4-9-2019

Establishing an Injury Determinant Framework in NCAA Division I Soccer

Ryan M. Curtis
University of Connecticut - Storrs, ryan.curtis@uconn.edu

Follow this and additional works at: https://opencommons.uconn.edu/dissertations

Recommended Citation
https://opencommons.uconn.edu/dissertations/2121
The purpose of this work was to 1) examine injury risk, rates and physical and psychological wellbeing; 2) identify risk factors for injury; 3) investigate mechanistic pathways for changes in perceived fatigue and 4) investigate the ability of supervised machine learning techniques to predict injury in women and men’s student-athletes competing in national collegiate athletics association (NCAA) division I soccer.

Injuries, workload, psychological well-being, sleep characteristics and physical activity disablement was longitudinally assessed for 256 athletes from 12 separate NCAA division I teams. Absolute injury risk and injury rates were calculated. Multi-level models were used to 1) assess differences in sleep and wellness inventories 2) identify injury risk factors, and 3) investigate causal pathways (moderators and mediators) of perceived fatigue. Supervised learning techniques were used to predict subsequent injury and area under the receiver operator characteristics curve (AUC) was used to evaluate model performance. Women’s collegiate soccer players experienced 2.05 (95%CI 1.20-3.51, p<0.001) times the rates of overuse injury, higher levels of global sleep dysfunction (b=0.99, p<0.001, ES=0.52), sports-related anxiety (b=3.9, p<0.001, ES=0.67), physical activity disablement (b=8.5, p<0.001, ES=0.87) and 38% less non-contact time-loss injury rates when compared with men’s soccer (IRR: 0.62, 95%CI 0.39-0.98, p=0.03). Relative workloads, chronic workloads, workload monotony, season type, session type, days relative to a match, session congestion, days off, weekly sleep latency and weekly sleep quality were identified as risk factors of non-contact injury. Men’s soccer players responded with higher fatigue levels when sleep duration, sleep quality, and chronic workload were low relative to baseline and when relative workload and perceived stress were higher compared to baseline. Logistic regression (AUC[95%CI]: 0.74[0.62-0.87]) and naïve bayes (AUC[95%CI]: 0.73[0.61-0.87]) performed equally as well as more
complex algorithms such as a support vector machine (radial basis) (AUC[95%CI]: 0.74[0.62-0.86]) and random forests (AUC[95%CI]: 0.74[0.62-0.85]). Prediction ability was improved with non-contact muscle strain injuries when compared with all non-contact injuries. Multi-team prospective cohort studies involving workload, wellness and sleep monitoring allow for the modeling of multiple injury risk factors in sport. Developing a multi-factorial view of determinants is vital for context when trying to understand complex phenomena such as injury.
Establishing an Injury Determinant Framework in NCAA Division I Soccer

Ryan Matthew Curtis

B.S., Culver-Stockton College, 2006
M.S., Kansas State University, 2009

A Dissertation
Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy
at the
University of Connecticut

2019
Copyright by

Ryan Matthew Curtis

2019
Acknowledgments

To my doctoral advisor, Dr. Douglas Casa:

Thank you very much for this opportunity and for the mentorship you have provided during my time with KSI and UCONN. You have been an outstanding role model and incredibly gracious with sharing your knowledge and time. Additionally, thank you for bringing on the large-scale project that made this dissertation possible and for all your efforts in bringing research that supports your doctoral students interests. Your energy, passion and giving-spirit is amazing. Thank you.

To my doctoral committee, Dr. Craig Denegar, Dr. Lindsey Lepley, Dr. Robert Huggins, Mr. John Wilson, Dr. Tania Huedo-Medina:

Many thanks for sharing your expertise and perspective throughout this process. You have all been a tremendous resource for which I am very grateful for. Dr. Denegar, your commitment to revealing real truths and meaning through research is inspiring. Thank you for always steering me towards practicality in my work and for making data analysis and statistics fun and applicable for the exercise/sports scientist and practitioner. Dr. Lepley, your expertise and support throughout my doctoral work has been a great blessing. Thank you for being so giving with you time and for helping me to focus my time and energy on work that matters. Dr. Huggins, you have been a great resource and confidant during my time at UCONN and KSI. Thank you for supporting my work and for sharing your knowledge freely. John W., thank you for introducing me and welcoming me into the world of analytics. You have been incredibly supportive of my efforts to grow in this area and I’m very grateful for your willingness to assist me. Dr. Tania Huedo-Medina, I’m very appreciative of the statistical assistance you provided for this work. It has been a great advantage having you on my committee and am so thankful for your guidance and time.
To my current and past colleagues:

Thank you to all my past and present KSI and UCONN HPL graduate colleagues. It has been my pleasure to journey down this path with you. Special thank you to Courteney Benjamin and Yasuki Sekiguchi for your dedication and service to the NCAA project and for being great friends. Additionally, special thank you to Dr. Dave Looney and Dr. Yuri Hosokawa for having a positive impact on me both personally and professionally. It was a great advantage to learn from you and my sincerest thanks for modeling discipline towards your craft and kindness to others.

To my family:

Thank you to my parents Greg and Theresa Curtis for providing incredible support throughout my life. You have always encouraged me to be the best that I can be, regardless of the situation. I am truly lucky to have such loving and supportive parents and am certain this would not have been possible without your dedication to my development. To my sisters Lauren and Stephanie, thank you for being so supportive. I am blessed to have you as younger sisters and proud of the women you have grown to be. Thank you to my grandparents Jim and Almeda Curtis for being a guiding and positive light throughout my life. Your unwavering faith and love for family is extraordinary, thank you for being positive role models.

To my wife and children:

To my daughter Estella, being your daddy is the greatest gift I have been given. It’s impossible to describe how much joy you bring and how grateful I am to have had you with me during this process, thank you for making each day fun and exciting. Even now it’s apparent how perceptive and intelligent you are, I hope you pursue your passions unapologetically.

To my unborn daughter, before you grace us with your presence or even have a name your spirit shines bright. Waiting and preparing for you with great anticipation has helped me more than you will know. Thank you for bringing so much happiness, I love you.
Finally, to my wife Ashley, you are my love and my rock. Thank you for being by my side throughout this journey. Without your support and sacrifice, this would not have been possible. It’s amazing to think of everything we have accomplished and experienced these past 4 years, I look forward to our future…
# Table of Contents

Chapter 1: Review of Literature ........................................................................................................... 1

Injury Etiology in Sport ........................................................................................................................... 12

Fatigue in Sport ....................................................................................................................................... 21

Workload in Sport .................................................................................................................................... 36

Sleep in Sport ........................................................................................................................................... 43

Monitoring in Sport ................................................................................................................................. 48

Injury Analytics ....................................................................................................................................... 63

REFERENCES ........................................................................................................................................ 86

Chapter 2: Injury and Psychological Wellbeing in Women’s and Men’s NCAA Division I Soccer .......... 115

ABSTRACT .............................................................................................................................................. 115

INTRODUCTION ..................................................................................................................................... 116

METHODS ............................................................................................................................................. 117

RESULTS ............................................................................................................................................... 120

DISCUSSION ......................................................................................................................................... 122

CONCLUSION ......................................................................................................................................... 127

REFERENCES ......................................................................................................................................... 128

TABLES/FIGURES ................................................................................................................................. 133

Chapter 3: Risk Factors for Non-Contact Injury in NCAA Division I Soccer ........................................ 143

ABSTRACT .............................................................................................................................................. 143

INTRODUCTION ..................................................................................................................................... 144

METHODS ............................................................................................................................................. 145

RESULTS ............................................................................................................................................... 150

DISCUSSION ......................................................................................................................................... 152
List of Figures

Figure 1 - The Incongruence of Stakeholder Interest ................................................................. 2

Figure 2 - The Injury Iceberg (Hanson 2005) .............................................................................. 4

Figure 3 - Reductionist vs Less Reductionist View of Injury Causality ......................................... 5

Figure 4 - Bittencourt et al. (2016) ............................................................................................... 6

Figure 5 - Meeuwisse (1994) .......................................................................................................... 12

Figure 6 - Bahr (2005) .................................................................................................................... 13

Figure 7 - Meeuwisse (2007) ........................................................................................................... 14

Figure 8 - Windt and Gabbett (2016) ........................................................................................... 15

Figure 9 - Biomechanical Model of Injury proposed by McIntosh 2005 .................................... 19

Figure 10 - Injury Prevention Research Sequence (Bahr and Krosshaug 2005) ......................... 19

Figure 11 - Taxonomy of Fatigue (Enoka 2016 – Originally Adapted from Kluger et al. 2013 ...... 25

Figure 12 - Wellbeing Continuum (Soligard 2017 [Adapted from Fry et al.1991]) ..................... 28

Figure 13 - Reilly, Drust and Clarke (2008) .................................................................................. 32

Figure 14 - Hans Selye's General Adaptation Syndrome Model ................................................. 38

Figure 15 - Acute:Chronic Ratio and Injury Risk (Gabbett 2016) ................................................. 39
Figure 16 - Stimulus-fatigue-recovery-adaptation Model .......................................................... 40

Figure 17 - Sleep Architecture .................................................................................................. 45

Figure 18 - Why Monitor Athletes? ......................................................................................... 50

Figure 20 - Association vs. Prediction (Adapted from McCall 2017) ........................................ 73

Figure 21 – ROC and Odds Ratio Relation (Pepe 2004) .......................................................... 74

Figure 23 - ML Hierarchy ......................................................................................................... 77

Figure 24 - Supervised vs. Unsupervised ML Models ............................................................... 79

Figure 25 - Moderation Model ................................................................................................. 81

Figure 26 - Mediation Model ................................................................................................... 82

List of Tables

Table 0-1 - Biomechanical Risk Factors (Adapted from Ashton-Miller 1999) ......................... 17

Table 0-2 - Relationship Between ARR, RRR and NNT ......................................................... 69

Table 0-3 - Exposures and Injuries Table .................................................................................. 70

Table 0-4 - Threats to Construct Validity .................................................................................. 82

Table 0-5 - Common Sport Science and Medical Calculations .................................................. 84

Table 0-6 - Injury and Exposure Definitions (Adapted from Fuller (2006)) ............................... 84
Chapter 1: Review of Literature

Introduction: Injury Causality and Complexity

Interests of key stakeholders (i.e., player, team, league and fan) are not always congruent (Figure 1). For instance, the player is interested in maximizing game performance and increasing worth, the team interested in improving winning percentage and increasing the value of its players, the league is concerned with generating more media revenue and enhancing team/owner value, while the fan wants continuous engagement and real-time information on their favorite team and players. However, there is one commonality that all stakeholders share. Reduction of injury. Injuries undermine team performance\(^1\)–\(^3\), pose a financial burden to various parties (e.g., players and organizations)\(^4\), and have the potential to threaten long-term athlete wellbeing.\(^5\) A reduction in injury incidence is likely to positively impact team performance, particularly since around 25% of athletes on one team may be injured at any one time.\(^6\) This association has been confirmed through research by Podlog et al.\(^7\) and Raysmith et al.\(^8\) on NBA and elite Australian track and field athlete-injuries, respectively, showing injury incidence is associated with competition wins. While introducing efficacious injury prevention and load management programs, which consequently impacting injury incidence rates, is a primary objective for key stakeholder, establishing accurate injury causation is a prerequisite.
Establishing Causality Framework

As mentioned, prior to establishing effective injury prevention strategies, injury causality must be understood. However, one’s perception of an injury is often limited to the event directly preceding the injury. Hulme and Finch (2005) suggest individuals develop tendencies towards “moncausality” in their daily thinking, which is typically directed by one’s experiences or habituation. In essence, individuals become conditioned to attribute isolated mechanisms to an outcome, usually the most noticeable event prior to an injury. However, it’s necessary to expand one’s thinking beyond basic and isolated cause-effect thinking in order to develop a deeper level of understanding of complex and multifactorial phenomena. This is an important concept for future sports injury research as establishing accurate causality requires an understanding of unique precipitating factors and mechanisms behind sports injury, and more importantly, the interrelatedness of injury determinants. Existing injury prevention frameworks (e.g., Translating Research into Injury Prevention Practice [TRIPP] and Sequence of Prevention Model) detail that execution of injury prevention interventions should not occur until risk factors and causal mechanisms of injury have been established.

Figure 1 - The Incongruence of Stakeholder Interest
Current Sport Injury Framework

Current injury etiology models follow predominately a biomedical archetype\textsuperscript{10}, meaning current frameworks are either biophysiological\textsuperscript{14-17} or biomechanical\textsuperscript{18,19} in nature. These models have origins in the individualistic paradigm established by early ‘medical models’ for disease.\textsuperscript{20,21} Due to the biological landscape of contemporary approaches, injury causation is narrowed down to the individual or intrapersonal level (i.e., behavior, biological, psychological). Subsequently, these approaches have reduced risk factors down to individual components such as; physical fitness, skill level, anthropometrics.\textsuperscript{14} This thought pattern suggests injury risk can be modified by educational, behavior or medically-oriented intervention.\textsuperscript{10} While this approach is useful in establishing modifiable risk-oriented prevention programs, assuming athletes always act independently and freely choose their own behavior constricts injury causality awareness. Perhaps this concept is best illustrated by Hanson’s Injury Prevention Iceberg (Figure 2).\textsuperscript{22} This conceptual image of the complex actuality of injury causality proposes the individual is only the noticeable ‘tip’ of an otherwise multifaceted metaphorical iceberg. Beyond interpersonal factors, Hanson et al. (2005) maintains latent and unobserved factors or factors deemed ‘below the waterline’ are acting on the individual’s injury risk profile in a complex and dynamical fashion.\textsuperscript{22} He defines multiple levels of causality as intrapersonal (i.e., behavior, biological, psychological), interpersonal (i.e., family and friends), organizational (i.e., school or occupational affiliations), community (i.e., social class) and society (i.e., infrastructure, education, government). Therefore, injury etiology research and subsequent prevention implementation would benefit from acknowledging the complexity inherent in injury causation.
Figure 2 - The Injury Iceberg (Hanson 2005)

The Problem with Reductionism

The idea of complexity is an important concept for making sense of behaviors which prove difficult to control or predict, such as the economy\textsuperscript{23}, weather\textsuperscript{24}, any living organism, or just about any assembly of people (e.g., family, organization or sports team).\textsuperscript{25} As alluded to previously, mostly reductionist approaches have been taken thus far in attempt to understand sports injury causality.\textsuperscript{26} This has entailed reducing components associated with injury into their most basic parts (e.g., sleep duration, aerobic fitness level, prior injury status, etc.) and then constructing speculative inferences to explain how these parts interact.\textsuperscript{25,26} This line of thinking is described by Newton’s “clockwork universe” logic, where big problems are divided into small ones, then deciphered by rationale deduction.\textsuperscript{25} For example, take the relationships between increased sleep quality and injury (unpublished data), as well as, increased high-speed distance and injury.\textsuperscript{27} Taken as individual parts of a whole (i.e., injury risk profile), one might assume these factors are acting independently on injury risk. However, if a mediating factor such as game exposure (Figure 3), which is explanatory of their interaction and association with injury is not accounted for (e.g., players increase sleep prior to a game therefore increasing sleep quality, players are exposed to
more total distance in games than training), complex causal mechanisms which are more explanatory of the phenomena may not be captured.

Figure 3 - Reductionist vs Less Reductionist View of Injury Causality

Although a number of other sports injury predictors have been revealed though univariate analysis such as, spikes in workload\textsuperscript{28–30}, aerobic capacity\textsuperscript{31}, and sleep quantity\textsuperscript{32}, very little work has been done to understand the complex interrelatedness of isolated injury risk predictors. Further, many of these precipitating factors have not been demonstrated as consistent predictors of injury across literature.\textsuperscript{33,34} This is likely due to a concept introduced by Meeuwisse et al. in 2007\textsuperscript{15}, in which susceptibility to injury is dynamic and responsive to recurrent exposures. Simply, injury risk is always changing because the state of the human body is always changing (i.e., positive or negative adaptation). Unfortunately, most techniques utilized to investigate injury risk factors have been linear and have investigated isolated predictors. Although narrative and an important step forward in injury research, previous models have failed to account for the interrelatedness of risk factors associated with injury.\textsuperscript{26,35} The multifactorial nature of phenomena such as injury may be better understood under Phlippe and Mansi’s framework\textsuperscript{11}, which is referred to as the ‘web of determinants’.\textsuperscript{26} This concept was introduced to
sport injury research by Bittencourt et al. (2016) in a narrative review discussing complex systems approach for injuries. They proposed, along with others, that to fully reveal the intricate landscape of sports injury etiology (Figure 4), a complex systems thinking was needed.

![Complex Systems Thinking and Chaos Theory](image)

**Figure 4 - Bittencourt et al. (2016)**

*Complex Systems Thinking and Chaos Theory*

While complex system thinking exists in many other research fields, only recently has this concept been introduced to biological and medical disease epidemiology. Complexity thinking stems from the fields of information and systems theory, as well as, cybernetics. As lengthily defined by Bittencourt et al., complex systems are “dynamic, open systems with inherent non-linearity due to the existence of recursive loops and complex interactions among units, which spontaneously organize themselves to generate emerging properties than cannot be deduced solely from their original properties (self-organization).” However, von Bertalanfly offers a simpler definition, with complex systems described as a “whole with units (parts) that interact with each other.” Rosen further expands that systems are complex because units are modulated by the interaction between other units, which sometimes result in
the unpredictable emergence of phenomena. In applying complex thinking concepts to sports injuries, the emergence of an athletic injury is complex, with the relationship between biomechanical, behavioral, physiological and psychological factors of an athlete (i.e., units) ultimately dictating his/her collective behavior.\textsuperscript{36} Seeking to understand how the athlete interacts and establishes relationships (regularities) with their environment provides a research focus for how complex phenomena such as injury occur. The inherent properties and laws which govern complex systems are described in greater detail.

\textit{Non-Linearity}

There are two principles which govern traditional linear systems, proportionality and superposition.\textsuperscript{43} Proportionality refers to an output which is directly relative to its input, while superposition is terminology to describe how the output of a linear system can be fully understood by its deconstruction into individual components and that the behavior of that system is the summation of its individual parts.\textsuperscript{11,43,44} However, complex systems do not shadow traditionally linear relationships where outputs are equivalent to individual parts or units, but are instead characterized by non-linear relationships where outputs are not proportional to inputs.\textsuperscript{37,44} Because relationships between individual parts are not proportional, small changes in one unit can have dramatic and unanticipated effects on the system.\textsuperscript{43} This phenomena is typically described as the “butterfly effect” and explained by chaos theory.

\textit{Chaos}

One can imagine rolling a snowball down a hill, where the initial input (small snowball) produces a much larger output (giant snowball). Simple cases such as this illustrate a phenomena whereby input-output relationships are exponential in nature. This effect was initially observed by Edward Lorenz in 1961 when producing mathematical models to predict weather patterns.\textsuperscript{24} Logically, Lorenz assumed that a small variation at the start of a calculation would produce a small deviation in the result, with the magnitude of difference directly proportional to the initial difference. In Lorenz’s prediction efforts, his computer program running the mathematical model truncated the initial 6-digit values down to three. Instead of the
results being slightly off from the previous 6-digit model, dramatic change in the prediction was displayed. Indeed, this result demonstrated that in a non-linear system, such as weather or the human system, differences in iterative functions can amplify differences in an exponential manner. Although first observed by Lorenz, it was Philip Merilees who organized a meteorological conference in 1972 and self-selected the title of Lorenz’s talk: ‘Predictability: does the flap of a butterfly’s wing in Brazil set off a tornado in Texas?’, which popularized the term “butterfly effect”.

**Self-Organization and Regularization**

Emergence is the term given for a particular output, property, or behavior of a system which has resulted from non-linear interactions among individual parts. When emergence is discussed in a sporting context, athletic injury or adaptation is the explained phenomena. Organization of a complex system is determined by the interaction and cooperation of individual parts within the system, which self-organize and operate within no particular structure other than staying within the confines of universal physical law. Therefore, an emergent phenomena is not proportional to individual part behavior, but resultant of patterns developed through self-organization of the system. Patterns or regularities can be seen when unit thresholds are attained and specific configurations of the system produce and emergent condition. This is why commonalities in injury causation are seen. While there are numerous precipitating factors which manifest into an emergence, there are often noticeable regularities within a system which precede it.

**Feedback Loops**

A complex system is by definition dynamic and one that evolves over time. This happens because complex systems have recurrent feedback loops in which output becomes the new input to the system. Ultimately, a systems output will influence future input into the system, subsequently altering the systems state. Meeuwisse et al. accounted for this characteristic in their most updated model of injury etiology. When an athlete is exposed to an event (e.g., training or match) they are either injured or not. Either way,
the output or work performed by the individual leads to an adaptation of the human body, altering the
individual and therefore intrinsic injury risk factors. For example, if an injury occurs, soft-tissue
restriction and altered joint mechanics can develop. If an injury does not occur, the athlete may be
transiently fatigue, however if loads are gradually increased over time, aerobic fitness and mechanical
load toleration can be improved. As noted by Bittencourt et al., acknowledging the existence of
recursive loops highlights that after an injury incident, a system may alter in an unpredictable way. Previous states of intrinsic risk factors (predictors) are changed and may no longer share the same
relationship with an emergence such as future injury.

**Uncertainty**

Living organisms (e.g., athletes) are open systems, meaning they interact with the external environment. That is, open systems exchange matter and energy with the environment without losing their identity. The fact that human cells interact with the environment but maintain dynamic equilibrium or homeostasis was first acknowledged by American physiologist Walter Cannon in his 1932 book *The Wisdom of the Body*. As described by Bittencourt and colleagues, the concept of equifinality (i.e. many diverse ways in which the same outcome can emerge) can help to explain sports injury, where various relationships between risk factors produce the same outcome (e.g., injury). Because athletes are open systems which fully interact with their environment and adapt over time, multiple pathways to the same emergence exist, which means cause and effect relationships can never be fully modeled. This concept is intellectualized by the George Box’s statistical aphorism, “All models are wrong; but some are useful”. However, complex systems have inherent regularities which are biological and social in nature (i.e., sleep cycles, circadian rhythms, variation in heart rate, workload norms for session types, training session structure) which relate to an emergence of phenomena such as injury. Establishing these regularities, or the interactions among risk factors, allows uncertainty in a model to be reduced. By establishing regularities and interaction (i.e., magnitude of mediation and moderation; dose-response relationships) among risk factors (i.e., fatigue, sleep architecture, neuromuscular control, workload, aerobic fitness) a more accurate prediction
of an emergence can occur (e.g., sports injury). Therefore, in establishing injury causality and establishing prevention programs, we should seek to understand interactions among determinants (i.e., ‘web of determinants’) rather than the determinants themselves.

**Future Directions**

Complex systems thinking has begun to seep into sports research\(^{49}\), however several inherent methodological implications and analytical barriers exist. The fundamental assumptions that are generally used in more orthodox statistical techniques are dissociated from complex systems analysis.\(^{10,50}\) For example, regression-based techniques are unable to account for system-wide occurrences resultant of adaptive feedback loops or effects which are time-distant form an injury emergence.\(^{51}\) However, complex systems approaches should not be viewed as a replacement for scientific reductionism or linear modeling, but rather as a supplementary method which may include traditional statistical approaches.\(^{10,52}\) As more modern systems-based analytical methods emerge such as System Dynamics\(^{52}\) and Agent Based Modeling\(^{53}\), computational system science may enhance current analytical frameworks. Acknowledged by Bittencourt et al. (2016), statistical learning techniques such as artificial neural networks (ANN) and classification and regression trees (CART) may be useful in uncovering non-linear interactivity.\(^{26}\) Indeed, these techniques have been used in the sports performance and injury arena successfully, as Pfeiffer and Hohmann found they could better predict talent development by non-linear (i.e., ANN) rather than linear methods (i.e., linear discriminant analysis).\(^{49}\) Additionally, Bittencourt et al. utilized recursive-partitioning CART techniques, which factor non-linear interactions among predictors, to predict knee valgus during landing following vertical jump.\(^{54}\)

Ultimately, for injury causality to be truly understood, it must be recognized that the emergence of injury is the result of complex interdependent processes and not isolated events within the human system.\(^{55}\) By investigating and accounting for more upstream activity, rather than the typical proximal mechanism, understanding of leverage points within complex systems may be strengthened, in turn progressing
preventative practices. Utilization of more modern statistical learning techniques may strengthen current frameworks in which injury causality is understood.
Injury Etiology in Sport

Biophysical Model of Injury Etiology

In efforts to better understand the phenomena of injury, models detailing causal factors have been proposed. Generally, these models portray the progression of the athlete from predisposition to susceptibility to injury. The foundation for our understanding of injury etiology was laid by the works of Meeuwisse et al. in 1994, who were the first to develop a conceptual framework describing sports injury causality. As seen in his initial model (Figure 5), prior to injury occurrence, the athlete is assumed to be predisposed in some intrinsic (internal) manner, presumably related to factors such as age, flexibility, and somatotype. Upon exposure to an external risk factor the athlete transitions from predisposition to susceptibility, in which injury is possible if an inciting event is experienced.

Figure 5 - Meeuwisse (1994)

Following up on Meeuwisse’s initial works, Bahr and colleagues in 2005 expanded upon the comprehensive model of injury causation (Figure X) by detailing examples of intrinsic (internal) and extrinsic (external) risk factors associated with injury risk. Secondly, they provided greater detail surrounding potential factors associated with the inciting event which could ultimately lead to injury. It was supposed that injury was the product of athlete susceptibility and an inciting event dictated by confounding factors such as playing situation, player characteristics and behavior, gross (whole body) and
A primary addition of the model and of value to the injury risk discussion was the need for complete description of both epidemiological and biomechanical factors in a comprehensive injury causation model. Working from a biomechanics-dominated model put forth by McIntosh et al.\(^\text{18}\) and the epidemiological model proposed by Meeuwise et al.\(^\text{14}\), Bahr and colleagues successfully merged the concepts into a more holistic model of injury causation.\(^\text{17}\)

![Figure 6 - Bahr (2005)](image)

In 2007, Meeuwise et al. proposed an update of his original injury causation model (Figure X) which acknowledged the ‘dynamic’ and ‘recursive’ nature of injury risk.\(^\text{15}\) Prior to this proposition, associations between precipitating factors and injury risk were largely considered linear, meaning injury risk was directly proportion to changes in intrinsic factors. Additionally, previous models didn’t account for the fact that injury may or may not occur, either way, injury risk would not be the same following an exposure. Meeuwise’s model emphasized injury risk is constantly being altered (dynamic) due to repeated exposure (recursive), and that if we are to truly understand injury causation we must look beyond risk factors directly preceding an injury and account for the consistent adaptation within the human system.\(^\text{15}\)

This has a number of important implications on the methodology and analysis strategies taken in investigating injury causation. Specifically, researchers are encouraged by this work to acknowledge that
sport exposure not only predisposes an athlete to injury, but additionally alters their injury risk profile for subsequent exposure.¹⁶

Figure 7 - Meeuwise (2007)

The most current update and proposed alteration to the injury-etiology model was offered recently (2016) by Windt and Gabbett.¹⁶ Their work acknowledges the key additions put forth by those who pioneered injury etiology research, however suggest that perhaps the most critical predisposing factor associated with exposure, workload during training and competition, were not properly credited in previous models. In light of workloads not being unequivocally mentioned as a risk factor for injury in previous models and due to their strong association with injury¹,²⁹,₅₈,₅⁹, Windt and Gabbett advise that their inclusion is essential for sports injury comprehension.
Biomechanical Model of Injury Etiology

As discussed, injury occurs from the combination of 3 factors, internal risk, external risk and an inciting event. Typically, in medical literature, the ‘inciting event’ is known as the injury mechanism. A mechanism of injury is therefore, “the fundamental physical process response for a given result”. One of the simplest explanations of a physical injury mechanisms is a transfer of energy to human tissue in excess of what it can tolerate. However, injury has also been likened to mechanical failure and described as “equivalent to the failure of a machine or structure”.

Tissue injury and dysfunction can result from excessive stress (i.e., force per unit area) and/or strain (i.e., the relative elongation of a given length of tissue) and that can result in inflammation, degeneration or disruptive changes. There are multiple biomechanical-specific etiologies of tissue injury, which will be described. Excessive stress or strain can result from an isolated mechanistic event such as planting, cutting or jumping. Single event, stretch-related injuries in tissue are caused by a mechanical-mediated event rather than a chemical or metabolic response to load. Generally, this involves a single rapid stretch to actively contracting muscle or a series of repetitive high-speed contractions.

Excessive
tissue strain can also result from an interaction with the environment such as landing from a jump or colliding with an object (i.e., goalie colliding with goal post). Perhaps the most modifiable injury mechanism is accumulated strain associated with lower-force, but repetitive, loading of tissue. Injury from these loads is resultant of training management error. Finally, excessive tissue strain can be caused by some combination of the two aforementioned events (i.e., mechanical event and interaction with environment) which are joined with a history of repetitive loading.\textsuperscript{67}

Tissue failure, as discussed, is influenced by both intrinsic and extrinsic factors, which are described in Table 1. Biomechanical and structural properties which are intrinsic such as tissue anatomy, physiology, state and functional patterns influence how the body reacts to a given physical load (Table 1).\textsuperscript{62} Extrinsic factors affecting tissue include variables such as magnitude, direction, duration, frequency and density of loading. A conceptual model proposed by Armstrong et al.\textsuperscript{68} has described biomechanical mechanisms of musculoskeletal disorders as it relates to work injuries, but this can be extrapolated to sports injury. They propose a dose-response model including concepts of exposure, dose, capacity and response which may provide a useful framework in which to understand injury. Firstly, exposure refers to the external factors (e.g., physical demands of sport) that produce the internal dose (e.g., tissue loads and metabolic demand). Exposure can be modulated by external factors such as dimensional aspects of play (e.g., small space [small sided games] = reduced space and increases mechanical loading frequency [more accelerations and deceleration] vs. large space = more high-speed distance/higher velocity loading [more extended runs and longer distance sprints]). Exposure can also be influenced by factors such as the environment (e.g., ambient temperature and altitude) or cofounding factors (e.g., coaches drill selection). Dose refers to features which can disturb the internal state of the athlete, which may be mechanical (e.g., muscular contraction), physiological (e.g., accumulation of metabolites), or psychological (e.g., anxiety).\textsuperscript{68} Response refers to the changes in the state of the individual (e.g., muscle temperature increase and accumulation of metabolites). However, the relation between dose and response is not so clear cut as a response can turn into a new dose, which produces another response. For example, repeated muscular
contractions (dose) produce accumulation of metabolites (response and then subsequent dose), which in turn produces discomfort (response). Responses that are a product of another response are referred to as secondary responses. To add another layer of complexity, the effect of a dose can occur immediately or it can require more extended periods of time to manifest. This concept has even been supported beyond the tissue level by Hulin et al. and Orchard et al. who found increased injury risk in a latent period following rapid workload increases.

A single force can result in an immediate deformation of tissue, repeated forces over a single session may lead to viscous deformation of tissue, and repeated forces of several sessions may result changes in the composition of tissue. The consequence of the aforementioned forces can lead to either desirable or undesirable effects. Specifically, positive tissue adaption can occur which increases dose tolerance or the changes can reduce tissue capacity (e.g., tissue restriction). Capacity refers to the ability of the tissue to resist deterioration or damage due to various doses. Physical capacities at the tissue level include resisting tissue degradation or excessive accumulation of metabolites.

Table 0-1 - Biomechanical Risk Factors (Adapted from Ashton-Miller 1999)

<table>
<thead>
<tr>
<th>Intrinsic Risk Factors</th>
<th>Extrinsic Risk Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tissue Anatomy - heritable factors, changes due to previous tissue injuries</td>
<td>Magnitude, direction, duration, rate, and lack of variability of workload-related external forces affecting tissue stress/strain history</td>
</tr>
<tr>
<td>Tissue Physiology - healing/remodeling potential; response to chronic loading</td>
<td>Frequency and density of loading</td>
</tr>
<tr>
<td>Tissue State - state of hypertrophy/atrophy/remodeling as it affects relevant tissue physical capacities</td>
<td>Postural regularities</td>
</tr>
</tbody>
</table>
As mentioned, if tissue is stressed beyond its load bearing capacity (a.k.a, ‘envelope of function’)\textsuperscript{70} or insufficient recovery is allowed between load cycles\textsuperscript{71}, micro-damage or injury can ensue.\textsuperscript{72} In an article discussing injury causality, McIntosh recognizes that previous biomechanically oriented injury prevention strategies attempt focus on modulating external and internal loads applied to the human system. These ‘neat’ models deduce that preventative interventions should be focused therefore in two distinct ways, 1) reducing physical loads to below injury tolerance levels or 2) increasing the body’s physical capacity to tolerate load (i.e., improving fitness).\textsuperscript{18} However, McIntosh’s mode (Figure 9) acknowledges the multifactorial nature of injury, while still keeping the a biomechanical focus on tissue properties and injury at its heart. McIntosh\textsuperscript{18} recognizes many inputs such as; behavior attitude, training, skills, equipment, coaching, other competitors and their environment which impact upon injury risk. However, all these aforementioned factors still feed into a biomechanically-mediated injury causality.
Establishing Injury Prevention Strategies

According to van Mechelen et al.\textsuperscript{13}, identifying efficacious injury interventions requires a sequential 4 step process (Figure 10). Prior to establishing preventative practices, the practitioner should first be aware of the magnitude of the problem. This is typically expressed as absolute injury risk (e.g., 61% of athletes sustain an injury on average) and rate statistics (7.6 injuries per 1000 athlete exposure-hours). Regardless of expression, defining injury magnitude should be in terms of incidence and severity so that injury prevention resources can be focused. The second step is deciphering injury risk factor and causality so that more targeted prevention can follow. The next step includes introducing practices that are likely to reduce risk of injury, which are based on established causative factors in the second step. Finally, the effect of the intervention must be assessed, which in the field usually involves comparing injury rate statistics from year to year. Although impractical in a high-performance athletic setting, it should be acknowledge that randomized control trials are preferable in assessing the efficacy of prevention programs.\textsuperscript{73}
**Injury in Soccer**

Soccer, being the world’s most popular sport, has received considerable attention with regard to the identification of risk factors associated with injury. A systematic review and meta-analysis by Silva et al. indicate several injury-related factors associated with soccer. Research on soccer injury prevalence has revealed between 65% and 91% of elite soccer players are likely to experience an injury throughout the course of a competitive soccer season, with 90% of all muscle injuries sustained localized to the lower limbs. In a study auditing professional soccer injuries, Hawkins et al. found that approximately 1.3 injuries per player season occur with 78% of injuries leading to at least one match missed. It’s also demonstrated that injury rate is significantly higher for matches compared with training. Interestingly, while overall injuries rates have stabilized since 2001, training-related hamstring injury rates have increased substantially. Silva et al. propose this may be resultant of increased match demands causing extended periods of residual fatigue or training prescription error. Relations between various risk factors and injury will be discussed in subsequent chapter.
Fatigue in Sport

Introduction

A reduction in fatigue and increase in vigor is essential for optimizing human performance and realizing human potential. In a model proposed by Banister et al, performance at any given time can be estimated by the difference between negative (fatigue) and positive (fitness) physiological responses to training. While conversations often focus around its reduction, fatigue is frequently a desired outcome of many training programs with the intention of progressively overloading an athlete and stimulating adaptation. Fatigue is accepted as a necessary means to maximizing one’s potential, being a natural and normal byproduct of the body’s response to an overload stimulus and a driver of physiological adaptation.

