. Positive relationship (green cells) indicates that the higher this value is, the higher the probability to be part of that specific class is, while vice versa for negative relationships. Specifically, a strong correlation is present in the relationship between Severe Fatigue status, sRPE and fatigue of the previous day and a moderate correlation between Severe Fatigue status and High stress and bad mood and $mood_{acute}$ underlining how an increase of these variables induce a growth in the fatigue status. Contrariwise, sRPE and fatigue of the previous day show a strong negative correlation with a Low Fatigue status together with a moderate negative correlation with $time_{acute}$ and $mood_{acute}$. The influence of each feature on Fatigue prediction slightly changes among soccer teams indicating that the fatigue status is affected differently in accordance with players', teams' and training schedules' characteristics. Even with this small difference between teams, the models could be considered generalizable due to the fact that the best model trained on each team is accurate to predict Fatigue classes in the others. As a matter of fact, accuracy of the best model trained on one team and tested on the other teams is strong (Table 3). However, due to small differences in feature, influence in the six soccer teams showed in Figure 3 does not permit the same prediction ability detected when the models are trained and tested on data derived from the same soccer team.

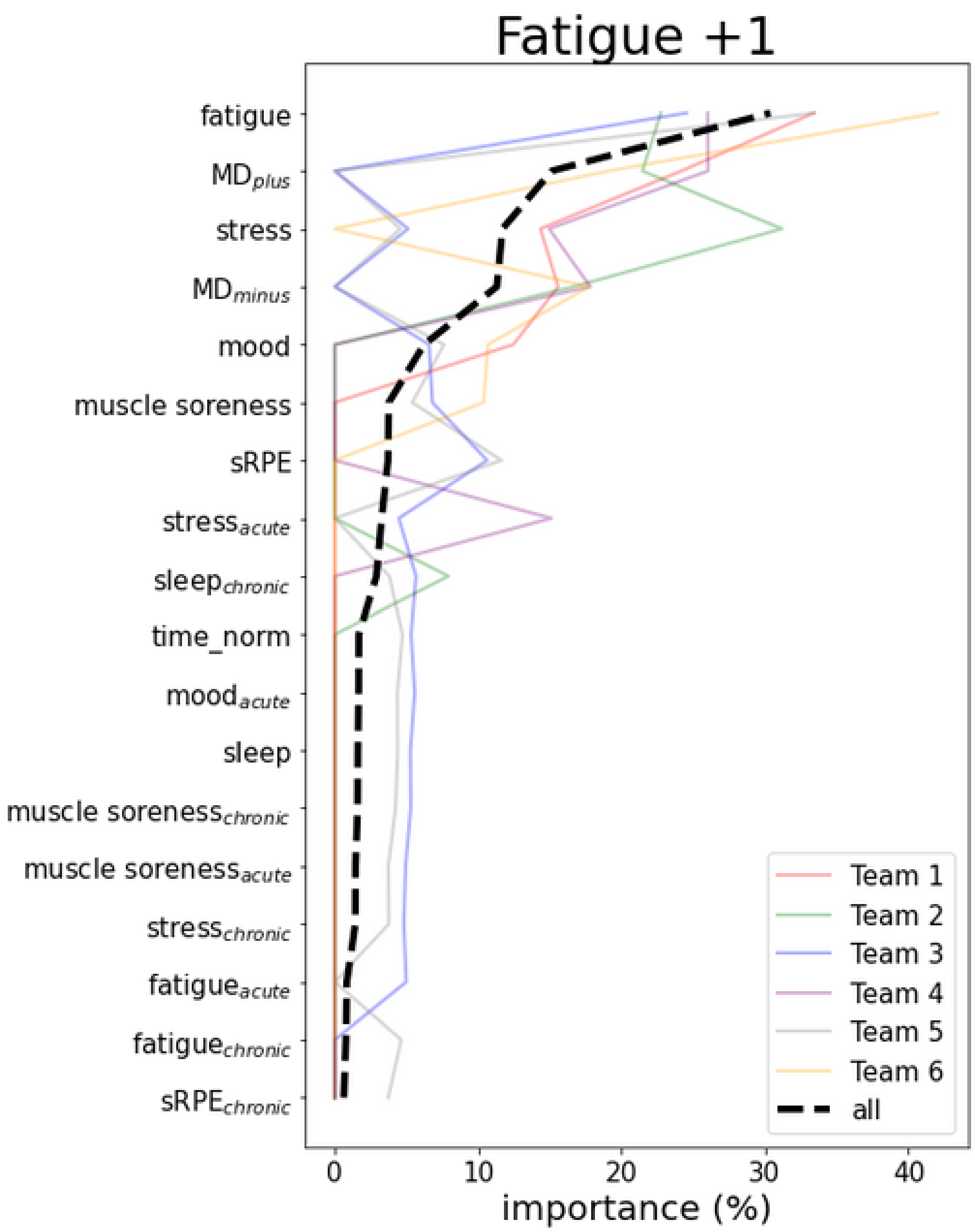



**Figure 2.** Feature importance for the best machine learning models for all the variables and Fatigue d+1 in all the teams.

**Table 3.** Feature importance for each team.

| | Team 1 | Team 2 | Team 3 | Team 4 | Team 5 | Team 6 | all | std |
|---|---|---|---|---|---|---|---|---|
| Fatigue | 33.41 | 22.76 | 24.56 | 26.03 | 33.41 | 42.05 | 30.37 | 6.65 |
| MD plus | 24.20 | 21.47 | 0.00 | 26.06 | 0.00 | 19.14 | 15.14 | 10.92 |
| Stress | 14.35 | 31.18 | 5.11 | 14.96 | 4.58 | 0.00 | 11.70 | 10.23 |
| MD minus | 15.59 | 16.67 | 0.00 | 17.83 | 0.00 | 17.74 | 11.30 | 8.03 |
| Mood | 12.45 | 0.00 | 6.57 | 0.00 | 7.62 | 10.69 | 6.22 | 4.80 |
| Muscle soreness | 0.00 | 0.00 | 6.81 | 0.00 | 5.40 | 10.38 | 3.76 | 4.05 |
| sRPE | 0.00 | 0.00 | 10.63 | 0.00 | 11.63 | 0.00 | 3.71 | 5.25 |
| Stress acute | 0.00 | 0.00 | 4.46 | 15.12 | 0.00 | 0.00 | 3.26 | 5.55 |
| Sleep chronic | 0.00 | 7.92 | 5.66 | 0.00 | 3.80 | 0.00 | 2.90 | 3.13 |
| Time | 0.00 | 0.00 | 5.32 | 0.00 | 4.72 | 0.00 | 1.67 | 2.37 |
| Mood acute | 0.00 | 0.00 | 5.58 | 0.00 | 4.36 | 0.00 | 1.66 | 2.37 |
| Sleep | 0.00 | 0.00 | 5.27 | 0.00 | 4.38 | 0.00 | 1.61 | 2.29 |
| Soreness chronic | 0.00 | 0.00 | 5.30 | 0.00 | 4.23 | 0.00 | 1.59 | 2.27 |
| Muscle soreness acute | 0.00 | 0.00 | 4.93 | 0.00 | 3.73 | 0.00 | 1.44 | 2.07 |
| Stress chronic | 0.00 | 0.00 | 4.81 | 0.00 | 3.77 | 0.00 | 1.43 | 2.04 |
| Fatigue acute | 0.00 | 0.00 | 4.98 | 0.00 | 0.00 | 0.00 | 0.83 | 1.86 |
| Fatigue chronic | 0.00 | 0.00 | 0.00 | 0.00 | 4.63 | 0.00 | 0.77 | 1.73 |
| sRPE chronic | 0.00 | 0.00 | 0.00 | 0.00 | 3.73 | 0.00 | 0.62 | 1.39 |

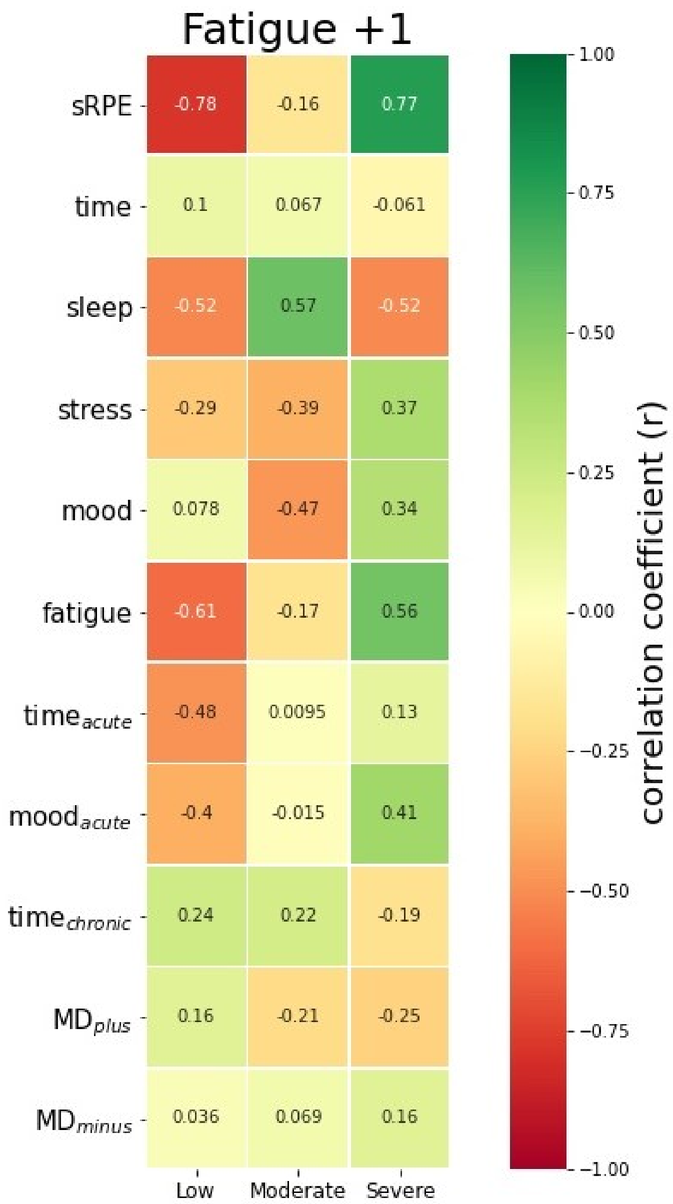



**Figure 3.** General influence of individuals' features on predictions. This plot shows the correlation coefficient between SHAP values (probability to be part of a fatigue class) and features' values. Green bars refer to a positive correlation, while the red ones show a negative relationship.

**Table 4.** Accuracy of the XGB model trained on one team and tested on the other teams.

| | | Training | | | | | |
|---|---|---|---|---|---|---|---|
| | | Team 1 | Team 2 | Team 3 | Team 4 | Team 5 | Team 6 |
| Test | Team 1 | --- | 0.68 | 0.61 | 0.64 | 0.63 | 0.65 |
| | Team 2 | 0.61 | --- | 0.63 | 0.69 | 0.65 | 0.67 |
| | Team 3 | 0.71 | 0.67 | --- | 0.64 | 0.66 | 0.69 |
| | Team 4 | 0.64 | 0.67 | 0.71 | --- | 0.69 | 0.71 |
| | Team 5 | 0.65 | 0.63 | 0.68 | 0.69 | --- | 0.69 |
| | Team 6 | 0.68 | 0.64 | 0.69 | 0.67 | 0.69 | --- |

**Mediation analysis**

A statistically significant total influence of TL on fatigue d+1 was detected (coefficient = -0.118 [-0.156, -0.081]; p-value < 0.001). The mediation analysis (Figure 4) shows a statistically significant direct influence (coefficient = -0.069 [-0.094, -0.042]; p-value = 0.55) of TL on Fatigue d+1. This relation is strongly mediated by muscle soreness (coefficient = -0.043 [-0.082, -0.047]; p-value < 0.001), sleep (coefficient = -0.019 [-0.022, -0.017]; p-value < 0.001) and stress (coefficient = -0.009 [-0.012, -0.007]; p-value < 0.001). The other perceived items do not mediate this relationship.

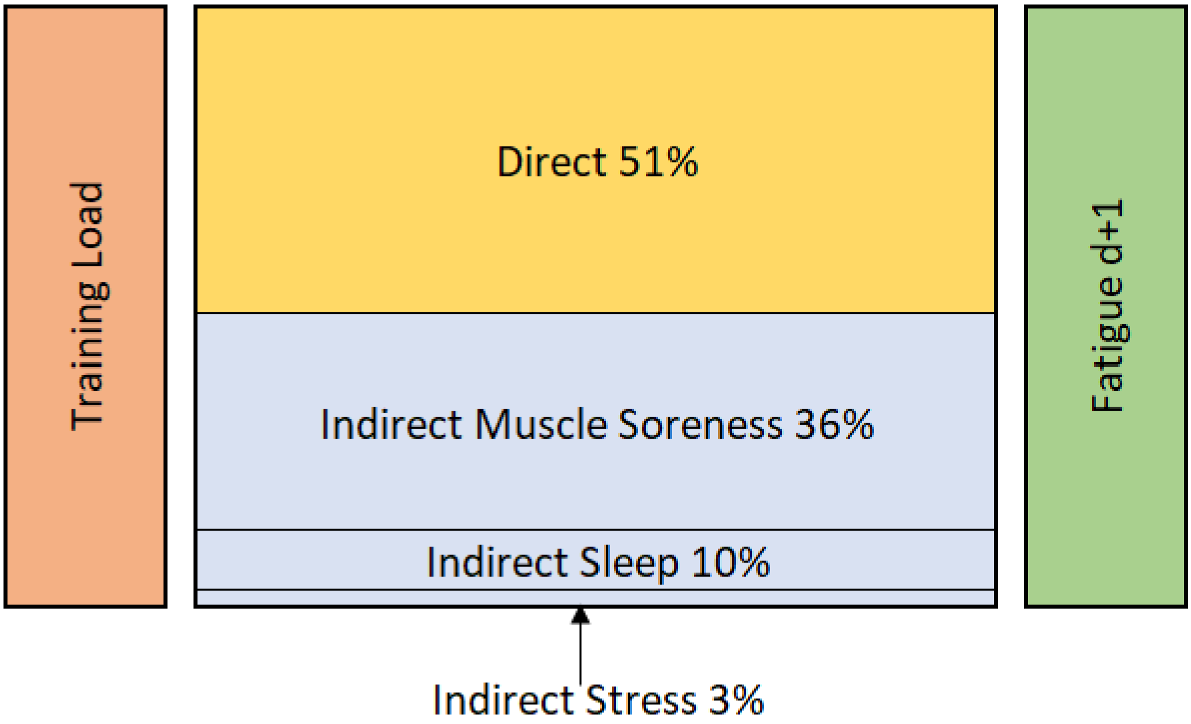



**Figure 4.** Mediation analysis on daily fatigue

## Match-day fatigue (B)

### Fatigue prediction

Table S2 in supplementary material provides the prediction goodness of the machine learning models. In particular for predicting match-day Fatigue, XGB shows the highest predictive goodness for Fatigue in all the teams (Team 1: accuracy = 0.70; Team 2: accuracy = 0.71; Team 3: accuracy = 0.72; Team 4: accuracy = 0.77; Team 5: accuracy = 0.82; Team 6: accuracy = 0.70). Supplementary material Table S2 shows that all the machine learning models are valid to estimate match-day Fatigue because the XGB model shows higher predictive performance compared to the baseline one ($B_s$).

Figure 5 shows the importance of each feature on the decision-making process for match-day Fatigue in all six soccer teams. All the variables of the day before the match are in the decision making to predict match-day Fatigue together with mood d-2. Specifically, stress d-1, mood d-1 and mood d-2 and sleep d-1 (9.30%, 9.40%, 7.61% and 8.53% respectively) are the most relevant variables in the prediction. Muscle soreness d-1, fatigue d-1, sRPE d-1 and time d-1 also have relevancy in the prediction model even if lower than the previous ones. All the other physical, psychological or environmental variables describe a *mare magnum* with little importance in the decision making of the process.

The general feature influence of every feature for all the teams was reported in Figure 6. The influence of each feature on match-day fatigue prediction slightly changes among soccer teams indicating that the fatigue status is affected differently in accordance with players, teams and training schedules characteristics. Even with this small difference between teams, the models could be considered generalizable due to the fact that the best model trained on each team is accurate to predict Fatigue classes in the others. As a matter of fact, accuracy of the best model trained on one team and tested on the other teams is strong (Table 6). Not surprisingly, in figure 6 is highlighted how every fatigue-1, sleep-1, stress-1 and mood-1 appear to have a strong negative correlation with Low Fatigue class, while muscle soreness-1 has a moderate negative correlation with Low Fatigue class. Contrarywise, sRpe-1, sleep -1, mood-1 and mood-2 appear to have a positive correlation with Severe Fatigue class.

