Objective quantification of physical athlete load in women's rugby sevens

by

Amarah Epp Stobbe
B.Sc. (Hons), University of Waterloo, 2014
M.Sc., Queen's University, 2016

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the School of Exercise Science, Physical and Health Education

We acknowledge and respect the Lək̓ʷəŋən (Songhees and Xʷsepsəm/Esquimalt) Peoples on whose territory the university stands, and the Lək̓ʷəŋən and W̱SÁNEĆ Peoples whose historical relationships with the land continue to this day.

Supervisory Committee

Objective quantification of physical athlete load in women's rugby sevens

by

Amarah Epp Stobbe
B.Sc. (Hons), University of Waterloo, 2014
M.Sc., Queen's University, 2016

Supervisory Committee

Dr. Marc Klimstra (School of Exercise Science, Physical and Health Education)
Supervisor

Co-Supervisor or Departmental Member

Dr. Ming-Chang Tsai
Departmental Member

Dr. Nick Clarke
Outside Member

Dr. David Clarke
Additional Member

Abstract

Supervisory Committee

Dr. Marc Klimstra (School of Exercise Science, Physical and Health Education)

Supervisor

Dr. Ming-Chang Tsai

Departmental Member

Dr. Nick Clarke

Outside Member

Dr. David Clarke

Additional Member

This dissertation investigates objective factors that objectively quantify physical athlete load in women's rugby sevens.

I examined the use of objective metrics to impute missing rating of perceived exertion (sRPE) data from elite international matches. Despite employing various machine learning models, the best-performing random forest classifier achieved 26.5% accuracy, indicating that sRPE, and potentially session rating of perceived exertion (sRPE-CL) may be unreliable measures of athlete physical workload, suggesting the need for alternative metrics.

I explored the relationship between contact events and sRPE, as contact has not been previously considered in workload quantification. A linear regression incorporating playing time and contact explained 30.6% of sRPE variance, indicating that contact is associated with perceived exertion.

Building on this knowledge, I developed model that included athlete mass, number of contacts, playing time, game number, speed, and acceleration data to assess athlete workload (sRPE-CL). This model accounted for 79.3% of sRPE-CL variance.

I analyzed the influence of match-specific factors on sRPE-CL a speed-deceleration contact (SDC) model, and mechanical work. Results showed that sRPE-CL is highly variable, influenced by contextual factors such as score differential, match outcome, match category (e.g., medal final vs. pool match), opposition, and player experience. Conversely, mechanical work and the SDC model provided more objective workload assessments, influenced by fewer external variables.

Finally, I assessed the feasibility of imputing sRPE-CL using machine learning. Results demonstrate that sRPE-CL can be estimated using objective metrics, including mechanical work, and the SDC model.

Overall, this research provides alternative, objective strategies for the monitoring of physical athlete loads.

Table of Contents

List of Tables

List of Figures

List of Equations

Acknowledgments

I would like to express my deepest gratitude to my doctoral supervisor, Dr. Marc Klimstra, for his unwavering support, insightful guidance, and thoughtful encouragement throughout this journey. Dr. Klimstra has supported me in more ways than I could have ever expected, encouraging me to pursue opportunities beyond academia and presenting me with experiences many students only dream of. Thank you for your generosity as a supervisor and your steadfast mentorship in my career.

I am also profoundly thankful to Dr. Ming-Chang Tsai, whose mentorship, significantly enriched knowledge and enthusiasm in the data science space. His perspectives and advice resonated loudly and were instrumental in shaping this dissertation Special thanks to Callum Morris for his diligent efforts in data collection, which were crucial to the success of this study. I extend my sincere appreciation to my committee members, Dr. Nick Clark and Dr. Dave Clarke, for their constructive feedback and thoughtful critiques, which greatly enhanced the quality of this work.

I am grateful to the coaches, staff, and athletes of Rugby Canada for initiating this research journey with their questions. I also extend my sincere appreciation to the athletes who participated, as well as, the staff of Rugby Canada and the Canadian Sports Institute Pacific, whose collaboration provided practical insights and facilitated various aspects of this research.
My heartfelt thanks to my colleagues at the Motion and Mobility Lab, especially Dan Geneau and Pat Cormier, for their coding support and camaraderie that enriched this journey.

This research was possible through the generous financial support from the Mitacs Accelerate PhD fellowship and Own the Podium's Innovation 4 Gold grants.

Finally, I wish to express my deepest gratitude to my family—Eleanor, David, & Tim—for their unwavering love, encouragement, and meals fueling me through this endeavour. Their belief in me has been a constant source of strength in setting a new family record for fastest doctoral degree.

Dedication

This work is dedicated to my mom, Eleanor (who became acquainted with objective metrics), my dad, David (who asked lots of questions about objective metrics), and my partner, Tim (who kept me sane as I explored objective metrics).

1 Introduction

1.1 Introduction

In high-performance sports, effective athlete load monitoring has become a crucial practice for optimizing performance and minimizing injury risk. Athlete workload is generally accepted to encompass both internal and external domains, reflecting the physiological and mechanical demands placed on athletes elicited by training and competition (Gabbett et al., 2017; Staunton et al., 2022). Internal factors, such as heart rate, blood lactate, and psychological perceptions, represent physiological and psychological responses to stress, while external factors, including speed, distance, and event duration, quantify the physical demands of sport (Gabbett et al., 2017; Impellizzeri et al., 2019).