Fatigue is often studied as either a mediator of injury/illness/maladaptation or response to some activity, however it’s definition often differs by the field discipline in which its being investigated. Physiologists may describe fatigue as a reduction in muscle force or power in response to an acute bout of exercise, psychologists might label fatigue as symptoms of tiredness or weakness, while an exercise scientist might describe it as an exercise-induced reduction in performance. In a sporting context, a coach might infer fatigue when her players make poor tactical decisions or a sport scientist may determine fatigue as the cause of a rapid decline in high-speed distance covered toward the end of competition. While many definitions of fatigue may be occurring when considered in each context, these isolated views often reduce fatigue causality down to an isolated factor, rather than acknowledging its multifactorial nature. Nevertheless, fatigue is generally associated with a diminishment of some aspect of physical and/or cognitive function. In an athletic context, fatigue has been described as the decrease in the pre-match/baseline psychological and physiological function of the athlete.

Fatigue as a Response to Stress

Deviations in factors contributing to fatigue are resultant of stress. Stress is broadly considered a disturbance of the body’s homeostatic state; therefore, a stressor is considered any influencer that
perturbs homeostatic state.\textsuperscript{48} French scientist Claude Bernard was the first to acknowledge that the human body operates independently of its external environment by preserving stability within its cells.\textsuperscript{48} Further work by Harvard physiologist Walter Cannon expanded this understanding by supporting the notion of a dynamic equilibrium within cells.\textsuperscript{48} He dismissed the view of complete constancy within the cell, instead suggesting a dynamic responsiveness by cells to perturbing stimuli. Cannon called this dynamic equilibrium \textit{homeostasis}, which is terminology accepted today for describing the body’s ability to maintain equanimity in the wake of perturbing stimuli.

When the body does experience stimuli which disturbs its internal equilibrium, the neural, endocrine and immune systems are all affected, with their responses being largely interrelated and coordinated.\textsuperscript{102} The central nervous system (i.e., brain and spinal nerves) senses homeostatic disturbance and responds with a cascade of hormonal activation, traditionally discussed as a \textit{hormonal axis}.\textsuperscript{48} There are two primary hormonal axes which react to stressful situation or environment; the sympathetic-adrenal-medullary axis (SAM) and hypothalamic-pituitary axis (HPA). These axes are initiated by the adrenal medulla and adrenal cortex, respectively. When the body enters a “fight or flight” mode or is stimulated in a sporting context, the SAM axis is galvanized by the sympathetic nerve branch and catecholamines are released (i.e., norepinephrine) from the adrenal medulla. Blood is diverted from internal organs to working muscle and cardiac output increases to compensate. To accommodate for the increased work demands and therefore metabolic fuel requirements, glucose and free fatty acid is mobilized and cellular metabolism stimulated.\textsuperscript{48} Meanwhile, a group of hormones known as corticosteroids is released by the adrenal cortex, most importantly cortisol, which stimulates conversion of non-carbohydrate sources to glucose (i.e., gluconeogenesis), reduces inflammation and suppresses the immune system.\textsuperscript{103} If the body incurs physical damage or injury, high cortisol serves the purpose of restraining the initial inflammatory and immune response, thereby blunting permanent damage.\textsuperscript{104}

Hormonal cascades from both axes are complimentary and aid in preparing the body for the physical or perceived stress. In addition to function of the HPA and SAM axes, it must be acknowledged
that other anabolic and catabolic hormones aid in physiologically preparing and restoring the body. Specifically, catabolic hormones such as growth hormone (in combination with norepinephrine, epinephrine and cortisol) act to mobilize energy stores. However, as catabolic hormones increase blood concentration, anabolic hormones are secreted to counterbalance. Anabolic hormones such as insulin, testosterone and estrogen help to build depleted energy stores and rebuild muscle tissue (i.e., protein synthesis).

**Fatigue Mechanisms**

Fatigue originates from different levels of the motor pathway and a key focus of mechanistic studies of fatigue have been on identifying physiological rate-limiting adjustments. Ultimately, production of force by the musculoskeletal system is reliant on contractile mechanisms, so any failure upstream of cross-bridge site can influence fatigability. Failure could be resultant in changes in nervous, ion, vascular and energy systems, or buildup/depletion of metabolic factors and fatigue mechanisms. Mechanistic research has led to a commonly recognize dichotomy of fatigue etiology, peripheral and central fatigue. Specifically, peripheral fatigue refers to force diminution resultant of changes distal to the neuromuscular junction, while central fatigue refers to decreases in neural drive to the muscle, originating from the central nervous system.

Recognized by Enoka and Stuart (1992), fatigue research has followed what is called ‘Mosso’s dichotomy’, in which force attenuation is examined separately or considered distinct from sensations (perceptions) of fatigue. However, this approach is considered a limitation as it’s not possible to differentiate force declines from sensations of fatigue. Sensory feedback (i.e., Group III and IV muscle afferents) can influence neural output and contribute to perception of pain and fatigue, thereby influencing force output.

According to a more recent description by Enoka et al., fatigue is a holistic and multi-factorial phenomena affecting various aspects of physical and cognitive function. Further, Enoka et al. acknowledges the
duality of fatigue manifestation (i.e., perceptual and performance), illustrated through an originally proposed Taxonomy of Fatigue (Figure 11) by Kluger et al.\textsuperscript{99,114} A distinctive feature of this model is the proposed inclusion of psychological factors contributing to fatigue. This awareness has led to revolutionized thinking in the determinants of fatigue, expanding thinking beyond mere physiology. The impact is shown in a recent article (2018) published by Smith et al. who was among the first to review the effects of psychological (mental) fatigue on a specific sport (soccer).\textsuperscript{115} According to Kluger et al. the concept of fatigue should acknowledge two separate attributes; 1) performance fatigability – decline in an objective measure of performance over a discrete period of time, and 2) perceived fatigability – changes in the sensations that regulate the integrity of the performer (Figure 11).\textsuperscript{99,114} The term fatigability is used to acknowledge relative individuality in fatigue response according to the demands of the task. Individuals inherently experience differing levels of fatigue for a given task, in addition to differing relative baselines at rest\textsuperscript{116,117} and rates of change in fatigue state during activity.\textsuperscript{118} As proposed, performance fatigability is reliant on contractile processes and adequate neural drive (i.e. muscle activation signal), while perceived fatigability is regulated by modulations in homeostasis and psychological state. Interestingly, perceived fatigability at rest is modulated by the state of homeostatic determinants such as internal body temperature, hydration, mood, arousal, while during activity, perceived fatigability responds to the rate of change in modulating factors of homeostasis and psychological state.\textsuperscript{99} These processes influence perceived fatigue and are utilized in governing the pace of activity.\textsuperscript{99} Performance fatigability is more a kin to a traditionalists central-peripheral dyad. For example, classic declines in voluntary muscle force (i.e., maximal voluntary contraction force [MVC]) resultant of low-intensity exercise are due to more central factors (i.e. activation signal), while declines in force from high-intensity exercise are due to diminished contractile function.\textsuperscript{112,119}

While the schema proposed by Kluger et al. recognized a dichotomy of perceived and performance fatigability, a key feature is the interactivity of the two domains.\textsuperscript{114} This is true of human voluntary action, which is governed by both physical and psychological aspects of fatigue, and should not be considered
separate but interconnected. The degree of fatigue to which an individual may experience depends on multiple modulating factors related to homeostatic and psychological perturbations, as well as, attenuated contractile and neural activation processes, as shown in Figure 11.99

Figure 11 - Taxonomy of Fatigue (Enoka 2016 – Originally Adapted from Kluger et al. 2013)

**Activity**

Rate-limiting adjustments in response to acute bouts of exercise are task-dependent.111 This has important implications for athletes as intensity, duration, frequency and density of activity differ by sport. Therefore, the mechanisms that cause fatigue will be inherently different by mode of exercise. Both high-intensity and prolonged-continuous exercise are considered. It should be noted that many team sports are intermittent-intensity in nature. Consequently, it intermittent-intensity sports involve fatiguing mechanisms inherent in both high-intensity and prolonged, continuous exercise.

**High-Intensity**

Rate-limiting adjustments to high-intensity exercise is fundamentally different than longer duration or sustained, low-intensity exercise. Rather than experiencing a diminished activation signal by the nervous system as seen in longer duration activity, limitations in high-intensity exercise are more often caused by
a reduction in contractile function.\textsuperscript{99,119} At its core, diminished function of the contractile properties typically has two broad causes; 1) accumulation of a substance or metabolite detrimental to contractile function or 2) depletion of a substance or metabolite necessary for contractile function. Short-term, high-intensity activity such as sprinting or heavy resistance training are limited by three primary causes of fatigue. Depletion of ATP and creatine phosphate (CP)\textsuperscript{120}, as well as, an increase in muscle acidity\textsuperscript{121}, resultant of an increase in hydrogen ions (H\textsuperscript{+}). ATP, essential for crossbridge cycling and force/power generation, can become depleted due to a mismatch of ATP utilization and regeneration. CP is the main energy source for ATP production during short-term, high-intensity exercise, therefore is reduction negatively impact ATP regeneration leading to fatigue and subsequent exhaustion. An increase in muscle acidity (metabolic acidosis) from an increase in hydrogen ions can interfere with Ca\textsuperscript{2+} role in crossbridge formation, therefore tension development in the muscle and force output is diminished. Hydrogen ion accumulation inhibits anaerobic glycolysis, which is a primary source of ATP generation during high-intensity exercise.

**Prolonged, Continuous**

In contrast to short-term, high-intensity activity, The primary causes of fatigue during lower intensity and longer term activity include 1) reductions in muscle and liver glycogen\textsuperscript{122-124}, 2) reduction in muscle Ca\textsuperscript{2+} \textsuperscript{107} and 3) an increase in body temperature. Glycogen is essential for replenishing ATP, therefore when exercise intensity and consequently ATP utilization increase, so does the demand for glycogen. When glycogen is not present in sufficient amounts to replenish ATP, intensity of activity must be reduced or discontinued.\textsuperscript{122} Both instances represent a fatigued state. During prolonged activity, the sarcoplasmic reticulum is repeatedly stimulated to release Ca\textsuperscript{2+} for cross-bridge formation and the development of muscular tension. However, a leaking of Ca\textsuperscript{2+} into extracellular fluid or uptake by the mitochondria during prolonged activity can result in a reduction in force and power output. Finally, an increase the temperature of the working muscles and body have been linked with fatigue.\textsuperscript{125} When internal body temperatures increase, a greater proportion of blood flow is diverted to the skin instead of the working muscles for
thermoregulation. Reductions in blood flow for working muscle can decrease delivery of oxygen and other substances critical for energy. A decrease in blood plasma volume (dehydration) can exacerbate fatigue as cardiac output, and therefore \( \text{O}_2 \) to working muscles is reduced.

**Fatigue Continuum**

An athlete’s physiological response to a given external or internal stimuli (e.g., physical activity, academic workload, perceived life-stress, etc.) complex and better understood as a state on a continuum (Figure 12) rather than a generalized or discrete classification.\(^{126, 127}\) Generally, if rest after a stimulus is not adequate, an athlete progresses down the continuum of fatigue, likely resulting in negative performance or health outcomes.\(^{127}\) Of importance, there is great variance both between and within individuals in their response to a given stimulus.\(^{128}\) Further, additional stresses beyond training, which are a natural part of the human experience, can impede the body’s ability to adapt positively to a given stimulus.\(^{129–132}\)
When athletes experience planned increases in frequency, density or intensity of fatiguing bouts of exercise, functional overreaching occurs. Functional overreaching is defined by a period of reduced performance due to a progressive overload stimulus. As mentioned previously, periods of functional overreaching are a normal part of the training process, with planned periods being during the offseason and preseason. Practitioners need to be aware of the negative impacts of performance during these periods and plan for adequate recovery following. In contrast to functional overreaching, non-functional overreaching occurs when the athlete experiences unplanned negative performance and fatigue accumulates in response to training overload. Non-functional overreaching can manifest not only as physical and performance decrements but clinical symptoms of chronic fatigue as well, which can take weeks to resolve. The final stage of the fatigue continuum is the overtraining syndrome which is characterized by performance decrements and psychological disturbances lasting from weeks to month, despite extended periods of rest.

**Fatigue and Injury**

Fatigue is an important risk factor for injury in sport. Fatigue secondary to training or competition load can cause damage at the tissue level or hamper decision-making ability, coordination and neuromuscular control. Reduced muscular force and contraction velocity subsequent to a session is a normal occurrence. As pointed out by Soligard et al., fatigue from training or completion has several deleterious effects including increased force on passive tissues, adversely altered kinetics, kinematics and neural feedback, and reduced joint stability. While fatigue is certainly linked to overuse injuries, such as those mentioned previously (i.e., bone stress, tendinopathy and patellofemoral pain), cumulative tissue fatigue due to repetitive loading has even been purported a mechanism for increased susceptibility to injurious events (i.e., inciting events) which are acute in nature, such as an anterior cruciate ligament tear. While this postulation needs further corroboration, the aforementioned factors contribute to increased risk of injury from an acute and/or residual fatigue effect.
Fatigue and Injury in Soccer

Increased injury incidence rates have been shown towards the end of halves (i.e., first and second) compared with the former stages in both professional and youth soccer, likely due to fatigue. Fatigue is inferred during this period of match-play as both total distance covered and total high-speed distance have shown to significantly decrease during the latter stages of halves. Evidence suggests locomotor efficiency is also diminished during the latter stages of each half, which offers supportive evidence of a fatigue-induced increase in injury risk. Soccer injury research also indicates injury risk is elevated during congestive competitive periods, suggesting a possible relationship between residual fatigue/under-recovery and injury. Interestingly, locomotor activities during match-play haven’t shown impairment during congested periods. Nevertheless, fatigue is recognized as a potent contributor to increased injury risk.

Fatigue Research in Soccer

There have been two articles which have extensively reviewed soccer-related fatigue manifestations and determinants, with several other reviews discussing applicable literature on the topics of fatigue and recovery in soccer. Additionally, a recent review by Smith et al. consider another key determinant, psychological factors and their association with fatigue in soccer. Evidence that fatigue occurs during a soccer match is substantial and the causal mechanisms underpinning fatigue are relatively well-understood. Generally, match fatigue has many potential causes, including dehydration, glycogen depletion, muscle damage, and mental fatigue.

At all levels of modern soccer, physical and psychological demand are increasing due to increases in the speed and intensity of play, as well as, the frequency of matches. Fatigue post-match presents in a variety of ways including decreased force production and physical abilities, accumulation of metabolites and alterations in psychological state, which typically linger for at least 2-3 days. Specifically, a meta-analysis and review by Silva et al. in 2017 showed soccer match-play results in acute
alterations in metabolic, biochemical, physical performance, technical and perceptual markers, with residual fatigue lasting around 72 hours for some markers. However, the time-course of recovery following soccer match-play is highly individual, with several intrinsic (i.e., aerobic capacity, neuromuscular strength and endurance, position) and extrinsic factors (e.g., level of opposition, tactical strategy, recovery day length) influencing loads which are sustained during a match.

**Work-rate**

There has been a great deal of attention to work-related fatigue due to the introduction of video-based time-motion analysis (TMA), followed more recently by microtechnology player tracking systems. Because of these technologies, practitioners can detail match-related declines in work or rate of performance. There is substantial research indicating work rate is diminished during a soccer match due to fatigue. This is reflected by reduced distances covered towards the end of halves, end of matches and after high-intensity periods.

**Muscle Damage**

Soccer is characterized by repeated intense activities such as sprints and moderate to high magnitude accelerations and decelerations. In conjunction with soccer-specific activities such as shots, tackles and contacts with opposing player, soccer presents a large stimulus for muscle damage. Locomotor activity in soccer fits within the mechanical stress model of exercise-induced muscle damage (EMID), with substantial impulse forces produced during high intensity movements, particularly eccentric muscle actions, leading to structural disturbances. Chemical responses to structural damage is evident following a soccer match. Indeed, moderate to very large (ES = 0.6-2.3) increases in direct (i.e., CK, Myoglobin LDH) and indirect (i.e., increased DOMS and decreased force production capacity) markers of muscle damage in the period directly preceding the match continuing to 72 hours are evident. Evidence also suggests very large inflammatory and immunological responses occur post-match which persist 72
hours afterwards. From a hormonal perspective, insulin reductions (ES = -1.0) and substantial increases in cortisol and testosterone levels are likely to occur post-match, although this hasn’t been a consistent finding in soccer.

**Glycogen Depletion**

Muscle glycogen is likely the most important substrate for energy production in soccer. Glycogen depletion is also thought to be associated with diminished high-intensity distance towards the end of a match and has been postulated a limiting factor in single and repeated high-intensity sprints. Early work by Saltin et al. (1973) involving 5 rested soccer player and 4 indicated soccer matches result in a significant decline in muscle glycogen stores, with simultaneous decreases in work-rate (Figure 13). Additionally, the players who had rest were found to have more than double the muscle glycogen concentration other those who trained prior to the match. Following the match, those who had rested prior to competition still had muscle glycogen stores in reserve, while those who had trained were found to have complete depletion of glycogen. Soccer results in glycogen depletion, therefore replenishing stores prior to subsequent competition is pertinent.
Figure 13 - Reilly, Drust and Clarke (2008)

**Technical Performance**

There is a lack of research investigating the effects of training interventions, structure and periodization on match-play characteristics, particularly examining whether fatigue can be delayed by certain interventions.\(^{160}\) Similarly, evidence of altered technical execution of tasks is lacking. However, works by Rampinini with youth soccer players indicate match-related fatigue negatively affect short-passing ability towards the end of matches.\(^{182}\) Further, a fatigue-related decline in technical proficiency for a given intensity was associated with the fitness level of the players.\(^{182}\)

**Heat**

Elite-standard soccer is typically played over the course of 8-10 months (English Premiere League, La Liga, Bundesliga, MLS, etc.), with teams experiencing a variety of environmental conditions depending
on the time of year. Youth and collegiate soccer, particularly in the states, have seasons broken into two cycles spanning from September through November and February through May in efforts to avoid colder environments. Despite the level of soccer participation, one challenge is consistent, which is performing in hot and/or humid environmental conditions. At the highest level of soccer, prime examples of this challenge were seen in the 2014 FIFA world cup and will again be seen in the Qatar 2022 World Cup, where temperatures may exceed 40 °C. A full understanding and subsequent acknowledgement of environmentally mediated running decrements and altered match-play (tactical) characteristics in soccer are vital for athlete, teams, medical practitioners and governing bodies so that adequate precaution and provision can be executed. From a performance perspective, negating or reducing the negative effects of hyperthermia have practically important implications on the degree of fatigue developed during training or competition. Whilst safety, the most important factor associated with any sport play, can be markedly increased when collective safeguards are considered.

*Effects of Heat on Soccer Match-Play*

Even in temperate environments, soccer has been shown to induce high internal temperatures (>39 °C), however performance is typically not altered unless other confounding factors are at play such as diminished fuel or body water. In contrast, more extreme heat environments have shown to induce substantially higher internal temperatures leading to marked decrements in running performance and altered match tactics. Internal body temperature responses appear to be also be dependent on level of competitive play, as elite-standard soccer players have been observed to have 0.4 °C higher rectal temperature than their sub elite-standard counterparts when competing in the heat. Naturally, more elite-standard players also present with higher aerobic fitness levels, which allow a higher work-rate and therefore greater metabolic heat production. The link between aerobic fitness and running performance (total distance and high-intensity distance) has been clearly established. From a running-based performance perspective, lowered total distance and high-speed distance (2.6-57%) appear to be the most frequently observed manifestation of performance decrement in soccer.
Although, it’s important to recognize that match running performance alone is a poor indicator of overall performance (wins vs. losses). Never the less, reduced exercise capacity is an expected expression of early onset of fatigue when playing soccer in the heat. From a match-play characteristic perspective, the effects of a hot environment on a soccer match appear to manifested by increased time of possession, higher percentages of successful passing and reduced player duels.

**Heat-related Mechanisms in Soccer**

Limiters of capacity or fatigue influencers are multifactorial, even in the heat. Generally speaking, physiological capacity is limited when the rate of heat production from activity exceeds that in which it is being dissipated. The magnitude of gain is dependent on numerous factors such as the intensity of exercise, environmental conditions, as well as duration and frequency of exposure. Cardiovascular strain and eventual fatigue are a direct result of competition for blood flow between skin and active skeletal muscle. Evaporation, the body’s primary heat loss mechanism, requires fluid diversion from the core and active skeletal muscle to the skin for cooling. Evaporation of fluid form the skin, in conjunction with other forms of heat loss such as convection (air or water flow over the skin), acts to cool the skin and reduce the gradient. Balancing heat generated by metabolic activity within the body, with adequate loss of heat is required to maintain healthy internal conditions. However, when fluid losses are not adequately replaced, rising body temperatures ensue. In an extensive review of literature pertaining to the effects of body mass loss on heat gain, Pryor et al. found that body temperature can increase by 0.21 °C for every 1% body mass lost during exercise. Indeed, Mohr et al. observed professional soccer athletes lost >2% of body mass in hot environmental conditions, with around 2% of body mass loss typical even in moderate conditions.

**Heat Acclimatization for Soccer**

Heat acclimatization is a natural process the body experiences when exposed to heat stress conditions. Of the potential mediators acknowledged, acclimatizing the body to stresses and therefore increasing its
capacity to withstand harmful conditions is perhaps one of the most potent form of protection available. Comparable to a proper physical preparation program, gradual increases in heat strain over the period of many days (10-14) will lead to natural physiological adaptations beneficial for heat loss potential. Short-term heat acclimatization (<7 days), although add some positive physiological benefit (plasma volume expansion), should be recognized as insufficient in allowing optimal physiological adaptation to the heat. Thermoregulatory adaptions such as earlier onset of sweating, increased sweat rate, and reduction in resting internal body temperature may not be realized until after 1-2 weeks after acclimatization begins. Of acknowledgment for soccer players and other intermittent-intensity sports is the barrier of in-season schedule congestion to heat acclimatization. As just recently discussed, full heat acclimatization protocols often require multiple weeks whilst most fixtures occur in less than or equal to a week. Yet, it’s been demonstrated in a group of semi-professional soccer players that a practical degree of positive physiological adaptation can occur while training for 6 days in a hot environment. This is an important consideration for practitioners preparing athletes for play in oppressive and/or extreme heat conditions. Firstly, a practical degree of acclimatization can occur by exposing athletes to warm conditions during normal training. Secondly, and more importantly from a practical standpoint, every effort should be made to fully acclimatize players prior to the season and subsequently maintain that acclimatization throughout the season.
Workload in Sport

Introduction

‘Load’ is somewhat nebulous without consensus definition across the literature. In a recent consensus statement released by the International Olympic Committee on load in sport and injury risk, an agreed upon definition of load was “the sport and non-sport burden (single or multiple physiological, psychological or mechanical stressors) as a stimulus that is applied to a human biological system (including subcellular elements, a single cell, tissues, on or multiple organ systems, or the individual)’. Load was acknowledged as being applied to a biological system over a varying time periods (i.e., seconds, minutes, hours, weeks, month, years) and with varying magnitude (i.e., duration, frequency and intensity). Workload (i.e., training load, competition load) has been broadly refer to as ‘the cumulative amount of stress placed on an individual from multiple training sessions and games over a period of time’ by Gabbett et al. and ‘a combination of… [training and competition] intensity, duration, and frequency’ by Smith et al. With these suggestions in mind, for the purposes of this discussion, load will be referred to as stress placed on the body by a performed activity. Further, loads are typically referred to as either external or internal to the body, with external load (e.g., distance covered, pitches thrown, training hours) describing the quantification of work done by the body and internal load (e.g., heart rate [HR] or blood lactate [BLA] response, rating of perceived exertion [RPE], oxygen consumed) describing the physical loading experience by the body.

Workloads incurred during a match or game are due to the competitive demands of the sport, while practice or training workloads are utilized to promote positive physiological adaptation and performance improvements. This is an important point, because training workloads are modifiable and therefore can be structured in a way to promote optimal adaptation. Regulating the response of exercise stimuli (dose) requires coaches and practitioners to accurately titrate training workloads. If treated appropriately, workloads promote physiological adaptation and performance improvements through the acquisition of
tactical competence, technical skill, and psychological resilience. However, if the balance of workload and recovery are not managed properly, the athlete’s ability to positively adapt is diminished or worse, injury/illness occurs. This provides strong rationale for coaches to periodize the exercise stimulus to promote optimal adaptation and reduce the risks of maladaptation or injury/illness.

Several models have been proposed to explain the physiological response of the body to acute training stress such as the General Adaptation Syndrome Model (GAS), Fitness-Fatigue Model, and Stimulus-Fatigue-Recovery-Adaptation Model.

**General Adaptation Syndrome Model**

A fundamental concept introduced by Hans Selye’s GAS model (Figure 14) is that stress disrupts the body’s physiological state (i.e., homeostasis). An initial stimulus pushes the body into an alarm state (i.e., stage 1 – alarm), which is manifested by acute fatigue. This is a normal and essential part of physiological adaptation, particularly in training context. After the stimulus subsides, body resists the physiological perturbation of homeostasis (i.e., stage 2 – resistance). If adequate rest is allowed the body recovery beyond the original physiological state as an adaptive response (i.e., stage 3 – supercompensation). When adequate rest is received, the body supercompensates above the original baseline physiological state leading to increased physical capacity. If the stimulus is applied prior to allowing the body to return to its original physiological state, compounding fatigue occurs which if allowed can progress from exhaustion to overtraining (i.e., stage 4 – exhaustion). While GAS is a useful representation of the body’s response to an acute exercise bout and an excellent starting point for practitioners to understand the interplay between stimulus an response, adaption is a highly complex phenomena which is influenced by a multitude of factors such as sleep, academic stress, life events, etc.
Figure 14 - Hans Selye's General Adaptation Syndrome Model

**Fitness-Fatigue Model**

Early work by Banister and Calvert introduced the concept of a fitness and fatigue model. Their concept details a relationship between fitness (i.e., positive adaption) and fatigue (i.e., negative physiological function) with the interaction resulting in changes in performance after an exercise stimulus. While Banister’s model is simplistic and easily applied to a practical setting, likely individual components such as cumulated effect of load, recovery deficit and severity of fatigue symptoms explain an athlete's preparedness at any one time.

Based on Banister's model, Gabbett et al. introduced the concept of acute to chronic ratio (Figure 15). This measure gives a relative measure of load which has occurred in the previous week (i.e., acute load) compared to the rolling average of the previous 4 weeks (i.e. chronic load). As shown in Figure 15, the relationship between acute:chronic workload and injury risk has been found non-linear. According to the model, if the athlete has a high chronic load or high “fitness” and low acute load therefore low levels of “fatigue”, reduced injury risk is likely. However, as acute load spikes above chronic load tolerance, increased injury risk ensues. Indeed this model has been found significant in various contexts with a
range of 0.8-1.3 representing reduced injury risk and >1.5 representing increased risk of injury (i.e., danger zone).

![Diagram](image)

**Figure 15 - Acute:Chronic Ratio and Injury Risk (Gabbett 2016)**

**Stimulus-Fatigue-Recovery-Adaptation Model**

The stimulus-fatigue-recovery-adaptation model (Figure 16) offers another valid description of the biological response follow an acute bout of exercise. After a stimulus is applied in the form of workload in a sport context, fatigue ensue. The degree in which fatigue accumulates is directly proportional to the intensity and duration of the bout and is a key determinant in the length of the recovery-adaption period. No reductions in performance are seen at this time. As with other understandings, if the recovery period is sufficient, physiological homeostasis is returned and is followed by supercompensation. At this point, the athlete is “peaking” and is well-prepared for competition. If a new stimulus is not applied in an appropriate amount of time, preparedness will decline, which is referred to as involution in the model.
Workload and Injury

Extensive reviews of workload and injury risk in sport have shed light on several key relationships. Associations between workload and injury have been investigated in nearly all major sports including soccer, basketball, rugby, Australian football, and cricket. The strong association between workloads and injury has been demonstrated by numerous investigations showing links between poor aerobic capacity or low chronic workload and injury, as well as, “spikes” in workload or acute changes relative to the individual’s chronic baseline (i.e., acute chronic workload ratio) and injury. As described in several injury etiology models, there are innumerable precipitating factors which can lead to injury. However, regardless of explanatory model, exposure to workloads during practice and competition are a precondition for athletic injury to be sustained. While injury may or may not occur during an exposure, participation recurrently modifies subsequent risk.

Although injury causality is understood as multi-factorial, the way in which athlete loads are managed represents a major modifiable risk factor. Practitioners managing workloads for injury risk mitigation purposes usually track loads both external and internal to the body. External load, which refers to stress external stimuli which are applied to the body and result in a physiological response, the nature and magnitude of which can vary on influences such as environment or intrinsic biological factors.
Internal load refers to that physiological response. Tracking workloads sustained externally is important for profiling the capabilities and capacities of an athletes, while tracking internal load allows the ability to measure the biological stimulus provided. Poor managed workloads can influence injury risk in a negative way at either the whole-athlete level or tissue level.

As mentioned in Chapter 2, an important consideration when evaluation the relationship between workloads and injury, the effect of a dose can occur immediately or it can require more extended periods of time to manifest. Interestingly, field research with athletes by Hulin et al. and Orchard et al. have supported this concept on a broader scale, rather than at the tissue level. Hulin and colleagues found that if an athlete had a spike in workload (i.e. acute:chronic workload > 1.5 AU), injury risk was not effected in the same week as the spike but in the following week. Injury risk was on the magnitude of 2-4X higher the following week. Similarly, Orchard and colleagues rapid load increases didn’t manifest in increased injury risk for sometimes 3-4 weeks elite cricket fast bowlers.

**Workload Research in Soccer**

**Activity Demands**

Essential to any sport-specific recommendation is full clarity on the activity demands of the sport. Soccer (or football as it’s known more globally), being the world’s most popular sport, has been extensively studied and reviewed in the literature. Consequently, knowledge of the physical demands and limitations on performance in soccer are quite good. The normal duration of a typical soccer match is 90 minutes, with two 45 minute halves interrupted by a short 15-minute interlude between. Soccer is intermittent-intensity in nature and predominated by low-intensity movement (walking or jogging) alternated with short periods explosive, high-speed movements (high-speed running or sprinting). In line with other intermittent intensity sports, there are both anaerobic and aerobic aspects to fuel sourcing in soccer. Specifically, quick changes of direction and burst of activity require substantial anaerobic power, whilst the extended duration of play, as well as recovery between short bursts require adequate
aerobic contribution Reduction in the capacity of either system, whether through accumulation of metabolites or reduction in oxygen supply to working muscle can diminish soccer performance. Of additional consideration to the locomotor demands of soccer are the biomechanical stresses and forces applied to the player as sudden and repeated changes of direction (accelerations and decelerations) occur frequently while players maneuver for optimal tactical position. The increased metabolic cost\textsuperscript{233}, as well as heat production of intense accelerations and decelerations, in comparison with steady state locomotion, shouldn’t be discounted.

**Workload and Injury Risk in Soccer**

High absolute workload\textsuperscript{87,234} and abrupt changes\textsuperscript{27,235} (i.e., acute spikes) in workload have been associated with injury risk in soccer. However, in adolescent female soccer players, Clausen et al. reported that high workloads appeared to offer a protective effect from injury.\textsuperscript{236} According to a review by Soligard et al\textsuperscript{72}, six studies have investigated short\textsuperscript{81,157,237,238} and/or long \textsuperscript{81,85,158} periods of congestion in soccer. Although limited work has been done investigating the intricacies of between match recovery periods and injury, available data suggests congested match scheduling is associated with an increased injury risk. Most all research has investigated the effects of match schedule congestion on match injuries, however Dellal et al. did observed training injuries during congested time periods were either unaltered or reduced.\textsuperscript{158} In elite level soccer, matches are typically played weekly, therefore congestion alludes to greater than 1 match per week. However, some investigations have dichotomized between match recovery periods to assess relative risk. No differences were found for <3 days compared to >4 days of rest between matches\textsuperscript{81,238}, however significantly higher injury rates are observed for <3 days\textsuperscript{158,237} or <4 days\textsuperscript{81,157} compared to >6 days. Conflicting finds are probably resultant of contextual factors, such as individual team periodization structures. Its customary to decrease training loads, particularly in elite-sport, during highly congested periods.
**Sleep in Sport**

**Introduction**

In high-performance sport, recovery from congested playing periods and intense training or competition is critical and requires strategies to optimize performance reduce injury risk in subsequent sessions. Of the numerous recovery strategies and tools available to athletes such as active recovery, stretching, compression garments, and massage, sleep is regarded as the most powerful form of recovery available to athletes and understood as critical piece for physical and psychological well-being. This has been confirmed through both laboratory sleep loss research which has linked with poor cognitive function and performance and field research with athletes.