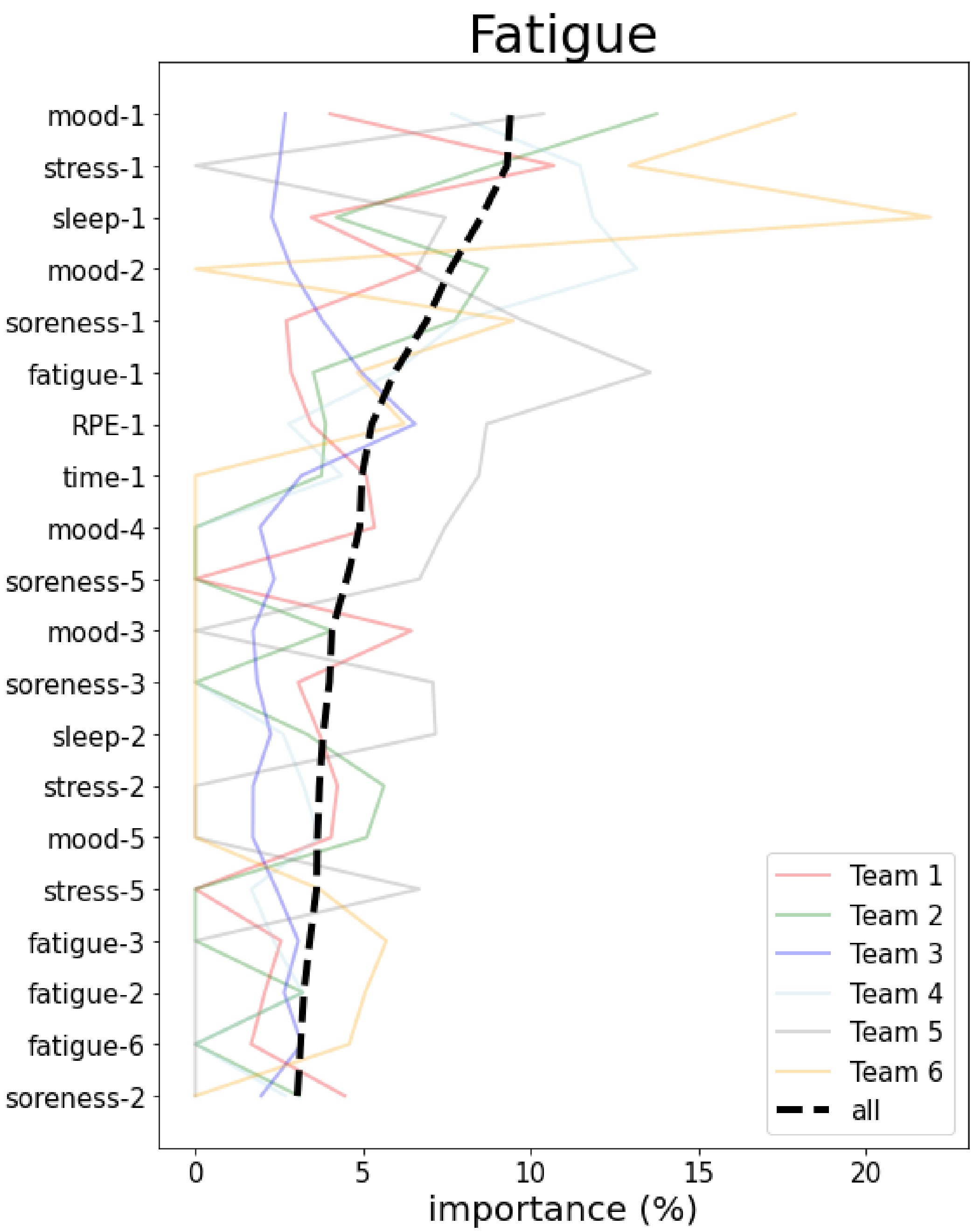



**Figure 5.** Feature importance for the best machine learning models for the perceptual variables in all the teams.

**Table 5.** Feature importance in each team.

| | Team 1 | Team 2 | Team 3 | Team 4 | Team 5 | Team 6 | all | Std |
|---|---|---|---|---|---|---|---|---|
| mood_1 | 4.01 | 13.77 | 2.68 | 7.66 | 10.38 | 17.90 | 9.40 | 5.31 |
| stress_1 | 10.70 | 8.86 | 2.50 | 11.48 | 0.00 | 12.95 | 9.30 | 3.65 |
| sleep_1 | 3.45 | 4.20 | 2.27 | 11.86 | 7.45 | 21.94 | 8.53 | 6.78 |
| mood_2 | 6.71 | 8.72 | 2.87 | 13.16 | 6.59 | 0.00 | 7.61 | 3.36 |
| soreness_1 | 2.71 | 7.74 | 3.79 | 7.89 | 9.78 | 9.48 | 6.90 | 2.70 |
| fatigue_1 | 2.85 | 3.52 | 4.95 | 5.93 | 13.58 | 4.84 | 5.95 | 3.56 |
| RPE_1 | 3.45 | 3.88 | 6.55 | 2.78 | 8.69 | 6.23 | 5.27 | 2.07 |
| time_1 | 5.10 | 3.76 | 3.16 | 4.37 | 8.45 | 0.00 | 4.97 | 1.86 |
| mood_4 | 5.34 | 0.00 | 1.92 | 0.00 | 7.45 | 0.00 | 4.90 | 2.28 |
| soreness_5 | 0.00 | 0.00 | 2.35 | 0.00 | 6.69 | 0.00 | 4.52 | 2.17 |
| mood_3 | 6.44 | 4.06 | 1.72 | 0.00 | 0.00 | 0.00 | 4.07 | 1.93 |
| soreness_3 | 3.06 | 0.00 | 1.84 | 0.00 | 7.09 | 0.00 | 4.00 | 2.24 |
| sleep_2 | 3.70 | 3.32 | 2.24 | 2.62 | 7.17 | 0.00 | 3.81 | 1.75 |
| stress_2 | 4.24 | 5.62 | 1.72 | 3.23 | 0.00 | 0.00 | 3.70 | 1.43 |
| mood_5 | 4.05 | 5.11 | 1.71 | 3.68 | 0.00 | 0.00 | 3.64 | 1.23 |
| stress_5 | 0.00 | 0.00 | 2.41 | 1.67 | 6.67 | 3.70 | 3.61 | 1.91 |
| fatigue_3 | 2.55 | 0.00 | 3.07 | 2.34 | 0.00 | 5.70 | 3.41 | 1.35 |
| fatigue_2 | 2.07 | 3.20 | 2.65 | 3.18 | 0.00 | 5.07 | 3.23 | 1.01 |
| fatigue_6 | 1.66 | 0.00 | 3.17 | 0.00 | 0.00 | 4.59 | 3.14 | 1.19 |
| soreness_2 | 4.45 | 3.08 | 1.96 | 2.67 | 0.00 | 0.00 | 3.04 | 0.91 |
| stress_6 | 0.00 | 2.55 | 2.14 | 0.00 | 0.00 | 4.26 | 2.98 | 0.92 |
| time_4 | 0.00 | 0.00 | 1.99 | 3.39 | 0.00 | 0.00 | 2.69 | 0.70 |
| stress_3 | 2.19 | 3.74 | 1.73 | 0.00 | 0.00 | 0.00 | 2.56 | 0.86 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| stress_4 | 1.41 | 3.65 | 2.11 | 0.00 | 0.00 | 0.00 | 2.39 | 0.93 |
| mood_6 | 2.29 | 2.94 | 1.48 | 2.63 | 0.00 | 0.00 | 2.33 | 0.54 |
| sleep_3 | 1.55 | 2.66 | 2.64 | 0.00 | 0.00 | 0.00 | 2.29 | 0.52 |
| soreness_6 | 2.16 | 2.69 | 1.82 | 0.00 | 0.00 | 0.00 | 2.22 | 0.36 |
| RPE_2 | 0.00 | 0.00 | 2.21 | 0.00 | 0.00 | 0.00 | 2.21 | 0.00 |
| time_3 | 0.97 | 0.00 | 2.33 | 2.02 | 0.00 | 3.36 | 2.17 | 0.85 |
| sleep_4 | 1.69 | 2.92 | 1.84 | 0.00 | 0.00 | 0.00 | 2.15 | 0.55 |
| soreness_4 | 1.75 | 0.00 | 2.40 | 0.00 | 0.00 | 0.00 | 2.08 | 0.32 |
| fatigue_4 | 1.59 | 0.00 | 2.56 | 0.00 | 0.00 | 0.00 | 2.07 | 0.48 |
| time_5 | 1.14 | 0.00 | 1.71 | 3.13 | 0.00 | 0.00 | 2.00 | 0.84 |
| RPE_3 | 0.00 | 0.00 | 1.99 | 0.00 | 0.00 | 0.00 | 1.99 | 0.00 |
| sleep_5 | 1.91 | 0.00 | 1.92 | 0.00 | 0.00 | 0.00 | 1.91 | 0.00 |
| RPE_5 | 0.00 | 0.00 | 1.87 | 0.00 | 0.00 | 0.00 | 1.87 | 0.00 |
| RPE_4 | 0.00 | 0.00 | 1.84 | 0.00 | 0.00 | 0.00 | 1.84 | 0.00 |
| RPE_6 | 0.00 | 0.00 | 1.80 | 0.00 | 0.00 | 0.00 | 1.80 | 0.00 |
| fatigue_5 | 1.30 | 0.00 | 2.28 | 0.00 | 0.00 | 0.00 | 1.79 | 0.49 |
| time_2 | 0.95 | 0.00 | 2.06 | 2.36 | 0.00 | 0.00 | 1.79 | 0.61 |
| time_6 | 1.18 | 0.00 | 2.18 | 1.95 | 0.00 | 0.00 | 1.77 | 0.43 |
| sleep_6 | 1.36 | 0.00 | 1.54 | 0.00 | 0.00 | 0.00 | 1.45 | 0.09 |

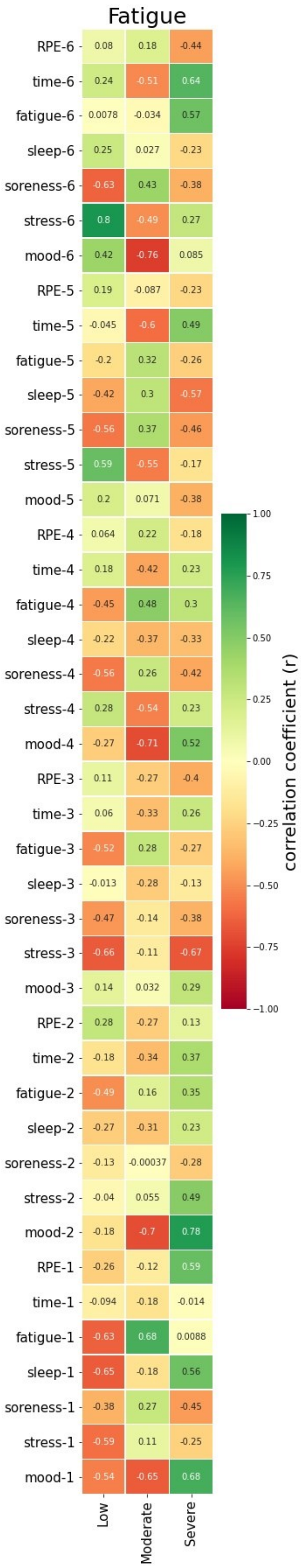

Fatigue
correlation coefficient (r)
Low
Moderate
Severe

**Figure 6.** General influence of individuals' features on all the variables and match-day fatigue predictions. This plot shows the correlation coefficient between SHAP values (probability to be part of a Fatigue class) and features' values. Green bars refer to a positive correlation, while the red ones show a negative relationship.

**Table 6.** Accuracy of the XGB model trained on one team and tested on the other teams.

| | | Training | | | | | |
|---|---|---|---|---|---|---|---|
| | | Team 1 | Team 2 | Team 3 | Team 4 | Team 5 | Team 6 |
| Test | Team 1 | --- | 0.58 | 0.56 | 0.54 | 0.52 | 0.53 |
| | Team 2 | 0.51 | --- | 0.53 | 0.56 | 0.57 | 0.59 |
| | Team 3 | 0.58 | 0.57 | --- | 0.51 | 0.53 | 0.51 |
| | Team 4 | 0.52 | 0.55 | 0.59 | --- | 0.56 | 0.59 |
| | Team 5 | 0.54 | 0.53 | 0.53 | 0.57 | --- | 0.52 |
| | Team 6 | 0.57 | 0.56 | 0.60 | 0.53 | 0.55 | --- |

**Mediation analysis**

Figure 7 shows a statistically significant total influence of TL d-1 on match-day fatigue was detected (coefficient = 0.0088 [0.006, 0.010]; p-value < 0.001). The mediation analysis (Figure 7) shows a statistically significant direct influence (coefficient = 0.006 [0.004, 0.007]; p-value < 0.001) of TL on Fatigue. This relation is mediated by Muscle Soreness (coefficient = 0.002 [0.001, 0.003]; p-value < 0.001), Mood (coefficient = 0.001 [0.001, 0.002]; p-value < 0.001). The other items do not mediate this relationship.

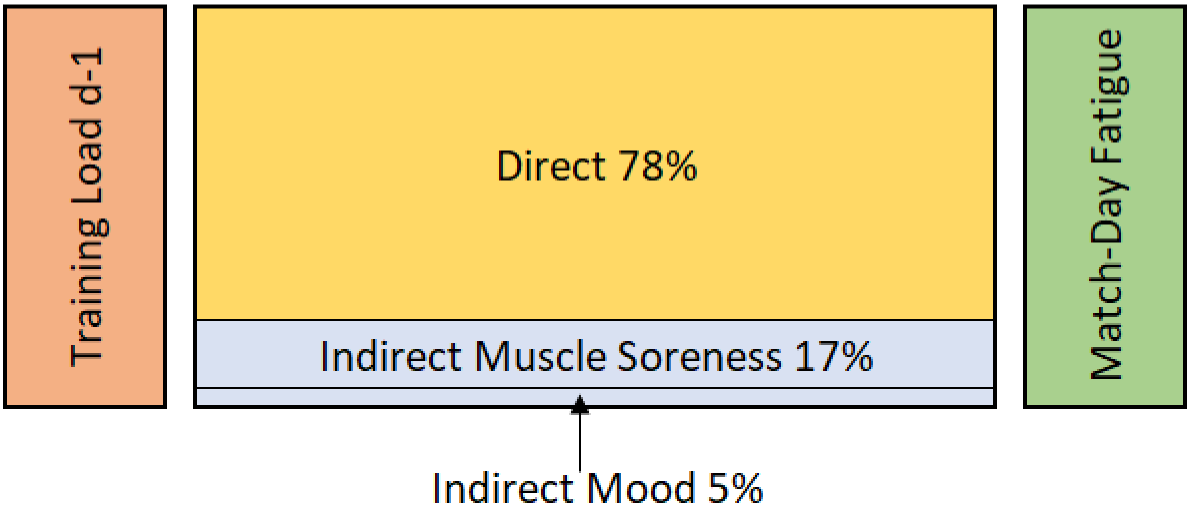



**Figure 7.** Mediation analysis on match-day fatigue

## Discussion

In a real-world scenario the importance of predicting the fatigue status of a single player is paramount for training prescription and periodization in any day of the season. The aim of the present study was to explore the influence of sRPE and time of each session and/or match and self-reported perceived items' scores (i.e., fatigue, muscle soreness, stress, mood and sleep quality) on next day subjective fatigue and match-day subjected fatigue in professional soccer players. The goal is to provide coaches a useful tool to control training stimuli and prevent states of overreaching and/or overtraining.

### Daily fatigue

Analysing the data, the most surprising finding is that training duration and intensity (i.e. time and sRPE variables) have low importance in the prediction of daily fatigue. When interpreting these findings, it is important to remember that our fatigue ratings refer to the fatigue experienced the morning after the training session or match and, thus, several hours of recovery. However, with regards to daily fatigue, the aim of our study was to identify the predictors of the fatigue state experienced by the player before the training session in order to modulate the training load.

Another explanation for the relatively low importance of training duration/intensity in predicting daily fatigue is that factors other than the training load of have a stronger influence on subjective feelings of fatigue measured in the morning. Specifically, our findings suggest that both mood and stress influence the state of fatigue experienced by the player. This interpretation of our correlative evidence is supported by both clinical and experimental observations that negative mood and stress are strongly associated with subjective feelings of fatigue and reduced performance in athletes[11,27].

An additional potential explanation for the relatively low importance of training duration/intensity in predicting daily fatigue is that modulation of training duration/intensity is already accounted for by the match-day plus and match-day minus variables. Actually, practitioners define the schedule training programme of the week in accordance with intra-match duration. The distance from the previous and to the next match respectively, permits to perform specific training tasks related to recovery strategy (rest time from previous match and recovery period to the next match) to reach the best recovery status in the match day. For example, resistance and intensity training are performed usually in MD+2-5 and MD+4-3, respectively. Hence the duration and intensity of the training change in accordance to training objectives resulting in different players' response from a similar TL. For this reason, knowing the distance from and to the previous and next matches, respectively, permits to have an idea of the typologies of training that may result in different fatigue response.

## Match-day fatigue

Contrary to daily fatigue, the relatively low importance of fatigue, muscle soreness, and training duration/intensity in predicting match-day fatigue is not surprising because coaches prescribe a moderate-to-low training load the days before the match with the goal of having no or low fatigue on the day of the match. In our study, the most important factors predicting fatigue on the day of the match are stress, mood and sleep quality in the day(s) preceding the match. We tentatively interpret these correlative findings to suggest that the residual fatigue before the match is mental rather than physical. Given the experimental evidence that mental fatigue reduces soccer performance[11,27], our findings have two practical implications. First, coaches and other staff (e.g. the team psychologist) should try managing the psychological load in the day(s) before the match in addition to the standard tapering of the physical load. Examples of such interventions are relaxation strategies and sleep hygiene. Given that psychological load is not always easy to manage, coaches and other staff should also employ other strategies to reduce mental fatigue on the day of the match such as alimentary (e.g., caffeine), behavioural (e.g., listening to music) and psychological (e.g., extrinsic motivation) countermeasures[28].

XGB model trained on a team is accurate to predict fatigue classes in the other ones highlighting the fact that the models developed to predict Fatigue are generalizable (Table 6). Obviously, lower prediction ability was detected in models trained in a different soccer team compared to the one trained on the same team because it is less personalized. The differences detected in features importance in the six teams explain the reduction in prediction performance when the models are not trained with the same team. However, this result indicates that the players from different teams have a similar response to training load and players' health features on fatigue status. Hence, training a machine learning algorithm in a big dataset with several teams and/or several seasons could increase the generalizability of the models.

Evaluating the prediction goodness in a real scenario gives important insights about the usability of this framework of big data analytics. This validation approach permits testing the prediction goodness of the machine learning algorithms as the season goes by and assessing the change in feature importance and influence in each training week. Actually, RFC shows a cumulative performance goodness at the end of the season between 78-81%. The accuracy of the prediction models in all the teams is similar during all the seasons indicating that 10 weeks of the machine learning model training is enough to create accurate models for fatigue prediction. However, the influence of each feature changes as the seasons go by. This is due to the fact that different competition demands and players' physical status in different parts of the season is related to a different perception of fatigue.

## Limitations

The current research presents some weaknesses. Firstly, as already repeatedly discussed, these data are correlattive and, therefore, do not prove a cause-and-effect relationship between fatigue and its predictors. Secondly, the models discussed here are specific to the teams that were monitored. Although the items adopted in this study represent custom measures used widely in practice, they haven't been rigorously evaluated for their validity and reliability. Moreover, the current algorithm cannot be extended to all soccer teams. However, the model provided needs to be "trained and fitted", for at least 10 weeks, to have a reliable assessment on fatigue prediction for each team.

## Conclusion

In the context of the current training and match scheduling practices of Italian professional soccer, our findings suggest that both daily and match-day fatigue can be predicted with reasonable accuracy using big data analytics of practical measures of training load and players subjective status. The prediction of daily fatigue can assist coaches in the planning of training sessions day by day following the individual responses of each player. The prediction of match-day fatigue can help in the preparation leading up to the match in understanding, for example, which player can be more suitable for the subsequent match.

Not surprisingly, the most important predictor of daily fatigue is the fatigue experienced the previous day. Distance of the training session from the matches is also important. Interestingly, the stress experienced by the athlete the day before contributes significantly to the prediction of daily fatigue. Psychological variables like stress, mood and sleep quality are also the most important predictors of match-day fatigue. Together with the results of experimental studies showing a negative impact of mental fatigue on soccer performance[11,27], our correlative findings suggest that interventions to reduce stress/negative mood, and improve sleep quality may improve the match performance of soccer players. Our correlative findings also suggest that the stress perceived by soccer players need to be considered in the day-to-day modulation of training load based on the current fatigue state.

In conclusion, athletic trainers and coaches can use the framework of big data analytics proposed in this paper to predict their players' fatigue status, understanding the influence of other perceptions, judgements and training load characteristics in order to improve the decision-making process when designing a training plan. In particular, field experts could maximize the training effect by controlling the fatigue status of the soccer players simulating the scheduled training programme. Finally, this approach can be very useful for practitioners of amateur and grassroots with a limited budget. Both status and internal load measures' data collection, as shown above, is virtually free and it can be taken into account in assessing fatigue by our model.