**Sleep in Athletes**

The current sleep recommendation from the Mayo Clinic indicates that generally, individuals should be obtaining at least 7-9 hours of good sleep each night, regardless of athlete status. Sadly, athletes have shown to attain less sleep than current recommendations, although this isn’t a consistent finding. Duration is only one component of sleep, however, and the importance of sleep quality has been increasingly recognized as a vital element of overall health and well-being. Unfortunately, the quality of athletes sleep has additionally been shown inferior to non-athletic populations, particularly after competition. There are many factors that affect the sleep quantity and quality of the athlete. Among those reasons are competition (e.g. night matches with late kick-off; congested scheduling), travel schedules, and self-imposed negative sleep practices or socialization. A study of 890 South African athlete showed that around 75% average around 6-8 hours, with 11% reporting less than 6 hours. Pre-competition anxiety has also been found a significant barrier for quality sleep prior to important competitions. Similarly to reports of South African athletes. Juliff et al. found that 64% of a group of Australian athletes (n = 283) reported poor sleep quality the night preceding a competition. Results were accredited to nervousness and/or mental stress.
In an investigation of Olympic athletes using wrist-worn actigraphy, Leeder et al.\textsuperscript{246} found lower sleep duration and poorer sleep quality when compared with matched-controls. In one of the few studies to utilize the gold-standard for sleep, Taylor et al.\textsuperscript{256} found a significant effect of training volume on physical movement during sleep time, indicating restless sleep. Sleep has shown affected by overall training load\textsuperscript{258} and the time in which training occurs.\textsuperscript{242}

Of particular importance to the recovery process of athletes is both the quality and quantity of sleep, which is well accepted as critical for the optimization of physiological state.\textsuperscript{205,259} Although athletes regard sleep as essential for both recovery and performance\textsuperscript{252}, evidence suggest elite athletes demonstrate less than optimal sleep characteristics when compared with normal, healthy individuals.\textsuperscript{246} Coupled with evidence suggesting sleep may be disrupted by exercise load in a dose-response manner\textsuperscript{256,260}, athletes may be at particular risk for compounding fatigue throughout an intense and congested schedule if sufficient rest is not realized. Recent investigation into the sleep characteristics of elite soccer\textsuperscript{250} and AFL athletes\textsuperscript{244} have shown novel insights into the effects of scheduled match time (day vs night)\textsuperscript{250} and location (home vs. away)\textsuperscript{244,250} on sleep quantity and quality, with both aspects generally reduced after matches. This is of particular concern as sleep deprivation following increased exercise load can impede recovery and adversely affect performance.\textsuperscript{249,259}

**Function of Sleep**

Frank and Bennington identify that sleep 1) restores the immune and endocrine system, 2) assists in the recovery of the nervous and metabolic cost imposed during the wake and 3) allows cognitive development which is necessary for learning, memory and neural plasticity.\textsuperscript{205,261} Sleep is composed of typically four to five 90 minute cycles, which rotate thought periods of non-rapid eye movement (NREM) and rapid eye movement (REM).\textsuperscript{262} Figure 17 illustrates that NREM is further divided into four other stages (e.g., 1-4). Stages 1 and 2 are referred to as ‘light sleep’, whereas stage 3 and 4 are deep sleep
(a.k.a. slow wave sleep [SWS]). During sleep the body attains its most metabolically inactive point, which is characterized by slow breathing, low heart rate and low cerebral blood flow.\textsuperscript{263} During this time, the body releases anabolic hormones, predominately growth hormone, which aids in physiological recovery.\textsuperscript{264} Release of anabolic hormones plays a pivotal role in stimulating protein synthesis and mobilizing free fatty acids to reduce amino acid catabolism.\textsuperscript{265} These deeper stages of sleep, especially SWS, are critical for tissue repair and athletes recovery from exercise induced muscle damage (EMID). Sleep has also been implicated as important for motor learning and memory consolidation, which can occur in both NREM and REM states.\textsuperscript{266}

![Sleep Architecture Diagram](image)

**Figure 17 - Sleep Architecture**

**Sleep Disruption**

Sleep plays a pivotal role in many physiological and cognitive functions, with sleep loss or restriction showing many deleterious impacts. The has been demonstrated in sleep deprivation and chronic sleep loss research by diminished cognitive functioning, learning and memory, reaction time, auditory vigilance and mood.\textsuperscript{267–269} Additionally, heightened states sleepiness, depression, and confusion are linked with poor
sleep. Specifically, research suggests that when sleep is reduced below 7 hours in adults, several executive function and decision-making tasks and mood can be compromised. From a mechanistic perspective, reduced cerebral metabolism in the thalamus, cerebellum, and prefrontal, posterior parietal, and temporal cortices subsequent to sleep quantity and quality reduction has been proposed. This has been supported though correlative analysis, with reduced metabolic rates in the aforementioned regions and decreased cognitive functioning showing relation.

From a physiological perspective, disrupted sleep (i.e., 3 hour sleep loss) has been associated with increased heart rate, oxygen consumption, plasma lactate concentration during both submaximal and maximal sleep. These physiological responses are explained in part by increased metabolic demand, hormonal stimulants (i.e., catecholamines epinephrine and norepinephrine) and perceived effort, in addition to the influence of exercise stress on physiology. Regarding substrate repletion, 30 hours of sleep deprivation has shown to prohibit complete restoration of muscle glycogen stores in team sport athletes.

**Sleep and Injury**

Unfortunately, research on the relationships between sleep and injury in athletes is limited despite sleep being recognized as critical to the recovery process. To date, only a handful of studies have assessed sleep and injury in elite sport, however as sleep monitoring technologies become less cumbersome and more insightful, research in this area is expected to increase. Likely the most prominent investigation of the association of sleep and injury was conducted by Milewski et al. who found that sleep was and independent predictors of injury in adolescent athletes. The primary practical finding from Milewski was that athletes who reported sleeping under 8 hours were 1.7 times more likely to sustain an injury compared to those who reported sleeping more than 8 hours.

Similar to Milewski et al., von Rosen et al. found sleep loss was a risk factor for injury in adolescent athletes. This study was additionally questionnaire based, as was Milewski and colleagues. Specifically,
von Rosen found that athletes who slept more than 8 hours a day during the weekdays, in conjunction with meeting appropriate nutrition recommendations, decreased their risk of injury. Finally, Dennis et al. used both actigraphy and sleep diaries in a study of injury risk factors with Australian football players. This study is unique in that it’s the only study to date which has assessed the association between objective measures of sleep and injury risk in an athletic population. Although lower sleep duration and quality were hypothesized, they found no association between either sleep quality or quantity and injury risk. While sleep and injury investigations are lacking in high-performance sport, it is likely that more research will focus in this area as validated sleep tracking devices become more prevalent.

Sleep Research in Soccer

As discussed, the requirements of elite soccer may negatively influence athletes ability to achieve optimal sleep and therefore recovery. However, there is a lack of quality research specific to soccer. Yet, soccer athletes are generally assumed to be have healthy sleep patterns during “normal days.” Soccer athlete at both elite and sub-elite levels deal with confounding factors which can negatively affect sleep behavior such as night matches, travel, and congested schedules. Works by Fullager et al. found that compared with training days and afternoon match days, night matches resulted in significantly less sleep duration. Sleep loss following night matches also resulted in significant reduction in perceptual recovery compared with the other conditions. Along the same lines, Fowler et al. showed substantial reduction in sleep following a night match in elite soccer athletes, which resulted in impaired stress-recovery balance. In a recent study on soccer players competing in the Portuguese First League (Liga NOS) and UEFA Champions League, Carrico et al. found night matches resulted in later bed times compared with normal training days, which also resulted in a significant reduction in sleep duration. However, it should be acknowledged that these findings are not consistent. Both Roach et al. and Robey et al. found no effect of early evening match or late evening high intensity exercise on sleep in youth soccer athletes, respectively.
**Monitoring in Sport**

*Introduction*

Athlete monitoring is quickly becoming standard practice for maximizing player performance\(^{296,297}\), reducing injury risk\(^{58,298,299}\), and optimizing competition readiness\(^{300}\). For high-performance programs, monitoring load-performance and load-injury relationships are useful for providing insight into how athletes are responding to stresses incurred during and outside of training and competition. There are numerous benefits to monitoring athletes such as gathering scientific explanations for changes in performance or injury risk, enhancing coach and practitioner confidence when manipulating training loads, and boosting athlete-coach-practitioner relationships.\(^{287}\) Athletes can feel empowered during the monitoring process as they are not only reminded of their importance to the program, but additionally gain insight into their body’s responses and adaptations to stress.\(^{287}\) This involvement encourages ownership, accountability to teammates, and can drive excellence.

*Assessing Readiness*

Assessing athletes’ wellness, hydration, and fatigue status is essential to ensuring readiness to optimally perform. Consistent monitoring of wellness through subjective questionnaires can provide insight into athletes’ stress, soreness, and motivation levels. Monitoring training loads can additionally verify an appropriate taper prior to competition\(^{296}\), which can be confirmed through player wellness reports. Also, objective internal measures such as heart rate recovery and variability metrics can provide insight into autonomic nervous system status\(^{301}\), while measures of urine concentration and color can detail hydration status\(^{104,302}\).

*Peaking*

It’s widely accepted, particularly in team sports that it’s not possible to peak for every competition or important event given the congested seasonal scheduling\(^{303}\). However, by capturing physical loads...
coaches and practitioners can be sure a scientific approach to training periodization is employed. This is essential in allowing the athlete to peak at the right time (i.e. post-season, national or international competitions, or important rivalries). Monitoring loads and wellness allows the practitioners to realize athlete supercompensation (i.e., positive adaptation to stress) and reduce the risk of non-functional overreaching (long-lasting fatigue).

**Mitigating Risk**

A growing body of literature confirms a meaningful relationship between training load and injury risk. Monitor load-injury relationships can help identify and manage risk factors (i.e. acute:chronic workload, high-intensity running distance, body load, mean running speed) relevant to the team and individual athlete. While injury is complex and difficult to predict, gathering insight into important confounding factors such as environmental conditions, fatigue status, mood or sleep disturbances, or stress can provide insight to the practitioner for potential load management or athlete educational intervention.

**Return to Play**

Medical and fitness practitioners play a vital role in providing the safest environment possible for athletes to return to play from injury or illness. In preparing athletes to return, its critical for practitioners to feel confident that their load manipulations are both optimizing adaptation and reducing the risk of re-injuring. Additionally, confirmation that physical loads and capacities meet or exceed those expected during competition are highly useful in ensuring athletes are ready to be reintroduced to full play.
Introduction

The progression from healthy and normal adaptation to maladaptation from stimuli (training stress, life stress, or likely a mixtures of many factors) can be either gradual or sudden which is dependent on a myriad of complexities. As discussed in Chapter 1, the body is complex and dynamically responding to stimuli external to and within the body. Therefore, coaches, sport scientists and practitioners need tools and a good understanding of how to monitor the health and fatigue status of their athletes. Equally important, ensuring athletes are adequately and safely progressing towards predefined goals. The primary purpose of training is to introduce a stimulus that develops performance capacity and ability. To ensure athletes are progressing in a healthy manner, workloads and responses to workloads should be tracked.

Quantifying the stress imposed on athletes by training or competition requires measuring outputs such as frequency, intensity, duration and mode of exercise. However, for simplicity sake, workload is the product of session duration and intensity (i.e., workload = session duration X session intensity). While
there are numerous ways of tracking workloads, all measures are categorized as either internal or external to the body.

**Measures of Internal Load**

Internal load refers to the stresses experienced within the body or the physiological and psychological responses to stimuli, such as increases in heart rate, oxygen consumption, or perceived exertion. Practitioners and sport scientists often prefer tracking internal loads as they give a true reflection of the stresses incurred by the body, which can respond differently to similar or the same task.\(^{136}\)

**Perceptual Methods**

Capturing athletes perceived effort is perhaps the most used method of assessing exercise intensity in athletes.\(^{307,308}\) The degree to which an athlete experiences stress is influenced by a myriad of external (e.g., environment and activity demand) and internal characteristics (e.g., hormone and substrate concentration, psychological and personality characteristics). Athletes rating of their perceived exertion (i.e., RPE) is captured by a number of scales including the Borg 6-20, category ratio (CR)-10 or the Borg CR-100, which originated from Gunnar Borg.\(^{307}\)

The most widely used monitoring tool in high-performance sport is the session RPE method.\(^{137}\) This method, proposed by Carl Foster in 2001, utilizes the basic question of “How hard was your session?” ranked on a scale from 1-10 (i.e., 1 = very, very easy; 10 = maximal).\(^{309}\) From this simple rating, session load can be assessed, which is the product of RPE and session duration (i.e., Session Load [AU] = session RPE X session duration [min]). Session RPE has been validated by many works in sports and various activities and found consistent with many other markers of internal load, including muscle damage biomarkers and heart rate.\(^{309-315}\)

**Heart Rate-Based Methods**
Owing to the strong linear relationship between heart rate and oxygen uptake\textsuperscript{208}, monitoring heart rate measures is a popular method of quantifying training intensity and load. In fact, Akenhead and Nassis found 40 out of 41 high performance soccer clubs were using heart rate (and GPS) data to monitor their athletes.\textsuperscript{316} Heart rate can be assessed with chest, wrist or smartphone devices, with only the chest device being accurate and valid enough for monitoring of high-performance athletes.\textsuperscript{317}

One of the most common ways of utilizing heart-rate based methods is the calculation of training impulse or TRIMP, which is a score which represents the total workload incurred by the athlete during a single bout.\textsuperscript{47,318} There are a few different ways to calculate training impulse with one using a mathematical model (i.e. Banister’s TRIMP) and one a simple function of duration in 5 separate relative heart rate zones (i.e., Edwards TRIMP).\textsuperscript{309} The weighting factor incorporated into Banister’s model of TRIMP was used to accommodate for the stresses which higher-intensity exercise invokes, therefore the curve mimics blood lactate responses to incremental increases in exercise intensity.\textsuperscript{319} In addition, another methods which factors both ventilator threshold (i.e., sudden increase in breathing) and the respiratory compensation point (i.e., onset of hyperventilation) was developed by Lucia and colleagues.\textsuperscript{320}

Although heart monitoring provides a non-invasive and objective tool of assessing training intensity and load, it is not without limitation. Firstly, heart rate monitors do not adequately respond to intermittent exercise, with frequent delays in heart rate response.\textsuperscript{317} This hampers the practitioners ability to quantify heart rate-based internal load during interval-type training or strength training. Additionally, athletes may feel uncomfortable or restricted while wearing a chest belt. Finally, analyzing heart rate often requires technical ability to interpret data. Nevertheless, heart rate monitoring offers the best non-invasive, objective assessment of workload available.

**Measures of External Load**

External load is defined as the work completed by the athlete irrespective of internal characteristics.\textsuperscript{287} Tracking external loads such as distance, time and pace or average speed are very traditional approaches.
to quantifying workload. However, advances in microtechnology and cloud computing has expanded the ability to track detailed aspects of motion and provide information real-time. This has revolutionized the way in which training is managed. By detailing the intricacies of workload completed by the athlete, the coach and practitioner develops a better understanding of the demands of the sport and the individual capacities of the athlete.²⁸⁷

**Time-Motion Analysis**

Time-motion data can be collected from a broad range of technologies including pedometers, accelerometers, global positioning systems (GPS), digital video analysis (e.g., Prozone Amisco), and microelectromechanical sensor (MEMS) technologies.¹³⁷,²⁸⁷ Pedometers record simple steps taken by quantifying the times a vertical perturbation (i.e., ambulation) produces an oscillation that exceeds a predefined threshold. Basic accelerometers are also a popular sensor integrated into most modern wearable motion devices. Tri-axial accelerometers measure the magnitude of motion in 3 planes (i.e., up-down, forward-backward, right-left). Although a popular inclusion in wearable technologies available to the public (e.g. commercial wrist-based wearables), these technologies are not commonly utilized in a high-performance environment. Also, very little work has been done to validate commercial devices against acceptable research methods.³²¹,³²² However, GPS devices with embedded MEMS technology have become almost standard monitoring technology for most high-performance programs. Taylor and colleagues report 43% of the high-performance surveyed indicated they were using player tracking technology with GPS.³²³ This trend seems to be much more prevalent in soccer though, as a survey of elite European, United States and Australian soccer clubs indicate 40 out of the 41 surveyed were using GPS-enabled devices for every player during every session.³¹⁶

**Player Tracking Technologies**

Modern player tracking technologies are integrated with both GPS and MEMS technology, making it possible to track a wide range of metrics from distance in speed zones to mechanical stress of impacts and
jumps. Additionally, devices are quite small and unobtrusive, therefore there utilization has gained favor with organizations looking to gather quantitative information on athletic performance, positional demands of the sport and player movements and positioning during training and competition. \(^{137,287,324}\) From a training management standpoint, one can imagine the benefit of real-time information on the physical stresses incurred by the athlete. Insight into workloads sustained allows coaches to structure training during, intensity and density, as well as, potentially resting players when predefined workload thresholds are met. Additionally, detailed insight into the physical demands of play allow practitioners to prepare injured players for return to play or sub-elite players for safe integration into progressively more elite play. Indeed, a number of investigations into the physical demands of play have been conducted for sports such as soccer, American football, Australian rules football, rugby, and cricket – to name a few.

GPS-enabled devices function by receiving a continuous signal from at least 4 separate satellites orbiting earth, which provide information on distance and time which is used to calculate positioning. \(^{331}\) Metrics such as distance and velocity are delivered by GPS while more detailed information on the magnitude of acceleration and its relationship with the earth’s magnetic field is delivered by integrated inertial sensors. These integrated inertial sensors are referred to as a microelectromechanical device (i.e., accelerometer, magnetometer and gyroscope). The integration of MEMS technology allows the quantification of mechanical load (e.g., Catapult’s PlayerLoad or GPSports’ New Body Load), magnitude of collisions (e.g., g-force classification of impact), or estimations of metabolic load. \(^{233,334,335}\)

**Monitoring Wellness and Fatigue**

**Introduction**

Measuring fatigue can be difficult due to its multifactorial nature. It’s important to acknowledge there is no single, universal assessment or marker which can differentially diagnose fatigue or a maladaptive state. While there has been substantial work investigating the underlying mechanisms of fatigue, scientists and
practitioners are more concerned with identifying fatigue state rather than understanding the intricacies of its development.

There are a number of methods for quantification of fatigue in sport\(^\text{100,287}\), with both objective and subjective means having been studied. From an objective performance perspective; average or peak power, force (e.g., counter-movement jump, cycle ergometer), total work, or time/speed (e.g., linear or multidirectional movement test) are traditional measures in high-performance testing. Regarding subjective methods, perceptual indices (e.g., stress, fatigue, soreness and more extensive wellness inventories (e.g., RESTQ, DALDA) are standard practices. Perceptual scales can range from Borg 6-20 scales, to 10 or 100-point category ratio scale.\(^\text{137}\)

**Subjective**

Psychological measures of wellness can be used to gauge both acute and chronic fatigue states, with the intention of avoiding non-functional overreaching or overtraining in athletes. Some wellness questionnaires have shown good reliability with training load fluctuation\(^\text{336}\) and have been used to detect maladapted states such as overreaching and overtraining.\(^\text{336,337}\) Generally, practitioners should seek a multi-faceted wellness monitoring approach with questionnaires that ideally detect a broad range self-reported measures but ask few number of questions\(^\text{336,338}\). No single questionnaire should guide interventional strategies; however, practitioners need to consider the time and resources needed to collect, analyze and provide feedback to athlete and coaches. Although current works have shown technology can assist with implementing wellness questionnaires\(^\text{339}\), and specifically apps and smartphones may lessen the burden of monitoring health and wellness.\(^\text{316,340,341}\) However, ‘buy-in’ from the athletes and organization should predicate use of subjective monitoring tools.

**Periodic Wellness Inventory**

Wellness inventories include questionnaires such as Recovery Stress Questionnaire for Athletes (RESTQ-Sport)\(^\text{342}\), Profile of Mood States (POMS)\(^\text{343}\), Daily Analysis of Life Demands (DALDA)\(^\text{344}\), and
Training Distress Scale (TDS)\textsuperscript{345}. The RESTQ-Sport is a commonly utilized wellness inventory\textsuperscript{342} which has been found response to fatigue induced by both acute and chronic workloads.\textsuperscript{336} Concerns regarding the length, and therefore time commitment of the RESTQ-Sport have led to the development of a shortened version.\textsuperscript{323} The short REST-Q has been subsequently validated.\textsuperscript{346} The POMS is a validated 65-item questionnaire which has shown sensitivity to training load changes and associated altered mood states.\textsuperscript{127,347} The POMS assessed six mood or ‘feeling’ states: tension, anger, vigor, depression and fatigue and is robust enough for examination of individual mood states. Construct validity has been explored by Terry and colleagues for its use with both adults and adolescents.\textsuperscript{348,349} The DALDA is unique in that each item of the inventory is ranked according to the athletes norm. Specifically, athlete rank whether the item is worse than normal, better than normal, or normal. Finally the TDS is an inventory for assessing athlete readiness to perform.\textsuperscript{345} Grove et al. have showed the TDS inventory is a valid tool for assessing performance readiness in athletes in both a laboratory and field setting.\textsuperscript{345}

**Daily Wellness Inventory**

Perceptual measures of wellness taken daily may assist in early identification of non-functionally overreached states or confirming intentional, functionally overreached states. Multiple investigations have shown their sensitivity to changes in stress and fatigue in athletes.\textsuperscript{336,350,351} Common daily measures include gathering perceived stress, fatigue, muscle soreness, and general well-being. Generally speaking, daily wellness measures are typically less time consuming than more extensive inventories, easy to implement prior to training, and inexpensive.\textsuperscript{137} Coaches have shown favor to short daily perceptual measure as an assessment of current monitoring trends taken by Taylor and colleagues show 80\% of high-performance clubs use their own crafted questionnaires.\textsuperscript{323} A good example of these questionnaires are ratings of fatigue, stress muscle soreness and sleep on a scale for 1 to 7, which was first implemented by Hooper and colleagues and therefore known as the *Hooper Index*.\textsuperscript{352} Interestingly, Impellizerri et al. applied this same logic to tracking soreness in different parts of the lower limbs, adding an extra level of specificity to the monitoring program.\textsuperscript{353}
Monitoring ‘recovery’ status has also been found a useful means of early identification of maladaptive states and making training program. A questionnaire which measures daily recovery called the Total Quality Recovery Scale has been used previously, however limited research exists on its efficacy of the assessment tool.\textsuperscript{354} Unique to this scale, aspects of recovery such as nutrition, sleep, relaxation, hydration, active recovery and emotional support are rated to indicate daily status. Kentta and Hasssmen did review its use in overtraining and recovery, with a score of less than 13 indicating inadequate recovery.\textsuperscript{128} Similarly, the Perceived Recovery Status Scale has been used to track changes in performance\textsuperscript{355} and following heavy resistance training.\textsuperscript{356} From a practitioners perspective, this particular scale may be much more intuitive and easy to implement as scores range from 0 to 10 with scores below 2 suggested as an indicator of underrecovery\textsuperscript{355}, however more research is needed to confirm its efficacy in the field.\textsuperscript{357}

**Neuromuscular Performance**

Neuromuscular performance tests are a common tool in athlete monitoring\textsuperscript{287,316,358–360}, with fatigue generally manifesting in a decrement of force, power, velocity or displacement.\textsuperscript{137,361} In fact, Taylor et al. found that 54\% of high-performance organizations were assessing neuromuscular fatigue via vertical jump testing.\textsuperscript{323} While the methods of assessing neuromuscular performance are many, fatigue is typically assessed via a jumping protocols or muscular strength and power assessment.\textsuperscript{137} As previously noted, assessing vertical jump is perhaps the most popular methods of assessing neuromuscular fatigue, likely owing to its non-fatiguing nature and the minimal time commitment needed from athletes.\textsuperscript{137} Due to time efficiency, single jumps are more popular than repeated jumps.\textsuperscript{362} Additionally, assessment of vertical jump doesn’t require expensive equipment; a simple tape measure may be used to assess peak height. Although, jump height has been suggested as an insensitive measure of fatigue following competition\textsuperscript{363} and during intensified training periods.\textsuperscript{305,364} Interestingly, smartphone apps have been developed and validated which can assess vertical jump metrics.\textsuperscript{365} However, monitoring in high-performance sport typically involves utilization of more sophisticated equipment such as force
plates, linear position transducers or contact mat. Basic measures of jump height, average and peak power and velocity are often utilized\textsuperscript{323}, however time-related jump metrics (e.g., flight time to contraction time ratio) have been proposed as a more sensitive measure of fatigue.\textsuperscript{366} Neuromuscular fatigue monitoring with vertical jumps has been used extensively across multiple sports\textsuperscript{363,366–370}, however there is still no consensus regarding the most sensitive metric. This is likely due to inconsistencies across research studies (e.g., jump protocols, warm-ups, technology, equipment, etc.). Although there is a lack of consensus regarding a single jump metric to track, evidence suggests both concentric and eccentric metrics, along with the complete force-time curve relationships may provide practitioners with more detailed view of fatigue status.\textsuperscript{371,372}

Other key attributes of assessing neuromuscular performance through jumps is the ability to profile force production capacity\textsuperscript{260} and assess potential asymmetries.\textsuperscript{362} Jimenez-Reyes et al. used jump height to predict 1RM squat\textsuperscript{373}, while Bailey and colleagues found an association between movement force-production asymmetry and strength levels in university athletes.\textsuperscript{374} Utilization of certain technologies which provide real-time feedback such as a linear position transducer has also shown to lead to greater training gains than control conditions.\textsuperscript{375} Although other measures of strength and power are available for tracking strength such as isometric (e.g, iso mid-thigh pull, iso squat, iso bench press)\textsuperscript{376–379}, repetition maximum and dynamometry, there has been little research into their efficacy or utility in assessing fatigue.\textsuperscript{380}

**Heart Rate-Based Methods**

When monitoring fatigue via heart rate-based methods, resting heart rate (RHR), heart rate variability (HRV) and heart rate recovery (HRR) are most commonly utilized.\textsuperscript{301} Heart rate (HR) measures have shown utility in assessing both fitness and fatigue status\textsuperscript{381} and are reflective of autonomic nervous system (ANS) status.\textsuperscript{301} Although many internal or physiological markers are used to assess fatigue status such as blood biomarkers and saliva biomarker, HR is well-accepted due its ease of implementation and relative inexpensive capture.
RHR is an attractive fatigue monitoring tool because it doesn’t require sophisticated technology and software to analyze. RHR has been defined as the lowest measure of HR taken from a 10-min lying position or upon awakening. Chronic decreases in RHR are typically associated with an increase in aerobic fitness, while an acute decrease in RHR is generally associated with greater parasympathetic drive. RHR has been shown sensitive to short-term fatigue, but is likely not useful in assessing chronic fatigue or maladaptation such as overtraining. It should be noted, as with other fatigue assessment tools, RHR should not be used in isolation but as a supplement to other monitoring practices.

HRV is likely the most popular contemporary, objective, non-invasive, physiological marker of readiness available to sport scientists and practitioners. Simply, HRV is the variability in heart beat-to-beat intervals. As outlined previously, HRV is modulated by an intricate interplay between the sympathetic and parasympathetic nervous system. While there are several indices of HRV including both time-domain (e.g., root mean square of successive difference [rMSSD]) and frequency-domain (e.g., low frequency [LF] and SD1), the most popular and reliable measure is the natural logarithm of the square root of successive differences between adjacent normal RR interval (i.e., Ln rMSSD). Generally, low HRV is associated with a state of fatigue or sympathetic dominance, however it should be noted that this has not been shown consistently. It should be noted that HRV has no utility as a single, isolated marker. Rather, recording day-to-day variation or at least 3 day a week is useful in tracking meaningful differences and therefore detecting under-recovered states. HRV has shown to increase chronically with increases in aerobic training, therefore rolling or exponentially-weighted moving average of HRV may account for chronic adaptation. From a monitoring perspective, HRV values responds differently to sitting vs. standing postures, therefore nightly measures during the last slow-wave stage of sleep have shown the most reliable means of capturing HRV. In any case, care should be taken in reducing environmental stimuli as HR has shown highly sensitive to noise, light temperature, etc. HRV is highly individualistic, therefore training phase and history should be considered when interpreting.
Following exercise, heart rate slows due to inhibited sympathetic drive and parasympathetic reactivation. Tracking the recovery period immediately following exercise has been suggested a viable means of assessing ANS status and therefore recovery status, with decrements indicating fatigue or underrecovery, performance declines, and potentially overreaching. In contrast, improvements in HRR over time has been associated with improvements in aerobic fitness. HRR is assessed by quantifying the decline in beats per minute immediately following exercise for a predefined period. A variety of recording periods have been utilized, ranging from 30 sec to 5 min. However, a HRR does require a submaximal test prior to recording, which must be standardized and conditions held consistent. Unfortunately, the magnitude of error in submaximal tests has been shown high, therefore practitioners should take caution in using tests that cannot be easily replicated. A typical HRR protocol consists of 5 min of standardized submaximal testing at a fixed intensity (e.g., cycle ergometer at a standard power output and cadence) followed by 5 minutes of continuous heart rate monitoring in room free of external stimuli. HRR is expressed as the absolute decline in beats per minute or relative decline (e.g., average HR towards the end of exercise relative to heart rate 60 sec after completion of exercise). In soccer, Dellal et al. has concluded HRR is a relevant tool for tracking recovery following soccer-specific activities such as small sided games (SSG), repeated sprint ability (RSA) and high-intensity training (HIT). This conclusion has not been consistent across the literature, with some authors finding no association.

**Hematological and Biochemical Markers**

Several field-based studies have investigated hormonal and biochemical marker responses to training stress in athletes, with a range of blood, saliva and urine analysis being used. Indeed, biological markers can provide indication of acute training stress and chronic adaptation, although its generally accepted that biological responses to training stress are highly individual and can be influenced by many external (e.g., environment and training program) and internal factors (e.g., age, gender, and psychological state).
While the clinical usefulness of biomarker analysis is accepted, organizations must understand and accommodate for key limitations which exist with biological monitoring. Specifically, examining hormonal and biochemical responses to training stress can be expensive, results require time (i.e., days) to turn around, and analysis require extensive expertise. For these reasons, biological analysis is not as common in high-performance. In an analysis of monitoring practices of high-performance athletics organizations, Taylor and colleagues found only 8% of organizations were implementing hormonal and biochemical analysis, while Akenhead and Nassis found a slightly higher proportion (i.e., 24%) in high-performance football clubs. While many lab-based research studies of athletes conduct blood analysis, non-invasive measures such as saliva and urine may offer greater practicality to applied fatigue monitoring in the field. With that, saliva measures have shown strong correlations between blood serum measures and saliva measures of cortisol (i.e., $r = 0.90-0.93$) and testosterone (i.e., $r > 0.89$) have been found. Also, Hakkinen and colleagues used urinary markers of cortisol and cortisone to assess adaptation of HPA response to progressive strength training.

In general, hormonal and biochemical markers can be used to assess acute training response and chronic biological adaptation, accumulation of metabolites, and an athlete’s health status. Common hormonal markers include testosterone, cortisol, testosterone to cortisol ratio (T:C ratio), catecholamines (i.e., norepinephrine and epinephrine), growth hormone, and insulin-like growth factors. From a biochemical marker perspective, markers of muscle damage such as creatine kinase (CK) are common, as well as, red blood cell markers (e.g., leukocyte, hematocrit, hemoglobin, blood cell count).

**Immunological Markers**

Longitudinal research into the effects of prolonged and/or intense training on immune system function is not prevalent and requires future attention. Nevertheless, it’s been observed that athletes who sustain high-loads for long durations show immune system suppression. This was supported by Hausswirth and colleagues who observed 5 out of 9 non-functionally overreached athletes develop upper respiratory illness symptoms. Reid et al. investigated an association between chronic fatigue and recurrent
infections in a group of athletes.\textsuperscript{422} They found 68\% of athletes had an underlying condition with the potential to cause fatigue and/or recurrent infections. They identified the most common were partial humoral immune deficiency (28\%) and unresolved viral infections (27\%). A suggested association between overtraining and infection (i.e., immunosuppression) has also been hypothesized.\textsuperscript{423} Despite the lack of research in this area, it’s probable that an athlete will experience immunological suppression when overreached or overtrained.

**Monitoring Sleep**

As discussed in Chapter 5, sleep is recognized as a vital physical and psychological recovery tool\textsuperscript{424} and suggested as the most important method of recovery available to athletes.\textsuperscript{425} Poor sleep has been linked with injury\textsuperscript{32,285}, illness\textsuperscript{421} and overreaching/overtraining states. As noted in Chapter 5, Milewski et al.\textsuperscript{32} found the relative risk of injury was increased by 1.7 times when sleep was less than 8 hours compared with more than 8 hours in adolescent athletes. As noted previously, Hausswirth and colleagues noticed upper respiratory illness in 5 non-functionally overreached athletes, who additionally showed substantial decreases in sleep duration (-7.9\%) and sleep efficiency (-1.6\%).\textsuperscript{421} Poor sleep is also a common complaint of overreached and/or overtrained athletes.\textsuperscript{426} Interestingly, multiple sources have shown sleep disturbance is a common symptom of an overreached or overtrained state.\textsuperscript{127,427} Killer et al. found that as few as 9 days of intensified training was sufficient to decrease sleep quality, mood state and maximal performance in well-trained cyclist.\textsuperscript{428} Halson and colleagues observed similar reductions in sleep efficiency in female sprint cyclists who had been underperformance and persistently fatigue for several months (i.e., overtrained).
**Injury Analytics**

**Introduction**

In an elite-sport setting, players, coaches, and administrators need accurate information from scientist and analysts in which to make important decision. In the modern technological age, sports science and medicine practitioners have access to a wealth of tool and technologies in which to inform key stakeholders and establish programs for maximizing player performance\(^{296,297}\), reducing injury risk\(^{58,298,299}\), and optimizing competition readiness\(^{300}\). Establishing a monitoring system where key performance indicators (KPI) and injury risk factors are tracked should in theory provide an advantage. However, prior to providing insightful information to key stakeholders, practitioners and scientists must have a firm understanding of appropriate practices for collecting, analyzing and interpreting data which is being collected.\(^{429}\) This is particularly important where athletes are competing on a national or global level, where inaccuracies in information provided can lead to potentially burdensome and destructive consequences. Take for example an athlete whom displays multiple signs of overreaching (e.g., elevated perceived stress, excessive ‘spike’ in workload, biomarkers indicate accumulation of metabolites) but is not appropriately advised to rest and recover, in turn leading to subsequent injury. This result has a host of negative sequela, particularly in elite sport, where injuries undermine team performance\(^1\)–\(^3\), financially burden parties (e.g., players and organizations)\(^4\), and may threaten long-term wellbeing of the athlete.\(^5\) Additionally, providing inaccurate information or failing to provide key information to stakeholders can result in a loss of credibility for sport science and medicine practitioners and undermine future efforts.\(^{430}\)

**Defining Outcomes of Interest**

Establishing a clear and concise definition of an outcome of interest is perhaps the most important precondition when studying sports related phenomena. However, defining primary outcomes of interest such as performance and injury is often a difficult task for science and medical practitioners as these
terms are vague and multifactorial in nature. Operational differences in outcome definition can lead to anomalies when attempting to determine causality, benchmark the outcome, and establish tactics and strategies which optimize the outcome of interest. From a sport injury epidemiological perspective, most organizations are “concerned with quantifying injury occurrence with respect to who is affected by injury, where and when injuries occur and what is their outcome—for the purposes of explaining why and how injuries occur and identifying strategies to control and prevent them”.

As an example of the inherent difficulty with defining vague terms, consider ‘performance’, which is determined by a multitude of complex and interrelated tasks of which also depend on the sporting context. While a simple metric such as ‘Win or Loss’ can be used in both team and individual sport, ‘performance’ is dictated by physical, psychological, technical and tactical aspects of play. Further, the assessment of each aspect of “performance” will be dependent on the method used.

**Defining Injury**

Between study comparisons of injury statistics can be difficult with disparities in definitions, study design and methodology, and exposure quantification. As noted by Finch (1997), “the success of any sports injury surveillance system and its wide scale applicability is dependent upon valid and reliable definitions of sports injury, injury severity and sports participation”. Naturally, injury surveillance can lead to deep level questioning such as; what is a meaningful measure of exposure to injury risk?, what is a sports injury?, what threshold best represents a meaningful injury for surveillance and incidence quantification purposes? However, the fundamental question that must be addressed in research which investigates incidence, nature, causation, treatment and prevention of injuries sustained during sport and exercise is ‘what constitutes an injury?’ The answer to these questions directly depends on the definition of injury adopted. According to Fuller, operational definitions of sport injury should contain four criteria: 1) conditions which should be counted as an injury, 2) how the severity of the injury will be measured, 3) classification of injury location and pathology and 4) detailing the underlying mechanism.
**Injury Classification**

While disagreement between sports injury research studies have sparked discussion by the sports medicine community for some time, it’s only been in the past decade that large steps forward have been taken to construct global consensus of sport injury definition. This has resulted in multiple, sport-specific injury consensus’ including soccer (association football)\(^4\), rugby union\(^5\), rugby league\(^6\), cricket\(^7\) and the Olympic Games\(^8\). By in large, global consensus has gravitated towards soccer injury definitions. Therefore, the following operation definition of injury, adapted from Fuller et al.\(^4\), is largely accepted:

“Any physical complaint (caused by a transfer of energy that exceeds the body’s ability to maintain its structural and/or functional integrity) sustained by an athlete during completion or training directly related to the sport or exercise activity investigated, irrespective of the need for medical attention or time-loss from athlete activity. An injury that results in a player receiving medical attention is referred to as a “medical attention” injury and an injury that results in a player being unable to take a full part in future football training or match play as a ‘time loss injury’.”

While establishing a consensus injury definition is important, organizations are less concerned with the theoretical conceptualization of injury and more with a standard definition for comparison across investigations or years. This is particularly important for organizations (e.g., team, institution, governing body) which are investigating the effectiveness of prevention programs. Unfortunately, the quantification of injury statistics is often used for justification of job services as well (e.g., justification of medical and fitness practitioners), therefore if consensus definitions are not set, recording of injury data may be manipulated.

**Injury Severity**

Organizations investigating injury must take into consideration the severity of injury for both classification and analysis purposes. Firstly, not all occurrences which require medical attention will constitute an injury, likewise, not all injuries will be reported to medical professionals. Mostly through,
the primary concern is tracking injuries which result in the athlete losing training or competition time (i.e. time-loss). Therefore, the following definition of injury severity is accepted:

“The number of days elapsed from the date of injury to the date of the athletes return to full participation in training and availability for competition.”

Season-Ending Injuries

Season-ending, career-ending, non-fatal catastrophic and fatal injuries present some difficulties in analyzing sports injuries. As one can imagine, the inclusion of such cases would misrepresent injury severity averages if grouped in the same analysis as time-loss injuries. Regarding career-ending injuries, it is not possible to anticipate the length of an athlete’s career or speculate on the time lost due to the injury. Therefore, it’s best to include these rare cases in a separate analysis.

Injury Causation

Injuries should be categorized as either acute onset or gradual onset. Acute onset or traumatic injuries are injuries which are caused by a single, identifiable event. Gradual-onset injuries are caused by repeated micro-trauma without evidence of a single, identifiable event. The term-gradual onset is preferred to ‘overuse’ as it cannot always be established that the injury occurred due to true overuse or if it is related to a level of inactivity (i.e., underuse) and therefore a result of being under-prepared.

Injury Analysis Methods

Many statistical methods have been used in sport science literature to assess risk factors (e.g., workload, sleep, wellness, schedule congestion etc.) of injury including: mean difference (injured vs. non-injured groups), correlation, linear regression, logistic regression, multiple regression, and general and generalized linear mixed effect regression modeling (LMER), and generalized estimating equations (GEE). In addition, simple estimates of injury incidence are often derived from counts of injured and uninjured athletes. Likely the most popular methods of assessing risk factor association are use of
statistical procedures such as logistic regression, Poisson regression and proportional hazards (i.e., Cox) regression, which assess risk factor associations in the form of odds ratios, rate ratios, and hazard ratios, respectively. However, the nature of the data being analyzed and the questions being answered will ultimately determine the analytical procedure used. As discussed in Chapter 1, complex phenomena such as injury are resultant of many interrelated and dynamical factors, described by Bittencourt et al. as a ‘web of determinants’. Practitioners should take caution in assuming causality, especially when a single risk factor is assessed. Further, the meaningfulness and effect of that risk factor must be clearly established. However, little consensus exists regarding the appropriate methods by which injury causality should be established.

Prevalence vs. Incidence

In researching and reporting sport injuries to key stakeholders, it’s important to differentiate between prevalence and incidence. *Prevalence* can be thought of as the proportion of athletes on a team who have an injury during at a discrete time point. For example, we could report to the coach that 3 out of 23 players or approximately 13% of athlete are injured currently. In contrast, *incidence* refers to the number of new injury occurrences during a predefined period (e.g., day, week, month, season). A simple reporting of injury incidence is telling the coach “7 new injuries occurred over the preseason”. Although it’s important to differentiate between incidence and prevalence, most research is centered around reporting measures of incidence.

Assessing Injury Risk and Odds

Injury *risk*, *rate*, *odds* and *hazards* are proportional measures which are used to assess factors which are associate with injury and can be presented either absolutely (i.e., incidence or injured athlete proportion) or relatively (i.e., difference or ratio). As described by Hopkins et al., injury risk is calculated to identify what proportion of a group (e.g., team, starters, midfielders) has sustained an injury or to assess the probability of an athlete sustaining an injury within that group. Risk is typically expressed as either a
decimal fraction or percent of athletes which have sustained an injury. Risk is often an important metric the public and specifically parents who would be interested in knowing the risk of their child sustaining a serious injury if participating in a sport. For example, if in a youth soccer league consisting of 300 athletes, a total of 24 athletes sustain a serious injury over the course of the season, the reported risk of is \( \frac{24}{300} = 0.08 \) 8%.

Odds are somewhat less intuitive. The odds of injury is defined as the probability of an injury occurring divided by the probability of an injury not occurring. In the example provided, the number of injured athletes (i.e., 24) is divided by the number of athletes who did not sustain an injury (i.e., 276), therefore the odds of sustaining an injury is 0.087.

**Absolute vs. Relative Risk**

In the abovementioned scenario, suppose the parent hears that another youth soccer league in town implements individualized warm-up programs and wants to compare the risk and/or odds that his/her child while sustain an injury in this league compared with the other. The parent finds out that in the other local soccer league of 300 athletes, a total of 19 athletes sustained a time-loss injury last year. Therefore, she determines the risk of injury is 6.3% and the odds of sustaining a serious injury are \( \frac{19}{300-19} = 0.067 \). So, its concluded that the warm-up regimen reduces the absolute risk of injury by 2.4%, when compared to no warm-up. When expressed another way, the relative risk of injury is 72% or nearly \( \frac{3}{4} \) of the risk when compared with no-warm up.

Relative risk reduction is a measure of how much risk is reduced due to the intervention. When results are expressed relatively, it’s easy to overvalue the efficacy of an intervention. Take for example the 3rd row in Table 3. The relative risk reduction is 50%, which can be quite misleading given the absolute risk reduction was a mere 1%. As suggested by Akobeng and colleagues, absolute risk reduction is a more useful tool than relative risk reduction when assessing the efficacy of an intervention. This is particularly important to consider in injury research, as injury risks can be quite low. Although a relative reduction
can be quite large in some instances, these can be deceiving and can negatively impact a cost-benefit analysis, when assessing injury prevention strategies, absolute risk reduction must be considered.

**Table 0-2 - Relationship Between ARR, RRR and NNT**

<table>
<thead>
<tr>
<th>Risk of Injury</th>
<th>Absolute Risk Reduction (ARR)</th>
<th>Relative Risk Reduction (RRR)</th>
<th>Numbers Needed to Treat (NNT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>League 1 [control]</td>
<td>League 2 [warmup]</td>
<td>Control Risk- Intervention Risk</td>
<td>ARR/control group risk</td>
</tr>
<tr>
<td>8.7%</td>
<td>6.3%</td>
<td>8.7%-6.3% = 2.4%</td>
<td>6.3%/8.7% = 72%</td>
</tr>
<tr>
<td>72%</td>
<td>40%</td>
<td>72-40% = 32%</td>
<td>40%/72% = 55%</td>
</tr>
<tr>
<td>2%</td>
<td>1%</td>
<td>2%-1% = 1%</td>
<td>1%/2% = 50%</td>
</tr>
</tbody>
</table>

**Assessing Injury Rate and Hazards**

While risk and odds statistics alone can be useful, comparisons between risk factor groups (i.e., starter vs. reserve) require some manipulation to the denominator as calculation can be difficult when exposure times are different. Consider the comparison of injury risk of multiple teams throughout a single season. Clearly, exposure times will vary greatly between teams over the course of a season due to differing roster depths and training times. Calculating injury rates, which factor the number of injuries divided by a total exposure for a given time-period, are a way to overcome this limitation. As shown in Table 4, which
displays injury and exposure data for 7 different NCAA teams, exposure times are quite different between teams. Take for example the comparison of Team 1 and Team 3. Team 3 has 17 more athletes being assessed than Team 1, resulting in more than double the amount of session exposures (i.e., 987 vs. 2,089 athlete-exposures) and exposure hours (i.e., 1,872 vs. 4,409 athlete exposure-hours). By standardizing the denominator to a common expression such as 1,000 athlete-exposures or 1,000 athlete exposure-hours, the researcher can make comparisons across groups. Common comparisons between groups come in the form of difference or ratio functions, which allow the researcher to determine the effect of risk factors.

For example, the injury rate difference between Team 1 and 3 is 8 injuries for every 1000 athlete-exposures (i.e., 33.4/1000 – 25.4/1000 = 8/1000 athlete-exposures) or 5.6 injuries for every 1000 athlete exposure-hours (i.e., 17.6/1000 – 12/1000 = 5.6/1000 athlete exposure-hours). If expressed as a ratio, the rate of 33.4 and 25.4 injury incidences per 1000 athlete-exposures results in a rate ratio of 1.31 (i.e., 33.4/25.4 = 1.31). Therefore, the rate of injury is 1.31 times greater for Team 1 than Team 3 over the course of 1 season.

**Table 0-3 - Exposures and Injuries Table**

<table>
<thead>
<tr>
<th></th>
<th>Team 1</th>
<th>Team 2</th>
<th>Team 3</th>
<th>Team 4</th>
<th>Team 5</th>
<th>Team 6</th>
<th>Team 7</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Athletes</td>
<td>18</td>
<td>20</td>
<td>35</td>
<td>20</td>
<td>21</td>
<td>19</td>
<td>20</td>
<td>153</td>
</tr>
<tr>
<td>Exposures</td>
<td>987</td>
<td>1752</td>
<td>2089</td>
<td>1073</td>
<td>1630</td>
<td>1066</td>
<td>1085</td>
<td>9682</td>
</tr>
<tr>
<td>Exposure Hours</td>
<td>1871.78</td>
<td>3714.48</td>
<td>4409.34</td>
<td>2314.11</td>
<td>4343.83</td>
<td>1911.11</td>
<td>1740.36</td>
<td>20305</td>
</tr>
<tr>
<td>Total Injuries</td>
<td>33</td>
<td>64</td>
<td>53</td>
<td>22</td>
<td>23</td>
<td>32</td>
<td>15</td>
<td>242</td>
</tr>
<tr>
<td>Time Loss</td>
<td>7</td>
<td>9</td>
<td>20</td>
<td>20</td>
<td>22</td>
<td>6</td>
<td>10</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>55</td>
<td>33</td>
<td>2</td>
<td>1</td>
<td>26</td>
<td>5</td>
<td>148</td>
</tr>
</tbody>
</table>
While rate ratio are helpful in assessing relative differences over a typical length of time such as 1 season, hazards are calculated to express injury rate over shorter periods of time such as days or hours. By manipulating the time period by which an injury rate is expressed, a rate becomes known as a hazard. This is also known as the instantaneous risk ratio and is represented as the number of injuries that will be sustained over a predefined unit of time. Expressing injury rates through shorter periods of time can become use as suppose a coach want to be informed on the average injuries per month or year.

**Association vs. Prediction**
Differentiating between association and prediction has important implications for reporting and interpretation of sports injury data. Often, these terms cause confusion and are misused in sport science and medical literature. In association research, the primary objective is to confirm a hypothesis that the identified risk factor is associated with an outcome (i.e., injury, disease, performance outcome). In contrast, prediction is utilized to forecast future events and is often used for making practical decision in the field. For example, predicting injury or performance outcome from one or many precipitating factors. It’s important to acknowledge that association analysis infers on a population level and therefore is normally not particularly useful in decision-making at the individual level. Association measures such as correlation, odds ratio’s and relative risk infer upon the strength and nature of an association, while predictive outcomes such as ROC analysis (i.e., sensitivity, specificity, positive prediction value, negative prediction value) and likelihood ratio may assist in making practical decisions on an individual level. For example, if modifiable risk factors such as a workload metric relative to the athlete norm (e.g., acute to chronic workload ratio (ACWR) or exponentially-weighted moving average (EWMA) of 4-week chronic baseline) offers sufficient injury predictive value (i.e., adequate sensitivity and specificity), that metric may be used to alter training prescription. However, simply showing an association between a risk factor and outcome, especially of low-magnitude, offers little to individual decision-making ability.
The idea of identifying markers which are strongly associated with an outcome of interest, using the marker to identify individuals at risk, and then implementing preventative programs or countermeasures is appealing. As noted above, with the wealth of technologies and tools available to quantify stress and response, establishing individualize risk factors and subsequent preventative programs is a real prospect. However, sport science and medicine practitioners must acknowledge that a strong association between a risk factor and an outcome, such as injury, is a necessary condition for successful prediction, but is not necessarily a sufficient one. For a score, test, factor, biomarker to have ‘predictive power’ or ‘classification value’, a remarkably strong association is essential. According to some, for a test to have real predictive power, which is represented by adequate sensitivity and specificity, an odds ratio would need to be somewhere in the order of 25 to 100. This of course is rarely (if ever) seen with one single marker due to the inherent multi-causality and complexity surrounding this phenomenon. This concept was demonstrated by Pepe et al. who addressed the limitation of odds ratios in assessing performance of diagnostic, prognostic and screening markers. They demonstrated that with an odds
ratio as high as 3 (i.e., odds are 3 times higher) and a false positive rate of just 10% resulted in only identifying 25% of true cases positively (Figure 21). In essence, when assessing effect statistics such as odds and risks it’s important to acknowledge very large ratios are necessary for a risk factor to discriminate between injury outcomes. According to Pepe et al. an odds ratio of >16 would be necessary to determine a risk factor has an acceptable level of accuracy.

Predicting Injury

As statistician George Box once remarked, “all models are wrong, but some are useful.” This is a fundamental principle that sport scientist should consider when interpreting model outputs and defining the diagnostic utility of a marker or test. It is very unlikely that a monitoring tool or assessment will attain perfect predictive accuracy of the outcome of interest. With that being said, sport scientist must be able to interpret diagnostic utility so that confident recommendations can be made to decision-makers.
“All models are wrong, but some are useful” – George Box

As mentioned previously, there are a number of classification models which are currently used to predict injury (binary outcome), however logistic regression\textsuperscript{58} is the most prominent method. However, other classification methods such as multiple logistic regression\textsuperscript{88}, generalized linear mixed effect regression modeling (LMER)\textsuperscript{222}, and generalized estimating equations (GEE)\textsuperscript{30,31,459} are used. A common method of assessing model accuracy or diagnostic utility of a marker or test is the construction of a confusion matrix, which accounts for predicted vs actual outcomes. Common values that can be of use to practitioners include sensitivity, specificity, positive and negative predictive value and positive and negative likelihood ratio. Sensitivity is defined as the ability of a test or marker to correctly identify an outcome, while specificity refers to the ability of a test or marker to correctly identify no outcome.\textsuperscript{430} A test or marker with 100% sensitivity correctly identifies all athletes with an injury. If for example, a particular test only accurately predicts 70% of actual injuries (true positives), 30% of injuries go undetected (false negatives).\textsuperscript{467} Similarly, a marker with a high specificity value correctly identifies athletes who do not sustain an injury. If a test or marker has high sensitivity but low specificity, athletes who are not actually at risk of injury may receive inappropriate intervention. Positive predictive value refers to how likely an athlete will sustain an injury given a positive predicted injury occurrence.\textsuperscript{467} The value is useful because it informs about the likelihood of an athlete sustaining an injury if an injury is predicted. In contrast, negative predictive value refers how likely the athlete is to not sustain an injury given the predicted value being no injury.\textsuperscript{467} Ultimately, it is up to the high-performance team to decide acceptable levels of accuracy. The usefulness of a test or assessment will be likely context-specific.

Emerging Trends in Analytics

Machine Learning Techniques and Uses

Defined by James and colleagues\textsuperscript{468}, statistical learning or machine learning (ML) refers to a techniques utilized by a broader field, called artificial intelligence. Artificial intelligence is defined as “a branch of
computer science dealing with the simulation of intelligent behavior in computers. As shown in Figure 23, ML is a type of artificial intelligence. ML has received a great deal of attention in the past decade due to advances in computational power.

There are two main reasons to use statistical or machine learning techniques, prediction or inference. Most of the time, the function which connects the output to the input in a model is not known, therefore the relationship must be estimated. When predicting and outcome, the primary interest is not in understanding how the input variables are connected to the output variable, but instead in generating the most accurate estimation of the output (e.g., injury). When modeling for inference, the concern is understanding how the output changes in response to input variables. Understanding the relationship between the output and input and specifically how the output changes as a function of the input is the primary goal. When modeling for inference, we usually want to establish which input variables are associated with the output, specifically identifying the stronger predictors out of a potentially large number of possible parameters. We may want to investigate the nature of the relationship between the output variable and the input. This is particularly useful in sport injury research, as it is important to establish the nature of relationships between predictors (i.e., linear or non-linear). Generally, the question generated by the team or organization governs whether prediction or inference is sought, and therefore types of machine learning algorithms which are used. For example, if the purpose of modeling is to understand injury causality to further develop more robust injury prevention practices, inferential modeling is preferred. However, if an organization is merely concerned with predicting the likelihood of injury during a given time period, maximizing predictive accuracy is the primary concern. This impacts upon the type of model used, as all models range from flexible (i.e., often times non-linear models) to (i.e. typically linear models). Practitioners must weigh the pro’s and con’s when selecting models as more rigid models such as linear regression are limited in detecting more complex relationships, especially if their nature is non-linear, however rigid models are easy to interpret (i.e., beta coefficient). By contrast, flexible models such as support vector machines and artificial neural
networks are highly flexible and can model non-linear data structures, however the relationships between the input and output variable can be very difficult or impossible (e.g. black box algorithms) to interpret. Therefore, these models are not useful in drawing practical insights.

Supervised vs. Unsupervised

Machine learning tools are typically classified as either supervised or unsupervised. Supervised ML algorithms predict or estimating an output based on one or more inputs, that is for every given predictor observation there is an associated outcome observation. For example, supervised learning would consist of using a set of variables pertaining to athlete workload (e.g., total distance and average speed) and sleep (e.g., sleep duration and efficiency) to predict injury (i.e., output). Supervised learning algorithms are best for modeling inference, so when we want to understand the relationships between predictor and response variables. Examples of classic supervised models that are supervised and easy to interpret are simple linear and logistic regression. As mentioned, less flexible models which are more difficult to interpret, yet still supervised are algorithms such as support vector machines.

Figure 21 - ML Hierarchy
Unsupervised learning techniques do not contain outputs in the model, therefore the structure of input variables is modeled. In using unsupervised methods, we are interested in the relationships between variables or observations since a response variable is not present. Typical unsupervised models include clustering (e.g. k-means clustering or ward hierarchical clustering) and dimensionality reduction (e.g., principle component analysis) techniques. For example, modern player tracking technologies export sometimes 30-50 dimensions (i.e., variables, parameter, attributes) pertaining to an athlete’s workload during a session (e.g., distance in speed zones [walk, jog, run, sprint], velocity in speed zones, count of accelerations and decelerations in speed zones, etc.). The sheer volume of dimensions can become overwhelming to the sports scientist and especially for coaches and other key stakeholders. Therefore, unsupervised dimensionality reduction techniques can be used to find hidden structures in the data and reduce parameters. Reducing dimensions is necessary to condense dimensions which are alike in nature as a more effective model can be generated.468
Classification vs. Regression

Variables or dimensions are characterized as either quantitative (continuous) or qualitative (categorical). Problems which are regression-based will have quantitative variables as the response. Examples may be predicting metabolic cost of activity or number of goals scored in a season. In contrast, classification problems are those which have a qualitative or categorical variable as the response. Examples may be predicting if an athlete sustains an injury (yes or no), if a team will win (win vs. loss).

Relationships Between Risk Factors
An important step in guiding preventative practices is first understanding sports injury causality, which as discussed earlier, is highly complex and is exemplified by a multi-causal ‘web of determinant’. The notion of structuring sports injury prevention practices around single risk factors is naïve. In fact, sports injury researchers should seek to establish how risk factors interrelate, which can uncover underlying mechanisms which contribute to injury susceptibility. Although several analytical methods exist (e.g., multi-level modeling, statistical learning techniques, structural equation modeling), establishing the interrelatedness of factors in relation to an outcome is generally known as mediation and moderation analysis. Model illustration of moderation and mediation can be found in Figure 26. According to Armstrong et al., mediators are defined as biological, social and psychological modifiers that act on stressor to alter the level of physiological strain experienced. Although the definition presented by Armstrong is in the context of homeostatic perturbation resultant of stress, it can be further extrapolated to not just alterations in physiology, but an alteration of sports injury risk profile.

In a commentary by Windt et al., mediators and moderators were described more simply through an analogy of a ‘domino’ or ‘dimmer switch’, respectively. In this analogy, mediating or ‘domino’ factors are intermediary, in that they explain the association between a predictor and an outcome. Examples of mediators which are biological in nature may include neuromuscular fatigue, which often precedes injury. Social mediators may include situations such as encouragement from parents or an aggressive tone from a coach. Psychological mediators refer to factors such as mood state or arousal level, which can be affected by factors such as negative life events or academic stress. In contrast, factors which may moderate or ‘dim’ the effect of a risk factor on injury outcome are referred to as moderators, interactions, or effect modifiers. Simple examples of factors which may moderate the effect of increased physical demands of a match on injury risk are physical fitness factors such as aerobic endurance and strength or behavioral factors such as hydration or sleep behaviors. Perhaps one of the most apparent and studied moderating factors in sport science literature currently is the moderating effect of chronic load or aerobic fitness plays on acute spikes in workload (e.g., match or
marathon). Although coaches have understood for ages that increasing physical capacity can reduce its likelihood, researchers have begun to show this relationship statistically in team sports. While some have indirectly measured the moderating effect of aerobic fitness on injury, high chronic workload (i.e., 4-week average workload) is often analogous to a high physical capacity, although this represents an inferential leap. Nevertheless, improving causal mechanistic understanding has the potential to drive more targeted and efficacious preventative programs. By conceptualizing factors associated with injury as mediators and moderators and then quantifying their effect on direct causal factors (i.e., neuromuscular fatigue), practitioners and coaches can prescribe interventions based on physical screening or monitoring practices. For example, establishing the moderating effect of sleep characteristics (i.e., sleep duration and aspect of sleep quality) on injury risk during high-load sessions can inform targeted sleep prescription strategies the night prior to the session or allow the coach to modify training intensity real-time.

Figure 23 - Moderation Model
Threats to Construct Validity

Construct validity is a principal consideration for sport scientist conducting research in the field as it represents the ability of the measurements being taken to represent the underlying construct or theory developed to explain observations. Construct validity refers to overall validity, or the extent to which the test measures what it was designed to measure. To be valid, physical performance tests, physiological measurements, questionnaires, as well as, player tracking and sleep technology should measure abilities important in the sport, produce repeatable results, and be meaningful. For example, the researchers are interested in assessing the effect of workload on injury risk. This is a very common research question in contemporary sport science research. However, ‘workload’ or ‘training load’ is used synonymously with a range of workload constructs such as total distance covered, session load (Session Load [AU] = session RPE [AU] x session duration [min]) and mechanical load. Also, a key barrier in sport injury research is an inadequate operational definition of an injury. Consider a scenario where injury classification has not been adequately defined to practitioners recording injury characteristics. In this instance, the lack attention to constructs of interest is a threat to the validity of the research and any claims associated with it. Table 5 identifies 9 separate threats to construct validity, defines the threat and offers an example which might occur in the field of sport science.

Table 0-4 - Threats to Construct Validity

<table>
<thead>
<tr>
<th>Threat</th>
<th>Definition</th>
<th>Sport Science Example</th>
</tr>
</thead>
</table>

Figure 24 - Mediation Model
<table>
<thead>
<tr>
<th>Inadequate Preoperational Explication of Constructs</th>
<th>A threat to validity that occurs because researcher did an inadequate job of defining the construct idea.</th>
<th>Injury Definition – time loss vs. non-time loss, severity, type (e.g., lower extremity vs. upper extremity), location (e.g., shoulder, thigh, ankle)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mono-Operation Bias</td>
<td>A threat to construct validity that occurs when the researcher relies on only a single implementation an independent variable, program or treatment in the study.</td>
<td>Inferring a cold-water immersion intervention is effective in reducing muscle soreness, when only that specific cold water immersion protocol (e.g., 8 min at 10°C) was effective.</td>
</tr>
<tr>
<td>Mono-Method Bias</td>
<td>A threat to construct validity that occurs because the researcher used only a single method of measurement.</td>
<td>Using a single measure of fatigue such as athlete-reported fatigue rather than including other aspects such as biomarkers, neuromuscular or physical performance</td>
</tr>
<tr>
<td>Interaction of Different Treatments</td>
<td>Threat to validity caused by experiences outside of researchers study that might have contributed or influenced the study/measurement outcome</td>
<td>Injury rate reduction determined to be resultant of pre-match warm-up intervention, however, athletes also significantly improved sleep characteristics during the same period.</td>
</tr>
<tr>
<td>Restricted Generalizability Across Constructs</td>
<td>Threats to validity caused by unintended consequences that researchers were not prepared to measure.</td>
<td>Recovery intervention improves next-day perceived fatigue, however recovery protocol improved sleep quality rather than having direct effects on measure of fatigue.</td>
</tr>
<tr>
<td>Confounding Constructs with Levels of Constructs</td>
<td>A threat to validity associated with using the wrong dose (intensity/frequency) of intervention.</td>
<td>Defining injury as the outcome, when only a specific level of injury (e.g., time-loss injuries only) is being assessed</td>
</tr>
<tr>
<td>Hypothesis Guessing</td>
<td>A threat to validity caused by study participants trying to guess the study purpose and thereby change their behavior based on their guess.</td>
<td>Athlete’s increased sleep reduces injury risk, therefore increases sleep based on guess</td>
</tr>
<tr>
<td>Evaluation Apprehension</td>
<td>A threat to validity caused by the participant's anxious response to being tested.</td>
<td>Athlete's don’t report injuries because they are apprehensive about perceived negative repercussions</td>
</tr>
<tr>
<td>Researcher Expectancies</td>
<td>A threat to validity caused by the experimenter's bias being</td>
<td>Sport scientist believes recovery intervention will be effective and therefore talks about positive aspects of</td>
</tr>
</tbody>
</table>
injected into to some aspect
of the study.  
the recovery intervention. In turn, the
athlete perceives improved outcomes 
and reports less pain and/or soreness.

### Table 0-5 - Common Sport Science and Medical Calculations

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Calculation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Game Exposures</strong></td>
<td>Game Exposure = Games (n) × Players (n)</td>
</tr>
<tr>
<td><strong>Practice Exposures</strong></td>
<td>Practice Exposure = Practice Session (n) × Players (n)</td>
</tr>
<tr>
<td><strong>Match Exposure Time (hr)</strong></td>
<td>Match Exposure (hr) = Matches (n) × Players (n) × Duration of Matches (hr)</td>
</tr>
<tr>
<td><strong>Practice Exposure Time (hr)</strong></td>
<td>Practice Exposure (hr) = Practice Session (n) × Players (n) × Duration of Training Session (hr)</td>
</tr>
<tr>
<td><strong>Incidence Rate Ratio (Game vs. Practice)</strong></td>
<td>( IRR = \frac{\sum \text{Total Game Injuries}}{\sum \text{Game AE}} \times \frac{\sum \text{Practice Injuries}}{\sum \text{Practice AE}} )</td>
</tr>
<tr>
<td><strong>Risk Ratio (Game vs. Practice)</strong></td>
<td>( RR = \frac{\sum \text{Athletes with Game Injuries}}{\sum \text{Games}} \times \frac{\sum \text{Athletes with Practice Injuries}}{\sum \text{Practices}} )</td>
</tr>
<tr>
<td><strong>Game Availability</strong></td>
<td>Game Availability (%) = 100 – ((# of games absent/Total n of games) X 100)</td>
</tr>
<tr>
<td><strong>Practice Availability</strong></td>
<td>Practice Availability (%) = 100 – ((# of practices absent/Total n of practice sessions) X 100)</td>
</tr>
</tbody>
</table>

### Table 0-6 - Injury and Exposure Definitions (Adapted from Fuller (2006))

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Injury</strong></td>
<td>Any physical complaint sustained by a player that results from a match or training, irrespective of the need for medical attention or time-loss from football activities. An injury that results in a player receiving medical attention is referred to as a “medical-attention” injury and an injury that results in a player being unable to take a full part</td>
</tr>
</tbody>
</table>
in future football training or match play as a “time-loss” injury.

**Recurrent Injury**

An injury of the same type and at the same site as an index injury and which occurs after a player’s return to full participation from the index injury. A recurrent injury occurring within 2 months of a player’s return to full participation is referred to as an “early recurrence”; one occurring 2 to 12 months after a player’s return to full participation as a “late recurrence”; and one occurring more than 12 months after a player’s return to full participation as a “delayed recurrence.”

**Injury Severity**

The number of days that have elapsed from the date of injury to the date of the player’s return to full participation in team training and availability for match selection.

**Match Exposure**

Play between teams from different teams.