# References


1. Bartley SH, Chute E. A preliminary clarification of the concept of fatigue. *Psychological Review*. 1945;52(3):169-174. doi:10.1037/h0059244
2. Krustrup P, Mohr M, Steensberg A, Bencke J, Kjaer M, Bangsbo J. Muscle and blood metabolites during a soccer game: implications for sprint performance. *Med Sci Sports Exerc*. 2006;38(6):1165-1174. doi:10.1249/01.mss.0000222845.89262.cd
3. Mohr M, Krustrup P, Bangsbo J. Fatigue in soccer: A brief review. *Journal of Sports Sciences*. 2005;23(6):593-599. doi:10.1080/02640410400021286
4. Mohr M, Krustrup P, Nybo L, Nielsen JJ, Bangsbo J. Muscle temperature and sprint performance during soccer matches – beneficial effect of re-warm-up at half-time. *Scandinavian Journal of Medicine & Science in Sports*. 2004;14(3):156-162. doi:10.1111/j.1600-0838.2004.00349.x
5. Rahnama N, Reilly T, Lees A, Graham-Smith P. Muscle fatigue induced by exercise simulating the work rate of competitive soccer. *Journal of Sports Sciences*. 2003;21(11):933-942. doi:10.1080/0264041031000140428
6. Russell M, Sparkes W, Northeast J, et al. Changes in Acceleration and Deceleration Capacity Throughout Professional Soccer Match-Play. *Journal of Strength and Conditioning Research*. 2016;30(10):2839-2844. doi:10.1519/JSC.0000000000000805
7. Sun H, Soh KG, Mohammadi A, Wang X, Bin Z, Zhao Z. Effects of mental fatigue on technical performance in soccer players: A systematic review with a meta-analysis. *Frontiers in Public Health*. 2022;10. Accessed October 18, 2023. https://www.frontiersin.org/articles/10.3389/fpubh.2022.922630
8. Russell M, Benton D, Kingsley M. The Effects of Fatigue on Soccer Skills Performed During a Soccer Match Simulation. *International Journal of Sports Physiology and Performance*. 2011;6(2):221-233. doi:10.1123/ijspp.6.2.221
9. Rampinini E, Impellizzeri FM, Castagna C, Azzalin A, Ferrari Bravo D, Wisløff U. Effect of match-related fatigue on short-passing ability in young soccer players. *Med Sci Sports Exerc*. 2008;40(5):934-942. doi:10.1249/mss.0b013e3181666eb8
10. Impellizzeri FM, Rampinini E, Maffiuletti NA, Castagna C, Bizzini M, Wisløff U. Effects of aerobic training on the exercise-induced decline in short-passing ability in junior soccer players. *Appl Physiol Nutr Metab*. 2008;33(6):1192-1198. doi:10.1139/H08-111
11. Smith M, Coutts A, Merlini M, Deprez D, Lenoir M, Marcora S. Mental Fatigue Impairs Soccer-Specific Physical and Technical Performance. *Medicine and science in sports and exercise*. 2015;48:267-276. doi:10.1249/MSS.0000000000000762
12. Coutinho D, Gonçalves B, Travassos B, Wong DP, Coutts AJ, Sampaio JE. Mental Fatigue and Spatial References Impair Soccer Players' Physical and Tactical Performances. *Frontiers in Psychology*. 2017;8. Accessed October 18, 2023. https://www.frontiersin.org/articles/10.3389/fpsyg.2017.01645
13. Gantois P, Caputo Ferreira ME, Lima-Junior DD, et al. Effects of mental fatigue on passing decision-making performance in professional soccer athletes. *European Journal of Sport Science*. 2020;20(4):534-543. doi:10.1080/17461391.2019.1656781
14. Smith MR, Zeuwts L, Lenoir M, Hens N, De Jong LMS, Coutts AJ. Mental fatigue impairs soccer-specific decision-making skill. *Journal of Sports Sciences*. 2016;34(14):1297-1304. doi:10.1080/02640414.2016.1156241

15. Buchheit M, Racinais S, Bilsborough JC, et al. Monitoring fitness, fatigue and running performance during a pre-season training camp in elite football players. *Journal of Science and Medicine in Sport*. 2013;16(6):550-555. doi:10.1016/j.jsams.2012.12.003

16. Fields JB, Merigan JM, Gallo S, White JB, Jones MT. External and Internal Load Measures During Preseason Training in Men Collegiate Soccer Athletes. *The Journal of Strength & Conditioning Research*. 2021;35(9):2572. doi:10.1519/JSC.0000000000004092

17. Selmi O, Marzouki H, Ouergui I, BenKhalifa W, Bouassida A. Influence of intense training cycle and psychometric status on technical and physiological aspects performed during the small-sided games in soccer players. *Research in Sports Medicine*. 2018;26(4):401-412. doi:10.1080/15438627.2018.1492398

18. Selmi O, Ouergui I, Castellano J, Levitt D, Bouassida A. Effect of an intensified training period on well-being indices, recovery and psychological aspects in professional soccer players. *European Review of Applied Psychology*. 2020;70(6):100603. doi:10.1016/j.erap.2020.100603

19. Selmi O, Ouergui I, E Levitt D, et al. Training, psychometric status, biological markers and neuromuscular fatigue in soccer. *bs*. 2022;39(2):319-327. doi:10.5114/biolsport.2022.104065

20. Haddad M, Chaouachi A, Wong DP, et al. Influence of fatigue, stress, muscle soreness and sleep on perceived exertion during submaximal effort. *Physiology & Behavior*. 2013;119:185-189. doi:10.1016/j.physbeh.2013.06.016

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

22. Winder N, Russell M, Naughton RJ, Harper LD. The Impact of 120 Minutes of Match-Play on Recovery and Subsequent Match Performance: A Case Report in Professional Soccer Players. *Sports (Basel)*. 2018;6(1):22. doi:10.3390/sports6010022

23. McLean BD, Coutts AJ, Kelly V, McGuigan MR, Cormack SJ. Neuromuscular, endocrine, and perceptual fatigue responses during different length between-match microcycles in professional rugby league players. *Int J Sports Physiol Perform*. 2010;5(3):367-383. doi:10.1123/ijspp.5.3.367

24. Perri E, Simonelli C, Rossi A, Trecroci A, Alberti G, Iaia FM. Relationship Between Wellness Index and Internal Training Load in Soccer: Application of a Machine Learning Model. *International Journal of Sports Physiology and Performance*. 2021;16(5):695-703. doi:10.1123/ijspp.2020-0093

25. Borg G. *Borg's Perceived Exertion And Pain Scales*.; 1998.

26. Staunton CA, Abt G, Weaving D, Wundersitz DWT. Misuse of the term 'load' in sport and exercise science. *Journal of Science and Medicine in Sport*. 2022;25(5):439-444. doi:10.1016/j.jsams.2021.08.013

27. Van Cutsem J, Marcora S, De Pauw K, Bailey S, Meeusen R, Roelands B. The Effects of Mental Fatigue on Physical Performance: A Systematic Review. *Sports Med*. 2017;47(8):1569-1588. doi:10.1007/s40279-016-0672-0

28. Proost M, Habay J, De Wachter J, et al. How to Tackle Mental Fatigue: A Systematic Review of Potential Countermeasures and Their Underlying Mechanisms. *Sports Medicine*. 2022;52. doi:10.1007/s40279-022-01678-z

# Supplementary material

## S1. Cross-correlation analysis

In order to assess the time series relationship of TL to fatigue, a cross-correlation analysis was conducted. Cross-correlation is a measure of similarity of two series that permits to objectively determine how well they match up with each other and at what point the best match occurs. The correlation coefficient between the two times series could be ranged from -1.0 to +1.0. The closer the cross-correlation value is to 1, the more the two times series are similar.

TL is strongly related to fatigue of the next day in all the soccer teams. As a matter of fact, Figure S1 shows the highest cross-correlation values on lag +1 (Team 1: r=0.89; Team 2: r=0.90; Team 3: r=0.88; Team 4: r=0.92; Team 5: r=0.89; Team 6: r=0.92). This result indicates that the TL of the current day affects the perceived fatigue of the following days.

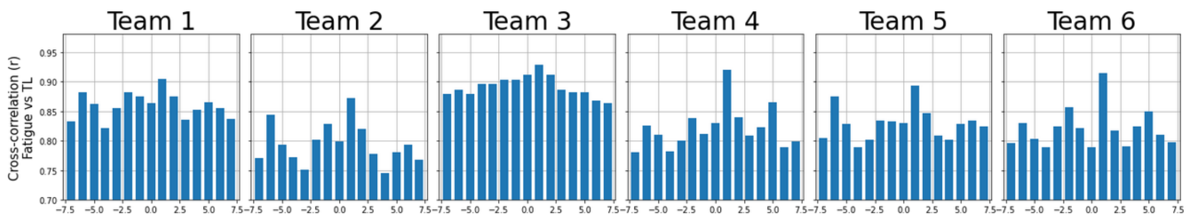



**Figure S1.** Cross-correlation analysis between Fatigue and TL.

**Table S1 - 1st part.** Prediction goodness of Fatigue in daily approach

| Model | classes | Team 1 | | | | Team 2 | | | | Team 3 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | prec | rec | f1 | acc | Prec | rec | f1 | Acc | Prec | rec | f1 | acc |
| DTC | Low | 0.82 | 0.70 | 0.77 | 0.72 | 0.82 | 0.79 | 0.80 | 0.72 | 0.85 | 0.79 | 0.82 | 0.71 |
| | Moderate | 0.62 | 0.64 | 0.63 | | 0.54 | 0.58 | 0.56 | | 0.35 | 0.43 | 0.38 | |
| | Severe | 0.50 | 0.55 | 0.53 | | 0.61 | 0.62 | 0.62 | | 0.37 | 0.44 | 0.40 | |
| XGB | Low | 0.82 | 0.80 | 0.81 | 0.81* | 0.85 | 0.87 | 0.86 | 0.80* | 0.86 | 0.91 | 0.87 | 0.80* |
| | Moderate | 0.75 | 0.73 | 0.74 | | 0.68 | 0.68 | 0.68 | | 0.52 | 0.41 | 0.46 | |
| | Severe | 0.72 | 0.57 | 0.63 | | 0.74 | 0.69 | 0.71 | | 0.60 | 0.47 | 0.53 | |
| RFC | Low | 0.86 | 0.88 | 0.87 | 0.80 | 0.85 | 084 | 0.85 | 0.78 | 0.86 | 0.90 | 0.88 | 0.79 |
| | Moderate | 0.70 | 0.76 | 0.73 | | 0.63 | 0.67 | 0.65 | | 0.49 | 0.42 | 0.45 | |
| | Severe | 0.93 | 0.57 | 0.63 | | 0.69 | 0.67 | 0.68 | | 0.55 | 0.48 | 0.51 | |
| LR | Low | 0.86 | 0.70 | 0.77 | 0.62 | 0.84 | 067 | 0.75 | 0.64 | 0.88 | 062 | 0.73 | 0.59 |
| | Moderate | 0.42 | 0.57 | 0.49 | | 0.31 | 0.49 | 0.38 | | 0.22 | 0.45 | 0.29 | |
| | Severe | 0.48 | 0.62 | 0.54 | | 0.54 | 0.65 | 0.59 | | 0.28 | 0.51 | 0.36 | |
| $B_s$ | Low | 0.65 | 0.65 | 0.65 | 0.48 | 0.61 | 0.62 | 0.61 | 0.45 | 0.75 | 0.74 | 0.75 | 0.59 |
| | Moderate | 0.20 | 0.20 | 0.20 | | 0.14 | 0.13 | 0.13 | | 0.07 | 0.07 | 0.07 | |
| | Severe | 0.19 | 0.19 | 0.19 | | 0.20 | 0.20 | 0.20 | | 0.12 | 0.13 | 0.13 | |

prec, rec, f1 and acc refer to precision, recall, f1-score and accuracy, respectively. * indicates the model with higher prediction ability for each team

**Table S1 - 2nd part.** Prediction goodness of Fatigue in daily approach

| Model | classes | Team 4 | | | | Team 5 | | | | Team 6 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **prec** | **rec** | **f1** | **acc** | **prec** | **rec** | **f1** | **acc** | **prec** | **rec** | **f1** | **acc** |
| DTC | Low | 0.86 | 0.75 | 0.80 | | 0.86 | 0.86 | 0.86 | | 0.81 | 0.86 | 0.83 | |
| | Moderate | 0.48 | 0.61 | 0.54 | 0.70 | 0.59 | 0.60 | 0.59 | 0.79 | 0.89 | 0.94 | 0.87 | 0.85 |
| | Severe | 0.13 | 0.57 | 0.49 | | 0.60 | 0.64 | 0.62 | | 0.91 | 0.77 | 0.83 | |
| XGB | Low | 0.85 | 0.88 | 0.87 | | 0.88 | 0.95 | 0.91 | | 0.95 | 0.98 | 0.97 | |
| | Moderate | 0.65 | 0.58 | 0.61 | 0.79* | 0.76 | 0.61 | 0.68 | 0.86* | 0.88 | 0.81 | 0.85 | 0.87* |
| | Severe | 0.63 | 0.58 | 0.60 | | 0.82 | 0.64 | 0.72 | | 0.92 | 0.85 | 0.88 | |
| RFC | Low | 0.86 | 0.86 | 0.86 | | 0.86 | 0.95 | 0.91 | | 0.87 | 0.87 | 0.87 | |
| | Moderate | 0.60 | 0.59 | 0.60 | 0.78 | 0.76 | 0.56 | 0.65 | 0.85 | 0.88 | 0.77 | 0.81 | 0.80 |
| | Severe | 0.59 | 0.60 | 0.59 | | 0.86 | 0.62 | 0.72 | | 0.89 | 0.62 | 0.73 | |
| LR | Low | 0.89 | 0.65 | 0.75 | | 0.78 | 0.94 | 0.85 | | 0.69 | 0.84 | 0.79 | |
| | Moderate | 0.38 | 0.57 | 0.46 | 0.64 | 0.34 | 0.07 | 0.12 | 0.75 | 0.39 | 0.15 | 0.22 | 0.66 |
| | Severe | 0.38 | 0.68 | 0.49 | | 0.59 | 0.38 | 0.46 | | 0.47 | 0.09 | 0.15 | |
| $B_s$ | Low | 0.72 | 0.71 | 0.72 | | 0.72 | 0.72 | 0.72 | | 0.65 | 0.65 | 0.65 | |
| | Moderate | 0.15 | 0.16 | 0.15 | 0.55 | 0.12 | 0.13 | 0.12 | 0.56 | 0.21 | 0.20 | 0.20 | 0.48 |
| | Severe | 0.09 | 0.09 | 0.09 | | 0.16 | 0.15 | 0.16 | | 0.15 | 0.16 | 0.16 | |

prec, rec, f1 and acc refer to precision, recall, f1-score and accuracy, respectively. * indicates the model with higher prediction ability for each team

**Table S2 - 1st part.** Prediction goodness of Fatigue in match related approach

| Model | classes | Team 1 | | | | Team 2 | | | | Team 3 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **prec** | **rec** | **f1** | **acc** | **Prec** | **rec** | **f1** | **Acc** | **prec** | **rec** | **f1** | **acc** |
| DTC | Low | 0.74 | 0.79 | 0.76 | 0.66 | 0.70 | 0.71 | 0.70 | 0.64 | 0.84 | 0.80 | 0.82 | 0.69 |
| | Moderate | 0.57 | 0.56 | 0.56 | | 0.63 | 0.61 | 0.62 | | 0.23 | 0.28 | 0.25 | |
| | Severe | 0.32 | 0.24 | 0.28 | | 0.51 | 0.49 | 0.50 | | 0.33 | 0.38 | 0.36 | |
| XGB | Low | 0.74 | 0.92 | 0.82 | 0.71* | 0.70 | 0.86 | 0.77 | 0.70* | 0.83 | 0.96 | 0.89 | 0.80* |
| | Moderate | 0.66 | 0.58 | 0.62 | | 0.75 | 0.56 | 0.64 | | 0.40 | 0.17 | 0.23 | |
| | Severe | 0.51 | 0.10 | 0.17 | | 0.64 | 0.40 | 0.49 | | 0.60 | 0.34 | 0.44 | |
| RFC | Low | 0.74 | 0.80 | 0.77 | 0.64 | 0.69 | 0.73 | 0.71 | 0.65 | 0.81 | 0.98 | 0.89 | 0.79 |
| | Moderate | 0.57 | 0.56 | 0.56 | | 0.64 | 0.59 | 0.61 | | 0.62 | 0.13 | 0.22 | |
| | Severe | 0.33 | 0.24 | 0.28 | | 0.52 | 0.49 | 0.50 | | 0.74 | 0.29 | 0.41 | |
| LR | Low | 0.66 | 0.99 | 0.79 | 0.46 | 0.62 | 0.93 | 0.74 | 0.62 | 0.80 | 0.97 | 0.88 | 0.68 |
| | Moderate | 0.67 | 0.09 | 0.17 | | 0.82 | 0.24 | 0.37 | | 0.69 | 0.10 | 0.16 | |
| | Severe | 0.00 | 0.00 | 0.00 | | 0.54 | 0.17 | 0.26 | | 0.55 | 0.30 | 0.39 | |
| $B_s$ | Low | 0.64 | 0.64 | 0.64 | 0.42 | 0.58 | 0.58 | 0.58 | 0.42 | 0.74 | 0.76 | 0.75 | 0.59 |
| | Moderate | 0.17 | 0.18 | 0.17 | | 0.17 | 0.17 | 0.17 | | 0.07 | 0.06 | 0.07 | |
| | Severe | 0.17 | 0.17 | 0.17 | | 0.24 | 0.24 | 0.24 | | 0.11 | 0.10 | 0.10 | |

prec, rec, f1 and acc refer to precision, recall, f1-score and accuracy, respectively. * indicates the model with higher prediction ability for each team

**Table S2 - 2nd part.** Prediction goodness of Fatigue in match related approach

| Model | classes | Team 4 | | | | Team 5 | | | | Team 6 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **prec** | **rec** | **f1** | **acc** | **prec** | **rec** | **f1** | **acc** | **prec** | **rec** | **f1** | **acc** |
| DTC | Low | 0.78 | 0.87 | 0.82 | 0.72 | 0.84 | 0.82 | 0.93 | 0.72 | 0.67 | 0.70 | 0.69 | 0.59 |
| | Moderate | 0.53 | 0.50 | 0.51 | | 0.25 | 0.32 | 0.28 | | 0.46 | 0.46 | 0.46 | |
| | Severe | 0.19 | 0.08 | 0.11 | | 0.46 | 0.44 | 0.45 | | 0.46 | 0.43 | 0.44 | |
| XGB | Low | 0.78 | 0.93 | 0.85 | 0.76* | 0.83 | 0.94 | 0.89 | 0.80* | 0.71 | 0.83 | 0.76 | 0.66* |
| | Moderate | 0.65 | 0.51 | 0.57 | | 0.32 | 0.24 | 0.28 | | 0.58 | 0.46 | 0.51 | |
| | Severe | 0.53 | 0.08 | 0.22 | | 0.66 | 0.49 | 0.56 | | 0.58 | 0.45 | 0.51 | |
| RFC | Low | 0.77 | 0.88 | 0.82 | 0.71 | 0.82 | 0.98 | 0.89 | 0.82 | 0.68 | 0.71 | 0.69 | 0.60 |
| | Moderate | 0.55 | 0.48 | 0.51 | | 0.88 | 0.05 | 0.09 | | 0.47 | 0.46 | 0.46 | |
| | Severe | 0.18 | 0.07 | 0.09 | | 0.76 | 0.42 | 0.54 | | 0.46 | 0.44 | 0.45 | |
| LR | Low | 0.75 | 0.96 | 0.84 | 0.70 | 0.79 | 0.98 | 0.88 | 0.58 | 0.64 | 0.88 | 0.74 | 0.56 |
| | Moderate | 0.59 | 0.28 | 0.39 | | 0.00 | 0.00 | 0.00 | | 0.48 | 0.32 | 0.38 | |
| | Severe | 0.00 | 0.00 | 0.00 | | 0.60 | 0.17 | 0.27 | | 0.38 | 0.16 | 0.22 | |
| $B_s$ | Low | 0.71 | 0.73 | 0.72 | 0.57 | 0.77 | 0.75 | 0.76 | 0.50 | 0.56 | 0.56 | 0.56 | 0.42 |
| | Moderate | 0.17 | 0.17 | 0.17 | | 0.10 | 0.11 | 0.10 | | 0.19 | 0.19 | 0.19 | |
| | Severe | 0.13 | 0.11 | 0.12 | | 0.10 | 0.10 | 0.10 | | 0.26 | 0.26 | 0.26 | |

prec, rec, f1 and acc refer to precision, recall, f1-score and accuracy, respectively. * indicates the model with higher prediction ability for each team

# CHAPTER 3: Correlates of subjective recovery in professional soccer players: a machine learning and mediation approach

This paper was presented as an oral at the WCSF 2023 – World Congress of Science in Football in Groningen (NL). 24-26 may and as a poster presentation (with mention) at XIII national SISMES congress in Milan, November 4-6, 2022

# Correlates of subjective recovery in professional soccer players: a machine learning and mediation approach

Carlo Simonelli[1,2], Alessio Rossi[3], Samuele M. Marcora[1]

[1] Department of Life Quality Studies (QuVi), University of Bologna, Bologna, Italy.

[2] Vero Volley Research Center, Monza, Italy.

[3] Independent researcher.

## Abstract

**Purpose** - Coaches often ask players to judge their recovery status (i.e. subjective recovery). The aim of this study was to explore the factors that correlate with subjective recovery in a group of professional soccer players.

**Methods -** 101 male players from four professional Italian soccer clubs competing in Serie C were recruited for this study. A complete season was recorded for each team for a total of 16,989 training sessions and matches. Every morning, the players rated fatigue, sleep quality, muscle soreness, stress and mood. At the same time, they formally judged their recovery using the Total Quality Recovery questionnaire (TQR). After each training session or match, the session Rating of Perceived Exertion (sRPE) was obtained and multiplied by duration to calculate the Training Load (TL). A framework of data analytics of time series was employed to detect the factors associated with subjective recovery.

**Results** - Machine learning and mediation analyses suggest that TQR scores are primarily associated with ratings of fatigue and muscle soreness at the time of such judgments, and that these factors mediate most of the relationship between training load of the previous day and subjective recovery.

**Conclusion -** These findings suggest that, in order to maximize subjective recovery of professional soccer players, coaches and other relevant support staff should implement strategies to minimize fatigue and muscle soreness. From a programming perspective, reducing the training load of the previous day seems to be the most effective strategy. Future experimental studies are required to confirm these correlational findings.

**Keywords:** TQR; fatigue; football; muscle soreness; training load; wellness.

## Introduction

Recovery has been defined as a multifaceted restorative process relative to time [1]. It is widely known that an optimal balance between training load and recovery is required to adapt to training and perform optimally [2]. Poor recovery between training sessions and matches is also associated with increased incidence of injury, illness, and overtraining [3,4].

Because of its multifaceted nature, recovery has been quantified in many ways including performance tests such as CMJ [5] and biochemical markers like creatine kinase (CK) concentration [6]. A more practical and commonly used way to asses recovery is to ask athletes to judge their recovery status (i..e. subjective recovery), either informally in daily conversations or formally using specific scales or questionnaires like the Total Quality of Recovery (TQR) scale [7]. The validity of subjective measures of recovery has been supported by their associations with performance decrements, injury rates, and biochemical markers of recovery [1,2,8–15].

Because of its relevance to athletes' performance and health, it is important to explore the determinants and correlates of subjective recovery. Short-term intensified training studies have provided experimental evidence that training load and time to recover are key determinants of subjective recovery [16]. Small correlational studies also suggest that subjective judgments of recovery status may be influenced by feelings of fatigue and delayed onset muscle soreness (DOMS) [5,17]. Moreover, it is believed that factors such as stress, mood and sleep quality affect the recovery process in athletes[18–21].

The main aim of the present study was to further explore potential correlates of subjective recovery (training load, fatigue, sleep quality, muscle soreness, stress and mood) using machine learning analysis on a large data set collected in the field during an entire professional soccer season. Moreover, we explored whether the hypothesised relationship between training load and subjective recovery is mediated by fatigue, sleep quality, muscle soreness, stress and mood.

# Methods

## Subjects

101 players (23.5 ± 4.0 yrs, 73.8 ± 7.4 kg, 180 ± 5 cm) from four professional Italian soccer clubs competing in the third Italian division (Serie C) were recruited from this study. The data recording was conducted on a complete championship season for each team, from summer camp on. A total of 16,989 training or match sessions (113 ± 52 per player) were recorded. The data recording was conducted between July 16th 2020 and May 11th 2023. After a detailed description of the procedure and possible risks, players voluntarily decided to participate by signing an informed consent. The project was conducted according to the Declaration of Helsinki and was approved by the Bioethics Committee of the University of Bologna.

## Data recording

Every day, usually in private and at a consistent time in the morning, in any case before each training session or match, the players filled six scales. First, the players provided subjective ratings of fatigue, sleep quality, muscle soreness, stress and mood using the Wellness Questionnaire (WQ) [22]. This questionnaire consists of five single items with a 5-point Likert scale where 1 and 5 indicate the highest and lowest values of wellness for each item, respectively [23,24]. The sixth scale the players had to fill in was the Total Quality Recovery (TQR) scale. This instrument evaluates the players' recovery status [7] by a self-reported single-item scale ranging between 6 and 20. Values of about 6 refer to no recovery at all, while 20 means that the athlete fully recovered. Finally, about 30 minutes after the end of each training session or match, the players rated how hard the training session or match was (sRPE) by using the 10-point scale (CR-10 Borg' scale), where 0 refers to resting state and 10 to maximal effort [25]. The Training Load (TL) for each training session or match was computed as the product between sRPE and the duration in minutes (time) and, in each analysis, refers to the load of the training session or match performed the day before the subjective ratings of recovery, fatigue, sleep quality, muscle soreness, stress and mood were provided. However, we decided to include sRPE and time separately in the machine learning approach to differentiate between the intensity and duration of the training session/match. Training load can be simply described as a label attributed to a higher-order construct overarching other interrelated sub-dimensions such as sRPE and time specifically [26].

## Data preprocessing

Data preprocessing permits to aggregate the data time series to create indexes that provide more details about players' history. In particular, two types of aggregations were computed for each independent feature (i.e., fatigue, sleep quality, muscle soreness, stress, mood ratings of the current day, sRPE and time of the previous day): i) exponential weighted moving average of past 7 days (Acute); ii) exponential weighted moving average of past 28 days (Chronic). The weight of both Acute and Chronic aggregation methods was computed with a specified decay in terms of span (2/(SPAN+2)) [27].

The data preprocessing was made by using Python 3.9 language programming.

## Machine learning approach

Extreme Gradient Boosting Regressor (XGBR) model was fitted on 21 independent features (i.e., fatigue, sleep quality, muscle soreness, stress, mood ratings of the current day, sRPE and time of the previous day, and their acute, and chronic aggregations) with the aim of predicting the players' recovery status (TQR score, i.e. recovery perception). The XGBR model was trained on 70% of the total observations (train set) where the best subset of features was obtained by a Recursive Feature Elimination with 10 Cross-Validation (RFECV) based on the XGBR model. In the train set were also fitted the best hyperparameters of XGBR by a grid search approach. The grid search parameters tested are: learning rate = 0.01, 0.03, 0.05, 0.07, 0.09; max depth = 5, 10, 15, 20, None; colsample by tree = 0.1, 0.3, 0.5, 0.7, 0.9; number of estimators = 100, 500, 1000.

The predictive performance was assessed on the test set (30% of observations not included in the train set) in order to avoid any overfitting problem. Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Pearson's correlation coefficient (r) were used in order to evaluate the predicting performance of the XGBR model. Shapley Additive exPlanations (SHAP) method was used to explain the decision-making process created by XGBR to predict TQR. To explain the output of the machine learning models, SHAP uses a game theory approach. By giving a pre-trained regressor, SHAP explainer creates a value for each independent feature that represents how much they influenced the regressor's decision-making process by providing a global and a local explanation of the machine learning model.

The machine learning analysis was conducted by using Python 3.9 language programming.

## Mediation analysis

A correlation analysis for repeated measures based on an autoregressive model was conducted on the entire dataset between the TQR ratings and the following variables:

TL, $TL_{acute}$ and $TL_{chronic}$;

Fatigue, sleep quality, muscle soreness, stress and mood rated at the same time as the TQR ratings.

Secondly, where the assumptions of statistically significant correlation were met, a mediation analysis was conducted with the aim of detecting if the hypothesised relationship between training load variables and TQR was mediated by one or more of the five WQ variables listed above. In order to take in consideration the assumption of related pairs, a repeated measures mediation analysis was performed. In particular, Marginal Regression Model using Generalised Estimating Equations based on Autoregressive covariance structure was used to meet both time series and repeated measures assumptions. Specifically, mediation analysis was conducted for TL and its acute and chronic aggregations if the assumption of statistically significant correlations were met (i.e., the mediation model was built on variables that are both correlated with dependent and independent features). The Root Mean Squared Error (RMSE) was used as indicators of model goodness. Moreover, in the context of GEE, where observations may be correlated within clusters or subjects, in order to evaluate the strength of the relationship it is more appropriate to use a correlation-like measure that accounts for the correlation structure. One such measure is the Intraclass Correlation Coefficient (ICC), which quantifies the proportion of total variance that is due to between-cluster (or between-subject) variability. The ICC ranges between 0 and 1. Higher values of ICC indicate a stronger correlation within clusters (higher inter-cluster variability) relative to the total variability, while lower values indicate a weaker correlation within clusters (lower inter-cluster variability) relative to the total variability. An ICC close to 0 suggests that observations within clusters are almost independent, while an ICC close to 1 suggests high correlation between observations within clusters.

The mediation analysis was conducted by using Python 3.9 language programming.

# Results

## Machine learning approach

Thirteen out of 21 independent features were selected by the RFECV approach. Table 2 shows the features selected and their importance to predict TQR. Fatigue is the most important feature (importance = 37.20%) followed by muscle soreness (importance = 12.00%), $mood_{chronic}$ (importance = 7.93%) and $fatigue_{chronic}$ (importance = 7.42%). The other daily features (i.e. sRPE and sleep) show a cumulative importance of about 5.64% resulting in a total importance of about 54.84%. Otherwise, the weekly aggregations (i.e. $fatigue_{acute}$ and $sleep_{acute}$) show a cumulative importance of 6.26%, while the chronic aggregations (i.e. $mood_{chronic}$, $fatigue_{chronic}$, $stress_{chronic}$, muscle $soreness_{chronic}$, $sleep_{chronic}$, $sRPE_{chronic}$ and $time_{chronic}$) describe an overall importance of 38.91%. The best XGBR hyperparameters selected with the grid search approach were: learning rate = 0.03; max_depth = 10; callsample by tree = 0.7; number of estimators: 1000. Figure 1 shows a strong correlation and a small error between TQR observed and predicted (r = 0.81, MSE = 1.27, RMSE = 1.13). Figure 2 shows the global explanation (indications of the relationship between the value of a feature and the impact on the prediction) of the XGBR decision-making process.

## Mediation analysis

Descriptive statistics of all the features included in this study was provided in Table 1. Significant relationships with TQR were detected for both TL and WQ variables (Table 2). In order to detect if the relationship between TL and TQR is mediated by the WQ variables, a mediation analysis was conducted. The analysis was performed for TL, $TL_{acute}$, $TL_{chronic}$. A statistically significant relationship of TL on TQR was detected (beta coefficient = 0.002 [0.001, 0.002]; p-value < 0.001; RMSE = 1.96; ICC = 0.22). The mediation analysis (Figure 3) shows a statistically significant direct effect (beta coefficient = 0.001 [0.001, -0.002]; p-value <0.001; RMSE = 1.86; ICC = 0.28) of TL on TQR. This relation is mediated by muscle soreness (beta coefficient = 0.0006 [0.0003, 0.0008]; p-value < 0.001; RMSE = 1.93; ICC = 0.24) and fatigue (beta coefficient = 0.0005 [0.0002, 0.0007]; p-value < 0.001; RMSE = 1.87; ICC = 0.27). The other WQ variables (sleep quality, stress, mood) did not qualify as mediators of the relationship between TL and TQR.

The relationships between $TL_{acute}$ (beta coefficient = 0.0003 [-0.001, 0.002]; p-value = 0.70; RMSE = 1.94; ICC = 0.21) and $TL_{chronic}$ (beta coefficient = 0.0007 [-0.003, 0.004]; p-value = 0.69; RMSE

= 1.94; ICC = 0.21) with TQR do not appear significant and, therefore, no further mediation analysis was required because the assumption for conducting the mediation analysis was not met.

**Table 1.** Descriptive statistics.

| Features | mean | SD | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|
| sRPE | 4.14 | 1.91 | 0.5 | 3 | 4 | 5 | 10 |
| time | 73.5 | 23.3 | 31 | 60 | 73 | 87 | 170 |
| TL | 320 | 208 | 18 | 180 | 280 | 413 | 1360 |
| fatigue | 2 | 0.75 | 1 | 2 | 2 | 2 | 5 |
| sleep | 1.91 | 0.78 | 1 | 1 | 2 | 2 | 5 |
| muscle soreness | 1.82 | 0.75 | 1 | 1 | 2 | 2 | 5 |
| stress | 2.17 | 0.77 | 1 | 2 | 2 | 3 | 5 |
| mood | 2.05 | 0.81 | 1 | 1 | 2 | 3 | 5 |
| TQR | 15.81 | 1.93 | 6 | 15 | 16 | 17 | 20 |

**Table 2.** Feature importance (SHAP mean) of the best subset of the features selected by RFECV for XGBR.

| Feature | Importance (%) |
|---|---|
| fatigue | 37.20 |
| muscle soreness | 12.00 |
| $mood_{chronic}$ | 7.93 |
| $fatigue_{chronic}$ | 7.42 |
| $stress_{chronic}$ | 5.37 |
| muscle $soreness_{chronic}$ | 5.25 |
| $sleep_{chronic}$ | 4.47 |
| $sRPE_{chronic}$ | 4.43 |
| $time_{chronic}$ | 4.04 |
| sRPE | 3.79 |
| $fatigue_{acute}$ | 3.14 |
| $sleep_{acute}$ | 3.12 |
| Sleep | 1.85 |

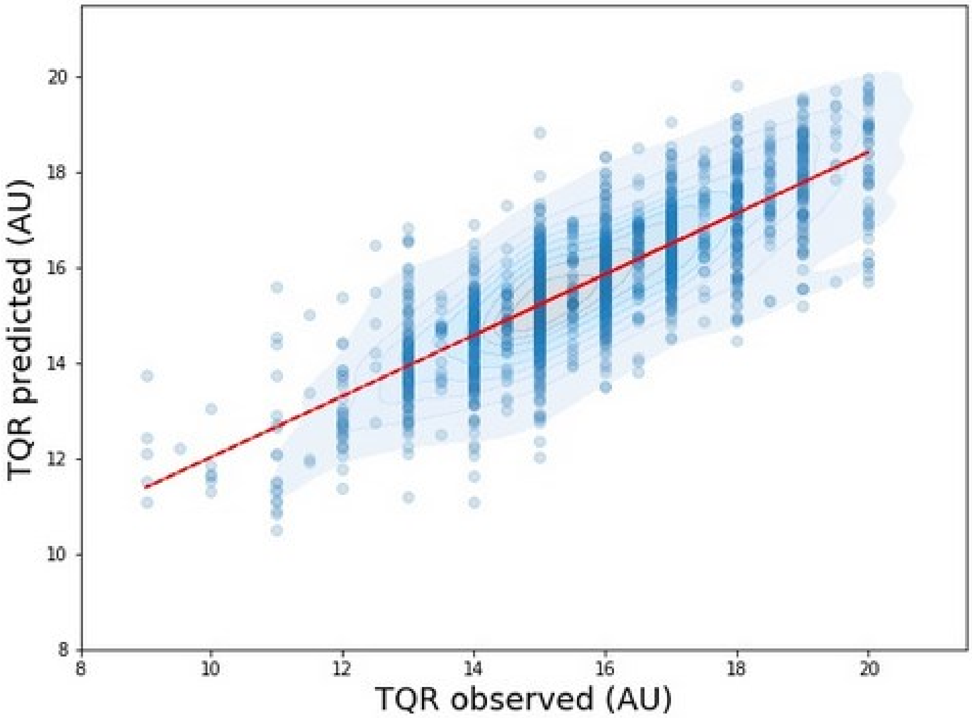



**Figure 1.** Scatter density plot of the TQR observed and predicted.

**Table 3.** Autoregressive model for repeated measures correlation analysis.

| model parameters | coefficient | SD | z-score | p-value |
|---|---|---|---|---|
| Intercept | 18.51 | 0.41 | 45.35 | <0.001 |
| fatigue | -1.35 | 0.15 | -9.11 | <0.001 |
| Intercept | 16.97 | 0.33 | 50.86 | <0.001 |
| Sleep | -0.61 | 0.12 | -5.14 | <0.001 |
| Intercept | 17.81 | 0.35 | 50.30 | <0.001 |
| muscle soreness | -1.10 | 0.14 | -7.77 | <0.001 |
| Intercept | 16.66 | 0.39 | 43.19 | <0.001 |
| stress | -0.39 | 0.14 | -2.72 | <0.01 |
| Intercept | 16.49 | 0.41 | 40.75 | <0.001 |
| mood | -0.33 | 0.16 | -2.05 | <0.05 |
| Intercept | 15.27 | 0.21 | 73.90 | <0.001 |
| TL | 0.00 | 0.00 | 6.13 | <0.001 |

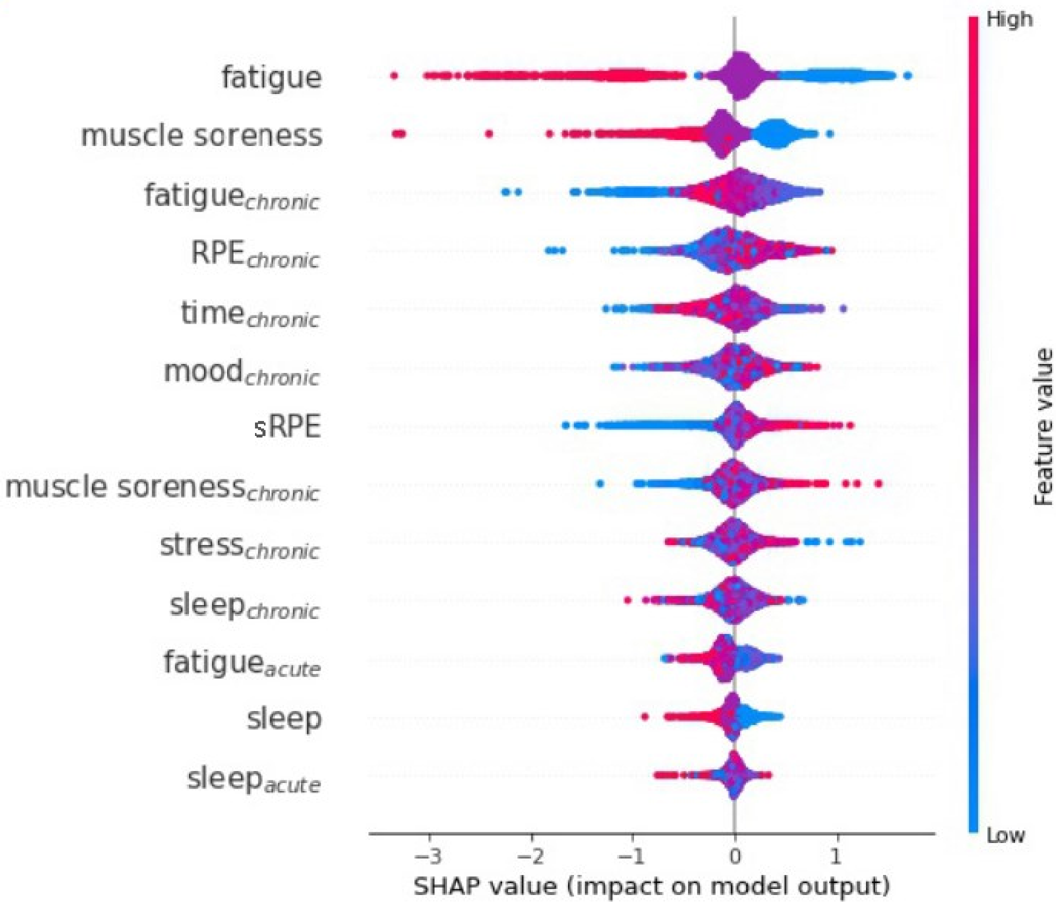



**Figure 2.** Global explanation of the fitted XGBR model based on SHAP analysis. Influence of independent features on subjective recovery (TQR). The predictors are ranked in descending order in accordance with the features importance. Each point on this plot is a Shapley value for a feature and an instance. The position on the y-axis is determined by the feature and on the x-axis by the Shapley value. The colour represents the value of the feature from low to high.

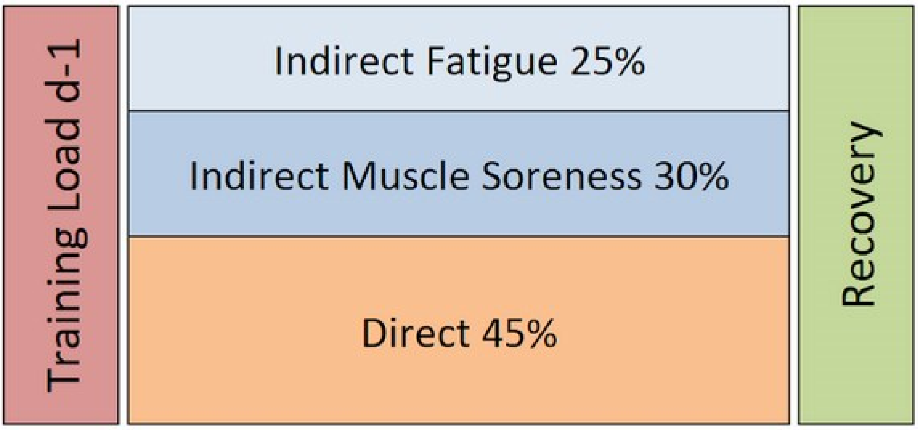



**Figure 3.** Mediation analysis.

## Discussion

The correlative evidence provided by this study suggests that the subjective judgements of recovery status by soccer players are primarily influenced by the feelings of fatigue and muscle soreness at the time such judgements are made. These feelings also seem to mediate most of the relationship between training load of the previous day and subjective recovery. Interestingly and confirmed by mediation analysis, training load of the previous week and training load of the previous four weeks were not found to be associated with subjective recovery because it seems to be strongly mediated by the perception of fatigue.

An important result of this study is that it is possible to predict the players' subjective recovery based on TL components (sRPE and time) and ratings of fatigue, sleep quality, muscle soreness, stress and mood. As a matter of fact, the machine learning approach shows a very low error (MSE = 1.27 and RMSE = 1.13) suggesting that the model is very accurate in estimating the TQR scores (Figure 1). Although other factors may also play a role, the high accuracy of our model suggests that the combination of training load, fatigue, sleep quality, muscle soreness, stress and mood can explain a lot about the subjective recovery of professional soccer players.