**Training Exposure**

Team-based and individual physical activities under the control or guidance of the team’s coaching or fitness staff that are aimed at maintaining or improving players’ football skills or physical condition.
REFERENCES


240. Why is Sleep Important? Available at: https://www.nhlbi.nih.gov/node/4605.


Chapter 2: Injury and Psychological Wellbeing in Women’s and Men’s NCAA Division I Soccer

ABSTRACT

Purpose: To examine injury risk, rate, and characteristics, as well as, physical and psychological wellbeing in women and men’s student-athletes over the course of a national collegiate athletics association (NCAA) soccer season. Methods: Injuries, mood, sports-related anxiety, sleep characteristics and physical activity disablement was longitudinally assessed for 256 athletes from 12 separate NCAA division I teams. Injury incidences, rates, percentages, and athlete risk was assessed by sex, session type and team. Multi-level models were used to assess seasonal and sex differences in psychological and physical wellness inventories. Results: Women’s collegiate soccer players experienced 2.05 (95%CI 1.20-3.51, p<0.001) times the rates (per 1000 exposure-hours) of overuse injury, more average nightly sleep disturbances (b=0.49, p=0.01, ES=0.37), higher levels of global sleep dysfunction (b=0.99, p<0.001, ES=0.52), sports-related anxiety(b=3.9, p<0.001, ES=0.67) and physical activity disablement (b=8.5, p<0.001, ES=0.87) over a season compared to their male counterparts. The rate of non-contact time-loss injury for women’s soccer was 38% less when compared with men’s soccer (IRR: 0.62, 95%CI 0.39-0.98, p=0.03). 48% of all injuries were non-time-loss. Total mood disturbance (Profile of Mood States) was significantly elevated at time points when athletes were in school (b=5.76-7.99, ES=0.26-0.36) and 47% of athletes were classified as poor sleepers (PSQI>5). Conclusions: Seasonal and sex differences in injury rate and characteristics, as well as, physical and psychological wellbeing are apparent in collegiate soccer. Athlete surveillance practices should seek to bolster traditional epidemiological injury research with physical and psychological assessment, providing a more detailed and complete view of athlete wellbeing.
INTRODUCTION

Player availability is critical for team performance as high match availability and lower time-loss injury burden is associated with team’s competitive success. While optimal team performance is reliant on player availability, optimal player performance requires both good physical and psychological condition. Soccer demands not only physical fitness and skill but a high degree of perceptual-cognitive ability, with athletes needing to dynamically react to their environment and integrate technical skill and tactical strategies under time constraint. Physical fatigue and maladapted psychological states due to mismanaged stressors can have detrimental effects on injury risk in sport.

Collegiate athletes are at considerable risk because of participation in a multitude of team and school-related functions (e.g. course-work, practice, competition, strength and conditioning, film study, etc.), which if not properly managed can contribute to fatigue accumulation over the course of a season and potentially lead to maladaptive physical and psychological states. This is a topic of interest for both sports medicine and performance researchers and practitioners, as well as, governing organizations.

Little is known regarding the psychological and physical health of collegiate soccer players over a competitive season. Evidence suggests that negative states or traits such as anxiety can play a role in sport injury and that employing strategies to combat negative emotional states can be an integral part of preparing athlete’s for optimal performance. The mechanistic understanding of the link between psychological stress and injury is a bidirectional interaction between attentional (e.g., narrowed visual field, distractibility) and somatic (e.g., muscle tension or fatigue, reduced coordination) aspects.

Injury and illness surveillance is recognized as an important initial step in establishing risk in sport. Detailed and consistent classification of injury allows comparison within and between organizations and sports, which is vital for moving the field forward with respect to understanding sports injury burden and implementing best-practices. Epidemiological research is an important part of determining the extent of the injury in sport, with information such as injury incidence rates, athlete risk,
severity and outcomes of injury offering detail on overall injury burden. However, a comprehensive view of seasonal burden mandates the investigation of athlete wellness, which can further elucidate both the physical and psychological impact of sport.

The seasonal impact of NCAA division I soccer should be understood by coaches, medical and performance practitioners to foster athlete-centered care and evidence-based practices. Comprehensive assessment of the effect of seasonal stressors on all dimensions of health including physical and psychological well-being is essential for guiding organizational best-practices. Therefore, the aim of the study was to examine injury risk, rate, and characteristics, as well as, health-related quality of life outcomes (i.e., sleep, mood, anxiety, physical activity disablement) in women and men’s student-athletes over the course of a national collegiate athletics association (NCAA) soccer season.

**METHODS**

**Participants.** A prospective cohort study of 256 NCAA division I athletes from 12 separate university teams was conducted over the 2016 (2 teams), 2017 (6 teams) and 2018 (4 teams) seasons. One-hundred and thirty-nine of the participants were female (age, 20±1 y; body mass, 64.7±6.1 kg; height, 166.8±6.2 cm; VO$_{2\text{max}}$, 46.8±4.0 ml·kg$^{-1}$·min$^{-1}$) and one-hundred and seventeen were male soccer players (age, 20±2 y; body mass, 77.4±5.1 kg; height, 179.9±6.5 cm; VO$_{2\text{max}}$, 53.8±4.1 ml·kg$^{-1}$·min$^{-1}$). All participants were medically cleared for physical activity by their respective university’s sports medicine department and free of any debilitating musculoskeletal injuries or contraindicated medical conditions. Institutional review board (IRB) and ethics approval was obtained from all institutions, with primary oversight and coordination provided by the University of <blinded for review> (IRB Approval ID: H17-134). All participants provided written informed consent prior to the season. When the participant was under the age of 18, parental consent was obtained.
**Injury/Exposure Classification.** Injuries were diagnosed by each team’s sports medicine staff and recorded by a single medical staff member (i.e., certified athletic trainer). Injuries were classified according to the current consensus statement on recording of soccer injuries, which states that an injury is “any physical complaint sustained by a player that results from a football (soccer) match or football (soccer) training, irrespective of the need for medical attention or time loss from football (soccer) activities.” Therefore, a range of injury classifications were considered including; medical attention, time-loss, non-contact, non-contact time-loss, overuse and illnesses. In addition to incidence, other pertinent information such as incidence type, severity, location and mechanism were recorded. Each team’s medical staff member recording injuries was supplied with both an injury record template and thorough instruction on injury classification practices prior to the start of data collection. All Injuries were documented daily and de-identified data were transferred to researchers. An athlete-exposure (AE) was defined as “1 student–athlete participating in 1 NCAA-sanctioned practice or competition in which he or she was exposed to the possibility of athletic injury, regardless of the time associated with that participation.” Non–time-loss injuries were those which were evaluated or treated by the medical provider but did not result in restriction from participation for more than 1 day. Time loss injuries were defined as an injury that “(1) occurred as a result of participation in an organized intercollegiate practice or competition, (2) required attention from an AT or physician, and (3) resulted in restriction of the student–athlete’s participation for 1 or more days beyond the day of injury.” Overuse injury is defined as “an injury caused by repeated micro-trauma without a single, identifiable event responsible for the injury.”

**Psychological Wellness.** Two-hundred and thirty participant’s mood, sports-related anxiety, sleep quality and physical activity disablement was longitudinally assessed during 6 distinct time points throughout the season. Inventories were administered by a member of each research team prior to preseason (baseline), directly after preseason (start of in-season), at week 4 and 8 of the in-season, end of regular season (start of postseason) and end of postseason play (when applicable). The POMS is a validated 65-item
The POMS assessed six mood or ‘feeling’ states: tension, anger, vigor, depression and fatigue and is robust for examination of individual mood states. Construct validity has been explored by Terry and colleagues for its use with both adults and adolescents. The Sport Anxiety Scale-2 (SAS-2) was used to measure cognitive and somatic trait anxiety. The SAS-2 is a 15-item inventory assessing anxiety, worry and concentration disruption with a total score ranging from 15-60. The SAS-2 has been previously used in women’s collegiate soccer to assess the effectiveness of mental skill training. PSQI has been used to assess perceived sleep quality in the collegiate student-athlete population and elite athletes. The PSQI consists of 19 items assessing subjective sleep quality, latency, efficiency, duration, and disturbances. The scoring for each component is combined for a Global Sleep Quality Score ranging from 0-21, with >5 being indicating general poor sleep quality. Wellness-Injury associations were assessed by investigating each global scores association with injury in the time period directly following assessment. The Disablement of the Physically Active (DPA) scale is a 16-item inventory assessing quality of life, impairment, disability and functional limitations. Good validity and reliability has been reported with the DPA instrument (intraclass correlation coefficient [ICC] = 0.94). The total DPA score assesses over disablement and is the sum of each of the subscales minus 16. DPA total scores range from 0-64, with higher scores corresponding with higher levels of disablement.

**Sleep Data.** Daily sleep behavior was assessed via the Karolinska Sleep Diary. The KSD is an eleven item questionnaire used to evaluate several facets of sleep, including quantity and aspects of perceived quality, such as ease of falling asleep, ease of awakening, overall perception of sleep quality, sleep disturbances, sleep calmness, and feeling of rest. The KSD questionnaire was electronically distributed daily and participants were encouraged to complete the questionnaire upon waking.

**Statistical Analysis.** Injury incidence rates were calculated by dividing the number of injuries by total exposures or exposure hours and reported as rate per 1000 exposures or hours. Coefficient of variation
(CV) within sexes was calculated for between team variability in injury rates by dividing the standard deviation of all injury rates by the mean of all injury rates. Sleep and wellness data are presented as mean ± standard deviation (SD). PSQI, SAS2, POMS and DPA changes over the seasonal time points and differences between sex were assessed using multi-level modeling. Univariate models were constructed with time-point only, sex only and time-point-sex interaction entered as fixed effects. Model random effects were clustered by player. Time point differences from baseline were divided by the between-subjects model standard deviation to determine a standardized effect size (ES). ES was interpreted according to the following thresholds: < 0.2 = trivial, 0.2–0.6 = small, 0.7–1.1 = moderate, 1.2–2.0 = large, and > 2.0 = very large. Statistical analyses and model plotting were conducted in R Studio (Version 3.2.5, R Core Team) under the lme4, jtools and ggplot2 packages.

**RESULTS**

*Injury Risk.* Of the 256 players included in this study, a total of 74.2% had a medical attention injury, 45.0% time-loss injury, 47.7% non-contact injury, 27.0% non-contact time-loss injury, 22.2% overuse injury and 11.7% incurred illness over a season.

*Overall Incidence Rates.* A total of 372 medical attentions were recorded from 12 team-seasons, with 154 time-loss injuries, 191 non-contact injuries, 87 non-contact time-loss, 81 overuse and 34 illnesses. Table 1 displays incidence rates for all injury classifications. There were no differences in incidence rates between women’s and men’s players for all injury classifications except overuse and non-contact time-loss injuries. Overuse injury rates were 1.87 (95%CI 1.10-3.20, p=0.01) times higher for females compared with males when exposures were considered, 2.05 (95%CI 1.20-3.51, p<0.001) times higher when exposure hours were considered. The rate of non-contact time-loss injury for women’s soccer was 37% less when compared with men’s soccer (RR: 0.62, 95%CI 0.39-0.98, p=0.03).
Incidence Rates by Session Type. Table 2 displays injury rates and rate ratios comparing match to training. Medical attention, time-loss, non-contact, and non-contact time-loss injury rates were 1.8-3.6 times higher during matches as compared to training for overall injuries and men’s injuries. Only medical attentions (IRR: 2.55 (1.94-3.34), p<0.001) and time-loss (IRR: 3.02 (1.86-4.93) injury rates were significantly higher during matches compared to training for women’s soccer.

Incidence Rates by Team. Table 3 displays injury rates (per 1000AE and 1000AEH) by team and the variability in injury rates between teams. Variability for all injury rates between teams for all classifications was higher for women’s soccer than men’s soccer. Additionally, between-team injury rates had higher variability when expressed per 1000 AEH compared to AE.

Incidence Proportions. Incidence proportions by injury type, mechanism of injury (MOI), severity, session type and body part are shown in Table 4. Contusions represented the largest proportion of injuries for women’s soccer (20.25%), while muscle strains represented the largest proportion of injury for men’s soccer (33.94%). Acute non-contact account for 38.79% of injuries for men while contact with another person represented 32.23% of injuries in women’s soccer. The highest proportion of injuries occurred in matches for both men’s (56.36%) and women’s soccer (47.93%). A large majority of incidences were ‘transient’ in nature with 45% of incidences for women’s soccer and 52% of incidences for men’s soccer non-time-loss. Both men’s and women’s soccer had larger incidence proportions for lower extremity areas such as hip/groin, upper leg/thigh, lower leg, ankle, and foot/toes. Illnesses represented ~10% of all incidences recorded for women’s soccer, while men’s soccer experienced ~5%.

Sleep, Anxiety, Mood and Disablement. Descriptive statistics for all wellness total scores and sleep diary responses are shown in table 5. Women’s soccer players had significantly higher average PSQI (b=0.99, p<0.001, ES=0.52), SAS-2 (b=3.9, p<0.001, ES=0.67) and DPA (b=8.5, p<0.001, ES=0.87) total scores when compared with men’s soccer players. Of note, 46.7% of athletes averaged greater than a global score of 5 on the PSQI. There were no differences in sleep diary measures except for sleep disturbances, with women reporting significantly higher nightly disturbances (b=0.49, p=0.01, ES=0.37) than men.
Descriptive statistics for wellness inventories and daily sleep diary are shown in table 5. Seasonal changes in PSQI, SAS2, POMS and DPA total scores by sex are shown in Figures 1-4. There were significant reductions in PSQI total score (improved sleep characteristics) at end of preseason (b=-.92, p<0.001, ES=-0.47) and end of postseason (b=-.82, p<0.001, ES=-0.42) time points compared to baseline. SAS-2 total score was significantly reduced at week 8 (b=-1.23, p=0.01, ES=-0.20), end of in-season (b=-2.04, p<0.001, ES=-0.36) and end of postseason (b=-1.75, p<0.001, ES=-0.29) time points compared with baseline. POMS total mood disturbance score was significantly elevated at all time points (b=5.76-7.99, ES=0.26-0.36) beyond baseline and preseason. DPA total score was significantly elevated at week 8 (b=2.42, p=0.02, ES=0.22) compared with baseline. There was a significant interaction between time point and sex for PSQI, with week 4 (b=-1.29, p=0.01, ES=-0.68) and end of in-season (b=-1.03, p=0.04, ES=-0.55) time points being lower for men as compared to women’s soccer players.

**DISCUSSION**

This study investigated injury alongside health-related quality of life in women and men’s NCAA division I soccer players over 12 team-seasons. Our findings present several key considerations for NCAA division I soccer stakeholders in addition to presenting several novel athlete surveillance considerations. A key finding of this study was that women’s collegiate soccer players experienced significantly higher rates of overuse injury, more nightly sleep disturbances, higher levels of global sleep dysfunction, sports-related anxiety and physical activity disablement over a season as compared to their male counterparts. Additionally, there is a noteworthy amount of variability in injury rates between teams with trends indicating 1) injury rate variability is higher when expressed as exposure hours as compared with exposures, 2) higher for women’s soccer teams as compared with men’s teams, and 3) markedly higher for overuse injuries as compared with all other injury classifications.

There is much debate over injury surveillance methods and injury classification. A substantial proportion of injury surveillance studies have focused on time-loss injuries only, with
reports also investigating injuries under the broader definition of ‘medical attention’. There is also consideration of an anatomical tissue injury classification defined as tissue damage caused by sporting activity, regardless of time-loss or medical attention. On the other end of the spectrum, there are strong calls for a more conservative injury classification of ‘missed-matches only’. Injury surveillance systems have an inherent paradox whereby conservative injury definitions improve reporting reliability allowing for better comparison across studies, however suffer from incomplete capture of injury burden in sport. Rather than limiting our investigation to one injury classification, our study took the novel approach of considering 6 separate injury classifications (medical attention, time-loss, all non-contact, non-contact time-loss, overuse and illness) and 2 different reporting methods (per exposure and per exposure-hour). In doing so, we allow for direct comparison with a range of injury epidemiological literature and provide a detailed investigation of injury burden in NCAA division I collegiate soccer.

Injury Rate

Time-loss injury incidence rates have been reported between 2.0-19.4 per 1000AEH for youth soccer and between 2.5-9.4 per 1000 AEH in professional soccer. Our findings indicate collegiate soccer injury rates are around the middle of these ranges with over time-loss injury rate of 4.1 per 1000AEH (Women: 3.69, Men: 4.57). Injury rates in the current investigation are substantially higher than a previous report of total medical attention (22.63-22.78 vs. 8.07-8.44/1000AE) and time-loss injuries (7.30-8.46 vs. 4.09-4.28/1000AE) in a large cohort (167 team-seasons) of collegiate soccer players. However, the rates presented in this investigation more closely align with previous reports in collegiate soccer from 1988/1989-2003/2004 time period. Nevertheless, expressing injury rates in only exposures can give an imprecise indication of true burden of injury as exposure duration can vary markedly.

Consistent with a prior report on NCAA division I soccer, non-time-loss injury represented a substantial portion of all recorded injuries. We found ~48% (Women: 44.96%; Men: 52.12% of all injuries did not cause time-loss of greater than or equal to 1 session. This is similar to reports by Roos et
al.\textsuperscript{25} on collegiate soccer athletes and Clausen et al.\textsuperscript{29} on adolescent female soccer athletes who report non-time-loss accounting for ~50\% and 36.6\%, respectively.

Differences between men’s and women’s injury rates were consistent with previous reports indicating there are no differences between sexes for medical attention injuries.\textsuperscript{25} However, when expanding injury classification, we did find significant differences in both non-contact time-loss injuries and overuse injuries. Consistent with prior investigations\textsuperscript{25,27}, injury rates were elevated in matches as compared with training for both medical attention and time-loss injury classifications, regardless of sex. Time-loss injuries have been reported in the range of 3.3 to 15.3 times higher during matches than during training in professional soccer and 2.3 to 4.9 times higher in youth soccer.\textsuperscript{27} We found collegiate soccer time-loss injury rates to be on the lower end of this spectrum with matches having a 2.22 time higher rate than training (Women: 1.94; Men: 2.54).

Interestingly, non-contact and non-contact time-loss injuries were not elevated in matches compared with training for women’s soccer. To our knowledge, these injury classifications have not been considered in previous comparisons therefore comparison are not possible. Further research should report a range of injury classifications, including all medical attention, non-contact, non-contact time-loss and overuse so that more detailed comparisons across studies are possible.

Of note, our results highlight that between-team variability in injury rates are 1) higher when expressed as exposure hours as compared with exposures, 2) higher for women’s soccer teams as compared with men’s teams, 3) substantially higher for overuse injuries as compared with all other injury classifications. These are important findings to consider when comparing injuries across studies and additionally suggests multi-team, multi-year studies are needed to capture true injury burden.\textsuperscript{8,24} Single or small team studies will be limited in generalizability.

Our findings are consistent with reports indicating lower extremity injuries assume the vast majority of injuries in collegiate soccer.\textsuperscript{22,25,28} Noteworthy, the proportion of muscular strains for men’s
soccer was double that of women’s soccer (33.94% vs. 16.53%). Although not directly assessed, it could be speculated this difference is likely related to the 38% higher rate of non-contact time-loss seen in men’s soccer as compared with women’s soccer. Similarly, women’s soccer had substantially higher proportions of soft-tissue inflammation and muscle spasm as compared with men’s soccer (28.51% vs. 4.24%). This is likely contributing to the significant differences noted in overuse injury rates, with women’s soccer 1.87-2.05 times as likely to incur an overuse injury as compared men’s soccer players.

**Physical and Psychological Wellbeing**

A key finding of our study is that women’s soccer athletes experienced higher levels of sports-related anxiety, sleep dysfunction and physical activity disablement over the course of the season compared with men’s soccer athletes. Additionally, substantial increases in total mood disturbance can be seen after preseason time point, regardless of sex. Support exists for a direct positive relation between injury and mood states of tension, anxiety, hostility and anger. Of note, both the baseline and preseason measurement time points occur before semester coursework has started. Although not investigated, academic demand may be a confounding factor affecting mood alterations throughout a competitive soccer season. Further research on the relationship between academic workload and psychological wellbeing is warranted.

Previous works have investigated sleep characteristics in a range of student-athletes and sports from one NCAA university. In agreement with Mah et al. our results indicate a substantial portion (42.4%) of collegiate student-athletes are poor sleepers, with 47% of athletes reporting higher than a PSQI global score of 5 (sleep dysfunction cutoff). Further, our results suggest the women’s NCAA soccer players on average are poor sleepers (PSQI=5.4), despite reporting average sleep durations of 7.8 hours per night. There were no differences found in men’s and women’s soccer daily sleep diary responses with respect to sleep duration and aspects of sleep quality, however women’s soccer players did report significantly higher rates of night sleep disturbance. This finding appears consistent with findings of global sleep dysfunction from the PSQI.
Competitive anxiety has received the most attention with regard to psychological wellbeing and has been the most consistent variable associated with sport injury occurrence. Our results indicate there is a steady decline in sports-related anxiety over the season, regardless of sex, with women’s soccer athletes experiencing significantly higher levels of anxiety compared to males throughout the season. These findings suggest strategies to address anxiety and stress management may be useful during early competition season and dedicated attention to women’s soccer athletes may be warranted. Cognitive-behavioral stress management training has been found to reduce injury and illness incidence (~50%) in 40 collegiate rowers, additionally resulting in a reduction in lost training time.

Previous works by Hoch et al. utilizing the DPA inventory to establish minimal detectable change (MDC) scores for the inventory and additionally track women’s soccer disablement over the course of collegiate spring soccer season. They found a change of at least 13 points on the DPA scales was needed for classification of a clinically meaningful change in disablement. Our results indicate a significant spike in disablement at week 8 of the season for both men’s and women’s soccer, however this increase was well below the previously established MDC by Hoch et al. and additionally was of trivial difference from baseline (ES=0.22).

Although there were no differences seen in overall medical attention or time-loss injury rates between men’s and women’s soccer, women’s soccer athletes reported significantly higher physical activity disablement over a college soccer season. This finding suggests physical activity disablement perception is closely linked with chronic injury as women’s soccer saw 2-fold greater rates of overuse injury than men’s soccer, with illnesses also being twice as high in women’s soccer than men’s soccer. This is an important consideration for medical providers as isolated medical attention or time-loss injury rates may not provide the best indication of overall physical wellbeing.

Although this study presents a comprehensive view of injury rates and characteristics over a range of injury classifications, alongside measures of wellbeing in collegiate soccer player, it is not without limitation. Firstly, only one medical provider reported injuries from each team therefore inter-
rater reliability could not be considered in this study. There are legitimate concerns over reporting reliability with studies utilizing broader injury classification terms such as this one, where reporters classification biases or motivation influence reporting and overall higher levels of subjectivity are introduced into the system.\(^2\)\(^1\) To address this, we presented a range of injury classifications common in injury research.

**CONCLUSION**

While injury surveillance is a vital piece of an injury prevention paradigm,\(^3\)\(^3\) monitoring practices should seek to bolster epidemiological injury research with psychological wellbeing assessment. Athlete surveillance should be attune to injuries but should additionally be sensitive to clinical symptoms experienced by athletes such as mood alteration, disablement and anxiety.\(^4\) Additionally, there is an inherent paradox between reliable injury classification and comprehensive capture of sports injury burden.

Acknowledgement:

The authors would like to express their sincerest thanks to the athletes, coaching staff, sports medicine staff and strength and conditioning staff of each participating institution for their time and commitment during the study. Additionally, the authors would like to gratefully acknowledge participating research personnel from all institutions involved. <contributors will be acknowledged after review>

**Figure Legend:**

Figure 1a – Pittsburgh Sleep Quality Index (PSQI) over time. * = significant (p<0.05) difference from baseline.

Figure 1b – Pittsburgh Sleep Quality Index (PSQI) over time by sex (men’s soccer vs. women’s soccer). * = significant (p<0.05) difference from baseline. $ = significant interaction between sexes

Figure 2a – Sports Anxiety Scale (SAS2) over time. * = significant (p<0.05) difference from baseline.
Figure 2b – Sports Anxiety Scale (SAS2) over time by sex (men’s soccer vs. women’s soccer). * = significant (p<0.05) difference from baseline. $ = significant interaction between sexes

Figure 3a – Profile of Mood States (POMS) over time. * = significant (p<0.05) difference from baseline.

Figure 3b – Profile of Mood States (POMS) over time by sex (men’s soccer vs. women’s soccer). * = significant (p<0.05) difference from baseline. $ = significant interaction between sexes

Figure 4a – Disablement of the Physically Active (DPA) over time. * = significant (p<0.05) difference from baseline.

Figure 4b – Disablement of the Physically Active (DPA) over time by sex (men’s soccer vs. women’s soccer). * = significant (p<0.05) difference from baseline. $ = significant interaction between sexes

Declaration of Interest:

The authors report no conflicts of interest. The authors alone were responsible for the content and writing of this article. They are thankful for the financial support for this research provided in-part by the National Collegiate Athletics Association (NCAA).

REFERENCES


**TABLES/FIGURES**

Table 1. Overall Injury/Illness Rates by Athlete Exposures and Exposure Hours

<table>
<thead>
<tr>
<th>Injury Definition</th>
<th>Overall IR per 1000 AE</th>
<th>Women’s IR per 1000 AE</th>
<th>Men’s IR per 1000 AE</th>
<th>IRR (95%CI) Women’s v Men’s</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Injury - Medical Attention</td>
<td>22.63</td>
<td>22.78</td>
<td>22.45</td>
<td>1.01 (0.82-1.24)</td>
<td>0.88</td>
</tr>
<tr>
<td>Injury – Time Loss</td>
<td>8.46</td>
<td>7.30</td>
<td>9.92</td>
<td>0.73 (0.52-1.03)</td>
<td>0.07</td>
</tr>
<tr>
<td>Injury – Non-Contact</td>
<td>10.22</td>
<td>9.59</td>
<td>11.02</td>
<td>0.87 (0.64-1.18)</td>
<td>0.37</td>
</tr>
<tr>
<td>Injury – Non-Contact Time Loss</td>
<td>4.56</td>
<td>3.60</td>
<td>5.78</td>
<td><strong>0.62 (0.39-0.98)</strong></td>
<td><strong>0.03</strong></td>
</tr>
<tr>
<td>Injury - Overuse</td>
<td>3.89</td>
<td>4.91</td>
<td>2.62</td>
<td><strong>1.87 (1.10-3.20)</strong></td>
<td><strong>0.01</strong></td>
</tr>
<tr>
<td>Injury - Illness</td>
<td>1.76</td>
<td>2.18</td>
<td>1.24</td>
<td>1.75 (0.80-3.86)</td>
<td>0.15</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Injury Definition</th>
<th>Overall IR per 1000 AEH</th>
<th>Women’s IR per 1000 AEH</th>
<th>Men’s IR per 1000 AEH</th>
<th>IRR (95%CI) Women’s v Men’s</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Injury - Medical Attention</td>
<td>10.96</td>
<td>11.5</td>
<td>10.34</td>
<td>1.11 (0.90-1.36)</td>
<td>0.31</td>
</tr>
<tr>
<td>Injury - Time Loss</td>
<td>4.10</td>
<td>3.69</td>
<td>4.57</td>
<td>0.80 (0.57-1.12)</td>
<td>0.20</td>
</tr>
<tr>
<td>Injury – Non-Contact</td>
<td>4.95</td>
<td>4.84</td>
<td>5.08</td>
<td>0.95 (0.70-1.29)</td>
<td>0.76</td>
</tr>
<tr>
<td>Injury – Non-Contact Time Loss</td>
<td>2.21</td>
<td>1.82</td>
<td>2.67</td>
<td>0.68 (0.43-1.07)</td>
<td>0.10</td>
</tr>
<tr>
<td>Injury - Overuse</td>
<td>1.89</td>
<td>2.48</td>
<td>1.21</td>
<td><strong>2.05 (1.20-3.51)</strong></td>
<td><strong>0.01</strong></td>
</tr>
<tr>
<td>Injury - Illness</td>
<td>0.85</td>
<td>1.10</td>
<td>0.57</td>
<td>1.92 (0.87-4.23)</td>
<td>0.10</td>
</tr>
</tbody>
</table>
Table 2. Overall Injury Rate Comparison between Match and Training for Athlete Exposures

<table>
<thead>
<tr>
<th>Injury Type</th>
<th>Match</th>
<th>Training</th>
<th>Rate Ratio IRR (95%CI)</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Injury - Medical Attention</td>
<td>42.02</td>
<td>14.22</td>
<td>2.95 (2.41-3.63)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Injury - Time Loss</td>
<td>17.09</td>
<td>4.71</td>
<td>3.63 (2.58-5.10)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Injury – Non-Contact</td>
<td>14.88</td>
<td>8.2</td>
<td>1.81 (1.34-2.46)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Injury – Non-Contact Time Loss</td>
<td>7.64</td>
<td>3.23</td>
<td>2.37 (1.50-3.72)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Injury - Overuse</td>
<td>4.22</td>
<td>3.75</td>
<td>1.12 (0.67-1.90)</td>
<td>0.668</td>
</tr>
<tr>
<td>Injury - Illness</td>
<td>1.21</td>
<td>2.01</td>
<td>0.60 (0.24-1.48)</td>
<td>0.262</td>
</tr>
<tr>
<td>Women’s Soccer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Injury - Medical Attention</td>
<td>38.45</td>
<td>15.10</td>
<td>2.55 (1.94-3.34)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Injury - Time Loss</td>
<td>13.26</td>
<td>4.39</td>
<td>3.02 (1.86-4.93)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Injury – Non-Contact</td>
<td>10.94</td>
<td>8.93</td>
<td>1.22 (0.79-1.89)</td>
<td>0.357</td>
</tr>
<tr>
<td>Injury – Non-Contact Time Loss</td>
<td>4.64</td>
<td>3.09</td>
<td>1.50 (0.75-2.99)</td>
<td>0.245</td>
</tr>
<tr>
<td>Injury - Overuse</td>
<td>4.64</td>
<td>5.03</td>
<td>0.92 (0.49-1.73)</td>
<td>0.799</td>
</tr>
<tr>
<td>Injury - Illness</td>
<td>1.66</td>
<td>2.44</td>
<td>0.68 (0.24-1.87)</td>
<td>0.429</td>
</tr>
<tr>
<td>Men’s Soccer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Injury - Medical Attention</td>
<td>47.52</td>
<td>13.20</td>
<td>3.60 (2.64-4.91)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Injury - Time Loss</td>
<td>22.99</td>
<td>5.09</td>
<td>4.52 (2.80-7.28)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Injury – Non-Contact</td>
<td>20.95</td>
<td>7.35</td>
<td>2.85 (1.83-4.42)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Injury – Non-Contact Time Loss</td>
<td>12.26</td>
<td>3.39</td>
<td>3.61 (1.96-6.66)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Injury - Overuse</td>
<td>3.58</td>
<td>2.26</td>
<td>1.58 (0.62-4.02)</td>
<td>0.331</td>
</tr>
<tr>
<td>Injury - Illness</td>
<td>0.51</td>
<td>1.51</td>
<td>0.39 (0.04-2.71)</td>
<td>0.284</td>
</tr>
</tbody>
</table>
Table 3. Women's and Men's Soccer Injury Rates per Athlete Exposure and Exposure Hours by Team

<table>
<thead>
<tr>
<th>Injury Definition</th>
<th>Team 1</th>
<th>Team 2</th>
<th>Team 3</th>
<th>Team 4</th>
<th>Team 5</th>
<th>Team 6</th>
<th>CV</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Women's Soccer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Injury - Medical Attention</td>
<td>48.56</td>
<td>12.20</td>
<td>17.07</td>
<td>38.62</td>
<td>11.99</td>
<td>12.42</td>
<td>68%</td>
</tr>
<tr>
<td>Injury - Time Loss</td>
<td>17.45</td>
<td>7.95</td>
<td>9.65</td>
<td>4.83</td>
<td>1.41</td>
<td>8.97</td>
<td>64%</td>
</tr>
<tr>
<td>Injury - Non-Contact</td>
<td>25.80</td>
<td>3.71</td>
<td>5.20</td>
<td>16.90</td>
<td>4.94</td>
<td>8.97</td>
<td>80%</td>
</tr>
<tr>
<td>Injury - Non-Contact Time Loss</td>
<td>12.14</td>
<td>3.71</td>
<td>2.97</td>
<td>1.81</td>
<td>1.41</td>
<td>6.90</td>
<td>85%</td>
</tr>
<tr>
<td>Injury - Overuse</td>
<td>21.24</td>
<td>2.65</td>
<td>2.97</td>
<td>3.02</td>
<td>2.12</td>
<td>4.83</td>
<td>121%</td>
</tr>
<tr>
<td><strong>Men's Soccer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Injury - Medical Attention</td>
<td>23.52</td>
<td>3.44</td>
<td>6.84</td>
<td>10.62</td>
<td>4.96</td>
<td>3.76</td>
<td>86%</td>
</tr>
<tr>
<td>Injury - Time Loss</td>
<td>8.45</td>
<td>2.25</td>
<td>3.86</td>
<td>1.33</td>
<td>0.58</td>
<td>2.72</td>
<td>88%</td>
</tr>
<tr>
<td>Injury - Non-Contact</td>
<td>12.50</td>
<td>1.05</td>
<td>2.08</td>
<td>4.65</td>
<td>2.04</td>
<td>2.72</td>
<td>102%</td>
</tr>
<tr>
<td>Injury - Non-Contact Time Loss</td>
<td>5.88</td>
<td>1.05</td>
<td>1.19</td>
<td>0.50</td>
<td>0.58</td>
<td>2.09</td>
<td>108%</td>
</tr>
<tr>
<td>Injury - Overuse</td>
<td>10.29</td>
<td>0.75</td>
<td>1.19</td>
<td>0.83</td>
<td>0.88</td>
<td>1.46</td>
<td>148%</td>
</tr>
</tbody>
</table>

**Per 1000 AEH**

<table>
<thead>
<tr>
<th>Injury Definition</th>
<th>Team 1</th>
<th>Team 2</th>
<th>Team 3</th>
<th>Team 4</th>
<th>Team 5</th>
<th>Team 6</th>
<th>CV</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Women's Soccer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Injury - Medical Attention</td>
<td>6.13</td>
<td>3.73</td>
<td>5.49</td>
<td>10.38</td>
<td>16.59</td>
<td>15.19</td>
<td>56%</td>
</tr>
<tr>
<td>Injury - Time Loss</td>
<td>4.38</td>
<td>1.81</td>
<td>1.90</td>
<td>8.02</td>
<td>4.06</td>
<td>3.68</td>
<td>57%</td>
</tr>
<tr>
<td>Injury - Non-Contact</td>
<td>2.63</td>
<td>2.13</td>
<td>2.35</td>
<td>4.72</td>
<td>9.22</td>
<td>5.98</td>
<td>61%</td>
</tr>
<tr>
<td>Injury - Non-Contact Time Loss</td>
<td>2.19</td>
<td>0.96</td>
<td>1.34</td>
<td>4.25</td>
<td>2.58</td>
<td>1.38</td>
<td>57%</td>
</tr>
<tr>
<td>Injury - Overuse</td>
<td>0.44</td>
<td>0.00</td>
<td>0.22</td>
<td>4.25</td>
<td>2.58</td>
<td>4.60</td>
<td>103%</td>
</tr>
</tbody>
</table>
### Table 4. Injury Proportion by Sex for Injury Type, MOI, Severity and Session Type