In this study, a machine learning model was employed to gain insights into the impact of subjective ratings (i.e., fatigue, sleep quality, muscle soreness, stress and mood) on the recovery construct, rather than the development of a predictive model. The evaluation of the model's quality was essential to confirm the presence of a cause-effect relationship in the underlying dynamics. The machine learning approach (Table 3) suggests that players' judgement of their recovery is mainly affected by the fatigue and muscle soreness experienced when providing the TQR rating (49.20%). Moreover, except for sleep that shows a minimal importance (1.85%), the other WQ variables (mood and stress) do not significantly contribute to the prediction of subjective recovery. Additionally, even if the chronic aggregation (previous four weeks) of each WQ variable and TL components shows a low importance to predict subjective recovery, the sum of these chronic parameters reaches an overall importance of 38.91%. This finding suggests that chronic stress, mood and sleep quality, together with the training history of the players, affect their subjective recovery. Conversely, the acute aggregations (previous week) of fatigue and sleep contribute only marginally to the prediction of subjective recovery, and the other acute aggregations (muscle soreness, stress, mood and TL components) do not contribute at all.

Interestingly, the machine learning approach only shows a marginal importance for sRPE (3.79%) and time of the previous day did not contribute significantly to the prediction of subjective recovery. There are two possible explanations for such findings. The first is that the training load of the previous day has only a small effect on subjective recovery. The second, and more likely, explanation is that the effect of training load of the previous day is mediated to a large extent by its effects on fatigue and muscle soreness. To test this hypothesis and explore further the relationships between training load of the previous week and previous four weeks.

The first step of this mediation approach showed that neither $TL_{acute}$ and or $TL_{chronic}$ correlates with subjective recovery. Only the training load of the previous day is significantly correlated with subjective recovery. Further analysis of the factors mediating such correlation revealed that fatigue (25%) and muscle soreness (30%) experienced when providing the TQR scores mediated most of the relationship between training load of the previous day and subjective recovery. Our findings based on mediation analysis of a large set of ecological data are corroborated by smaller intervention studies. For example, Selmi et al. [16] found that intensified training in a group of 15 professional soccer players increases subjective feelings of fatigue and muscle soreness, and reduces TQR scores. Furthermore, these authors found that TL correlates positively with subjective feelings of fatigue and muscle soreness, and negatively with TQR scores. Interestingly, as in our study, Selmi et al. [16] found no correlations of TL with subjective ratings of sleep and stress.

Although fatigue and muscle soreness experienced when providing the TQR scores mediated most of the relationship between training load of the previous day and subjective recovery (indirect effect), there was also an important direct effect of TL. One possible explanation for this finding is that there are other important mediators that were not measured in the present study. However, a true direct effect of the training load of the previous day on the players' judgement of their recovery status is also plausible. For example, strong feelings of fatigue and muscle soreness after a day of rest may be interpreted more negatively in terms of recovery (i.e. poor recovery) compared to the same feelings experienced the day after an intense training session (i.e. expected recovery).

## Limitations of the study

Although a very large set of ecological data from professional soccer players was collected and a complex statistical analysis was performed, the main limitation of this study is its correlative nature. Therefore, causality cannot be established. Another important limitation is the use of the WQ to measure fatigue, sleep quality, muscle soreness, stress and mood. Although widely used in both research and practice, the five single-item scales of the WQ have not been validated against criterion measures of the same constructs. Furthermore, our measures of TL refers only to training duration and sRPE which is a subjective measure of training intensity. Other training characteristics that we did not measure may affect muscle soreness and, thus, subjective recovery. These include the prevalence of eccentric vs concentric muscle contractions or the accustomization of the player to a specific training modality.

Finally, the current algorithm cannot be directly extended to all soccer teams; the model provided needs to be “trained” specifically for each team, due to the fact that every team has different characteristics such as, for example, different players' physical fitness and training scheduled programmes.

## Conclusion and practical application

This study suggests that the players’ judgement of their recovery relies primarily on the feelings of fatigue and muscle soreness experienced when making such judgement. From a programming perspective, our study suggests that reducing the internal load of the training session performed the previous day is the most effective strategy if the goal is to maximise subjective recovery. Indeed, this practice is already commonly used by coaches before an important match [28]. Reducing the internal training load of the whole week before a match does not seem so important in terms of subjective recovery.

In addition to internal training load manipulations, feelings of fatigue and muscle soreness may also be affected by factors other than training. For example, cognitive load can increase feelings of fatigue in soccer players [29]. Although the evidence for efficacy is not strong, muscle soreness may also be improved by physiotherapic, pharmacological and nutritional interventions[30]. Therefore, employing specific interventions to reduce mental fatigue and muscle soreness may also improve subjective recovery of soccer players. Further experimental studies should be employed to verify these hypotheses and establish more strongly a link between fatigue, muscle soreness, subjective recovery and, ultimately, soccer performance.

## References

1. Kellmann M, Bertollo M, Bosquet L, et al. Recovery and Performance in Sport: Consensus Statement. *International Journal of Sports Physiology and Performance*. 13(2):240-245. doi:10.1123/ijspp.2017-0759

2. Nedelec M, McCall A, Carling C, Legall F, Berthoin S, Dupont G. The influence of soccer playing actions on the recovery kinetics after a soccer match. *J Strength Cond Res*. 2014;28(6):1517-1523. doi:10.1519/JSC.0000000000000293

3. Gabbett TJ, Whyte DG, Hartwig TB, Wescombe H, Naughton GA. The Relationship Between Workloads, Physical Performance, Injury and Illness in Adolescent Male Football Players. *Sports Med*. 2014;44(7):989-1003. doi:10.1007/s40279-014-0179-5

4. Prevention, Diagnosis, and Treatment of the Overtraining Syndrome: Joint Consensus Statement of the European College of Sport Science and the American College of Sports Medicine. *Medicine & Science in Sports & Exercise*. 2013;45(1):186-205. doi:10.1249/MSS.0b013e318279a10a

5. Freitas VH, Nakamura FY, Miloski B, Samulski D. Sensitivity of Physiological and Psychological Markers to Training Load Intensification in Volleyball Players. *J Sports Sci Med*. 2014;13(3):571-579. Published 2014 Sep 1.

6. Lazarim FL, Antunes-Neto JMF, da Silva FOC, et al. The upper values of plasma creatine kinase of professional soccer players during the Brazilian National Championship. *J Sci Med Sport*. 2009;12(1):85-90. doi:10.1016/j.jsams.2007.10.004

7. Kenttä G, Hassmén P. Overtraining and Recovery: A Conceptual Model. *Sports Med*. 1998;26(1):1-16. doi:10.2165/00007256-199826010-00001

8. Bessa AL, Oliveira VN, Agostini GG, et al. Exercise Intensity and Recovery: Biomarkers of Injury, Inflammation, and Oxidative Stress. *J Strength Cond Res*. 2016;30(2):311-319. doi:10.1519/JSC.0b013e31828f1ee9

9. Brink MS, Visscher C, Arends S, Zwerver H, Post W, Lemmink KA. Monitoring stress and recovery: New insights for the prevention of injuries and illnesses in elite youth soccer players. *British Journal of Sports Medicine*. 2010;44(11):809-815. doi:10.1136/bjsm.2009.069476

10. Nédélec M, McCall A, Carling C, Legall F, Berthoin S, Dupont G. Recovery in soccer: part I - post-match fatigue and time course of recovery. *Sports Med*. 2012;42(12):997-1015. doi:10.2165/11635270-000000000-00000

11. Nédélec M, McCall A, Carling C, Legall F, Berthoin S, Dupont G. Recovery in soccer : part ii-recovery strategies. *Sports Med*. 2013;43(1):9-22. doi:10.1007/s40279-012-0002-0

12. Nédélec M, Halson S, Abaidia AE, Ahmaidi S, Dupont G. Stress, Sleep and Recovery in Elite Soccer: A Critical Review of the Literature. *Sports Med*. 2015;45(10):1387-1400. doi:10.1007/s40279-015-0358-z

13. Andersson HM, Raastad T, Nilsson J, Paulsen G, Garthe I, Kadi F. Neuromuscular fatigue and recovery in elite female soccer : effects of active recovery. *Medicine & Science in Sports & Exercise*. 2008;40(2):372-380.

14. Rampinini E, Bosio A, Ferraresi I, Petruolo A, Morelli A, Sassi A. Match-Related Fatigue in Soccer Players. *Medicine and science in sports and exercise*. 2011;43:2161-2170. doi:10.1249/MSS.0b013e31821e9c5c

15. Hausswirth C, Mujika I, *Recovery for Performance in Sport*.Champaign, IL: Human Kinetics; 2013.

16. Selmi O, Ouergui I, Castellano J, Levitt D, Bouassida A. Effect of an intensified training period on well-being indices, recovery and psychological aspects in professional soccer players. *European Review of Applied Psychology*. 2020;70(6):100603. doi:10.1016/j.erap.2020.100603

17. Howle K, Waterson A, Duffield R. Recovery profiles following single and multiple matches per week in professional football. *European Journal of Sport Science*. 2019;19(10):1303-1311. doi:10.1080/17461391.2019.1601260

18. Wiese-Bjornstal DM. Psychology and socioculture affect injury risk, response, and recovery in high-intensity athletes: a consensus statement. *Scandinavian Journal of Medicine & Science in Sports*. 2010;20(s2):103-111. doi:10.1111/j.1600-0838.2010.01195.x

19. Heidari J, Beckmann J, Bertollo M, et al. Multidimensional Monitoring of Recovery Status and Implications for Performance. *International Journal of Sports Physiology and Performance*. 2019;14(1):2-8. doi:10.1123/ijspp.2017-0669

20. Vitale KC, Owens R, Hopkins SR, Malhotra A. Sleep Hygiene for Optimizing Recovery in Athletes: Review and Recommendations. *Int J Sports Med*. 2019;40(08):535-543. doi:10.1055/a-0905-3103

21. Samuels C. Sleep, recovery, and performance: the new frontier in high-performance athletics. *Neurol Clin*. 2008;26(1):169-180; ix-x. doi:10.1016/j.ncl.2007.11.012

22. Perri E, Simonelli C, Rossi A, Trecroci A, Alberti G, Iaia FM. Relationship Between Wellness Index and Internal Training Load in Soccer: Application of a Machine Learning Model. *Int J Sports Physiol Perform*. 2021;16(5):695-703. doi:10.1123/ijspp.2020-0093

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

24. McLean BD, Coutts AJ, Kelly V, McGuigan MR, Cormack SJ. Neuromuscular, endocrine, and perceptual fatigue responses during different length between-match microcycles in professional rugby league players. *Int J Sports Physiol Perform*. 2010;5(3):367-383. doi:10.1123/ijspp.5.3.367

25. Borg G. *Borg's Perceived Exertion and Pain Scales*. Human Kinetics; 1998:viii, 104.

26. Staunton CA, Abt G, Weaving D, Wundersitz DWT. Misuse of the term 'load' in sport and exercise science. *Journal of Science and Medicine in Sport*. 2022;25(5):439-444. doi:10.1016/j.jsams.2021.08.013

27. Rossi A, Pappalardo L, Cintia P. A Narrative Review for a Machine Learning Application in Sports: An Example Based on Injury Forecasting in Soccer. *Sports*. 2022;10(1):5. doi:10.3390/sports10010005

28. Impellizzeri FM, Rampinini E, Coutts AJ, Sassi A, Marcora SM. Use of RPE-based training load in soccer. *Med Sci Sports Exerc*. 2004;36(6):1042-1047. doi:10.1249/01.mss.0000128199.23901.2f

29. Smith MR, Coutts AJ, Merlini M, Deprez D, Lenoir M, Marcora SM. Mental Fatigue Impairs Soccer-Specific Physical and Technical Performance. *Med Sci Sports Exerc*. 2016;48(2):267-276. doi:10.1249/MSS.0000000000000762

30. Heiss R, Lutter C, Freiwald J, et al. Advances in Delayed-Onset Muscle Soreness (DOMS) - Part II: Treatment and Prevention. *Sportverletz Sportschaden*. 2019;33(1):21-29. doi:10.1055/a-0810-3516

# CHAPTER 4: Validation of the single-item scales of the Wellness Questionnaire in soccer players

A part of this paper was presented as a poster at the WCSF 2023 – World Congress of Science in Football in Groningen (NL). 24-26 may 2023.

# Validation of the single-item scales of the Wellness Questionnaire in soccer players

Carlo Simonelli[1,2], Alessio Rossi[3], Alessandro Quartiroli[4,5], Franco M. Impellizzeri[6], Samuele M. Marcora[1]

[1] Department of Life Quality Studies (QuVi), University of Bologna, Bologna, Italy.

[2] Vero Volley Research Center, Monza, Italy.

[3] Independent researcher.

[4] University of Wisconsin - La Crosse, Wisconsin, USA.

[5] University of Portsmouth, Portsmouth, UK.

[6] Faculty of Health, School of Sport, Exercise and Rehabilitation, University of Technology Sydney, Sydney, NSW, Australia.

## Abstract

**Background/aim:** Athlete self-reported outcome measures are widely used both by practitioners and researchers to monitor team sport athletes. Whilst there are many measurement tools available, some of the most commonly used are the so-called Wellness (single-item) scales. However, these scales have yet to be validated when used as single-item. Therefore, the aim of this study was to test the construct validity of these scales in soccer players.

**Methods:** A sample of 186 italian soccer players volunteered to take part in this cross-sectional web-based study to investigate the construct validity of 5 single-item scales investigating Fatigue, Sleep Quality, Muscle Soreness, Stress and Mood. Moreover, 106 players completed the survey for an entire competitive season and 5269 pairs of training resulting in the same conditions was used to investigate reliability. Data was collected and managed between July 2021 and May 2023 using Qualtrics. The survey was advertised through social media and industry contacts. The survey was voluntary, with no monetary incentives.

**Results:** Almost all of the Wellness single-items revealed no clear evidence of construct validity and do not support the use of these scales in Italian soccer population. No statistical difference was detected between the two similar training sessions ($p>0.05$) for each item, revealing a very good reliability.

**Conclusions:** These findings question the so popular use of similar items in absence of consistent evidence of their validity.

## Introduction

Athlete-reported outcome measures are questionnaires with both mono- and multi- dimensional structures commonly used by sport practitioners and researchers to assess, evaluate, and monitor a variety of athletes' performance and performance related outcomes[1].

Among these instruments is not uncommon to find the so-called Wellness Questionnaires. A survey of high-performance sport revealed that 84% of responders used self-report questionnaires, 80% of which were customised designs including 4 or more (up to 12) single-items such as fatigue, sleep quality, muscle soreness, stress and mood[2]. The relationship between training or match load and the subsequent perceptions or judgements are typically explored by these customised psychometric tools[3]. The included items are used both as a total wellness score and as a single-item. With regards to the overall score, the wellness construct, as a complex interaction of various mind and body perceptions, is a very challenging task to be clearly assessed[4]. A lack of consistency was underlined by many researchers in the definition of wellness[5–9]. The absence of clarity and consistency has led to the proliferation of numerous wellness assessments, with the vast majority of them being tailored compositions incorporating a diverse range of elements[10]. In fact, there's no agreement upon which elements comprise wellness[7–9]). This absence of clarity leads to many issues concerning the wellness construct. An issue with wellness measures can be found everyday both by practitioners and researchers is that the same wellness score could represent very different conditions. For example, high physical stress or high psychological stress could be represented by the same total wellness score (e.g., 13 could be based on a double 5 in fatigue and muscle soreness and a triple 1 in the other items or on a double 5 in stress and mood), describing two totally different scenarios. For this reason, many practitioners and researchers tend to use it as single-item scales instead of an aggregate. Actually, various academic disciplines have employed single-item measures across diverse research domains[11–14]. Practitioners lean toward using single-item questionnaires due to their practical usefulness for daily monitoring of athletes, requiring minimal time and effort to be completed, especially when seamlessly integrated into daily surveys[15,16]. Furthermore, these questionnaires exhibit strong face validity and minimal criterion contamination, offering respondents immediate clarity compared to multi-item counterparts. Although, if with a superficial level these single-items may have some face validity, this is not sufficient to establish the validity of an instrument.[17]. Furthermore, the absence of item redundancy or repetitive similar elements minimizes participant boredom, fatigue, and frustration. While these advantages lead sport practitioners to prefer single-item instruments, researchers seem to continue to rely on multi-item assessment instruments as, according to the Classical Testing Theory (CTT)[18], item redundancy is a strength[15]. However, while single-item scales are believed to lack content validity and struggle to assess the facets of multi-dimensional constructs comprehensively, multiple-item scales do not necessarily outperform single-item[16,19].

Single-item questionnaires are widely used in sport practice with a clear utility, their psychometric strength is often not clearly developed, as many of them have not been properly validated or validated at all, reducing their ability to offer accurate, reliable, and useful information about the athletes taking them[1].

Given the widespread utilization of these wellness questionnaires by practitioners to gauge athletes' subjective general well-being, alongside its already proven versatility and utility it seems necessary to explore and confirm the psychometric properties of all the single-items, to ensure the reliability and validity of this instrument. In fact, according to CTT an instrument must be reliable, valid and useful[20].

The aim of the study was to test the construct validity of the 5 single-item scales (Fatigue, Sleep Quality, Muscle Soreness, Stress and Mood) comprising the Wellness Questionnaire. There is no defined standard, in fact, over time, it has been refined and adapted based on practical experience in the field, in which wellness questionnaires are commonly used also as single-items. We decided to developed a rationale on the questionnaires developed by Hooper[21], McLean[22] and their modifications by Perri[23], Gastin[24], Lathlean[25], and Crowcroft[26]. for fatigue, muscle soreness and sleep quality, and from for stress and mood. Over time, it has been refined and adapted based on practical experience in the field, in which wellness questionnaires are commonly used also as single-items. For this study we adopted the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN)[27–29]. The primary objectives encompass examining construct validity (Study A) — convergence through hypothesis testing — and assessing the reliability (Study B) of its individual components (Fatigue, Sleep Quality, Muscle Soreness, Stress, and Mood) in the Italian version in professional soccer players.

# Study A - Construct validity

## Methods

### Study Design and Recruitment of Participants

Italian soccer players (mainly national level) Swann[30] volunteered to take part in this cross-sectional web-based study. The survey was advertised primarily through industry contacts. The survey was voluntary, with no monetary incentives.

After the ethical approval by the University, we collected between January 2023 and May 2023 using Qualtrics (2023 Qualtrics; Provo, UT). In the first page of the survey, participants were provided with an information page describing length of survey (< 8 min to complete), purpose of the study and data security. Exclusion criteria were < 18 years and not currently competing in soccer.

Before being able to move forward with the survey, participants had to agree to the informed consent. While those who agreed were invited to the demographic page of the survey, those who did not agree to participate were thanked for their consideration and sent directly to the end of the survey.

After the demographic information the questionnaire was created as follows: 5 single-item scales of the Wellness Questionnaire, Italian version of the stress subscale of the Italian Depression Anxiety Stress Scales Short Version (DASS-21), Italian version of the Brunel Mood Scales (ITAMS), Groningen Sleep Quality Scale, Numerical Pain Rating Scale.

Groningen Sleep Quality Scale and Numerical Pain Rating scale were back translated in italian following a translation-back translation methodology[31,32] by a group of bilingual (i.e., Italian, English) experts in sport and clinical psychology. All of the questionaires wiil be described below.

The COSMIN guidelines and reporting of results of e-surveys were followed[33].

For the construct validity investigation, a total of 186 participants completed the survey in its entirety. All partial responses were excluded from the analyses.

96.7% of the total eligible participants were males (n=180). The mean age was 27.2 (±9.2) years (range 18–60). The main level of sport was national (68.8%, n=128), followed by regional (15.5%, n=29). Almost the whole population of the respondents (99.6%) competed in their sport for 6 years or more.

The 95.2% of all the participants trained and played their sport for more than 6 training a week (match included), and the 80.2% trained for 6 hours a week or more.

Full participant characteristics are provided in Table 1.

**Table 1.** Description of the respondents.

| Category | Group | Frequency (n) | Percentage (%) |
|---|---|---|---|
| Age | 18-20 | 43 | 23.1 |
| | 21-25 | 57 | 30.6 |
| | 26-30 | 40 | 21.5 |
| | 31-35 | 19 | 10.2 |
| | 36-40 | 3 | 1.6 |
| | 41-45 | 6 | 3.2 |
| | 46-50 | 3 | 1.6 |
| | 51-55 | 0 | 0.0 |
| | 56-60 | 15 | 8.1 |
| Gender | Female | 1 | 0.7 |
| | Male | 180 | 96.7 |
| | Do not answer – apply | 5 | 2.6 |

| | | | |
|---|---|---|---|
| Education level | Primary level | 1 | 0.7 |
| | Middle school | 21 | 11.2 |
| | High school | 111 | 59.7 |
| | Bachelor degree | 27 | 14.5 |
| | Master degree | 20 | 10.7 |
| | PhD | 6 | 3.2 |
| Working type | Professional athlete | 116 | 62.4 |
| | Student | 20 | 10.7 |
| | Unemployed | 3 | 1.6 |
| | Worker | 47 | 25.4 |
| Sport level | Recreational | 11 | 5.9 |
| | Local | 8 | 4.3 |
| | Regional | 29 | 15.5 |
| | National | 128 | 68.8 |
| | International | 10 | 5.3 |
| Years of playing | <5 | 1 | 0.4 |
| | 6-10 | 10 | 5.3 |
| | 11-15 | 59 | 31.7 |
| | 16-20 | 66 | 35.5 |
| | >21 | 50 | 26.9 |

| | | | |
|---|---|---|---|
| N° of training per week | 1-3 | 4 | 2.2 |
| | 4-5 | 5 | 2.6 |
| | >6 | 177 | 95.2 |
| Hours of training per week | <5 | 37 | 19.9 |
| | 6-10 | 51 | 27.4 |
| | 11-15 | 77 | 41.3 |
| | 16-20 | 12 | 6.5 |
| | >21 | 9 | 4.8 |

## Instruments

### Wellness single-item scales

Participants completed the 5 single-item measures for Fatigue, Sleep Quality, Muscle Soreness, Stress and Mood. The single-item measures were a simple statement of "Descrivi come ti senti in questo momento (rispetto alla seguenti variabili) - Please describe how you currently feel (according to the following items)". All items were rated on a five-point response option from one to five. Specifically, the verbal anchors were 1 = "nessuna fatica - no fatigue" 5 ="molto stanco - very tired" for Fatigue (item 1), 1 = "ottima - very good" 5 ="pessima - very bad" for Sleep Quality (item 2), 1 ="nessuno - none" 5 ="molto indolenzito - very sore" for Muscle Soreness (item 3), 1 = "molto rilassato - very relaxed" 5 ="molto stressato - very stressed" for Stress (item 4), 1 ="molto positivo - very positive" 5 ="molto negativo - very negative" for Mood (item 5).

### Groningen Sleep Quality Scale (GSQS)

The Groningen sleep quality scale (GSQS)[34] was used to assess sleep quality. The 14-items GSQS specifically covered various sleep complaints: 3 items about general sleep quality, 3 items about insufficient sleep, 2 items about troubles with falling asleep, 1 item about tossing and turning, 3 items about trouble with sleeping on, 2 items about waking up unrested.

The 14 items of the scale fit a one dimensional scaling model as proposed by Mokken and Lewis[35], which represents a probabilistic version of the Gutman-procedure (Cronbach's coefficient is $\alpha = 0.89$). The scale was back translated in italian.

### Muscle Soreness Numerical Pain Rating scale

A single-item measure of pain as proposed by Williamson[36] was used in the study. Specifically, a 11-point numerical pain rating scale ranging from 0 (no pain) to 10 (worst imaginable pain) was used in this study answering the statement "Descrivi il tuo indolenzimento muscolare percepito durante la fase discendente durante la fase discendente di un mezzo squat - Please rate the lower limb muscle soreness perceived during the descending phase of a half squat movement". Bijur[37] found a significant correlation between the VAS and the NRS (r=0.94, 95% CI= 0.93–0.95). They also found that the slope of the regression line was 1.01 (95% CI=0.97–1.06) indicating a strong level of agreement between the two tools.

## ITAMS

The italian version of the Brunel Mood Scale (ITAMS[38]) consists of 6 dimensions with 4 descriptors each, specifically 4 related to anger, 4 related to confusion, 4 to depression, 4 to fatigue, 4 to tension and 4 to vigor (for a total of 24).

Participants rated their feelings on a 5-point response scale ranging from 0 (not at all) to 4 (extremely) rating "how you feel right now" for each mood descriptor, in fact ratings reflected their moods at the time of evaluation. Total scores were calculated for each of the six subscales, with lower scores represented weaker endorsement of each specific dimension.

3 of the 4 items in the Fatigue subscale were also used for the Multicomponent Training Distress Scale[39] (MTDS; Main and Grove, 2009) Fatigue subscale.

The internal consistency values (Cronbach's α) of all 6 dimensions and the total scale were all greater than 0.76 (Rohlf)[40], while in Brandt[41] the total scale was 0.841 (Tension 0.736; Depression 0.891; Anger 0.856; Vigour 0.789; Fatigue 0.654; Mental confusion 0.541).

## DASS-21

The Italian Depression Anxiety Stress Scales Short Version (DASS-21)[42–44] questionnaire was used. Each construct is assessed by seven items rated on a 4-point Likert scale. Only the DASS-21 Stress subscale was used in the present study.

Perceptions were indicated in reference to the previous week. Participants rated their feelings on a 4-point response scale ranging from 0 (did not apply to me at all) to 3 (applied to me very much or most of the time). Higher scores indicated higher levels of stress.

The reliability of the DASS-21 scales were .88 for Depression, 0.82 for Anxiety, 0.90 for Stress, and 0.93 for the Total scale as described by Henry[42]. There is no absolute criterion for the reliability of an instrument. However, as a rule of thumb, Anastasi[45] has suggested that a should be at least 0.85 if an instrument is to be used to draw inferences concerning an individual. By this criterion the Depression, Stress, and Total scales can be viewed as possessing adequate reliability, while the Anxiety scale fell below this criterion, although only marginally. Moreover, α is strongly affected by the number of items (the smaller the number of items, the lower alpha is). Therefore, a particular alpha value needs to be interpreted relative to the number of items, not as an absolute figure. Given that the DASS-21 was designed to provide brief measures of broad constructs, these values are satisfactory. Were the alphas much higher than those observed then they could be taken as an indication that the scales have insufficient bandwidth[46].

## Statistical Analysis

In order to assess the construct validity of the Wellness Single-Items we adopted COSMIN guidelines.

First of all, we investigated the goodness of the constructs with the COSMIN criteria to confirm factor structure (RMSEA<0.06, CFI>0.95, TLI>0.95, SRMR<0.08). Specifically, we performed the following analyses:

- Confirmatory Factor Analysis (CFA) with all ITAMS subscales but vigour, to confirm the Negative Affect construct;
- CFA with all 6 ITAMS subscales to confirm TMD construct;
- CFA with all the 5 single-item subscales to confirm the Wellness construct.

Construct validity is defined by COSMIN as the degree of which the scores of a measurement instrument are consistent with the hypotheses, with regard to internal relationships, relationships with scores of other instruments or differences between relevant groups[28,29]. The guidelines from COSMIN state that each hypothesis is formed a priori with specific and clearly defined direction, magnitude, and rationale[27]. Specifically, from one to nine independent hypotheses, dependent on the construct, were developed a priori. For each of these formulated hypotheses, the correlation direction, criteria, and rationale for which the hypothesis is based were clearly defined. The relationship between each single-item and validated questionnaire for the same construct was assessed by Kendall $\tau$ b correlation coefficient.

The hypotheses were defined as follows:

- single item of fatigue would reflect the construct of fatigue measured with fatigue in the ITAMS (and in the MTDS) Lower Limit CI>0.49 (same construct)
- single item of fatigue would reflect the construct of fatigue measured with vigor in the ITAMS CI>-0.30 (related but different construct)

- single item of fatigue would reflect the construct of fatigue measured with stress in the DASS-21 CI>0.30 (related but different construct)
- **2 out of 3 satisfied to confirm construct validity**

- single item of sleep quality would reflect sleep quality measured with GSQS sleep quality scale Lower Limit CI>0.49 (same construct)
- **1 out of 1 satisfied to confirm construct validity**

- single item of muscle soreness would reflect the construct of muscle soreness measured with the NPRS Lower Limit CI>0.49 (same construct)
- **1 out of 1 satisfied to confirm construct validity**

- single item of stress would reflect the multifactorial construct of stress measured with the DASS stress subscale. Lower Limit CI>0.49 (same construct)
- single item of stress would reflect the multifactorial construct of stress measured with the ITAMS tension subscale. Lower Limit CI>0.30 (related but different construct)
- single item of stress would reflect the multifactorial construct of stress measured with the ITAMS confusion subscale. Lower Limit CI>0.30 (related but different construct)
- single item of stress would reflect the multifactorial construct of stress measured with the ITAMS anger subscale. Lower Limit CI>0.30 (related but different construct)
- single item of stress would reflect the multifactorial construct of stress measured with the ITAMS depression subscale. Lower Limit CI>0.30 (related but different construct)
- single item of stress would reflect the multifactorial construct of stress measured with the ITAMS fatigue subscale. Lower Limit CI>0.30 (related but different construct)
- single item of stress would reflect the multifactorial construct of stress measured with the ITAMS vigor subscale. Lower Limit CI>-0.30 (related but different construct)
- **5 out of 7 satisfied to confirm construct validity**

- single item of mood would reflect the multifactorial construct of mood measured with ITAMS Total Mood Disturbance (TMD). Lower Limit CI>0.49 (same construct)
- single item of mood would reflect the multifactorial construct of mood measured with ITAMS Negative Affect subscales (NA). Lower Limit CI>0.49 (same construct)
- single item of mood would reflect the multifactorial construct of mood measured with the ITAMS tension subscale. Lower Limit CI>0.30 (related but different construct)
- single item of mood would reflect the multifactorial construct of mood measured with the ITAMS confusion subscale. Lower Limit CI>0.30 (related but different construct)
- single item of mood would reflect the multifactorial construct of mood measured with the ITAMS anger subscale. Lower Limit CI>0.30 (related but different construct)
- single item of mood would reflect the multifactorial construct of mood measured with the ITAMS depression subscale. Lower Limit CI>0.30 (related but different construct)
- single item of mood would reflect the multifactorial construct of mood measured with the ITAMS fatigue subscale. Lower Limit CI>0.30 (related but different construct)
- single item of mood would reflect the multifactorial construct of mood measured with ITAMS vigor subscale. Lower Limit CI>-0.30 (related but different construct)
- Ordinary regression with ITAMS (hypothesis, significant coefficients for all subscales as independent variables)
- **6 out of 9 satisfied to confirm construct validity**

Within our hypotheses, we included criteria that extended beyond mean estimates, encompassing confidence intervals. These confidence intervals were constructed as compatibility intervals—encompassing the range of effects wherein, if postulated as the true, unknown parameter value, the statistical test based on the "null" hypothesis would remain unaltered at the .05 significance level (95% confidence interval). Hence, this approach permits to be more confident about our results.

A Kendall $\tau$ b correlation coefficient was assessed in order to detect the relationship between each single-item in order to evaluated if the subjects provided the similar information of different scores.

# Results

## Confirmatory factor analyses

The results of the CFAs performed (Table 2) show poor structural validity of the Wellness construct. However, CFAs show moderate structural validity for the TMD and NA construct. Therefore, we decided to include TMD and NA for the validation of the Mood single-item.

**Table 2.** Confirmatory Factor Analysis (CFA) for each construct. COSMIN guideline references in brackets

| | CFI | TLI | RMSEA |
|---|---|---|---|
| Wellness | (>0.95) 0.856 | (>0.95) 0.711 | (<0.06) 0.186 |
| TMD | (>0.95) 0.938 | (>0.95) 0.897 | (<0.06) 0.198 |
| NA | (>0.95) 0.949 | (>0.95) 0.944 | (<0.06) 0.23 |

## Item correlations

To understand if some of the single-item proposed was redundant we employed a correlation matrix between the items as shown in Figure 1. None of the items seems to be very strongly correlated to another one. Actually, a relation was detected between Stress and Mood ($\tau$ = 0.58), while weak correlation was found among all the other single-items' pairs. Due to the fact that no very strong correlation was detected ($\tau$>0.7), the single-items likely measure different aspects of the players status.

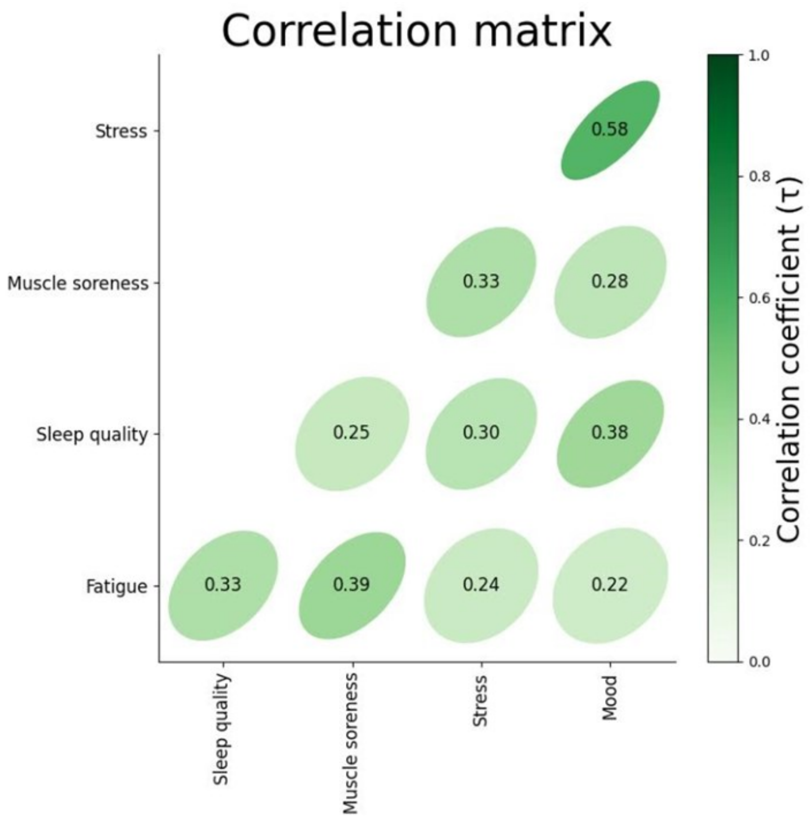



**Figure 1.** Correlation matrix between the items. The intensity of color green represents the intensity of the correlation and the direction of the major axis of the ellipses represents the direction of the relation.

## Validity analysis

Results for the validity hypotheses tested for each item are presented in Table 3. The COSMIN validity criteria recommend that around 75% of the hypothesis developed a priori should be satisfied. The number of hypotheses to be confirmed are shown in parenthesis for each single-item. The criteria to confirm, reject or partially support the hypotheses are included in Table 3 for each item. Of note is that we considered the uncertainty (CI) instead of the point estimate only, which is stricter than the COSMIN criteria.

### Fatigue (2/3)

Three out of 3 hypotheses tests were not confirmed for the single-item fatigue scale. Results do not support validity of the item.

**Sleep Quality (1/1)**

One out of 1 hypothesis tests was not confirmed for the single-item sleep quality scale. Results do not support validity of the item.

**Muscle Soreness (1/1)**

One out of 1 hypothesis tests was not confirmed for the single-item Muscle Soreness scale. Results do not support validity of the item.

**Stress (5/7)**

Three out of 7 hypotheses tests (ITAMS tension, depression and anger) were partially confirmed for the single-item Stress scale. The other 4 hypothesis were not confirmed. Results do not support validity of the item.

**Mood (6/9)**

Three out of 9 hypotheses tests (ITAMS depression, tension and confusion) were partially confirmed for the single-item Mood scale. The other 6 hypothesis were not confirmed. Results do not support validity of the item.

**Table 3.** Construct validity for each item.

| Hypothesis | Criteria | Results / Confirmed |
|---|---|---|
| | ***Fatigue*** | |
| Positive correlation with ITAMS and MTDS fatigue scores (two hypotheses were combined since MTDS is calculated using 3 out of 4 items of the BRUMS). | Lower limit for confidence intervals > 0.49 indicating strong evidence, and > 0.69 suggesting very strong evidence of convergent validity. If point estimate is > 0.49 and the lower limit is <0.49, the hypothesis is considered partially confirmed. If the upper limit of CI > 0.49 and the point estimate <0.49 the hypothesis is considered not confirmed with uncertainty. | No<br><br>ITAMS: $\tau$ = 0.35 (95% Cl, 0.2 to 0.48)<br><br>MTDS: $\tau$ = 0.34 (95% Cl, 0.19 to 0.47) |
| Positive correlation with DASS-21 stress scale | Lower limit of the confidence intervals at least > 0.30. If point estimate is > 0.30 and the lower limit is <0.30, the hypothesis is considered partially confirmed. If the upper limit of CI > 0.30 and the point estimate <0.30 the hypothesis is considered not confirmed with uncertainty. | No<br><br>$\tau$ = 0.12 (95% Cl, 0.01 to 0.23) |
| Negative correlation with the ITAMS vigor scale. | Lower limit of the confidence intervals at least <-0.30. If point estimate is <-0.30 and the lower limit is >-0.30, the hypothesis is considered partially confirmed. If the upper limit of CI <-0.30 and the point estimate >-0.30 the hypothesis is considered not confirmed with uncertainty. | No<br><br>$\tau$ = -0.10 (95% Cl, -0.26 to -0.05) |

| | | |
|---|---|---|
| | ***Sleep Quality*** | |
| Positive correlation with Groningen Sleep Quality Scale | Lower limit for confidence intervals > 0.49 indicating strong evidence, and > 0.69 suggesting very strong evidence of convergent validity. If point estimate is > 0.49 and the lower limit is <0.49, the hypothesis is considered partially confirmed. If the upper limit of CI > 0.49 and the point estimate <0.49 the hypothesis is considered not confirmed with uncertainty. | No<br>$\tau$ = 0.29 (95% Cl, 0.15 to 0.42) |
| | ***Muscle soreness*** | |
| Positive correlation with the Numerical Pain Rating Scale | Lower limit for confidence intervals > 0.49 indicating strong evidence, and > 0.69 suggesting very strong evidence of convergent validity. If point estimate is > 0.49 and the lower limit is <0.49, the hypothesis is considered partially confirmed. If the upper limit of CI > 0.49 and the point estimate <0.49 the hypothesis is considered not confirmed with uncertainty. | No<br>$\tau$ = 0.29 (95% Cl, 0.14 to 0.43) |
| | ***Stress*** | |

| | | |
|---|---|---|
| Positive correlation with the DASS-21 stress scale | Lower limit for confidence intervals > 0.49 indicating strong evidence, and > 0.69 suggesting very strong evidence of convergent validity. If point estimate is > 0.49 and the lower limit is <0.49, the hypothesis is considered partially confirmed. If the upper limit of CI > 0.49 and the point estimate <0.49 the hypothesis is considered not confirmed with uncertainty. | No (uncertain)<br><br>$\tau$ = 0.41 (95% Cl, 0.28 to 0.53) |
| Positive correlation with the ITAMS fatigue scale. | Lower limit of the confidence intervals at least > 0.30. If point estimate is > 0.30 and the lower limit is <0.30, the hypothesis is considered partially confirmed. If the upper limit of CI > 0.30 and the point estimate <0.30 the hypothesis is considered not confirmed with uncertainty. | No (uncertain)<br><br>$\tau$ = 0.21 (95% Cl, 0.05 to 0.35) |
| Positive correlation with the ITAMS tension scale | Lower limit of the confidence intervals at least > 0.30. If point estimate is > 0.30 and the lower limit is <0.30, the hypothesis is considered partially confirmed. If the upper limit of CI > 0.30 and the point estimate <0.30 the hypothesis is considered not confirmed with uncertainty. | Partially confirmed<br><br>$\tau$ = 0.42 (95% Cl, 0.28 to 0.54) |
| Positive correlation with the ITAMS confusion scale. | Lower limit of the confidence intervals at least > 0.30. If point estimate is > 0.30 and the lower limit is <0.30, the hypothesis is considered partially confirmed. If the upper limit of CI > 0.30 and the point estimate <0.30 the hypothesis is considered not confirmed with uncertainty. | No (uncertain)<br><br>$\tau$ = 0.25 (95% Cl, 0.09 to 0.39) |