<table>
<thead>
<tr>
<th>Injury Type</th>
<th>Female</th>
<th>Male</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abrasion/Laceration</td>
<td>2.07%</td>
<td>2.42%</td>
<td>2.21%</td>
</tr>
<tr>
<td>Concussion</td>
<td>3.31%</td>
<td>5.45%</td>
<td>4.18%</td>
</tr>
<tr>
<td>Contusion</td>
<td>20.25%</td>
<td>21.82%</td>
<td>20.88%</td>
</tr>
<tr>
<td>Dislocation</td>
<td>0.83%</td>
<td>4.24%</td>
<td>2.21%</td>
</tr>
<tr>
<td>Dysfunction</td>
<td>2.48%</td>
<td>3.64%</td>
<td>2.95%</td>
</tr>
<tr>
<td>Fracture</td>
<td>0.83%</td>
<td>1.21%</td>
<td>0.98%</td>
</tr>
<tr>
<td>Illness</td>
<td>9.50%</td>
<td>4.24%</td>
<td>7.37%</td>
</tr>
<tr>
<td>Infection</td>
<td>0.41%</td>
<td>2.42%</td>
<td>1.23%</td>
</tr>
<tr>
<td>Soft Tissue Inflammation</td>
<td>18.18%</td>
<td>4.24%</td>
<td>12.53%</td>
</tr>
<tr>
<td>Spasm</td>
<td>10.33%</td>
<td>0.00%</td>
<td>6.14%</td>
</tr>
<tr>
<td>Sprain</td>
<td>15.29%</td>
<td>16.36%</td>
<td>15.72%</td>
</tr>
<tr>
<td>Strain</td>
<td>16.53%</td>
<td>33.94%</td>
<td>23.59%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MOI</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute – Non-Contact</td>
<td>20.25%</td>
<td>38.79%</td>
<td>27.76%</td>
</tr>
<tr>
<td>Contact - Apparatus</td>
<td>6.61%</td>
<td>2.42%</td>
<td>4.91%</td>
</tr>
<tr>
<td>Contact - Person</td>
<td>32.23%</td>
<td>29.70%</td>
<td>31.20%</td>
</tr>
<tr>
<td>Contact - Surface</td>
<td>5.37%</td>
<td>4.85%</td>
<td>5.16%</td>
</tr>
<tr>
<td>Contact - Unknown</td>
<td>0.41%</td>
<td>7.88%</td>
<td>3.44%</td>
</tr>
<tr>
<td>Illness</td>
<td>9.50%</td>
<td>4.85%</td>
<td>7.62%</td>
</tr>
<tr>
<td>Overuse</td>
<td>25.62%</td>
<td>11.52%</td>
<td>19.90%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Severity</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Did not interfere or Returned within the same session</td>
<td>33.88%</td>
<td>27.88%</td>
<td>31.45%</td>
</tr>
<tr>
<td>Returned within 24 hours</td>
<td>11.16%</td>
<td>24.24%</td>
<td>16.46%</td>
</tr>
<tr>
<td>Prevented participation for 1-6 days</td>
<td>26.86%</td>
<td>29.09%</td>
<td>27.76%</td>
</tr>
<tr>
<td>Prevented participation for 7-13 days</td>
<td>7.85%</td>
<td>10.91%</td>
<td>9.09%</td>
</tr>
<tr>
<td>Prevented participation for 14-29 days</td>
<td>3.31%</td>
<td>2.42%</td>
<td>2.95%</td>
</tr>
<tr>
<td>Prevented participation for 30+ days</td>
<td>0.41%</td>
<td>0.00%</td>
<td>0.25%</td>
</tr>
<tr>
<td>Body Part</td>
<td>2.07%</td>
<td>5.45%</td>
<td>3.44%</td>
</tr>
<tr>
<td>---------------------------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>Out for reminder of season</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Head/Face</td>
<td>8.68%</td>
<td>9.70%</td>
<td>9.09%</td>
</tr>
<tr>
<td>Cervical Spine/Neck</td>
<td>2.48%</td>
<td>0.00%</td>
<td>1.47%</td>
</tr>
<tr>
<td>Thoracic Spine/Upper Back</td>
<td>0.83%</td>
<td>0.00%</td>
<td>0.49%</td>
</tr>
<tr>
<td>Shoulder/Clavicle</td>
<td>1.65%</td>
<td>6.06%</td>
<td>3.44%</td>
</tr>
<tr>
<td>Elbow</td>
<td>0.83%</td>
<td>0.00%</td>
<td>0.49%</td>
</tr>
<tr>
<td>Hand/Wrist/Finger</td>
<td>0.83%</td>
<td>3.03%</td>
<td>1.72%</td>
</tr>
<tr>
<td>Trunk</td>
<td>3.31%</td>
<td>1.82%</td>
<td>2.70%</td>
</tr>
<tr>
<td>Lumbar Spine/Lower Back</td>
<td>4.55%</td>
<td>2.42%</td>
<td>3.69%</td>
</tr>
<tr>
<td>Hip/Groin</td>
<td>9.92%</td>
<td>15.76%</td>
<td>12.29%</td>
</tr>
<tr>
<td>Upper Leg/Thigh</td>
<td>14.05%</td>
<td>21.21%</td>
<td>16.95%</td>
</tr>
<tr>
<td>Knee</td>
<td>11.98%</td>
<td>7.27%</td>
<td>10.07%</td>
</tr>
<tr>
<td>Lower Leg/Achilles</td>
<td>8.68%</td>
<td>10.30%</td>
<td>9.34%</td>
</tr>
<tr>
<td>Ankle</td>
<td>13.64%</td>
<td>9.70%</td>
<td>12.04%</td>
</tr>
<tr>
<td>Foot/Toes</td>
<td>8.68%</td>
<td>8.48%</td>
<td>8.60%</td>
</tr>
<tr>
<td>General - Illness</td>
<td>9.92%</td>
<td>4.24%</td>
<td>7.62%</td>
</tr>
<tr>
<td>Wellness Inventory</td>
<td>Women's Soccer</td>
<td>Men's Soccer</td>
<td></td>
</tr>
<tr>
<td>------------------------------------</td>
<td>----------------</td>
<td>--------------</td>
<td></td>
</tr>
<tr>
<td>PSQI Global Score</td>
<td>5.4 (2.7)</td>
<td>4.5 (2.6)</td>
<td></td>
</tr>
<tr>
<td>SAS-2 Total</td>
<td>25.2 (6.2)</td>
<td>21.0 (7.3)</td>
<td></td>
</tr>
<tr>
<td>POMS Total Mood Disturbance</td>
<td>19.1 (25.4)</td>
<td>17.2 (28.7)</td>
<td></td>
</tr>
<tr>
<td>DPA Total</td>
<td>22.1 (11.8)</td>
<td>11.8 (13.0)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sleep Diary</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Time in Bed (hr)</td>
<td>8.0 (1.5)</td>
<td>8.1 (1.8)</td>
</tr>
<tr>
<td>Sleep Duration (hr)</td>
<td>7.8 (1.6)</td>
<td>7.8 (1.8)</td>
</tr>
<tr>
<td>Sleep Latency (hr)</td>
<td>0.3 (0.3)</td>
<td>0.3 (0.4)</td>
</tr>
<tr>
<td>Sleep Quality (1-5)</td>
<td>3.7 (1.0)</td>
<td>3.7 (1.0)</td>
</tr>
<tr>
<td>Refreshed (1-5)</td>
<td>3.2 (1.0)</td>
<td>3.2 (1.1)</td>
</tr>
<tr>
<td>Calm Sleep (1-5)</td>
<td>3.6 (1.0)</td>
<td>3.7 (1.0)</td>
</tr>
<tr>
<td>Sleep Planned Length (1-5)</td>
<td>3.7 (1.1)</td>
<td>3.6 (1.2)</td>
</tr>
<tr>
<td>Ease of Falling Asleep (1-5)</td>
<td>3.7 (1.1)</td>
<td>3.7 (1.1)</td>
</tr>
<tr>
<td>Ease of Awakening (1-5)</td>
<td>3.1 (1.1)</td>
<td>3.2 (1.1)</td>
</tr>
<tr>
<td>Dreaming (1-5)</td>
<td>2.3 (1.2)</td>
<td>2.5 (1.4)</td>
</tr>
<tr>
<td>Sleep Disturbances (count)</td>
<td>1.8 (2.0)</td>
<td>1.3 (2.1)</td>
</tr>
</tbody>
</table>
Figure 2
Figure 3

[Image: Graphs showing POMS - Time Association for different time points across baseline, preseason, inseason week 4, inseason week 8, end inseason, and end postseason. The graphs are split into two parts, comparing across different genders (female and male).]
Figure 4
Chapter 3: Risk Factors for Non-Contact Injury in NCAA Division I Soccer

ABSTRACT

Purpose: To examine athlete-specific, seasonal, congestion, workload, sleep and wellness injury risk factors associated with women and men’s collegiate over the course of a national collegiate athletics association (NCAA) soccer season. Methods: Injuries, workload, mood, sports-related anxiety, sleep characteristics and physical activity disablement was longitudinally assessed for 256 athletes from 12 separate NCAA division I teams. Injury risk factors were identified utilizing multi-level Poisson regressions to capture differences in injury rate. Results: Relative workloads (acute:chronic workload ratio), chronic workloads, workload monotony, season type, session type, days relative to a match, session congestion, days off, weekly sleep latency and weekly sleep quality were identified as risk factors of non-contact injury. Psychological wellbeing, chronic sleep behavior or weekly sleep duration was not associated with injury risk. Sleep duration or quality was not acutely altered prior to injury. Conclusions: Multi-team prospective cohort studies involving workload, wellness and sleep monitoring allow for the modeling of multiple injury risk factors in sport. Developing a multi-factorial view is vital for context when trying to understand complex phenomena such as injury.
**INTRODUCTION**

Injuries can negatively impact team performance\textsuperscript{1–3} and threaten long-term athlete wellbeing.\textsuperscript{4} A reduction in injury incidence is likely to positively impact team performance with basketball\textsuperscript{5} and track and field\textsuperscript{6} research indicating injury incidence is associated with competition wins. Implementing efficacious injury prevention programs is a primary objective for all stakeholders in sport. Prevention of injury has been described by van Mechelen\textsuperscript{7} as a process requiring the identification of risk factors contributing to injury occurrence. Soccer, being the world’s most popular sport\textsuperscript{8}, has received considerable attention with regard to the identification of risk factors associated with injury.\textsuperscript{10–27} It is well understood that sports injuries are multifactorial and no single risk factor is adequate to explain all injury occurrences. Multiple risk factors should be considered when investigating injury determinants in sport.

Evidence suggests athlete-specific (intrinsic)\textsuperscript{27} risk factors are important to consider such as previous injury, which has been previously identified as a risk factor for subsequent injury in soccer.\textsuperscript{11,28} Other factors such as collegiate playing experience or role on the team (starter vs. reserve) have potential to influence workload profiles and therefore might influence injury risk. Player position may also affect injury risk, as different positional roles experience different workload demands.\textsuperscript{29} Previous studies investigating the effect of position on injury risk are conflicting, with a systematic review on the topic finding five studies identifying an association between position and injury and six studies not.\textsuperscript{30}

Workload has been identified as a critical piece to understanding the injury etiology and strong evidence suggest that workloads are a primary modifiable risk factor for injury.\textsuperscript{31,32} In fact, exposure to workloads are a precondition for athletic injury.\textsuperscript{33} Risk factors such as low chronic workload\textsuperscript{34–43}, as well as, “spikes” in workload or acute changes relative to the individual’s chronic baseline (i.e., acute chronic workload ratio) have been identified in several athletic populations.\textsuperscript{1,44–46} Periods of workload intensification have additionally been identified as an important injury risk factor consideration.\textsuperscript{47} Intensified periods such as the preseason and calendar congested periods\textsuperscript{15,48} have shown elevated injury risk. Additionally, rapid increase in load from matches represents a significant risk factor (acute spike of
match workload relative to lower chronic training load). To our knowledge, no investigations have explored the relationship between load characteristics and injury in collegiate soccer players.

Sleep also represents an important behavior to consider from both an acute (i.e., fluctuations in sleep duration or quality) and chronic (i.e., normal sleep patterns) perspective. Sleep has been identified as a risk factor for injury in adolescent athlete populations and additionally in professional male soccer players. Further, a case study on 1 elite soccer athlete indicates sleep prior to injury may be compromised in elite male soccer players.

Psychological or emotional wellness may influence stress responses and can increase injury risk through attentional and somatic changes such as increased distractibility and peripheral narrowing, as well as muscle tension, fatigue and reduced timing/coordination. Prior works have found injury to be associated with altered psychological states or personality traits.

Research on injury risk factors has primarily focus on isolated subsets of workload, wellness, previous injury, congestion, sleep or demographics. There is a need to further examine these relationships collectively and in the same cohort. Specifically, non-contact injury is important to consider, which research indicates may be “preventable” or at least reduced with intervention-based exercise programs. Therefore, the purpose of this study was to investigate several potential injury risk factors related to workload, psychological wellness, sleep, previous injury, congestion, and athlete characteristics in NCAA collegiate soccer.

**METHODS**

**Participants.** A prospective cohort study of 256 NCAA division I athletes from 12 separate university teams was conducted over the 2016 (2 teams), 2017 (6 teams) and 2018 (4 teams) seasons. One-hundred and thirty-nine of the participants were female (age, 20±1 y; body mass, 64.7±6.1 kg; height, 166.8±6.2 cm; VO$_{2max}$, 46.8±4.0 ml·kg$^{-1}$·min$^{-1}$) and one-hundred and seventeen were male soccer players (age, 20±2 y; body mass, 77.4±5.1 kg; height, 179.9±6.5 cm; VO$_{2max}$, 53.8±4.1 ml·kg$^{-1}$·min$^{-1}$). All participants were
medically cleared for physical activity by their respective university’s sports medicine department and free of any debilitating musculoskeletal injuries or contraindicated medical conditions. Institutional review board (IRB) and ethics approval was obtained from all institutions, with primary oversight and coordination provided by the University of <blinded for review> (IRB Approval ID: H17-134). All participants provided written informed consent prior to the season. When the participant was under the age of 18, parental consent was obtained.

**Injury Classification.** Injuries were diagnosed and recorded by a single member of each team’s medical staff (i.e., certified athletic trainer). Injuries were recorded according to the current consensus statement on recording of soccer injuries, which clarifies that an injury is “any physical complaint sustained by a player that results from a football (soccer) match or football (soccer) training, irrespective of the need for medical attention or time loss from football (soccer) activities.” In addition to injury incidence, other pertinent information such as injury type, severity, location and mechanism were recorded. For this study, all non-contact injuries that required medical attention, irrespective of time loss were considered. Overuse injuries were included under the non-contact classification umbrella.

**Athlete, Session and Congestion.** Several athlete-specific, session-specific and seasonal congestions factors with the potential to influence injury risk either directly or indirectly were selected for analysis. To assess differences between player role within the team, athletes were classified as starters if they competed in greater than 60% of the total match time and started in greater than 60% of the total matches in the season, all other athletes were considered reserves. Athletes were additionally divided into position groups consisting of defenders, midfielders and forwards. Athletes were further grouped by the number of years they have been competing in intercollegiate athletics (range: 1-6). To examine the effect of season phase, injury risk during preseason, in-season and postseason were considered, with postseason referring to the period directly following the in-season where conference and NCAA tournament play occurs. All day-exposures were additionally classified by days relative to an upcoming match (match day minus [MD-]). Data were analyzed for one (MD-1), two (MD-2), three (MD-3), four (MD-4), five (MD-5)
and 6 or greater (MD-6+) days prior to a match. Further, a day-exposures were classified as either a
training or match day. The effect of session and match congestion on injury risk was determined by
grouping individual exposures by the number of sessions or matches completed by that individual in the
previous 7 days. Session congestion consisted of groupings <6 sessions or 6-7 session in the previous 7
days. Match congestion grouping consisted of either 0-1 or 2-3 calendar matches in the previous 7 days.
Number of off days (completely void of team-related activity) were additionally binned into groups of 0,
1 and >2 days off in the previous 7 days. To examine the effect of previous injury on injury risk, rather
than classify as injury vs. no injury which doesn’t consider the total number of injuries sustained
previously, a rolling cumulative sum was calculated for each player over the season.

Workload. Global positioning satellite (GPS) player tracking devices were used to capture workloads all
training sessions and matches (Polar Team Pro, Polar Electro, Lake Success, NY). The 10 Hz GPS player
tracking device has reported accuracy and reliability outdoors for 40 and 100 m total distances at four
separate movement (i.e., walk, jog, run, sprint) velocities (Mean Difference= -1.04 to -2.78m; CV=1.17-
3.16%) and during a team sport simulation circuit (Mean Difference=0.23m; CV=0.96%). Devices were
attached to the body via a chest strap before the start of each practice. To reduce inter-unit error, players
wore the same device for each training sessions. Players donned the player tracking device prior to the
beginning of the session warm up to the end of the last organized training activity. After each session was
completed, data were synced to a Polar Electro server and subsequently exported to Microsoft excel
spreadsheets (Microsoft Corporation, Redmond WA) for analysis.

For this investigation, all training and match exposures were considered. Workload metrics
comprised of total distance (TD) and total high-speed distance (HSD), which have been used
previously in workload-injury studies. Workload metrics were aggregated into daily sum totals and lagged
by one day so that injury risk was assessed based on prior workloads. Several workload features were
engineered from total distance (TD) and total high-speed distance (HSD), which have been used
previously in workload-injury research. Exponentially weighted moving averages (EWMA), which
account for the decaying effect of workload, were calculated for 3, 7 and 28 days of TD and HSD.
Research by Murray et al. suggests ACWR methods using EWMA’s instead of standard rolling average may be more sensitive to injury. Daily acute:chronic workload ratios (ACWR) by player for TD and HSD were calculated by dividing 7-day EWMA by 28-day EWMA. ACWR windows of 7 and 28-day windows were used as these are customary in workload-injury investigations. Both rolling 7-day means and rolling standard deviations of TD and HSD were computed to model workload monotony. Monotony was calculated by dividing each days’ rolling average of the previous 7 days by the rolling standard deviation of the previous 7 days. Training monotony has been previously linked with overtraining syndrome, with higher training monotony associated with increased illness.

Additionally, rolling 7-day and 28-day sums were computed to represent traditional acute and chronic workload, respectively.

**Psychological Wellness.** Participants’ mood, sports-related anxiety, sleep quality and physical activity disablement was longitudinally assessed during 6 distinct time points throughout the season. Inventories were administered by a member of each research team prior to preseason (baseline), directly after preseason, at week 4 and 8 of the in-season, end of regular season and end of postseason play (when applicable). The POMS is a validated 65-item questionnaire which has shown sensitivity to training load changes and associated altered mood states. The POMS assessed six mood or ‘feeling’ states: tension, anger, vigor, depression and fatigue and is robust for examination of individual mood states. Construct validity has been explored by Terry and colleagues for its use with both adults and adolescents. The Sport Anxiety Scale-2 (SAS-2) was used to measure cognitive and somatic trait anxiety. The SAS-2 is a 15-item inventory assessing anxiety, worry and concentration disruption with a total score ranging from 15-60. The SAS-2 has been previously used in women’s collegiate soccer to assess the effectiveness of mental skill training. PSQI has been used to assess perceived sleep quality in the collegiate student-athlete population and elite athletes. The PSQI consists of 19 items assessing subjective sleep quality, latency, efficiency, duration, and disturbances. The scoring for each component is combined for a Global
Sleep Quality Score ranging from 0-21, with >5 being indicating general poor sleep quality.72 Wellness-Injury associations were assessed by investigating each global scores association with injury in the time period directly following assessment.

Sleep Diary. Daily sleep behavior was assessed via the Karolinska Sleep Diary.73 The KSD is an eleven item questionnaire used to evaluate several facets of sleep, including quantity and aspects of perceived quality, such as ease of falling asleep, ease of awakening, overall perception of sleep quality, sleep disturbances, sleep calmness, and feeling of rest.73 The KSD questionnaire was electronically distributed daily and participants were encouraged to complete the questionnaire upon waking.

Statistical Analysis. Statistical analysis was conducted in R statistical programming language (R Core Team, Vienna, Austria).74 Injury incidence rates were calculated by dividing the number of injuries by total exposures or exposure hours and reported as rate per 1000 exposures or hours.

Daily workload-injury relationships were investigated using generalized multi-level regressions with a Poisson distribution, log link function and unstructured covariance matrix. Mixed effects modelling was used for its ability to handle unbalanced fix factors and to account for repeated measures75, which was seen with multiple exposures per player. During null model construction, both player sex and ID were entered as clustering variables, however ICC values for sex were 0, indicating no additional variance was being explained by this factor. Therefore, a random effect of player was included in all contextual and workload-injury modeling. In light of previous reports of non-linear relationships between workload variables (i.e., acute:chronic workload ratio) and injury76,77, both linear and non-linear workload-injury models were compared via chi-squared tests, which tests whether there is a statistically significant reduction in the residual sum of squares not. If there was no statistically significant difference between linear and quadratic model (2nd order polynomial), a linear model was used.

Sleep-Injury association was assessed via 3 separate analysis. Firstly, statistical differences were assessed via paired t-test between seasonal average sleep and the sleep directly preceding an injury, average of 3
days before an injury and average of 7 days before an injury. Secondly, seasonal average sleep, taken to be a representation of the participant’s chronic sleep habits over the season, was assessed as a potential risk factor for injury incidence over the season using logistic regression with a binary outcome distribution and logit link function. Finally, sleep measures were averaged by week and the likelihood of incurring an injury in the subsequent week was assessed via univariate generalized multi-level regressions with a Poisson outcome distribution, log link function and unstructured covariance matrix.

Psychological wellness was assessed for its relationship with injury in the subsequent measurement phase (i.e., preseason, week 1-week 4, etc.) via generalized mixed effects models with a binomial distribution and logit function. Outcome consisted of a binary indicator of injured vs not injured in the subsequent time grouping. A random effect of player id and time point were used to account for individual and seasonal changes in psychological wellness. Statistical significance level of p <0.05 was set a priori for all analysis.

**RESULTS**

**Overall.** Over the course of 12 team-season and 256 player-seasons there were a total of 372 medical attention injuries and 191 non-contact injury incidences. Medical attention injury incidence rates were 23.40/1000 athlete-exposures or 11.54/1000 exposure-hours and non-contact injury rates were 10.22/1000 athlete-exposures or 4.95/1000 exposure-hours.

**Athlete, Session and Congestion.** Associations between injury rate and athlete, season, calendar congestion and session-specific factors are displayed in Table 1. Results indicated athlete status and collegiate playing experience are significantly associated with medical attention injuries but not non-contact injuries, with starters and more senior-level players showing higher rates of medical attention than reserves and those with less collegiate playing experience, respectively. Previous injury was significantly associate with subsequent injury with incidence rates increasing by 2.23 times (95%CI: 2.05-2.42) for every additional medical attention injury (Figure 3). Season phase was also identified as a factor affecting
injury risk with both in-season and postseason incidence rates being reduced by 58% (IRR: 0.42, 95% CI: 0.31-0.57, \( p < 0.001 \)) and 52% (IRR: 0.48, 95% CI: 0.28-0.82), respectively. Injury incidence rates were significantly lower (36%) in training as compared with a match (IRR: 0.63; 95% CI: 0.47-0.86, \( p = 0.003 \)). Injury rates were significantly lower on MD-1 (IRR: 0.34, 95% CI: 0.21-0.55, \( p < 0.001 \)) and MD-2 (IRR: 0.35, 95% CI: 0.20-0.61, \( p < .001 \)) compared to a match. The rate of injury 4 days removed from a match was 2.24 times (95% CI: 1.49-3.38, \( p < 0.001 \)) the rate of injury in matches (Figure 3). Injury rates were higher when athletes had 6-7 sessions in the previous 7 days as compared with <6. When this relationship is explored in terms of number of off days, at least 1 day off in the previous 7 days reduced the rate of injury by 55-58%. There were no differences in injury rates between 1 day off in the previous 7 days and more than 1 day off. The differences in injury rate between 0-1 matches in the previous 7 days compared with 2-3 matches in the previous 7 days was not significant (\( p = 0.062 \)).

**Workload.** Quadratic modeling (2nd order polynomial) of ACWR-injury relationships for both TD and HSD showed significantly reductions in residual variance compared with linear models \( (X^2 = 6.37, \ p < 0.001) \), therefore quadratic functions were used to model this relationship. Residual variance was not statistically different between linear and quadratic models for monotony, acute load and chronic workload, therefore linear models were used.

Univariate multi-level Poisson regression results for workload-injury models are shown in Table 2. Workload-injury plots for ACWR and chronic load are shown in figure 1. Our findings indicated there was a significant association between ACWR and injury for both TD and HSD. Injury rates increased by 1.52 times (95% CI: 1.26 – 1.83, \( p < 0.001 \)) per unit increase in ACWR for TD and by 1.43 (95% CI: 1.20 – 1.71, \( p < 0.001 \)) per unit increase in ACWR for HSD. Chronic workload (28-day rolling sum) was negatively associate with injury rate, with the rate of non-contact injury decreasing by 6% \( (IRR: 0.94, 95\% CI: 0.90-0.98, \ p = 0.002) \) for per every 10km increase in TD and 2% \( (IRR: 0.98, 95\% CI: 0.88-0.97, \ p < 0.001) \) for per every 1km of HSD. Workload monotony was also positively associate with injury \( (IRR: 1.51, 95\% CI: 1.18-1.92, \ p < 0.01) \).
Sleep. Multinomial logistic regression results are presented in Table 3. Results indicate seasonal average sleep duration and aspects of sleep quality were not statistically associated with increased odds of sustaining an injury over the season (all p>0.05). Additionally, there were no significant differences found between the seasonal average sleep duration or quality and the night before an injury, the average of the 3 nights before an injury or the average of the 7 days before an injury (Table 4; all p>0.05). When sleep diary responses were grouped into weekly bins, there were significant relationships between weekly sleep and injury in the subsequent week. Specifically, sleep latency showed a positive association with increased injury incidence with 2.43 times (95% CI: 1.03-5.73, p=0.042) increase in injury incidence rate per 1-hr increase in time needed to fall asleep. Increases in sleep quality (IRR:0.59, 95% CI: 0.39-0.88, p=0.009), perceived calmness of sleep (IRR:0.57, 95% CI: 0.39-0.84, p=0.005) and ease of falling asleep (IRR:0.67, 95% CI: 0.46-0.98, p=0.041) were associated with decreased injury incidence rate per 1-unit increase by 41%, 43% and 33%, respectively. Weekly sleep and subsequent injury incidence rate ratios are shown in Table 5.

Wellness. Figure 2 displays PSQI, SAS2, POMS and DPA model of injury risk in the subsequent time-period. No significant association were found between wellness inventories and injury.

DISCUSSION

Injury risk factor identification is an important part of the prevention paradigm. Obtaining information on why athletes may be at risk in certain situations and developing a multi-factorial view of injury is essential understanding complex phenomena. Most of the work on injury risk factor identification has investigated isolated subsets of influencers, such as looking at workload-injury, sleep-injury, or calendar congestion-injury relationships in isolation. The novelty of this research was the implementation of a prospective cohort design to investigate a multitude of potential risk factors in the same cohort. In doing so, we found several risk factors associated with injury risk in collegiate soccer (Figure 3).
Athlete-Specific, Seasonal and Congestion Factors

Non-contact Injury rates were not influenced by player sex in the current study, which is supported by previous literature on epidemiological injury surveillance in collegiate soccer. Although non-contact injury was not directly investigated by Roos et al., they report no differences in injury rates between sexes for either time-loss or non-time loss injury in a large investigation of 167 collegiate team-seasons. Nevertheless, our findings support the fact that women’s and men’s soccer athletes are at equal risk of non-contact injury risk in collegiate soccer.

Since injury rates have consistently been found to be higher in matches compared to training, including in the current study, it is interesting that injury risk was not elevated for players assuming a majority of playing time throughout the season. Further works should look to investigate how starting status and session type interact to influence both workload characteristics over the season, as well as, injury risk. Since starter workloads are inherently elevated during matches, there is potential that reserves workload characteristics may be elevated during training to accommodate for lack of playing time.

Previous injury is a well-established risk factor for subsequent injury in soccer. Athlete with previous injuries have been found to have 4-7 times greater risk of subsequent injury. Consistent with previously identified risk factor, we found the number of prior injuries to be a significant risk factor for future injury with the relationship appearing exponential in nature. These finding are important because subsequent injury risk continues to elevate as more injuries are incurred throughout the season.

Most congestion studies have investigated match congestion, with congestion alluding to greater than 1 match per week (elite level soccer plays 1 match per week on average). Our investigation took a novel approach by examining overall session congestion, as well as, match congestion. Interestingly, we found overall session congestion, but not match congestions to be significantly associated with injury risk. Further, our results indicate that having at least 1 day off from training and matches in a 7-day period
may be beneficial in reducing player injury risk. Previous works investigating calendar congestion have been inconsistent. Dellal et al. observed training injuries during congested time periods were either unaltered or reduced.\textsuperscript{48} This is not surprising since it is customary to decrease training loads during highly congested periods. Other works have approached congestion by dichotomizing between match recovery periods to assess relative risk. No differences were found for <3 days compared to >4 days of rest between matches\textsuperscript{15,82}, however significantly higher injury rates are observed for <3 days\textsuperscript{48,83} or <4 days\textsuperscript{15,84} compared to >6 days. Conflicting finds are probably resultant of contextual factors, such as individual team periodization structures. Regardless, our results indicate overall session congestion, rather than match congestion, may be a more useful risk factor to moderate.

**Workload Factors**

Our findings are consistent with those finding low chronic workload\textsuperscript{34,36,39,40,43,60} and “spikes” in workload are associated with increased injury risk.\textsuperscript{1,44–46,60} Maintaining and attaining high chronic workloads may be protective of injury while low chronic workloads are typically associated with increased injury risk. However, not all studies support this as McCall et al. in a recent study of 5 professional soccer teams found no association or increased/decreased injury risk with chronic workload.\textsuperscript{85} While most evidence suggests chronic workloads are directly related to injury incidence, it is more likely that chronic workloads are dynamically interacting with acute workloads to influence injury risk. A multivariate analysis by Colby et al., found an interaction between a low chronic load and a very high distance (IRR=2.60, 95% CI=1.07-6.34) supporting this contention. This is the only study to-date which has modelled this interaction utilizing multivariate methods. Further research with adequate injury count should look to support this work. In addition to acute and chronic workloads, we also investigated the link between workload monotony and injury risk. Monotony was first studied by Foster in 1998, who showed that 77% of illnesses could be explained by a spike in training load monotony.\textsuperscript{65} This link has been supported by Brink et al. who found an increase in monotony (odds ratio [OR] = 2.59, 95% confidence interval [CI] 1.22–1.50) was significantly related to an increase in injury incidence.\textsuperscript{22} Our research further
supports these findings, highlight monotony as an important workload to consider when programming and altering athletes workload.

**Sleep Factors**

Survey-based research on sleep behavior has been investigated in adolescent athlete populations indicate chronic sleep behavior is associated with injury risk.\(^{50,51}\) Research by Milewski et al. found sleep duration to be a predictor of injury risk in adolescent athletes\(^{50}\), with youth athletes sleeping <8 hours per night being 1.7 times (95% confidence interval, 1.0-3.0; \(p = 0.04\)) more likely to incur an injury compared with athletes sleeping >8 hours. Additionally, von Rosen and colleagues found adolescent athletes getting more than 8 hours of sleep during weekdays reduced the odds of injury by 61% (OR, 0.39; 95% CI, 0.16–0.99).\(^{51}\) In contrast, we did not find an association between chronic sleep behavior and the odds of sustaining an injury over the course of a collegiate soccer season for a range of sleep duration and sleep quality measures. Interestingly, a case study on 1 elite soccer athlete indicates sleep prior to injury may be compromised in elite male soccer players.\(^{52}\) We also investigated this relationship with 91 separate non-contact injuries and found no disruption in sleep duration or quality in the night preceding injury or the average of the 3 and 7 nights preceding injury compared with an athletes chronic baseline sleep average. In contrast the Nedelec’s findings, \(^{52}\) our results suggest sleep disruption acutely preceding an injury may not be the norm. However, to investigate a potential lag effect of poor sleep characteristics, we also assessed whether poor sleep in any given week may be related to increased injury in the subsequent week. Indeed, we found aspects of sleep quality but not sleep duration to be associated with following week injury risk. These findings have important implications for the coach and practitioner. Our results indicate that although poor sleep may not directly influence injury risk in the subsequent few days, having a poor week of sleep may negatively influence an athlete’s injury risk profile the following week. Sleep hygiene strategies should be routinely implemented rather than at select instances (e.g., prior to or following matches), as there may be a lag effect whereby poor average weekly sleep leads to increased injury risk in the subsequent week. Additionally, sleep hygiene strategies should be focused on creating
an environment conducive to falling asleep and should also be promoting sleep quality, rather than extending sleep duration.