| | | |
|---|---|---|
| Positive correlation with the ITAMS anger scale. | Lower limit of the confidence intervals at least > 0.30. If point estimate is > 0.30 and the lower limit is <0.30, the hypothesis is considered partially confirmed. If the upper limit of CI > 0.30 and the point estimate <0.30 the hypothesis is considered not confirmed with uncertainty. | Partially confirmed<br><br>$\tau$ = 0.37 (95% Cl, 0.23 to 0.50) |
| Positive correlation with the ITAMS depression scale. | Lower limit of the confidence intervals at least > 0.30. If point estimate is > 0.30 and the lower limit is <0.30, the hypothesis is considered partially confirmed. If the upper limit of CI > 0.30 and the point estimate <0.30 the hypothesis is considered not confirmed with uncertainty. | Partially confirmed<br><br>$\tau$ = 0.35 (95% Cl, 0.21 to 0.48) |
| Negative correlation with the ITAMS vigor scale | Lower limit of the confidence intervals at least <-0.30. If point estimate is <-0.30 and the lower limit is >-0.30, the hypothesis is considered partially confirmed. If the upper limit of CI <-0.30 and the point estimate >-0.30 the hypothesis is considered not confirmed with uncertainty. | No<br><br>$\tau$ = -0.07 (95% Cl, -0.22 to -0.09) |
| | ***Mood*** | |
| Positive correlation with the ITAMS fatigue scale | Lower limit of the confidence intervals at least > 0.30. If point estimate is > 0.30 and the lower limit is <0.30, the hypothesis is considered partially confirmed. If the upper limit of CI > 0.30 and the point estimate <0.30 the hypothesis is considered not confirmed with uncertainty. | No (uncertain)<br><br>$\tau$ = 0.22 (95% Cl, 0.07 to 0.37) |

| | | |
|---|---|---|
| Positive correlation with the ITAMS tension scale | Lower limit of the confidence intervals at least > 0.30. If point estimate is > 0.30 and the lower limit is <0.30, the hypothesis is considered partially confirmed. If the upper limit of CI > 0.30 and the point estimate <0.30 the hypothesis is considered not confirmed with uncertainty. | Partially confirmed<br>$\tau$ = 0.34 (95% Cl, 0.20 to 0.48) |
| Positive correlation with the ITAMS confusion scale | Lower limit of the confidence intervals at least > 0.30. If point estimate is > 0.30 and the lower limit is <0.30, the hypothesis is considered partially confirmed. If the upper limit of CI > 0.30 and the point estimate <0.30 the hypothesis is considered not confirmed with uncertainty. | Partially confirmed<br>$\tau$ = 0.31 (95% Cl, 0.16 to 0.45) |
| Positive correlation with the ITAMS anger scale | Lower limit of the confidence intervals at least > 0.30. If point estimate is > 0.30 and the lower limit is <0.30, the hypothesis is considered partially confirmed. If the upper limit of CI > 0.30 and the point estimate <0.30 the hypothesis is considered not confirmed with uncertainty. | No (uncertain)<br>$\tau$ = 0.29 (95% Cl, 0.14 to 0.42) |
| Positive correlation with the ITAMS depression scale | Lower limit of the confidence intervals at least > 0.30. If point estimate is > 0.30 and the lower limit is <0.30, the hypothesis is considered partially confirmed. If the upper limit of CI > 0.30 and the point estimate <0.30 the hypothesis is considered not confirmed with uncertainty. | Partially confirmed<br>$\tau$ = 0.41 (95% Cl, 0.27 to 0.53) |

| | | |
|---|---|---|
| Positive correlation with the ITAMS vigor scale | Lower limit of the confidence intervals at least <-0.30. If point estimate is <-0.30 and the lower limit is >-0.30, the hypothesis is considered partially confirmed. If the upper limit of CI <-0.30 and the point estimate >-0.30 the hypothesis is considered not confirmed with uncertainty. | No<br><br>$\tau$ = -0.12 (95% Cl, -0.27 to -0.04) |
| Positive correlation with the TMD score | Lower limit for confidence intervals > 0.49 indicating strong evidence, and > 0.69 suggesting very strong evidence of convergent validity. If point estimate is > 0.49 and the lower limit is <0.49, the hypothesis is considered partially confirmed. If the upper limit of CI > 0.49 and the point estimate <0.49 the hypothesis is considered not confirmed with uncertainty. | No (uncertain)<br><br>$\tau$ = 0.41 (95% Cl, 0.27 to 0.53) |
| Positive correlation with the ITAMS NA score | Lower limit for confidence intervals > 0.49 indicating strong evidence, and > 0.69 suggesting very strong evidence of convergent validity. If point estimate is > 0.49 and the lower limit is <0.49, the hypothesis is considered partially confirmed. If the upper limit of CI > 0.49 and the point estimate <0.49 the hypothesis is considered not confirmed with uncertainty. | No (uncertain)<br><br>$\tau$ = 0.39 (95% Cl, 0.25 to 0.51) |
| Ordinary regression between ITAMS subscales and single item Mood | At least 75% of the coefficients (4 out of 6) for the subscales significant (p<0.05) | No (uncertain)<br><br>$r^2$=0.208, F(173.6)=7.585,<br><br>p=0.001 for vigor, p=0.017 tension |

| | | |
|---|---|---|
| | | p=0.306 for confusion<br>p=0.329 for anger<br>p=0.470 for depression<br>p=0.693 for fatigue |

# Study B - Consistency

## Participants and Data Collection

For consistency investigation a total of 68 national professional soccer players (Swann et al., 2015[30]), (age = 22.78±5.91 yrs, height = 181±4 cm, weight 78.94±4.5 kg) completed the survey for an entire competitive season (seasons 2019-2023). After a detailed description of the procedure and possible risks, players voluntarily decided to participate by signing an informed consent. The players were part of three professional teams reached by personal connections. The project was conducted according to the Declaration of Helsinki and was approved by the Bioethics Committee of the University of Bologna.

Every day, usually in private and at a consistent time in the morning, in any case before each training session or match, the players filled the Wellness Questionnaire. Moreover, at the end of each training or match, the players provide the rating of perceived exertion (CR-10 Borg Scale - RPE, where 0 refers to resting state and 10 to maximal effort) that permits to compute the training load (TL) of each session as the product between RPE and session duration. In order to assess the TL status of each player's session, the TL was aggregated in acute (mean of the past 7 days - $TL_{acute}$) and chronic (mean of the past 28 days - $TL_{chronic}$) status.

## Instruments

### Wellness Single-Item Scales

Participants completed the 5 single-item measures for Fatigue, Sleep Quality, Muscle Soreness, Stress and Mood. The single-item measures were a simple statement of "Descrivi come ti senti in questo momento (rispetto alla seguenti variabili) - Please describe how you currently feel (according to the following items)". All items were rated on a five point response option from one to five. Specifically, the verbal anchors were 1 = "nessuna fatica - no fatigue" 5 ="molto stanco - very tired" for Fatigue (item 1), 1 = "ottima - very good" 5 =”pessima - very bad" for Sleep Quality (item 2), 1 =”nessuno - none" 5 =”molto indolenzito - very sore" for Muscle Soreness (item 3), 1 = "molto rilassato - very relaxed" 5 =”molto stressato - very stressed" for Stress (item 4), 1 ="molto positivo - very positive" 5 =”molto negativo - very negative" for Mood (item 5).

## Statistical Analysis

In order to assess each item's reliability, a total of 17913 training and matches sessions were recorded. To detect the players' consistency to fill in the questionnaire, the item provided by players after a pair of similar training was compared. To define similar training the euclidean distance between each pair of records for each player was computed. The features used to obtain similarity were, time, $TL_{acute}$ (Training Load mean of the past week) and $TL_{chronic}$ (Training Load mean of the past month). Due to the fact that we compute euclidean distance the features were normalized between 0 and 1 for each player. A distance of 0.01 (variability about 1%) was used as thresholds for similarity. McNemar's test was used to detect differences between players' responses (paired nominal data).

## Results

5867 pairs of training sessions reach the criterion of inclusion (euclidean distance between training session characteristics lower than 0.05) in the reliability analysis. The consistency results are provided in Table 4. No statistical difference was detected between the two similar training sessions ($p>0.05$) for each single-item. In particular, mean differences were found between -0.03 and 0. For more details about consistency, the frequency analysis was provided in Figure 2. Figure 2 shows that most of the time the players provided the same single-item value in similar TL conditions. At major times, the difference between the two similar training sessions was 1 point. Only a few times the difference is higher than 1.

**Table 4.** Items consistency

| **Items** | **Day 1** | **Day 2** | **Differences** | **statistic*** |
|---|---|---|---|---|
| Fatigue | 2.10 ± 0.74 | 2.11 ± 0.73 | -0.01 ± 0.58 | p = 0.54 |
| Sleep Quality | 2.00 ± 0.61 | 2.01 ± 0.62 | -0.01 ± 0.42 | p = 0.68 |
| Muscle Soreness | 2.05 ± 0.64 | 2.06 ± 0.64 | -0.01 ± 0.35 | p = 0.78 |
| Stress | 2.18 ± 0.75 | 2.17 ± 0.72 | -0.00 ± 0.46 | p = 0.61 |

| | | | | |
|---|---|---|---|---|
| Mood | 2.16 ± 0.73 | 2.16 ± 0.79 | 0.01 ± 0.48 | p = 0.69 |

* McNemar's test

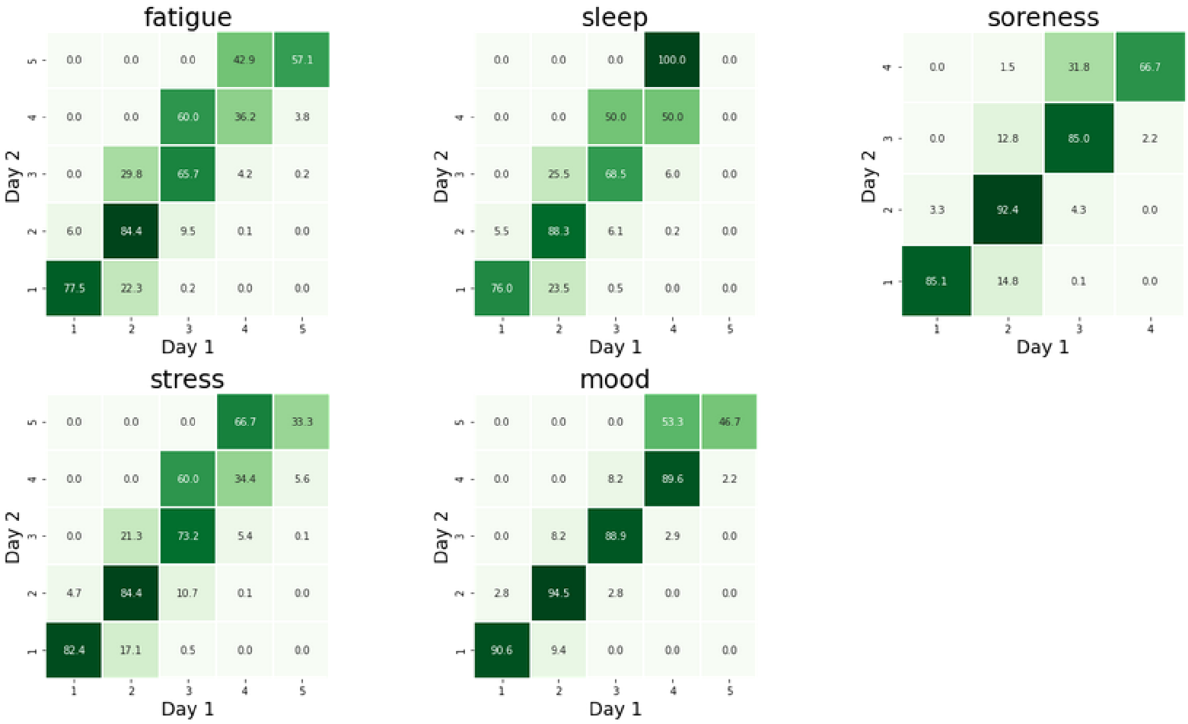



**Figure 2.** Frequency analysis for consistency. The values in the heatmap are expressed as percentage.

# Discussion

The aim of this study was to explore the validity and consistency of each single-item encompassed in the Wellness questionnaires. Those scales are commonly used by sports practitioners to monitor their athletes and by sport scientists in their research. Despite their widespread use, to the best of our knowledge these single-item scales have not been validated previously. Therefore, the degree to which these scales measure the constructs they purport to measure was unknown. This is particularly concerning when the constructs being measured with a single-item are broad constructs such as Mood and Stress.

## Validity

The results of our investigation do not support the validity of any of the five single-item scales in Italian soccer players. Specifically, the Fatigue and Muscle Soreness scales are commonly used as single-items in both practice and research, but surprisingly in our cohort, we did not find any substantial correlation with the criterion measures. This is surprising because other single-item scales for fatigue and muscle soreness have been validated and are commonly used in many fields such as occupational therapy[11] and/or medicine[47–49]. Therefore, the most likely explanation for this discrepant results is the way fatigue and muscle soreness were measured. In our study, we used a 5-point NRS ranging from 1 to 5. All the other validated fatigue and pain/muscle soreness scales are VAS and NRS normally ranging between 0 and 10, or 0 and 100. The validated scale with the least number of points is the muscle soreness scale developed by Vickers[47] which has 7 points. As shown by Dawes[50], 5-point scales have less variance than scales with more points. As a result, the 5-point Fatigue and Muscle Soreness scales used in our study may not be able to accurately capture the full range of fatigue and muscle soreness experienced by the respondent and reduced their validity compared to scales and questionnaires with better psychometric properties.

With regards to the Sleep Quality scale, it is important to note that sleep quality is a broad construct commonly measured by questionnaires referring to the previous week or previous month. To the best of our knowledge, the only validated instruments measures referring to the previous night was the GSQS that was used as criterion measure in our study because our Sleep Quality scale refers to the previous night. Nevertheless, we found no evidence of validity for such scale. Again, the most likely explanation is the narrow range of responses when using a single-item 5-point scale compared to a more psychometrically robust instrument like the GSQS[34] which consists of 14 items that are summed to provide an overall score of sleep quality.

Mood and stress are even more complex psychological constructs with a multifactorial nature that may be hard to capture with a single-item. To validate the single-item Stress Scale, the DASS-21 stress subscale (same construct) and all the ITAMS subscales (related but different construct) were used as a criterion measure. Compared to these subscales, the single-item Stress scale showed partial correlation with 3 of them, and no correlation with 4 of them. Hence, the single-item Stress scale provides no clear of evidence of validity in Italian soccer players.

In order to test the correlation between ITAMS and the single-item Mood scale, we analyzed all the ITAMS subscales plus 2 calculated a single score (Negative Affect and Total Mood Disturbance), and a ordinary regression between ITAMS subscales and single item Mood itself. Negative Affect score is given by the sum of the negative subscales (fatigue, depression, anger, confusion, tension) of the ITAMS and TMD equals the Negative Affect score minus the positive one (vigor). TDM was found to be elevated in periods of intensified training compared to baseline or periods of tapering in college athletes[51]. TMD directly relates with training intensity regardless to gender, in runners and college swimmers[52–54]. A significative difference in TMD was also found between abbreviated and normal sessions in swimmers. Specifically, during normal-distance practices, scores on TDM increased from pre- to post practice. In abbreviated practice sessions, athlete's scores on TMD showed no change from pre- to post practice. The mood changes related to practice distance were not influenced by the possible moderating factors of expectancy or performance times[55]. Single-item Mood Scale seems to provide a trivial assessment of mood when compared to the referring instruments. Compared to these subscales, the single-item Mood scale showed partial correlation with 3 of them, and no correlation with 6 of them.

Overall, these findings do not support the use of these single-item scales to measure mood and stress in soccer players. As for the other single-item scales, the poor validity of our Mood and Stress scales may be due to their narrow 5-point range. In addition, it is reasonable to think that a single-item scale cannot adequately capture and thus quantify the multi-factorial natures of complex constructs like mood and stress. However, there is some evidence from work psychology[56], health[57] and organizational psychology[19] that single-item scales with a broader range of points are able to provide reasonably valid measurements of complex constructs. For example, specifically, Matthews[19] investigated how well 91 single-items in organizational psychology performed in relation to multi-items. The vast majority (more than 70%) of the single-item measures demonstrated strong if not very strong content validity. Most of the single item measures were judged as reflecting the construct intended, and demonstrated good reliability over time. Therefore, given the practicality of single-item scales for day to day assessment of soccer players status, further research using single-item scales with more than 5-points are warranted.

Given that using the 5 items of the Wellness Questionnaire as single-items is not valid, one could argue for using the Wellness Questionnaire as originally intended, i.e as a sum of the 5 items to provide a score for the wellness construct. However our results argue also against such use. Indeed, the CFA do not support the existence of a wellness construct which also has a poor theoretical underpinning as discussed in the Introduction. Furthermore, the fact that the 5 single-items do not strongly correlate with each other suggests that, whatever they measure, they are not measuring similar or related constructs.

## Consistency

The consistency analysis results provided no statistically significant differences between ratings provided the morning after two similar training sessions ($p>0.05$) for each single-item (mean differences between -0.03 and 0) (Table 4). As shown in Figure 2, a high percentage of participants provided the same ratings when the TL of the previous day is controlled for. These results obtained with data collected in a real-life scenario underline the consistency of these 5 single-item scales.

## Limitations

A limitation could be the gender imbalance in the participant sample. We did not include gender differences in the study design and analysis. Future studies may seek to examine whether differences exist between sex, gender, or both.

Another limitation could be the use of a very specific participant sample. Thus, it should be investigated as well for different populations.

The use of 5 points single-item scales to monitor different aspects of the subjective status is very common in sport. Validated single-item scales normally have more than 5 points so, as underlined before, a more psychometrically sound version needs to be investigated. Sensitivity of change to physical or psychological load also need to be assessed. Floor and ceiling effect could also be examined. Future studies may seek to investigate these points.

We used the DASS-21 stress subscales separate from the whole instruments. However, the performance of a subscale may be different (eg. worse) when used in isolation, that is, not within the whole instrument.

# Conclusion and Practical Applications

These findings do not provide evidence for the validity of the 5 single-item scales derived from the Wellness Questionnaire and widely used in research and practice to monitor the status of soccer players daily. Furthermore, we provided empirical evidence (CFA) against the use of these scales to provide a total score for the Wellness Questionnaire as originally intended[21,22].

The use of single-item scales for daily monitoring in soccer is very common and frequent in all of the major soccer leagues in the world. in my experience, the use of sigle-item scales for daily monitoring is feasible and reasonably well tolerated by the athletes. However, without proven validity and reliability, practitioners cannot be confident that the data collected provide useful information. Given the potential of daily assessment for soccer players, we suggest performing the validation of single-items with broader range scales (e.g. 10 or 11 point scales).

These single-item scales are widely used by coaches of all major soccer leagues because they are simple and non-invasive tools to assess daily the status of soccer players. This information is then used to make decisions regarding the training or recovery sessions planned for that day. However, to be valid, such important decisions should be based on valid data. For the first time, we provided evidence that these single-item scales do not provide valid data regarding the constructs (fatigue, sleep quality, muscle soreness, stress and mood) they purport to measure. Therefore, we recommend coaches and sport scientists to stop using these scales. Although they are impractical for daily use, we recommend that soccer players are assessed weekly using validated scales and questionnaires such as those used as criterion measures in this study. From a research perspective, we recommend investigating the validity of single-items with broader range (e.g. 10 or 11 point scales) because the lack of validity of the scales currently used may be due to their narrow 5 point range.