**Psychological Wellbeing Factors**

A previous investigation on psychological wellbeing and injury found a positive relation between injury and mood states of tension, anxiety, hostility and anger.\textsuperscript{53} Psychological stress can increasing distractibility, peripheral narrowing, muscle tension, fatigue and lead to reduced timing/coordination, which can all negatively influence an athletes injury risk profile.\textsuperscript{49} Additionally, personality characteristics such as trait anxiety have been consistently associated with increased injury risk.\textsuperscript{54} Our results did not find overall sleep dysfunction, mood disturbance, sports related anxiety, or perceptions of physical activity disablement to be associated with injury in the 2-4-week period following assessment. However, a limitation of this analysis was the relative infrequency of measurement. Regardless, monitoring physical and psychological well-being at monthly time-periods throughout a collegiate soccer season did not offer value with respect to their relationship with injury. Further studies should investigate the usefulness of these inventories using a more frequent assessment period (e.g., weekly).

**CONCLUSION**

This investigation identified 11 separate non-contact injury risk factors in collegiate soccer which include: relative workloads (ACWR), chronic workloads, workload monotony, season type, session type, days relative to a match, session congestion, number of days off, sleep latency and sleep quality. We did not find a link between total sleep duration and non-contact injury risk indicating this measure may not be the most important aspect of sleep to monitor or to target with hygiene strategies. Further, seasonal average sleep duration or quality scores were not informative of injury risk in collegiate soccer. Also, sleep was not altered in the night prior, 3 nights prior or 7 night prior to an injury. Weekly fluctuations in sleep quantity and latency were informative of injury risk in the subsequent week. Multi-team prospective cohort studies involving workload, wellness and sleep monitoring allow for the modeling of multiple
injury risk factors in sport. Developing a multi-factorial view is vital for context when trying to understand complex phenomena such as injury.

**Acknowledgement:**

The authors would like to express their sincerest thanks to the athletes, coaching staff, sports medicine staff and strength and conditioning staff of each participating institution for their time and commitment during the study. Additionally, the authors would like to gratefully acknowledge participating research personnel from all institutions involved. <contributors will be acknowledged after review >

**Declaration of Interest:**

The authors report no conflicts of interest. The authors alone were responsible for the content and writing of this article. They are thankful for the financial support for this research provided in-part by the National Collegiate Athletics Association (NCAA).

**Figure Legend:**

Figure 1a – Association between acute:chronic workload ratio (ACWR) of total distance and subsequent session injury risk.

Figure 1b – Association between acute:chronic workload ratio (ACWR) of high speed distance (>14.4 km/h) and subsequent session injury risk.

Figure 1c – Association between chronic (28-day exponentially weighted moving average) total distance and subsequent session injury risk.

Figure 1d – Association between chronic (28-day exponentially weighted moving average) high-speed distance (>14.4 km/h) and subsequent session injury risk.

Figure 2 – Risk factors affecting injury risk in NCAA division I soccer
REFERENCES


Table 1. Athlete, Session and Congestion Factors and Injury

<table>
<thead>
<tr>
<th>Risk Factor</th>
<th>Medical Attention</th>
<th></th>
<th>Non-Contact</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IRR</td>
<td>CI</td>
<td>p</td>
<td>IRR</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>reference</td>
<td></td>
<td></td>
<td>reference</td>
</tr>
<tr>
<td>Male</td>
<td>0.96</td>
<td>0.73–1.27</td>
<td>.778</td>
<td>1.09</td>
</tr>
<tr>
<td>Status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reserve</td>
<td>reference</td>
<td></td>
<td></td>
<td>reference</td>
</tr>
<tr>
<td>Starter</td>
<td>1.42</td>
<td>1.08–1.88</td>
<td>.013</td>
<td>1.17</td>
</tr>
<tr>
<td>Playing Experience</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Per 1-Year Increase</td>
<td>1.19</td>
<td>1.05–1.34</td>
<td>.007</td>
<td>1.16</td>
</tr>
<tr>
<td>Position</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Defender</td>
<td>reference</td>
<td></td>
<td></td>
<td>reference</td>
</tr>
<tr>
<td>Forward</td>
<td>0.94</td>
<td>0.66–1.35</td>
<td>.741</td>
<td>0.93</td>
</tr>
<tr>
<td>Midfielder</td>
<td>0.83</td>
<td>0.60–1.15</td>
<td>.262</td>
<td>0.88</td>
</tr>
<tr>
<td>Previous Injury</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Per 1-Injury Increase</td>
<td>1.51</td>
<td>1.45–1.58</td>
<td>&lt;.001</td>
<td>2.23</td>
</tr>
<tr>
<td>Season Phase</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preseason</td>
<td>reference</td>
<td></td>
<td></td>
<td>reference</td>
</tr>
<tr>
<td>Inseason</td>
<td>0.70</td>
<td>0.56–0.88</td>
<td>.003</td>
<td>0.42</td>
</tr>
<tr>
<td>Postseason</td>
<td>0.67</td>
<td>0.46–0.99</td>
<td>.043</td>
<td>0.48</td>
</tr>
<tr>
<td>Session Type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Match</td>
<td>reference</td>
<td></td>
<td></td>
<td>reference</td>
</tr>
<tr>
<td>Training</td>
<td>0.37</td>
<td>0.30–0.46</td>
<td>&lt;.001</td>
<td>0.64</td>
</tr>
<tr>
<td>Day Relative to Match</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Match (MD)</td>
<td>reference</td>
<td>CI</td>
<td>reference</td>
<td>CI</td>
</tr>
<tr>
<td>------------</td>
<td>-----------</td>
<td>----------</td>
<td>-----------</td>
<td>----------</td>
</tr>
<tr>
<td>MD-1</td>
<td>0.18</td>
<td>0.12–0.26</td>
<td>&lt;.001</td>
<td>0.34</td>
</tr>
<tr>
<td>MD-2</td>
<td>0.29</td>
<td>0.20–0.42</td>
<td>&lt;.001</td>
<td>0.35</td>
</tr>
<tr>
<td>MD-3</td>
<td>0.40</td>
<td>0.27–0.59</td>
<td>&lt;.001</td>
<td>0.59</td>
</tr>
<tr>
<td>MD-4</td>
<td>0.96</td>
<td>0.69–1.35</td>
<td>.826</td>
<td>2.24</td>
</tr>
<tr>
<td>MD-5</td>
<td>0.71</td>
<td>0.42–1.21</td>
<td>.210</td>
<td>1.45</td>
</tr>
<tr>
<td>MD-6+</td>
<td>0.62</td>
<td>0.40–0.96</td>
<td>.031</td>
<td>0.80</td>
</tr>
</tbody>
</table>

**Session Congestion (7-Day)**

<table>
<thead>
<tr>
<th>&lt;6</th>
<th>reference</th>
<th>CI</th>
<th>reference</th>
<th>CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-7</td>
<td>1.44</td>
<td>1.08–1.91</td>
<td>.012</td>
<td>1.58</td>
<td>1.02–2.46</td>
</tr>
</tbody>
</table>

**Match Congestion (7-day)**

<table>
<thead>
<tr>
<th>0-1</th>
<th>reference</th>
<th>CI</th>
<th>reference</th>
<th>CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-3</td>
<td>1.28</td>
<td>1.04–1.57</td>
<td>.019</td>
<td>0.76</td>
<td>0.56–1.01</td>
</tr>
</tbody>
</table>

**# Off (7-day)**

<table>
<thead>
<tr>
<th>0</th>
<th>reference</th>
<th>CI</th>
<th>reference</th>
<th>CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.66</td>
<td>0.51–0.84</td>
<td>&lt;.001</td>
<td>0.45</td>
<td>0.32–0.63</td>
</tr>
<tr>
<td>&gt;1</td>
<td>0.64</td>
<td>0.48–0.85</td>
<td>.002</td>
<td>0.42</td>
<td>0.28–0.62</td>
</tr>
</tbody>
</table>

### Table 2. Workload Association with Injury Risk

<table>
<thead>
<tr>
<th></th>
<th>IRR</th>
<th>CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Distance</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACWR</td>
<td>1.52</td>
<td>1.26–1.83</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Monotony</td>
<td>1.51</td>
<td>1.18–1.92</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Acute Load (1-week sum)</td>
<td>0.93</td>
<td>0.78–1.09</td>
<td>.361</td>
</tr>
<tr>
<td>Chronic Load (4-week sum)</td>
<td>0.94</td>
<td>0.90–0.98</td>
<td>.002</td>
</tr>
<tr>
<td><strong>HSD</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACWR</td>
<td>1.43</td>
<td>1.20–1.71</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Monotony</td>
<td>1.47</td>
<td>0.73–2.97</td>
<td>.285</td>
</tr>
<tr>
<td>Acute Load (1-week sum)</td>
<td>0.91</td>
<td>0.78–1.05</td>
<td>.180</td>
</tr>
<tr>
<td>Chronic Load (4-week sum)</td>
<td>0.92</td>
<td>0.88–0.97</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>
Table 3. Association between Seasonal Average of Sleep Measures and Non-Contact Injury

<table>
<thead>
<tr>
<th>Sleep Measure</th>
<th>Estimate</th>
<th>SE</th>
<th>p-value</th>
<th>OR (95%CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sleep Duration (hr)</td>
<td>0.127</td>
<td>0.135</td>
<td>0.35</td>
<td>1.13 (0.87,1.48)</td>
</tr>
<tr>
<td>Sleep Latency (hr)</td>
<td>0.114</td>
<td>0.135</td>
<td>0.40</td>
<td>1.12 (0.86,1.46)</td>
</tr>
<tr>
<td>Sleep Quality (1-5)</td>
<td>-0.054</td>
<td>0.645</td>
<td>0.93</td>
<td>0.95 (0.27,3.35)</td>
</tr>
<tr>
<td>Calm Sleep (1-5)</td>
<td>0.557</td>
<td>0.685</td>
<td>0.42</td>
<td>1.74 (0.46,6.68)</td>
</tr>
<tr>
<td>Sleep Planned Length (1-5)</td>
<td>-0.063</td>
<td>0.377</td>
<td>0.87</td>
<td>0.94 (0.45,1.96)</td>
</tr>
<tr>
<td>Ease of Awakening (1-5)</td>
<td>0.000</td>
<td>0.318</td>
<td>1.00</td>
<td>0.99 (0.53,1.86)</td>
</tr>
<tr>
<td>Ease of Falling Asleep (1-5)</td>
<td>-0.663</td>
<td>0.506</td>
<td>0.19</td>
<td>0.52 (0.19,1.39)</td>
</tr>
<tr>
<td>Dream (1-5)</td>
<td>-0.004</td>
<td>0.190</td>
<td>0.98</td>
<td>0.99 (0.68,1.44)</td>
</tr>
<tr>
<td>Sleep Disturbances (count)</td>
<td>0.055</td>
<td>0.114</td>
<td>0.63</td>
<td>1.06 (0.84,1.32)</td>
</tr>
</tbody>
</table>

Table 4. Comparison of Prior Sleep Behavior with Seasonal Average for Non-Contact Injury Incidences (N=91 injury incidences)

<table>
<thead>
<tr>
<th>Sleep Measure</th>
<th>Season Average</th>
<th>1-Night Before</th>
<th>3-Nights Before (Avg)</th>
<th>7-Nights Before (Avg)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>vs. Season Avg</td>
<td>M (SD)</td>
</tr>
<tr>
<td>Sleep Duration (hr)</td>
<td>7.98 (1.15)</td>
<td>7.80 (1.87)</td>
<td>p=0.66</td>
<td>7.72 (1.68)</td>
</tr>
<tr>
<td>Sleep Quality (1-5)</td>
<td>3.58 (0.56)</td>
<td>3.54 (0.99)</td>
<td>p=0.69</td>
<td>3.53 (0.80)</td>
</tr>
</tbody>
</table>
Table 5. Association between Weekly Sleep Measures and Subsequent Week Non-Contact Injury

<table>
<thead>
<tr>
<th>Measure</th>
<th>IRR</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sleep Duration (hr)</td>
<td>1.22</td>
<td>0.94 – 1.59</td>
<td>.143</td>
</tr>
<tr>
<td>Sleep Latency (hr)</td>
<td>2.43</td>
<td>1.03 – 5.73</td>
<td>.042</td>
</tr>
<tr>
<td>Sleep Quality (1-5)</td>
<td>0.59</td>
<td>0.39 – 0.88</td>
<td>.009</td>
</tr>
<tr>
<td>Calm Sleep (1-5)</td>
<td>0.57</td>
<td>0.39 – 0.84</td>
<td>.005</td>
</tr>
<tr>
<td>Sleep Planned Length (1-5)</td>
<td>0.73</td>
<td>0.52 – 1.04</td>
<td>.081</td>
</tr>
<tr>
<td>Ease of Awakening (1-5)</td>
<td>1.10</td>
<td>0.74 – 1.64</td>
<td>.642</td>
</tr>
<tr>
<td>Ease of Falling Asleep (1-5)</td>
<td>0.67</td>
<td>0.46 – 0.98</td>
<td>.041</td>
</tr>
<tr>
<td>Dream (1-5)</td>
<td>0.90</td>
<td>0.67 – 1.21</td>
<td>.475</td>
</tr>
<tr>
<td>Sleep Disturbances (count)</td>
<td>0.95</td>
<td>0.78 – 1.14</td>
<td>.572</td>
</tr>
</tbody>
</table>
Table 6. ROC characteristics for univariate workload-injury prediction

<table>
<thead>
<tr>
<th>Threshold</th>
<th>AUC (95% CI)</th>
<th>Sensitivity/Recall</th>
<th>Specificity</th>
<th>Accuracy</th>
<th>NPV</th>
<th>PPV/Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance ACWR</td>
<td>1.57</td>
<td>0.54 (0.49-0.59)</td>
<td>0.27</td>
<td>0.87</td>
<td>0.86</td>
<td>0.99</td>
</tr>
<tr>
<td>Distance Monotony</td>
<td>1.30</td>
<td>0.52 (0.47-0.57)</td>
<td>0.33</td>
<td>0.73</td>
<td>0.72</td>
<td>0.99</td>
</tr>
<tr>
<td>Distance Acute Load</td>
<td>3.37</td>
<td>0.51 (0.47-0.56)</td>
<td>0.26</td>
<td>0.81</td>
<td>0.81</td>
<td>0.99</td>
</tr>
<tr>
<td>Distance Chronic Load</td>
<td>4.73</td>
<td>0.54 (0.49-0.59)</td>
<td>0.30</td>
<td>0.85</td>
<td>0.85</td>
<td>0.99</td>
</tr>
<tr>
<td>HSD ACWR</td>
<td>1.63</td>
<td>0.53 (0.49-0.58)</td>
<td>0.27</td>
<td>0.86</td>
<td>0.85</td>
<td>0.99</td>
</tr>
<tr>
<td>HSD Monotony</td>
<td>0.77</td>
<td>0.53 (0.48-0.58)</td>
<td>0.63</td>
<td>0.85</td>
<td>0.85</td>
<td>0.99</td>
</tr>
<tr>
<td>Distance Acute Load</td>
<td>1.34</td>
<td>0.52 (0.47-0.56)</td>
<td>0.52</td>
<td>0.54</td>
<td>0.54</td>
<td>0.99</td>
</tr>
<tr>
<td>HSD Chronic Load</td>
<td>5.60</td>
<td>0.58 (0.53-0.62)</td>
<td>0.67</td>
<td>0.47</td>
<td>0.47</td>
<td>0.99</td>
</tr>
<tr>
<td><strong>Women's Soccer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance ACWR</td>
<td>1.58</td>
<td>0.63 (0.57-0.69)</td>
<td>0.35</td>
<td>0.88</td>
<td>0.87</td>
<td>0.99</td>
</tr>
<tr>
<td>Distance Monotony</td>
<td>1.29</td>
<td>0.58 (0.52-0.65)</td>
<td>0.45</td>
<td>0.71</td>
<td>0.71</td>
<td>0.99</td>
</tr>
<tr>
<td>Distance Acute Load</td>
<td>3.53</td>
<td>0.53 (0.47-0.59)</td>
<td>0.20</td>
<td>0.89</td>
<td>0.89</td>
<td>0.99</td>
</tr>
<tr>
<td>Distance Chronic Load</td>
<td>4.73</td>
<td>0.58 (0.52-0.64)</td>
<td>0.37</td>
<td>0.85</td>
<td>0.85</td>
<td>0.99</td>
</tr>
<tr>
<td>HSD ACWR</td>
<td>1.53</td>
<td>0.62 (0.56-0.69)</td>
<td>0.38</td>
<td>0.85</td>
<td>0.84</td>
<td>0.99</td>
</tr>
<tr>
<td>HSD Monotony</td>
<td>0.77</td>
<td>0.60 (0.54-0.67)</td>
<td>0.82</td>
<td>0.35</td>
<td>0.36</td>
<td>1.00</td>
</tr>
<tr>
<td>Distance Acute Load</td>
<td>3.36</td>
<td>0.51 (0.45-0.57)</td>
<td>0.85</td>
<td>0.20</td>
<td>0.21</td>
<td>0.99</td>
</tr>
<tr>
<td>HSD Chronic Load</td>
<td>5.98</td>
<td>0.60 (0.54-0.66)</td>
<td>0.53</td>
<td>0.68</td>
<td>0.67</td>
<td>0.99</td>
</tr>
<tr>
<td><strong>Men's Soccer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance ACWR</td>
<td>0.99</td>
<td>0.58 (0.51-0.65)</td>
<td>0.61</td>
<td>0.57</td>
<td>0.57</td>
<td>0.99</td>
</tr>
<tr>
<td>Distance Monotony</td>
<td>1.14</td>
<td>0.56 (0.49-0.62)</td>
<td>0.76</td>
<td>0.38</td>
<td>0.39</td>
<td>0.99</td>
</tr>
<tr>
<td>Distance Acute Load</td>
<td>3.40</td>
<td>0.50 (0.43-0.56)</td>
<td>0.28</td>
<td>0.78</td>
<td>0.77</td>
<td>0.99</td>
</tr>
<tr>
<td>Distance Chronic Load</td>
<td>12.23</td>
<td>0.50 (0.43-0.58)</td>
<td>0.36</td>
<td>0.76</td>
<td>0.76</td>
<td>0.99</td>
</tr>
<tr>
<td>HSD ACWR</td>
<td>0.88</td>
<td>0.58 (0.51-0.65)</td>
<td>0.53</td>
<td>0.65</td>
<td>0.64</td>
<td>0.99</td>
</tr>
<tr>
<td>HSD Monotony</td>
<td>0.65</td>
<td>0.55 (0.48-0.63)</td>
<td>0.47</td>
<td>0.66</td>
<td>0.66</td>
<td>0.99</td>
</tr>
<tr>
<td>Distance Acute Load</td>
<td>1.31</td>
<td>0.54 (0.47-0.60)</td>
<td>0.85</td>
<td>0.25</td>
<td>0.25</td>
<td>0.99</td>
</tr>
<tr>
<td>HSD Chronic Load</td>
<td>0.91</td>
<td>0.55 (0.48-0.61)</td>
<td>0.23</td>
<td>0.86</td>
<td>0.86</td>
<td>0.99</td>
</tr>
</tbody>
</table>
Figure 1
Figure 2

Non-Contact Injury Incidence

- Perceived Sleep Quality
- Relative Workload
- Workload Monotony
- Chronic Workload
- Season Phase
- Session Type
- Day Relative to Match
- Session Congestion
- # Days Off
- # of Previous Injuries
- Sleep Latency
Figure 3
Chapter 4: Workload, Sleep, Fitness, and Wellness Factors Affecting Perceived Fatigue in Collegiate Soccer

ABSTRACT

Purpose: The aim of this investigation was to establish a causal mechanistic understanding of perceived fatigue in collegiate soccer utilizing workload, sleep, fitness, and wellness measures. Methods: Fitness (VO2max) was assessed preseason and 1-day acute workload, chronic workload (28-day accumulated distance), acute:chronic workload ratio (ACWR), indices of sleep (duration and quality), and perceived wellness (stress and fatigue) were collected daily from 107 collegiate male soccer athletes from five separate teams over the 2016 and 2017 seasons. Causal inferences were explored utilizing multi-level models and mediation analyses. Results: Fatigue (b=0.26), sleep quality (b=-0.15), stress (b=0.06), ACWR (b=0.24), session type (b=-0.45) and total distance (b=0.08) were significant predictors of perceived fatigue (all p<0.05). Session type was found to moderate the relationship between acute TL and fatigue. There was a significant indirect effect (b=0.005, p=0.002) of VO2max on fatigue, which was fully mediated by preceding day total distance. ACWR had both significant direct (b=0.239, p<0.001) and indirect (b=0.011, p=0.02) effects on fatigue, with 4.5% of the effect mediated by preceding day total distance. Conclusions: Players responded with higher fatigue levels when sleep duration, sleep quality, and chronic TL were low relative to baseline and when ACWR and perceived stress were higher relative to baseline per acute TL dosage. The type of session (match vs training) and chronic TL moderates the relationship between acute TL dosage and next day perceived fatigue response. Acute TL mediates a significant portion of the effect of aerobic fitness and workload spikes (ACWR) on next day perceived fatigue. Deriving causal inferences associated with report fatigue require accounting for moderating and mediating results of workload, wellness, sleep and fitness variables.
INTRODUCTION

Monitoring training load (TL) and fatigue state may aid practitioners in more effectively managing an athlete during periods of increased stress. Prescribing TL and assessing the impact of training represents an important step in the success of mitigating injury risk and encouraging positive physiological adaptations during a soccer season. Several means of assessing both external (distance covered, accelerations/decelerations, average speed) and internal (rating of perceived exertion, heart rate, oxygen uptake) aspects of training load exist. However, when selecting appropriate tools for monitoring the athlete’s responses to exercise load, time-efficient and non-invasive means are preferable. As noted by Thorpe et al, for a marker of fatigue to be valid it must be sensitive to variations in TL. Research has shown subjective ratings of wellness are sensitive to changes in exercise stress, in addition to being relatively easy to employ in a team setting.

In high-performance sport, recovery from congested playing periods and intense training or competition is critical and requires strategies to optimize performance reduce injury risk in subsequent sessions. Of the numerous recovery strategies and tools available to athletes such as active recovery, stretching, compression garments, and massage, sleep is regarded as the most powerful form of recovery available to athletes and understood as critical piece for physical and psychological well-being. This has been confirmed through both laboratory sleep loss research which has linked with poor cognitive function and performance and field research with athletes. Although athletes regard sleep as essential for both recovery and performance, evidence suggest elite athletes demonstrate less than optimal sleep characteristics when compared with normal, healthy individuals. Coupled with evidence suggesting sleep may be disrupted by exercise load in a dose-response manner, athletes may be at particular risk for compounding fatigue throughout an intense and congested schedule if sufficient rest is not realized. Taylor et al. found a significant effect of training volume on physical movement during sleep time, indicating restless sleep. Sleep has shown affected by overall training load and the time in which training occurs.
Improving understanding of the interrelatedness of fatigue determinants has the potential to drive more targeted and efficacious injury prevention programs. By conceptualizing both time-variant (e.g., acute TL, preceding night’s sleep quantity and quality) and time-invariant factors (e.g., aerobic fitness) as mediators and moderators of fatigue, practitioners and coaches may be able to prescribe interventions and more effectively appropriate risk management practices. For example, establishing the moderating effect of sleep characteristics (i.e. sleep duration and aspect of sleep quality) on the effect of acute TL on subsequent fatigue can help inform targeted sleep prescription strategies or allow the coach to modify training intensity real-time.

The aim of the present study was to establish the direct and indirect influence of workload, sleep and wellness on perceived fatigue in competitive soccer players. It is hypothesized that determinants of fatigue are many and that complex, interrelated pathways exist.

**METHODS**

**Subjects**

One-hundred and seven National Collegiate Athletics Association (NCAA) division I male collegiate soccer players (age, 20±2 y; body mass, 77.4±5.1 kg; height, 179.9±6.5 cm; %body fat, 9.9±2.4%; VO$_{2max}$, 53.8±4.1 ml·kg$^{-1}$·min$^{-1}$) from five separate universities participated in this study. All participants were medically cleared for physical activity by their respective university’s sports medicine department and free of any debilitating musculoskeletal injuries or contraindicated medical conditions. Institutional review board approval was obtained from all institutions and all participants provided written informed consent prior to the season. When the participant was under the age of 18, parental consent was obtained.

**Design**

This investigation was a prospective cohort study conducted with five NCAA Division I men’s soccer teams over the full 2016 (1 team) and 2017 (4 teams) NCAA soccer seasons (August to November).
Workload data are reported for field-based training and match sessions. A total of 6495 total sessions were recorded during the season (n = 4593 training sessions, n = 1902 match sessions).

**Procedures**

A heart rate (HR) and global positioning satellite (GPS) player tracking device was used to capture physical and physiological workloads during all training sessions and matches (Polar Team Pro, Polar Electro, Lake Success, NY). This device samples at a frequency of 10 Hz, which has been shown accurate and reliable outdoors. To avoid inter-unit error, players wore the same device for each training session. Players donned the player tracking device prior to the beginning of the session warm up to the end of the last organized training or match event. After each match or training session was complete, data were uploaded and subsequently exported to Microsoft excel spreadsheets (Microsoft Corporation, Redmon WA) for analysis.

A range of parameters were selected for analysis including total distance covered (i.e., Distance or Acute TL), total accumulated distance covered in the previous 28 days (i.e., Chronic TL) or acute to chronic workload ratio (ACWR). ACWR was computed using distance accumulated during the previous 7 days divided by the average distance accumulated over the previous 28 days.

Sleep was also assessed daily via a validated self-reported sleep diary, the Karolinska Sleep Diary (KSD). The KSD is an eleven item questionnaire used to evaluate several facets of sleep, including quantity and aspects of perceived quality, such as ease of falling asleep, overall perception of sleep quality, sleep disturbances, sleep calmness, feeling of rest, and ease of waking. The KSD questionnaire was distributed to each participant daily via each team’s designated athlete management system (CoachMePlus. Smartabase, Qualtics, Fitfor90). Participants were encouraged to complete the questionnaire upon waking. To reduce conceptual model dimension, self-reported sleep duration (SD) and sleep quality (SQ) were utilized in this analysis.
Perceived stress, fatigue and soreness were assessed prior to each training or match session. Participants were asked to assess their perceived wellness using their institutions monitoring practices. Either a 7 point Likert-type scale from –3 (worst) to +3 (best) using an online software program (fitfor90.com) or on a Likert-type scale from 1 (no stress, fatigue, soreness) to 10 (extreme stress, fatigue, soreness) was used. All wellness metrics were converted to within-athlete zscores to allow scale convergence and interpretability.

**Statistical Analysis**

Multi-level models were used to assess relationships between predictor, moderator, mediator and outcome variables. Mixed modelling was used for its ability to handle unbalanced fix factors and to account for repeated measures\textsuperscript{20}, which was seen with multiple players clustered within multiple teams. To account for inter-individual differences in workloads, sleep and wellness metrics, a multilevel random intercept was set for each player with an unstructured covariance matrix. Mediation and moderation relationships between variables were tested with conditional process modeling using bootstrapping methods. Each model testing for mediation between variables was simulated 1000 times and 95% confidence intervals were derived. Statistical analyses and plotting were conducted in R Studio (Version 3.5.2, R Core Team) with the “lme4”, “jtools” and “mediation” packages.

**RESULTS**

**Direct Effects on Distance and Next Day Fatigue**

Figure 1 displays the interrelatedness of workload, sleep, and aerobic fitness determinants on following day perceived fatigue. Table 1 displays mixed effects regression results for all predictor variables on total distance (Model 1) and next day perceived fatigue relative to baseline (Model 2). Model 1 indicates fatigue (b=-0.21, p=0.003), VO2max (b=0.09, p=0.001), stress (b=0.20, p=0.003), ACWR (b=0.78, p<0.001), chronic distance (b=0.02, p<0.001) and session type (b=-4.72, p<0.001) were significantly associated with total distance. Model 2 results indicate a significant direct effect of fatigue (b=0.26,
p<0.001), sleep quality (b=-0.15, p<0.001), stress (b=0.06, p=0.042), ACWR (b=0.24, p<0.001), session type (b=-0.45, p<0.001) and total distance (b=0.08, p<0.001) on next day perceived fatigue. VO2max (b=0.055, p=0.50), sleep duration (b=-0.026, p=0.410) and chronic TL (b=-0.003, p<0.810) were not significantly associated with next day perceived fatigue.

**Moderators of Distance on Next Day Fatigue**

All variables were assessed as potential moderators of acute TL and next day perceived fatigue. Results are displayed in Table 2. Session type was found to be a significant moderator (b=-0.050, p=0.012) of acute TL and next-day perceived fatigue, with higher fatigue scores reported following a match as compared to training (Figure 2). Although insignificant (p=0.096), chronic TL was identified as a potential moderator of the relationship between acute TL and next day perceived fatigue, with lower chronic TL values associated with higher perceived fatigue with increasing acute TL (Figure 2).

**Mediating Effects of Distance on Next Day Fatigue**

Results of the mediation analysis indicated the effect of VO2max on perceived fatigue is fully mediated by total distance covered in the preceding day. There was a significant indirect effect (b=0.005, p=0.002) of VO2max on perceived fatigue through total distance, with no significance found for either the direct (p=0.96) or total effect (p=0.416). The proportion of the effect mediated was 66%, however this was insignificant (p=0.414). Further mediation analysis indicates the effect of ACWR on next day perceived fatigue is partially mediated by previous day TL. The model indicates there is a significant indirect effect (b=0.011, p=0.02), direct effect (b=0.239, p<0.001), total effect (b=0.250, p<0.001) and proportion mediated (4.5%, p<0.001) of ACWR on next day perceived fatigue.

**DISCUSSION**

This study investigated direct, moderation and mediation pathways between workload, sleep aerobic fitness and perceived wellness in competitive soccer players. The presented analysis showed players
responded with higher levels of next day fatigue when sleep duration, sleep quality, and chronic TL were low and when ACWR and perceived stress were higher relative to baseline for a given acute TL dosage. Additionally, the type of session (match vs training) and a player’s 28-day accumulated TL was found to moderate the relationship between acute TL dosage and next day perceived fatigue response, with match days vs. training days and lower vs. higher chronic TL’s inducing higher fatigue levels relative to baseline, respectively. Acute TL was found to mediate a significant portion of the effect of aerobic fitness and workload spikes (ACWR) on next day perceived fatigue.

The idea of complexity is an important concept for making sense of occurrences which prove difficult to control or predict, such as the economy\(^{21}\), weather\(^{22}\), any living organism, or just about any assembly of people (e.g., organization or sports team).\(^{23}\) To date, mostly reductionist approaches have been taken in attempt to understand complex phenomena in sport such as fatigue in team-sports, which has entailed modeling isolated components (e.g., sleep duration, aerobic fitness level, prior injury status, etc.) on an outcome of interest and then constructing speculative inferences to explain how these parts interact.\(^{23,24}\) This study took a novel, integrative approach by investigating the interrelatedness of potential fatigue determinants. Our results indicate several measures of TL, sleep, wellness and aerobic fitness are both directly and indirectly related to perceived fatigue responses.

Perceptual measures of wellness taken daily may assist in early identification of non-functionally overreached states or confirming intentional, functionally overreached states. Multiple investigations have shown their sensitivity to changes in stress and fatigue in athletes\(^4,25,26\), as well as, their sensitivity to increased physical loading in soccer.\(^{27,28}\) Daily wellness measures are usually preferable as they are typically less time consuming than more extensive inventories, easy to implement prior to training, and inexpensive.\(^{29}\) Indeed, coaches have shown favor to short daily perceptual measure as an assessment of current monitoring trends taken by Taylor and colleagues show 80% of high-performance clubs use their own crafted questionnaires.\(^{30}\) In the current study, although perceived wellness measures of stress and fatigue were directly associated with following day perceived fatigue when accounting for acute TL.
effects, significant interactions between player well-being measures and acute TL were not found. Nevertheless, it is clear the increased levels of fatigue and stress are associated with higher levels of reported fatigue for a given acute TL dosage (Figure 2). These findings have important implications for practitioners and coaches as an imbalance TL and recovery can negatively affect player output and increase risk of maladaptation states such as non-functional overreaching or overtraining syndrome. These findings directly support the contention that compounding fatigue can occur when external workload stresses are not modulated when reported perceived-wellness rating are high.

The utilization of perceived ratings of wellness, such as stress, fatigue and soreness has also been used to quantifying responses to acute TL. Specifically, a number of studies have investigated the effect of pre-session perceived wellness on TL. Malone et al. found general measures of external, internal and external:internal workload ratios declined in response to reduced soccer player well-being. Similarly, Gallo et al. found reduced pre-training well-being indices corresponded to reductions in player load (-4.9 ± 3.1%) and external:internal workload measures. In the current study, we found both perceived fatigue and stress influenced workload output in the subsequent session. As expected, we found perceived fatigued was negatively associated with acute workload output (b=-0.21), indicating workload decreases with increased perceived fatigue. Interestingly, we found stress to have a positive association with acute TL, indicating higher stress levels were associated with greater TL (b=0.21). Although not assessed in the current study, these results may be due to various contextual factors such as session type or season phase.

Sleep is regarded as one of the most effective form of recovery available to athletes and an important method of promoting optimal psychological well-being. Indeed, in the current study both sleep quantity and quality directly affected following day perceived fatigue. To our knowledge, this is the first investigation to show measures of sleep quality are significantly associated with following day perceived fatigue in soccer players. Of note, previous studies have also investigated the association between sleep and subsequent TL, with Moalla et al, finding a significant negative association between acute TL and
perceived sleep quality (r =0.23). In contrast, Thorpe et al.\textsuperscript{32} did not find an association (b=-0.04, p=0.71) between perceived sleep and TL in elite soccer players over an in-season competitive phase. Our findings support those of Thorpe et al, in that no significant associations were found between either sleep duration or sleep quality changes and acute TL. The lack of association found between sleep changes and TL are likely due to the contextual factors affecting workloads such as session length, type, design and objectives.