# References


1. Jeffries A, Wallace L, Coutts A, McLaren S, Mccall A, Impellizzeri F. Athlete-Reported Outcome Measures for Monitoring Training Responses: A Systematic Review of Risk of Bias and Measurement Property Quality According to the COSMIN Guidelines. *International journal of sports physiology and performance*. 2020;15:1-13. doi:10.1123/ijspp.2020-0386
2. Taylor KL, Chapman D, Cronin J, Newton M, Gill N. Fatigue Monitoring in High Performance Sport: A Survey of Current Trends. *Journal of Australian Strength and Conditioning*. 2012;20:12-23.
3. Buchheit M, Racinais S, Bilsborough JC, et al. Monitoring fitness, fatigue and running performance during a pre-season training camp in elite football players. *Journal of Science and Medicine in Sport*. 2013;16(6):550-555. doi:10.1016/j.jsams.2012.12.003
4. Schonhardt S, Sullivan S, Marshall RS. A Focused Review of Multidimensional Well-Being Assessments. *Journal of Wellness*. 2023;4(2). doi:10.55504/2578-9333.1140
5. Charlemagne-Badal SJ, Lee JW, Butler TL, Fraser GE. Conceptual Domains Included in Wellbeing and Life Satisfaction Instruments: A Review. *Applied Research Quality Life*. 2015;10(2):305-328. doi:10.1007/s11482-014-9306-6
6. Conceição P, Bandura R. Measuring Subjective Wellbeing : A Summary Review of the Literature. In: ; 2008. Accessed October 27, 2023. https://www.semanticscholar.org/paper/Measuring-Subjective-Wellbeing-%3A-A-Summary-Review-Concei%C3%A7%C3%A3o-Bandura/177272a223411959e11966369c04b6f88a7b07c8
7. Cooke PJ, Melchert TP, Connor K. Measuring Well-Being: A Review of Instruments. *The Counseling Psychologist*. 2016;44(5):730-757. doi:10.1177/0011000016633507
8. Dodge R, Daly AP, Huyton J, Sanders LD. The challenge of defining wellbeing. *International Journal of Wellbeing*. 2012;2(3). Accessed October 27, 2023. https://www.internationaljournalofwellbeing.org/index.php/ijow/article/view/89
9. Linton MJ, Dieppe P, Medina-Lara A. Review of 99 self-report measures for assessing well-being in adults: exploring dimensions of well-being and developments over time. *BMJ Open*. 2016;6(7):e010641. doi:10.1136/bmjopen-2015-010641
10. Gallo TF, Cormack SJ, Gabbett TJ, Lorenzen CH. Pre-training perceived wellness impacts training output in Australian football players. *J Sports Sci*. 2016;34(15):1445-1451. doi:10.1080/02640414.2015.1119295
11. Van Hooff MLM, Geurts SAE, Kompier MAJ, Taris TW. "How Fatigued Do You Currently Feel?" Convergent and Discriminant Validity of a Single-Item Fatigue Measure. *Journal of Occupational Health*. 2007;49(3):224-234. doi:10.1539/joh.49.224
12. Milton K, Bull FC, Bauman A. Reliability and validity testing of a single-item physical activity measure. *British Journal of Sports Medicine*. 2011;45(3):203-208. doi:10.1136/bjsm.2009.068395
13. Mitchell AJ. Pooled results from 38 analyses of the accuracy of distress thermometer and other ultra-short methods of detecting cancer-related mood disorders. *J Clin Oncol*.

2007;25(29):4670-4681. doi:10.1200/JCO.2006.10.0438

14. Patrician PA. Single-item graphic representational scales. *Nurs Res*. 2004;53(5):347-352. doi:10.1097/00006199-200409000-00011
15. Wanous JP, Reichers AE, Hudy MJ. Overall job satisfaction: how good are single-item measures? *J Appl Psychol*. 1997;82(2):247-252. doi:10.1037/0021-9010.82.2.247
16. Measuring Global Self-Esteem: Construct Validation of a Single-Item Measure and the Rosenberg Self-Esteem Scale - Richard W. Robins, Holly M. Hendin, Kali H. Trzesniewski, 2001. Accessed October 27, 2023. https://journals.sagepub.com/doi/10.1177/0146167201272002
17. Impellizzeri F, Marcora S. Test Validation in Sport Physiology: Lessons Learned From Clinimetrics. *International journal of sports physiology and performance*. 2009;4:269-277. doi:10.1123/ijspp.4.2.269
18. DeVellis RF, Thorpe CT. *Scale Development: Theory and Applications*. SAGE Publications; 2021.
19. Matthews RA, Pineault L, Hong YH. Normalizing the Use of Single-Item Measures: Validation of the Single-Item Compendium for Organizational Psychology. *J Bus Psychol*. 2022;37(4):639-673. doi:10.1007/s10869-022-09813-3
20. Psychological Testing and Assessment: An Introduction to Tests and Measurement, 7th Edition.
21. Hooper SL, Mackinnon LT, Howard A, Gordon RD, Bachmann AW. Markers for monitoring overtraining and recovery. *Med Sci Sports Exerc*. 1995;27(1):106-112.
22. McLean BD, Coutts AJ, Kelly V, McGuigan MR, Cormack SJ. Neuromuscular, endocrine, and perceptual fatigue responses during different length between-match microcycles in professional rugby league players. *Int J Sports Physiol Perform*. 2010;5(3):367-383. doi:10.1123/ijspp.5.3.367
23. Perri E, Simonelli C, Rossi A, Trecroci A, Alberti G, Iaia FM. Relationship Between Wellness Index and Internal Training Load in Soccer: Application of a Machine Learning Model. *International Journal of Sports Physiology and Performance*. 2021;16(5):695-703. doi:10.1123/ijspp.2020-0093
24. Gastin PB, Meyer D, Robinson D. Perceptions of wellness to monitor adaptive responses to training and competition in elite Australian football. *J Strength Cond Res*. 2013;27(9):2518-2526. doi:10.1519/JSC.0b013e31827fd600
25. Lathlean TJH, Gastin PB, Newstead SV, Finch CF. A Prospective Cohort Study of Load and Wellness (Sleep, Fatigue, Soreness, Stress, and Mood) in Elite Junior Australian Football Players. *International Journal of Sports Physiology and Performance*. 2019;14(6):829-840. doi:10.1123/ijspp.2018-0372
26. Crowcroft S, McCleave E, Slattery K, Coutts AJ. Assessing the Measurement Sensitivity and Diagnostic Characteristics of Athlete-Monitoring Tools in National Swimmers. *International Journal of Sports Physiology and Performance*. 2017;12(s2):S2-95-S2-100. doi:10.1123/ijspp.2016-0406
27. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1147-1157. doi:10.1007/s11136-018-1798-3
28. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international

consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol*. 2010;63(7):737-745. doi:10.1016/j.jclinepi.2010.02.006

29. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res*. 2010;19(4):539-549. doi:10.1007/s11136-010-9606-8

30. Swann C, Moran A, Piggott D. Defining elite athletes: Issues in the study of expert performance in sport psychology. *Psychology of Sport and Exercise*. 2015;16:3-14. doi:10.1016/j.psychsport.2014.07.004

31. Brislin R. The wording and translation of research instruments. In: ; 1986. Accessed October 27, 2023. https://www.semanticscholar.org/paper/The-wording-and-translation-of-research-Brislin/5a63fa8802048b66007b70c129f457ab6add67af

32. Back-Translation for Cross-Cultural Research - Richard W. Brislin, 1970. Accessed October 27, 2023. https://journals.sagepub.com/doi/abs/10.1177/135910457000100301

33. Eysenbach G. Improving the Quality of Web Surveys: The Checklist for Reporting Results of Internet E-Surveys (CHERRIES). *Journal of Medical Internet Research*. 2004;6(3):e132. doi:10.2196/jmir.6.3.e34

34. Meijman TF, Thunnissen MJ, de Vries-Griever AGH. The after-effects of a prolonged period of day-sleep on subjective sleep quality. *Work & Stress*. 1990;4(1):65-70. doi:10.1080/02678379008256965

35. Mokkan RJ, Lewis C. A Nonparametric Approach to the Analysis of Dichotomous Item Responses. *Applied Psychological Measurement*. 1982;6(4):417-430. doi:10.1177/014662168200600404

36. Williamson A, Hoggart B. Pain: a review of three commonly used pain rating scales. *Journal of Clinical Nursing*. 2005;14(7):798-804. doi:10.1111/j.1365-2702.2005.01121.x

37. Bijur PE, Latimer CT, Gallagher EJ. Validation of a verbally administered numerical rating scale of acute pain for use in the emergency department. *Acad Emerg Med*. 2003;10(4):390-392. doi:10.1111/j.1553-2712.2003.tb01355.x

38. Quartiroli A, Terry PC, Fogarty GJ. Development and Initial Validation of the Italian Mood Scale (ITAMS) for Use in Sport and Exercise Contexts. *Frontiers in Psychology*. 2017;8. Accessed October 27, 2023. https://www.frontiersin.org/articles/10.3389/fpsyg.2017.01483

39. Main L, Grove R. A multi-component assessment model for monitoring training distress among athletes. *European Journal of Sport Science*. 2009;9:195. doi:10.1080/17461390902818260

40. Rohlfs RV, Weir BS. Distributions of Hardy–Weinberg Equilibrium Test Statistics. *Genetics*. 2008;180(3):1609-1616. doi:10.1534/genetics.108.088005

41. Brandt R, Herrero D, Massetti T, et al. The Brunel Mood Scale Rating in Mental Health for Physically Active and Apparently Healthy Populations. *Health*. 2016;8(2):125-132. doi:10.4236/health.2016.82015

42. Henry JD, Crawford JR. The short-form version of the Depression Anxiety Stress Scales (DASS-21): Construct validity and normative data in a large non-clinical sample. *British Journal of Clinical Psychology*. 2005;44(2):227-239. doi:10.1348/014466505X29657

43. Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the

Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behav Res Ther*. 1995;33(3):335-343. doi:10.1016/0005-7967(94)00075-u

44. Bottesi G, Ghisi M, Altoè G, Conforti E, Melli G, Sica C. The Italian version of the Depression Anxiety Stress Scales-21: Factor structure and psychometric properties on community and clinical samples. *Compr Psychiatry*. 2015;60:170-181. doi:10.1016/j.comppsych.2015.04.005
45. Anastasi A. *Psychological Testing*. 6th ed. Macmillan ; Collier Macmillan; 1988. Accessed October 27, 2023. http://www.gbv.de/dms/bowker/toc/9780023030208.pdf
46. Boyle GJ. Self-report measures of depression: Some psychometric considerations. *British Journal of Clinical Psychology*. 1985;24(1):45-59. doi:10.1111/j.2044-8260.1985.tb01312.x
47. Vickers AJ. COMPARISON OF AN ORDINAL AND A CONTINUOUS OUTCOME MEASURE OF MUSCLE SORENESS. *International Journal of Technology Assessment in Health Care*. 1999;15(4):709-716. doi:10.1017/S0266462399154102
48. Mattacola CG, Perrin DH, Gansneder BM, Allen JD, Mickey CA. A Comparison of Visual Analog and Graphic Rating Scales for Assessing Pain Following Delayed Onset Muscle Soreness. *Journal of Sport Rehabilitation*. 1997;6(1):38-46. doi:10.1123/jsr.6.1.38
49. Hjollund NH, Andersen JH, Bech P. Assessment of fatigue in chronic disease: a bibliographic study of fatigue measurement scales. *Health Qual Life Outcomes*. 2007;5(1):12. doi:10.1186/1477-7525-5-12
50. Dawes J. Five point vs. eleven point scales: Does it make a difference to data characteristics? *Australasian Journal of Market Research*. 10:39-47.
51. Hooper SL, Traeger Mackinnon L, Ginn EM. Effects of three tapering techniques on the performance, forces and psychometric measures of competitive swimmers. *Eur J Appl Physiol*. 1998;78(3):258-263. doi:10.1007/s004210050417
52. Raglin JS, Koceja DM, Stager JM, Harms CA. Mood, neuromuscular function, and performance during training in female swimmers. *Med Sci Sports Exerc*. 1996;28(3):372-377. doi:10.1097/00005768-199603000-00013
53. Morgan WP, Brown DR, Raglin JS, O'Connor PJ, Ellickson KA. Psychological monitoring of overtraining and staleness. *Br J Sports Med*. 1987;21(3):107-114.
54. Wittig AF, Houmard JA, Costill DL. Psychological effects during reduced training in distance runners. *Int J Sports Med*. 1989;10(2):97-100. doi:10.1055/s-2007-1024882
55. Berger BG, Prapavessis H, Grove JR, Butki BD. Relationship of Swimming Distance, Expectancy, and Performance to Mood States of Competitive Athletes. *Percept Mot Skills*. 1997;84(3_suppl):1199-1210. doi:10.2466/pms.1997.84.3c.1199
56. Fakunmoju SB. Validity of Single-item Versus Multiple-item Job Satisfaction Measures in Predicting Life: Satisfaction and Turnover Intention. *Asia-Pacific Journal of Management Research and Innovation*. 2020;16(3):210-228. doi:10.1177/2319510X21997724
57. Verster JC, Sandalova E, Garssen J, Bruce G. The Use of Single-Item Ratings Versus Traditional Multiple-Item Questionnaires to Assess Mood and Health. *European Journal of Investigation in Health, Psychology and Education*. 2021;11(1):183-198. doi:10.3390/ejihpe11010015

# CHAPTER 5: General discussion

## Subjective fatigue and recovery in professional soccer players

Chapter 2 and 3 describe two studies based on thousands of data collected over several years of practice as a strength and conditioning coach for various professional soccer teams. During these years, I followed the common practice of measuring the subjective status of my players using the 5 single-item scales derived from the Wellness Questionnaire[1]. Such practice is based on research using these scales[2–6] and it is widespread at all levels, including many major professional soccer leagues. This is not surprising because these are practical and non invasive tools that provide a more systematic way to gather information about the player daily status compared to the informal conversations and observations of coaches and other staff.

However, to the best of my knowledge, these 5 single-item scales measuring fatigue, sleep quality, muscle soreness, stress and mood had never been validated against more sophisticated and psychometrically sound measures of the same constructs. Therefore, as the final study of my PhD (Chapter 4), we conducted a validation study. As described in detail in that chapter and again later in this General Discussion, this study failed to provide evidence for convergent validity of these scales. Therefore, the conclusions reached in the first two studies (Chapter 2 and 3) cannot be considered valid and the related conclusions should be considered tentative at best.

Nevertheless, the studies described in Chapter 2 and 3 confirm that daily monitoring with simple single-item scales is feasible and well-tolerated by the players. It is unlikely that more complex multi-item questionnaires can be used for daily monitoring. Therefore, until valid single-item scales are provided, we recommend the use, once a week, of validated scales and questionnaires such as those used as "gold standards" in our validation study (Chapter 4). Although less ideal than daily monitoring, weekly monitoring of soccer players status, (together with valid data about their weekly training load using GPS technology and/or Session RPE) may nevertheless provide valid and therefore useful information to coaches and other staff such as physiotherapists and medical doctors.

A positive aspect of the studies described in Chapter 2 and 3 is that subjective data collected daily can be analysed and provide summary information using a big data analytic approach. Nowadays, in professional sports, daily monitoring of training loads and other variables is widespread. All this enormous amount of data is not easy to be handled. The big data analytics approached used in these two studies, applied through the lenses of a theoretical framework such as the one adopted for this thesis (Chapter 1), may have huge advantages in the development of accurate predictions once valid instruments to measure daily subjective status of soccer players are developed. Such predictions may then support the important decisions that coaches and support staff make every day with regards to training loads and other aspects of players management.

## Validity, consistency and feasibility of single-item scales for daily monitoring of professional soccer players

The use of single-item scales for daily monitoring in professional soccer is very common and frequent in all of the major soccer leagues in the world. During my experience, collecting everyday data during the whole competitive season, with 5 or 6 single-item scales is feasible and reasonably well tolerated by the athletes. However, without proven validity and reliability, practitioners cannot be confident that the data collected provide useful information. In Chapter 4 we, for the first time in sport science, have discussed this custom use putting it under the lens of a stricter psychometric approach.

Specifically, we have shown for the first time that the analysis of the validity of the so-called Wellness single-item scales do not clearly support their validity. Therefore, the idea of daily monitoring should be developed with a better psychometric approach in the development of the scales, for example encompassing a broder 10 or 11 point scale. Hence, the results of a validation process should make us more confident in research with these single-item scales and, together with the proven utility and feasibility of daily monitoring we expect a widespread use of big data analytics. Then, coaches and practitioners should be able to program the training sessions in accordance with the subjective status of the players.

## Directions for Future Research

Nowadays subjective monitoring is widely recognised as a reliable practice. There is also evidence that subjective measures are superior to objective measures in the monitoring of athletes[7]. Nevertheless, future research should investigate whether the predictive ability of the big data analytics approach can be improved by the inclusion of objective internal training load (e.g. heart rate (HR)) and external training load (e.g. GPS measures) that are nowadays commonly used in many top professional teams. Furthermore, as recently shown in cycling[8], the ratios between subjective measures like sRPE and either external training load or HR may provide even further information. This may especially be important in the prediction of fatigue because increases in these ratios could reflect progressive fatigue that cannot be clearly detected by changes in the individual measures of sRPE, external training load, and HR.

Previous research[9,10] also suggests that the combination of sRPE, HR and external training load may also provide useful information about the nature of fatigue. In fact, muscle fatigue increases both RPE and HR for the same external load, whereas mental fatigue increases only RPE for a given external load with HR staying the same or even decreasing. Therefore, the differential effects of muscle fatigue and mental fatigue on HR and RPE may provide to be useful in predictive models of fatigue which may be able to differentiate between physical and mental fatigue.

Future studies should also investigate further the relationship between subjective fatigue and performance. We should understand how performance is linked with subjective fatigue in a team sport like soccer where technical and tactical components are a very important part of the sport. Mental aspects of fatigue and recovery should be taken into account. Further evidence that subjective fatigue and recovery are associated with performance on the field and training responses may help educating coaches and players about the relevance and practical usefulness of the measurements and big data analytics approach described in this thesis.

## Summary and Conclusion

Based on the findings described in this thesis, the continued use of these 5 single-item scales should be discouraged as there is no evidence that they provide valid information about the subjective status of soccer players. Future studies should further investigate the validity of different single-item scales because daily monitoring of players is useful for coaches and other staff, and could be analysed effectively using the big data analytics approach described in this thesis. Until then, we recommend weekly monitoring of players using validated scales and multi-item questionnaires.

## References


1. Hooper SL, Mackinnon LT, Howard A, Gordon RD, Bachmann AW. Markers for monitoring overtraining and recovery. *Med Sci Sports Exerc.* 1995;27(1):106-112.
2. McLean BD, Coutts AJ, Kelly V, McGuigan MR, Cormack SJ. Neuromuscular, endocrine, and perceptual fatigue responses during different length between-match microcycles in professional rugby league players. *Int J Sports Physiol Perform.* 2010;5(3):367-383. doi:10.1123/ijspp.5.3.367
3. Gastin PB, Meyer D, Robinson D. Perceptions of wellness to monitor adaptive responses to

training and competition in elite Australian football. *J Strength Cond Res*. 2013;27(9):2518-2526. doi:10.1519/JSC.0b013e31827fd600

4. Lathlean TJH, Gastin PB, Newstead SV, Finch CF. A Prospective Cohort Study of Load and Wellness (Sleep, Fatigue, Soreness, Stress, and Mood) in Elite Junior Australian Football Players. *International Journal of Sports Physiology and Performance*. 2019;14(6):829-840. doi:10.1123/ijspp.2018-0372
5. Crowcroft S, McCleave E, Slattery K, Coutts AJ. Assessing the Measurement Sensitivity and Diagnostic Characteristics of Athlete-Monitoring Tools in National Swimmers. *International Journal of Sports Physiology and Performance*. 2017;12(s2):S2-95-S2-100. doi:10.1123/ijspp.2016-0406
6. Perri E, Simonelli C, Rossi A, Trecroci A, Alberti G, Iaia FM. Relationship Between Wellness Index and Internal Training Load in Soccer: Application of a Machine Learning Model. *International Journal of Sports Physiology and Performance*. 2021;16(5):695-703. doi:10.1123/ijspp.2020-0093
7. Saw AE, Main LC, Gastin PB. Monitoring the athlete training response: subjective self-reported measures trump commonly used objective measures: a systematic review. *Br J Sports Med*. 2016;50(5):281-291. doi:10.1136/bjsports-2015-094758
8. Sanders D, Heijboer M, Hesselink MKC, Myers T, Akubat I. Analysing a cycling grand tour: Can we monitor fatigue with intensity or load ratios? *Journal of Sports Sciences*. 2018;36(12):1385-1391. doi:10.1080/02640414.2017.1388669
9. Marcora SM, Bosio A, de Morree HM. Locomotor muscle fatigue increases cardiorespiratory responses and reduces performance during intense cycling exercise independently from metabolic stress. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*. 2008;294(3):R874-R883. doi:10.1152/ajpregu.00678.2007
10. Marcora S. Perception of effort during exercise is independent of afferent feedback from skeletal muscles, heart, and lungs. *Journal of Applied Physiology*. 2009;106(6):2060-2062. doi:10.1152/japplphysiol.90378.2008

# Appendix

## Wellness Questionnaire

Descrivi come ti senti in questo momento (rispetto alle seguenti variabili)

| | | | | | |
|---|---|---|---|---|---|
| **Fatica** | Nessuna fatica | | | | Molto stanco |
| | O | O | O | O | O |

| | | | | | |
|---|---|---|---|---|---|
| **Qualità del sonno** | Ottima | | | | Pessima |
| | O | O | O | O | O |

| | | | | | |
|---|---|---|---|---|---|
| **Indolenzimento muscolare** | Nessuno | | | | Molto indolenzito |
| | O | O | O | O | O |

| | | | | | |
|---|---|---|---|---|---|
| **Stress** | Molto rilassato | | | | Molto stressato |
| | O | O | O | O | O |

| | | | | | |
|---|---|---|---|---|---|
| **Stato d'animo** | Molto positivo | | | | Molto negativo |
| | O | O | O | O | O |