Conceived from Banisters original fitness-fatigue model\textsuperscript{37}, Gabbett et al. introduced the concept of acute to chronic workload ratio,\textsuperscript{38} which gives a relative measure of load which has occurred in the previous week (i.e., acute load) compared to the average of the previous 4 weeks (i.e. chronic load). Conceptually, if the athlete has a high chronic load or high “fitness” and low acute load therefore low levels of “fatigue”, reduced injury risk is likely. However, as acute load spikes above chronic load tolerance, increased injury risk ensues. Indeed ACWR has been associated with injury risk in various contexts.\textsuperscript{39,40} Novel to this study was the finding that ACWR possessed both direct and indirect effects on fatigue. Specifically, about 5% of its effect on perceived fatigue was mediated by the preceding day’s TL. These findings present key considerations for injury risk modeling, as fatigue is purported an important mediator to preventable injury.\textsuperscript{41}

**CONCLUSION**

This study demonstrates the complexity and interrelatedness of fatigue determinants in competitive soccer. These findings suggest a wide array of player evaluation and monitoring practices are necessary to understand determinants of perceived fatigue. Additionally, practitioners monitoring player fatigue should be aware that aspects of aerobic fitness, perceived wellness, acute sleep variables, workload, and contextual factors such as the type of session conducted will all have varying levels of influence on next day perceived fatigue.

Acknowledgement:
The authors would like to express their sincerest thanks to the athletes, coaching staff, sports medicine staff and strength and conditioning staff of each participating institution for their time and commitment during the study. Additionally, the authors would like to gratefully acknowledge participating research personnel from all institutions involved.

Declaration of Interest:

The authors report no conflicts of interest. The authors alone were responsible for the content and writing of this article. They are thankful for the financial support for this research provided by the National Collegiate Athletics Association (NCAA).

**Figure Legend:**

Figure 1 – Directed Acyclic Graph of Workload, Sleep, Wellness and Sleep Association

Figure 2 – Moderating effects of sleep duration, sleep quality, stress, fatigue, ACWR and chronic TL on the relationship between acute TL and next day perceived fatigue.

**REFERENCES**


### Table 1. Multivariate Linear Mixed Effect Regression Model for Next Day Fatigue and Distance

<table>
<thead>
<tr>
<th>IV</th>
<th>DV</th>
<th>Estimate</th>
<th>SE</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1 (Intercept)</td>
<td>Distance</td>
<td>2.03</td>
<td>1.39</td>
<td>0.151</td>
</tr>
<tr>
<td>Fatigue (zscore)</td>
<td>Distance</td>
<td>-0.21</td>
<td>0.07</td>
<td>0.003 **</td>
</tr>
<tr>
<td>VO2max</td>
<td>Distance</td>
<td>0.09</td>
<td>0.03</td>
<td>0.001 **</td>
</tr>
<tr>
<td>Sleep Duration (zscore)</td>
<td>Distance</td>
<td>-0.05</td>
<td>0.07</td>
<td>0.474</td>
</tr>
<tr>
<td>Sleep Quality (zscore)</td>
<td>Distance</td>
<td>-0.06</td>
<td>0.07</td>
<td>0.357</td>
</tr>
<tr>
<td>Stress (zscore)</td>
<td>Distance</td>
<td>0.20</td>
<td>0.07</td>
<td>0.003 **</td>
</tr>
<tr>
<td>ACWR</td>
<td>Distance</td>
<td>0.78</td>
<td>0.12</td>
<td>0.000 ***</td>
</tr>
<tr>
<td>Distance (28 day)</td>
<td>Distance</td>
<td>0.02</td>
<td>0.00</td>
<td>0.000 ***</td>
</tr>
<tr>
<td>Session Type (Training)</td>
<td>Distance</td>
<td>-4.72</td>
<td>0.16</td>
<td>0.000 ***</td>
</tr>
<tr>
<td>Model 2 (Intercept)</td>
<td>Next Day Fatigue (zscore)</td>
<td>-0.59</td>
<td>0.44</td>
<td>0.188</td>
</tr>
<tr>
<td>Fatigue (zscore)</td>
<td>Next Day Fatigue (zscore)</td>
<td>0.26</td>
<td>0.03</td>
<td>0.000 ***</td>
</tr>
<tr>
<td>VO2max</td>
<td>Next Day Fatigue (zscore)</td>
<td>0.01</td>
<td>0.01</td>
<td>0.516</td>
</tr>
<tr>
<td>Sleep Duration (zscore)</td>
<td>Next Day Fatigue (zscore)</td>
<td>-0.03</td>
<td>0.03</td>
<td>0.406</td>
</tr>
<tr>
<td>Sleep Quality (zscore)</td>
<td>Next Day Fatigue (zscore)</td>
<td>-0.15</td>
<td>0.03</td>
<td>0.000 ***</td>
</tr>
<tr>
<td>Stress (zscore)</td>
<td>Next Day Fatigue (zscore)</td>
<td>0.06</td>
<td>0.03</td>
<td>0.042 *</td>
</tr>
<tr>
<td>ACWR</td>
<td>Next Day Fatigue (zscore)</td>
<td>0.24</td>
<td>0.05</td>
<td>0.000 ***</td>
</tr>
<tr>
<td>Distance (28 day)</td>
<td>Next Day Fatigue (zscore)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.833</td>
</tr>
<tr>
<td>Session Type (Training)</td>
<td>Next Day Fatigue (zscore)</td>
<td>-0.45</td>
<td>0.11</td>
<td>0.000 ***</td>
</tr>
<tr>
<td>Distance (km)</td>
<td>Next Day Fatigue (zscore)</td>
<td>0.08</td>
<td>0.01</td>
<td>0.000 ***</td>
</tr>
</tbody>
</table>

Abbreviations:  IV = Independent Variable; DV = Dependent Variable; SE = Standard Error
Significance:  * denotes significant at p<0.05; ** denotes significant at p<0.01; *** denotes significant at p<0.001.
### Table 2. Moderation Models for Next Day Fatigue

<table>
<thead>
<tr>
<th>IV</th>
<th>Moderator</th>
<th>DV</th>
<th>Estimate</th>
<th>SE</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance</td>
<td>Fatigue (zscore)</td>
<td>Next Day Fatigue (zscore)</td>
<td>-0.006</td>
<td>0.010</td>
<td>0.562</td>
</tr>
<tr>
<td>Distance</td>
<td>Sleep Duration (zscore)</td>
<td>Next Day Fatigue (zscore)</td>
<td>0.015</td>
<td>0.010</td>
<td>0.126</td>
</tr>
<tr>
<td>Distance</td>
<td>Sleep Quality (zscore)</td>
<td>Next Day Fatigue (zscore)</td>
<td>-0.004</td>
<td>0.009</td>
<td>0.641</td>
</tr>
<tr>
<td>Distance</td>
<td>Stress (zscore)</td>
<td>Next Day Fatigue (zscore)</td>
<td>0.009</td>
<td>0.011</td>
<td>0.424</td>
</tr>
<tr>
<td>Distance</td>
<td>Session Type (Training)</td>
<td>Next Day Fatigue (zscore)</td>
<td>-0.050</td>
<td>0.020</td>
<td>0.012</td>
</tr>
<tr>
<td>Distance</td>
<td>VO2max</td>
<td>Next Day Fatigue (zscore)</td>
<td>0.004</td>
<td>0.002</td>
<td>0.125</td>
</tr>
<tr>
<td>Distance</td>
<td>ACWR</td>
<td>Next Day Fatigue (zscore)</td>
<td>0.004</td>
<td>0.008</td>
<td>0.623</td>
</tr>
<tr>
<td>Distance</td>
<td>Chronic TL</td>
<td>Next Day Fatigue (zscore)</td>
<td>0.000</td>
<td>0.000</td>
<td>0.096</td>
</tr>
</tbody>
</table>

**Abbreviations:** IV = Independent Variable; DV = Dependent Variable; SE = Standard Error

**Significance:** * denotes significant at p<0.05; ** denotes significant at p<0.01; *** denotes significant at p<0.001.
Figure 1
Figure 2
Chapter 5: Utilization of Machine Learning to Predict Injury in NCAA Division I Soccer

ABSTRACT

**Purpose:** Investigate the ability of supervised machine learning techniques to predict injury in collegiate soccer and explore processing, sampling and injury types that may influence prediction accuracy.

**Methods:** Non-contact injuries, overuse injuries and non-contact muscle strain injuries were recorded from 256 athletes from 6 men’s and 6 women’s NCAA division I teams. Metrics including; athlete-specific factors (sex, starter status, position), session-specific factors (season phase, session type, days relative to a match), congestion factors (session, match, training) in addition to GPS-derived and engineered workload factors (ACWR, monotony, strain, acute/chronic loads) were considered. Principal component analysis was used to address multicollinearity in predictors. Synthetic minority over sampling technique and down sampling were used to address class imbalance. Area under the receiver operator characteristics curve (AUC) was used to evaluate model performance. **Results:** Logistic regression (AUC[95%CI]: 0.74[0.62-0.87]) and naïve bayes (AUC[95%CI]: 0.73[0.61-0.87]) performed equally as well as more complex algorithms such as a support vector machine (radial basis) (AUC[95%CI]: 0.74[0.62-0.86]) and random forests (AUC[95%CI]: 0.74[0.62-0.85]). Prediction ability was improved with non-contact muscle strain injuries when compared with all non-contact injuries. **Conclusions:** Supervised learning approaches to predicting subsequent injury offer limited use as a daily decision tool in collegiate soccer. Predicting more specific injury classifications such as muscle strains may yield better performance than broader injury classifications (all non-contact). The utility of complex modeling such as machine learning in sports injury prediction require further investigation with more informative risk factors.
INTRODUCTION

Non-contact injuries in sport are of interest to stakeholders, which research indicates may be reduced with intervention-based exercise programs. Overuse injuries are especially important to consider as these injuries have been deemed resultant of ‘load management error’. While establishing isolated risk factors is an essential step forward in injury research, most injury modeling practices fail to factor the interrelatedness of injury risk predictors. The multifactorial nature of phenomena such as injury may be better understood under Philippe and Mansi’s framework, which is referred to as the ‘web of determinants’. This concept was introduced to sport injury research by Bittencourt et al. (2016) in a narrative review discussing complex systems approach for injuries. They proposed, along with others, that to fully reveal the intricate landscape of sports injury etiology, complex systems thinking was needed.

Complex systems thinking has begun to influence injury research, however several inherent methodological implications and analytical barriers exist. The fundamental assumptions that are used in more orthodox statistical techniques are dissociated from complex systems analysis. For example, regression-based techniques do not account for system-wide occurrences resultant of adaptive feedback loops or effects which are time-distant form an injury incidence. However, complex systems approaches should not be viewed as a replacement for scientific reductionism or linear modeling, but rather as a supplementary method which may include traditional statistical approaches.

Acknowledged by Bittencourt et al. (2016), statistical learning techniques such as artificial neural networks (ANN) and classification and regression trees (CART) may be useful in uncovering non-linear interactivity. Indeed, these techniques have been used in the sports performance and injury arena successfully, as Pfeiffer and Hohmann found they could better predict talent development by non-linear (i.e., ANN) rather than linear methods (i.e., linear discriminant analysis). Additionally, Bittencourt et al. utilized recursive-partitioning CART techniques, which factor non-linear interactions among predictors, to predict knee valgus during landing following vertical jump.
More recently, supervised learning approaches have been considered with an array of predictors such as those related to the athlete (e.g., age, body mass index, role), prior workloads (acute, chronic and relative), prior injury and strength.\textsuperscript{12–14} Supervised learning approaches offer the potential to progress the field of injury risk modelling by allowing pattern recognition and non-linear interaction between predictors.\textsuperscript{3,15} The majority of injury prediction attempts have been conducted with a single team.\textsuperscript{13,16} More large-scale investigations are needed to elucidate the usefulness of supervised learning techniques on injury prediction.

The purpose of this research was to investigate the ability of supervised machine learning techniques to predict non-contact injury risk and explore processing, sampling and injury types that may influence prediction accuracy. We use a variety of data pre-processing and sampling techniques to combat class imbalance and multicollinearity between injury predictors.

\textbf{METHODS}

\textbf{Participants.} Two-hundred and fifty-six NCAA division I athletes from 12 separate university teams participated in this study. One-hundred and thirty-nine of the participants were female (age, 20±1 y; body mass, 64.7±6.1 kg; height, 166.8±6.2 cm; VO$_{2\text{max}}$, 46.8±4.0 ml·kg$^{-1}$·min$^{-1}$), while one-hundred and seventeen were male soccer players (age, 20±2 y; body mass, 77.4±5.1 kg; height, 179.9±6.5 cm; VO$_{2\text{max}}$, 53.8±4.1 ml·kg$^{-1}$·min$^{-1}$). All participants were medically cleared for physical activity by their respective university’s sports medicine department and free of any debilitating musculoskeletal injuries or contraindicated medical conditions. Institutional review board (IRB) and ethics approval was obtained from all institutions, with primary oversight and coordination provided by the University of <blinded for review> (IRB Approval ID: H17-134). All participants provided written informed consent prior to the season. When the participant was under the age of 18, parental consent was obtained.
Injury Data Collection. Injuries were diagnosed and recorded by a single member of each team’s medical staff (i.e., certified athletic trainer). Injuries were recorded according to the current consensus statement on recording of soccer injuries,\textsuperscript{17} which clarifies that an injury is “any physical complaint sustained by a player that results from a football match or football training, irrespective of the need for medical attention or time loss from football activities.” \textsuperscript{17} In addition to injury incidence, other pertinent information such as injury type, severity, location and mechanism were recorded. For this study, three classifications of injury were considered including all non-contact injuries, non-contact lower-extremity muscle strain injuries and overuse injuries that required medical attention, irrespective of time loss. Overuse injury is defined as “an injury caused by repeated micro-trauma without a single, identifiable event responsible for the injury.”\textsuperscript{17} Muscle strain injuries were selected because of previous reports of muscle injuries incurring the highest injury incidence and injury burden in soccer.\textsuperscript{18,19} Injury incidence rates were calculated by dividing the number of injuries by total exposures or exposure hours and reported as rate per 1000 exposures or hours.

Predictors. A full list of features selected for this analysis are shown in Table 1. A host of athlete specific, session-specific, calendar congestion and workload variables were selected for inclusion based on previous connectivity with injury risk and potential for interdependent relationships (i.e. moderating effects). Importantly, all time-variant features (i.e., workload and congestion metrics) were lagged by 1 day to not predict same day occurrence.

Athlete, Session and Congestion. Several athlete-specific, session-specific and seasonal congestions factors with the potential to influence injury risk either directly or indirectly were selected for analysis. To assess differences between player role within the team, athletes were classified as starters if they competed in greater than 60% of the total match time and started in greater than 60% of the total matches in the season\textsuperscript{20}, all other athletes were considered reserves. Athletes were additionally divided into position groups consisting of defenders, midfielders and forwards. Athletes were further grouped by the number of years they have been competing in intercollegiate athletics (range: 1-6). To examine the effect of season phase, injury risk during preseason, in-season and postseason were considered, with postseason
referring to the period directly following the in-season where conference and NCAA tournament play occurs. All day-exposures were additionally classified by days relative to an upcoming match (match day minus [MD-]). Days relative to upcoming match was analyzed as a continuous variable. The effect of overall session, match and training congestion, a continuous variable was used indicating how many overall sessions, training session, and matches each athlete had participated in the in the previous 7 days (acute) and 14 days (chronic). To examine the effect of previous injury on injury risk, rather than classify as a binary variable (injury vs. no injury) which doesn’t factor the total number of injuries sustained previously, a rolling cumulative sum was used for each player over the season.

**Workload Data Collection.** For this investigation, only training and match exposures were considered. Global positioning satellite (GPS) player tracking devices were used to capture workloads all training sessions and matches (Polar Team Pro, Polar Electro, Lake Success, NY). The 10 Hz GPS player tracking device has reported accuracy and reliability outdoors for 40 and 100 m total distances at four separate movement (i.e., walk, jog, run, sprint) velocities (Mean Difference= -1.04 to -2.78m; CV=1.17-3.16%) and during a team sport simulation circuit (Mean Difference=0.23m; CV=0.96%). Devices were attached to the body via a chest strap before the start of each practice. To reduce inter-unit error, players wore the same device for each training session. Players donned the player tracking device prior to the beginning of the session warm up to the end of the last organized training activity. After each training session was completed, data were synced to a Polar Electro server and subsequently exported to Microsoft excel spreadsheets (Microsoft Corporation, Redmond WA) for analysis.

Several workload features were engineered from total distance (TD) and total high-speed distance (HSD), which have been used previously in workload-injury research. HSD was considered distance in meters covered >15 km/h for women’s soccer and >19.8 km/h for men’s soccer. HSD zones were selected based on previously used zones in women’s soccer and men’s soccer. Exponentially weighted moving averages (EWMA), which account for the decaying effect of workload, were calculated for 3, 7 and 28 days of TD and HSD. Research my Murray et al. suggests ACWR methods using
EWMA’s instead of standard rolling average may be more sensitive to injury.\textsuperscript{29} Daily acute:chronic workload ratios (ACWR) by player for TD and HSD were calculated by dividing 7-day EWMA by 28-day EWMA. ACWR windows of 7 and 28-day windows were used as these are customary in workload-injury investigations.\textsuperscript{24,26,30} Both rolling 7-day means and rolling standard deviations of TD and HSD were computed to model workload monotony. Monotony was calculated by dividing each day’s rolling average of the previous 7 days by the rolling standard deviation of the previous 7 days. Training monotony has been previously linked with overtraining syndrome, with higher training monotony associated with increased illness.\textsuperscript{31} Additionally, rolling 7-day and 28-day sums were computed to represent traditional acute and chronic workload, respectively. Acute and chronic workload have both been associated with injury risk in prior research. Strain was calculated by multiplying monotony by the sum of the previous 7-day workload (i.e., total distance or HSD).

\textbf{Sampling Techniques.} To combat class imbalance learning error, two sampling approaches were taken. Under sampling is a sampling technique in which the majority class observations (i.e., no injury) are randomly removed until balance is attained between classes. Synthetic minority oversampling technique (SMOTE) is a sampling method which synthesizes a new minority instance between a pair of one minority instance and one of its K nearest neighbors.\textsuperscript{32} This process is also combined with under sampling the majority class, creating balance between previous minority and majority classes. Predictive models in the current study were built using original unprocessed data, data processed with SMOTE and data which was under sampled.

\textbf{Dimensionality Reduction.} All algorithms were trained on unprocessed data and features which were extracted using principal component analysis (PCA). PCA is a dimensionality reduction technique used to eliminate multicollinearity, which can lead to instability in errors.\textsuperscript{13,33} All continuous predictors were scaled and centered before PCA was used. A percent cumulative variance threshold was used to extract the fewest number of components explaining at least 95% of the variance in data.\textsuperscript{34}
**Model Building.** Data for each injury classification were partitioned into training and testing sets using a 70-30 split. For models to be of practical use in the field, it is imperative that they be proven to generalize to new, untested data. For that reason, training model predictions were tested on a separate data set consisting of 30% of the original data (test set). Data partitioning was conducted using stratified sampling, whereby data where randomly partitioned and stratified by injury classification. All models were built using 10-fold cross-validation with 3 repeats. Model hyperparameters were tuned during the repeated k-fold cross validation process using a random grid search method. Several commonly used algorithms were included in this analysis including logistic regression (LR), naïve bayes (NB), decision trees (DT), random forests (RF), support vector machine with radial basis function (SVM-R) and neural networks (NNET). These algorithms have been used previously for other research investigating machine learning approaches to injury prediction. While it is beyond the scope of this article to describe the interworking of each model, they were chosen 1) for comparison with other studies and 2) to encompass a range of probabilistic and complex models commonly applied to binary outcome problems. Generally speaking, SVM’s, RF and NNETs tend to perform better when working with high dimensional data and continuous features, while logic-based models such as LR, NB and DT tend to perform better when dealing with discrete and categorical features. Further, more complex algorithms (SVM, NNET, RF) perform better when multicollinearity is present and nonlinear relationships exist between the input and output features. Probabilistic models have the advantage of potentially performing well on relative small dataset, whereas more complex models such as SVM and NNET require large sample sizes.

**Model Evaluation.** Due to inherent issues with accuracy when evaluation rare events data (i.e., predicting no injury leads to 99% accuracy in the current data set), receiver operator characteristics were chosen for model evaluation. Predicted probabilities of injury are modeled by each algorithm allowing for area under the ROC curve (AUC) to be computed. AUC scores range from 0.5 to 1 with 0.5 being equal to random chance and 1 being perfect prediction. To produce 95% confidence intervals around the modeling process, 30 resamples of each modeling procedure was conducted.
RESULTS

Overall. Injury counts and rates expressed per 1000AE are shown in Table 2. Injury rates demonstrated consistency across training and testing data sets for all injury classifications.

Principal component analysis. A substantial amount of multicollinearity existed between continuous predictors. Figure 1 displays a correlogram of all continuous predictors (factor variables were still included into the training of each model), with darker blue circles indicating higher positive correlation between predictors and darker red circles indicating higher negative correlation between predictors. To address this potential issue in the modeling process, we applied a PCA approach to reduce dimensionality in the feature space and develop a set of uncorrelated predictors (Figure 2). In doing so, we found 12 components (dimensions) explained greater than 95% of the variance in the data (Figure 3). Of note, the first 3 components explained ~64% of the variance in the data suggesting a small number of meaningful predictors can be derived from a multitude of workload-related variables. Contribution of each continuous predictor to each component are showed in Figures 4 (dimension 1-6) and Figure 5 (dimensions 7-12). Dimension 1 appeared to represent more volume based metrics (strain, monotony, chronic loading), dimension 2 was heavily loaded by relative workloads (ACWR), dimension 3 was workload intensity and calendar congestion (HSD variables and session congestion), while dimension 4 was primarily comprised of previous session workloads (i.e. TD and HSD).

Model Performance. Non-contact prediction performance is shown in Figure 6. Overall model performance on non-contact injuries shows primarily poor performance. Neural networks (AUC\text{range}:0.46-0.61) and decision tree (AUC\text{range}:0.49-0.66) showed the lowest AUCs over all sampling and processing techniques, with some showing AUC’s worse than chance (<0.5). The most complex algorithms, SVM and RF, seemed to perform the best of all algorithms and showed the highest AUC’s on unprocessed, down-sampled data (AUC[95\%CI]: 0.71[0.63-0.78]) and PCA-SMOTE sampled data (AUC[95\%CI]: 0.71[0.63-0.78]).
Model performances for overuse injuries are shown in Figure 7. There was a noticeable increase in confidence interval width from non-contact injury to overuse, which is likely an artifact of reduced injury occurrences seen with overuse injury. Simple LR performed the best on unprocessed data (AUC[95%CI]: 0.70[0.60-0.81]), however unprocessed methods were mostly poor for overuse injuries. Overall, PCA processed methods outperformed unprocessed methods, especially for SMOTE-sampled data with LR (AUC[95%CI]: 0.71[0.61-0.80]), RF (AUC[95%CI]: 0.71[0.61-0.81]) and SVM-R (AUC[95%CI]: 0.71[0.62-0.81]) all averaging above an AUC of 0.7.

Model performances for non-contact muscle strain injuries are shown in Figure 8. Overall AUCs for non-contact muscle strain injuries were elevated above both non-contact and overuse injuries. Unprocessed data with probabilistic models performed better with LR (AUC[95%CI]: 0.74[0.62-0.87]) and NB (AUC[95%CI]: 0.73[0.61-0.87]) showing some of the highest AUC’s. SVM-R performed well on both down (AUC[95%CI]: 0.74[0.62-0.87]) and SMOTE (AUC[95%CI]: 0.74[0.62-0.86]) sampling with unprocessed data. RF also performed well on unprocessed SMOTE-sampled data (AUC[95%CI]: 0.74[0.62-0.85]).

**DISCUSSION**

The aim of this investigation was to assess the ability of supervised machine learning techniques to predict injury risk in collegiate soccer and additionally explore processing, sampling and injury types that may influence prediction accuracy. This investigation included a range of factors including; athlete-specific factors (sex, starter status, position), session-specific factors (season phase, session type, days relative to a match), congestion factors (session, match, training) in addition to commonly used workload factors (ACWR, monotony, strain, acute/chronic loads). Our results indicate complex ML algorithms do not outperform probabilistic models such as simple logistic regression or naïve bayes at predicting injury risk in soccer. Collectively, supervised learning techniques showed minimal predictive power and use as a
clinical diagnostic tool. However, some of our models still showed superior performance to previously reported univariate\textsuperscript{38} or multivariate modeling\textsuperscript{39} methods.

Our results are similar to two other studies investigating predictive injury modeling in Australian football\textsuperscript{13,14} and one in professional rugby league.\textsuperscript{16} Carey et al.\textsuperscript{13}, Thornton et al.\textsuperscript{16}, Ruddy et al.\textsuperscript{14} and this investigation found average AUC’s below 0.7, indicating unusable performance as a clinical predictive tool. Like Carey et al., we found predictive capacity tended to increase with more specific injury classifications (i.e., hamstring injury vs all non-contact injuries). It should be noted that in this investigation, as well in other modeling hamstring injury\textsuperscript{13,14}, sample sizes of more specific injuries are much lower than a broader non-contact injury classification. Due to such low numbers in the testing set, improved predictive power with more specific injury classification may not hold constant when larger samples sizes are used. Larger scale investigations with more informative injury risk factors are needed.

Events which are highly rare such natural disasters, fraud and injury present a unique challenge to predictive modelling. Class imbalance is a statistical learning problem where the frequency of one outcome class outweighs the other. This leads to algorithms favoring prediction of the majority class (overfitting) and not generalizing well to new data.\textsuperscript{33} Like previous approaches\textsuperscript{12–14}, we used various sampling techniques (down and SMOTE) to balance classes and reduce classification bias. To date, these techniques have yielded only mildly better results and sometime no improvement. Future work in this area may benefit from exploring other approaches to modeling rare events such as cost-sensitive learning algorithms, which penalize false classifications during the model training phase.

The term ‘prediction’ is often misused in sports injury research. This fact was discussed by McCall et al. who point out the difference between measuring association between risk factors and injury and predicting injury.\textsuperscript{40} Very few sports injury studies have assessed prediction with a separate hold-out testing set. While cross-validation methods do offer some utility, the standard for injury prediction should be assessing model performance on unseen data.\textsuperscript{33}
As shown in Figure 1, several predictors (particularly workload metrics) were highly correlated, which have potential to negatively influence certain probabilistic learning techniques. Multicollinearity is a known issue surrounding modeling workloads as metrics derived from GPS technologies are often highly correlated. Our PCA results support contentions by Weaving et al. that workload variables are often presenting redundant information. We found uncorrelated dimension consisting of 4 main groupings: relative (ACWR), volume (accumulated distance), intensity/frequency (HSD and calendar congestion) and acute (previous session) workloads explained a substantial portion of the variation in workload-derived metrics.

The limited number of non-contact muscle strains and overuse injuries negatively affected the confidence in which models generalized to hold-out data. These injury incidences are relatively rare and require very large sample sizes to strengthen the confidence of predictive models. Small injury samples are a redundant issue in injury investigations in sport, more large-scale multi-team, multi-year studies are needed to develop more useful models. Additionally, it is unclear if predictive modeling of injury would have been improved by modelling sexes separately. Future work should look to elucidate this potential influence. Further, future works should look to use complex modeling approaches with other potential injury risk factors such as previous injury, playing surface, environmental factors (wet-bulb globe temperature), anthropometric measurements, and physical and psychological well-being factors. Other biological testing such as blood biomarker changes or indicators of neuromuscular performance decrement may offer additional information and improve predictive performance.

**CONCLUSION**

Our modeling performance, combined with previous modelling performances from others, indicate injury prediction using supervised learning techniques may not be useful as a daily decision tool. However, integration of advanced learning techniques such as machine learning are relatively new to the field and require further methodological development and testing. Further works are needed to investigate the
utility of supervised learning to predict more specific injury types. Additionally, larger injury sample sizes are needed to improve performance with more complex modeling approaches. Data provided from workload capture technology and metrics engineered from this type of data are often offering similar information. Sports scientist may benefit from considering injury in relation to subgroupings of workload including volume, intensity, frequency, and relative change.

Acknowledgement:

The authors would like to express their sincerest thanks to the athletes, coaching staff, sports medicine staff and strength and conditioning staff of each participating institution for their time and commitment during the study. Additionally, the authors would like to gratefully acknowledge participating research personnel from all institutions involved. <contributors will be acknowledged after review >

Declaration of Interest:

The authors report no conflicts of interest. The authors alone were responsible for the content and writing of this article. They are thankful for the financial support for this research provided in-part by the National Collegiate Athletics Association (NCAA).

**Figure Legend**

Figure 1 – Correlogram of all unprocessed continuous predictors

Figure 2 – Correlogram of principal component analysis (PCA) processed predictors

Figure 3 – Scree plot explaining variance explained for all dimensions selected using principal component analysis (PCA)

Figure 4 – Variance contribution by PCA dimensions 1-6

Figure 5 – Variance contribution by PCA dimensions 7-12
Figure 6 – Area under the curve and 95% confidence interval (CI) for machine learning algorithms used to predict all non-contact injury.

Figure 7 – Area under the curve and 95% confidence interval (CI) for machine learning algorithms used to predict all overuse injuries.

Figure 8 – Area under the curve and 95% confidence interval (CI) for machine learning algorithms used to predict all non-contact muscle strain injuries.

REFERENCES


**TABLES/FIGURES**

Table 1. Predictor Variables

<table>
<thead>
<tr>
<th>Demographic</th>
<th>Description of Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Male, Female</td>
</tr>
<tr>
<td>Status</td>
<td>Starter, Reserve</td>
</tr>
<tr>
<td>Position</td>
<td>Forward, Midfielder, Defender, Goalkeeper</td>
</tr>
<tr>
<td>Session</td>
<td></td>
</tr>
<tr>
<td>Season Phase</td>
<td>Preseason, Inseason, Postseason</td>
</tr>
<tr>
<td>Session Type</td>
<td>Match, Training</td>
</tr>
<tr>
<td>Day Relative to Match</td>
<td>Number of days until scheduled match</td>
</tr>
<tr>
<td>Congestion</td>
<td></td>
</tr>
<tr>
<td>Match Congestion</td>
<td>Count of match days in previous 7 and 14 days</td>
</tr>
<tr>
<td>Training Congestion</td>
<td>Count of training days in previous 7 and 14 days</td>
</tr>
<tr>
<td>Off Days</td>
<td>Count of days off in the previous 7 days</td>
</tr>
<tr>
<td>Workload</td>
<td></td>
</tr>
<tr>
<td>Distance</td>
<td>Distance in meters covered during the session</td>
</tr>
<tr>
<td>HSD</td>
<td>Distance in meters covered &gt;15 km/h for females and &gt;19.8 km/h for males during the session</td>
</tr>
<tr>
<td>Accumulated Distance</td>
<td>Sum of distance in meters covered in previous 3, 7 and 28 days</td>
</tr>
<tr>
<td>Accumulated HSD</td>
<td>Sum of HSD in meters covered in previous 3, 7 and 28 days</td>
</tr>
<tr>
<td>EWMA Distance</td>
<td>EWMA of distance covered in previous 3, 7 and 28 days</td>
</tr>
<tr>
<td>EWMA HSD</td>
<td>EWMA of HSD covered in previous 3, 7 and 28 days</td>
</tr>
<tr>
<td>ACWR Distance</td>
<td>ACWR of EWMA distance, 3:28 and 7:28</td>
</tr>
<tr>
<td>ACWR HSD</td>
<td>ACWR of EWMA HSD, 3:28 and 7:28</td>
</tr>
</tbody>
</table>
Monotony Distance  Ratio between the mean and standard deviation of distance covered in previous 7 days

Monotony HSD  Ratio between the mean and standard deviation of HSD covered in previous 7 days

Strain Distance  Monotony of distance multiplied by 7-day accumulated distance

Strain HSD  Monotony of HSD multiplied by 7-day accumulated HSD

Abbreviations: HSD, High-speed Running Distance; EWMA, Exponentially Weighted Moving Average; ACWR, Acute:Chronic Workload Ratio

Table 2. Injury Counts and Rates for Training and Testing Sets

<table>
<thead>
<tr>
<th>Injury Classification</th>
<th>Training Set</th>
<th>Testing Set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Injury Count (per 1000AE)</td>
<td>Injury Count (per 1000AE)</td>
</tr>
<tr>
<td>Non-Contact</td>
<td>119 (9.35)</td>
<td>50 (9.17)</td>
</tr>
<tr>
<td>Overuse</td>
<td>45 (3.54)</td>
<td>19 (3.48)</td>
</tr>
<tr>
<td>Non-Contact Muscle Strain</td>
<td>49 (3.85)</td>
<td>20 (3.67)</td>
</tr>
</tbody>
</table>
Figure 1
Figure 2
Figure 3
Variance Contribution by PCA Dimension 1-6

Figure 4
Figure 5

Variance Contribution by PCA Dimension 7-12

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Dim.7</th>
<th>Dim.8</th>
<th>Dim.9</th>
<th>Dim.10</th>
<th>Dim.11</th>
<th>Dim.12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day Relative to Match</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance 3day sum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance 7day sum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance 28day sum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance EWMA acute 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance EWMA acute 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance EWMA ACWR 3 28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance EWMA ACWR 7 28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance EWMA chronic 28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance monotony</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance strain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HSD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HSD 3day sum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HSD 7day sum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HSD 28day sum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HSD EWMA acute 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HSD EWMA acute 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HSD EWMA ACWR 3 28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HSD EWMA ACWR 7 28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HSD EWMA chronic 28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HSD monotony</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HSD strain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Match Congestion 7d</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Match Congestion 14d</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Off Day 7d</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Session Congestion 7d</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Session Congestion 14d</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training Congestion 7d</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training Congestion 14d</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 6
Figure 7
Figure 8

Non-Contact Muscle Strain Injuries

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Unprocessed</th>
<th>PCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>DT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NNET</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SVM-R</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sampling Technique

AUC

None | Under | SMOTE
None | Under | SMOTE