Given the wide range of both internal and external factors, numerous methods have emerged to quantify load. Internal workload measures like session rating of perceived exertion (sRPE-TL) and training impulse (TRIMP) are widely used due to their ease of application and accessibility (Impellizzeri et al., 2019). On the other hand, external workload measures, such as total distance covered and mechanical work, are often tracked using advanced tools like global navigation satellite system (GNSS) units and inertial measurement units (IMUs), which provide objective data on physical performance (Hulin et al., 2015; Bourdon et al., 2017). The interplay between internal and external measures is central to understanding factors affecting performance, athlete load, and effort. External physical outputs, such as distance and speed, act as primary determinants of internal physiological responses like heart rate and perceived exertion, making it essential to evaluate the associations between these external and internal measures in order to appreciate the athlete load of the performance (Impellizzeri et al., 2019).

In sport performance, athlete load varies significantly between individual and team sports, with unique challenges and considerations for each. Individual sports typically involve consistent, sustained efforts, in which load can be more easily quantified based on measurable outputs like distance or time (Halson, 2014). In contrast, team sports involve intermittent, high-intensity efforts that vary based on game dynamics, making load a more multi-faceted measure (Halson, 2014;

Boullosa et al., 2020). However, in competition settings, sport rules and regulations may limit the deployment of measurement tools (Kiely et al., 2020). Further, the feasibility of large-scale measurement strategies differs between individual sports where one person's performance is monitored and team sports where many athletes are involved and require monitoring. The complexity of monitoring load and the challenges of incomplete data necessitate advanced strategies for evaluating and imputing, or predicting, inputting load measures using statistical techniques and machine learning methods. This may enable comprehensive analysis even when data are missing or incomplete, improving our ability to monitor and optimize athlete performance (Waljee et al., 2013; Yin et al., 2019).

This dissertation will focus on women's rugby sevens, a sport that combines both high-speed running and physical collisions, making it uniquely demanding and potentially difficult to determine appropriate athlete load quantification. Given the intense physical demands and short recovery periods in women's rugby sevens tournaments, accurate load monitoring is essential for optimizing performance and preventing injury. The goal of the projects in this thesis is to attempt to determine objective, salient factors that quantify physical athletic load in women's rugby sevens, providing a foundation for more reliable and efficient load monitoring.

1.2 Terminology

As a means to enhance clarity, definitions for the following terms will apply to throughout the dissertation.

Athlete load / athlete workload – Sometimes referred to as "load", "workload" or with the determinants "athlete", "training" or "competition", athlete load or athlete workload reflects the physical, physiological, and psychological demands of performance in sport (Gabbett et al., 2017; Staunton et al., 2022). Notably, use of "load" or "workload" without determinants is disputed given the specific scientific metrics tied to these terms (Staunton et al., 2022). In this dissertation the terms "athlete load" and "athlete workload" will be used interchangeably with the focus of these terms on the physical demands of sport performance.

Competition load – A subset of athlete load, specifically referring to the competition demands experienced by athletes, in this dissertation the focus is on physical demands in rugby sevens matches.

Duration – The length of an activity, expressed in minutes (Hadded et al., 2017). In this dissertation, duration refers to the length of the rugby sevens competition session and is most easily understood as an athlete's playing time in a match.

Effort – A combination of objective (actual) and subjective (perceptual) dimensions of undertaking a task (Steele, 2020). Recognizing these two dimensions enables improved discourse around the methods of understanding effort, whereby *actual* effort is a reflection of what is *actually*, physically done in the task and *perceived* effort is a reflection of how the individual *perceives*, or feels, while performing the task (Steele, 2020). The focus of this dissertation is on actual effort.

Imputation – The process whereby missing data is replaced with estimated values (Kang, 2020).

Intensity – In exercise, generally considered the cause of "homeostatic disturbance" (Eston 2012).

Load – Operationalized in greater detail in section 1.3.

Player – Often refers to an athlete who is a part of a specific sport, in most cases a team sport. In this dissertation the term player appears and may be used interchangeably with athlete.

Rating of perceived exertion (RPE) – Historically, RPE reflected the rating given by an individual during exercise at any one point in time (Eston, 2012). More recently this term has expanded to encompass the score from a session or event of exertion (see Session rating of perceived exertion), however, this does create confusion in the application of the term (Eston, 2012).

Session rating of perceived exertion (sRPE) – This reflects the rating given that covers an entire event, or session, of exercise (Foster, 2001). More recently, this term has also been used to describe the training load measure whereby the rating from a training session is multiplied by a duration (Haddad et al., 2017). In this dissertation, use of sRPE refers strictly to the rating value given for a match and is not a reflection of athlete load (see Athlete load).

Session rating of perceived exertion training load (sRPE-TL) – This term reflects the product of the rating from a standalone session of exertion and the duration of that session (Haddad et al., 2017). This is considered a measure of athlete load (see Athlete load). In this dissertation, an adaptation of this acronym, sRPE-CL will be used in recognition that the data comes from competition and is therefore a reflection of competition load (see Competition load).

1.3 Athlete load

In high-performance sports, monitoring athlete workload has become an essential practice, offering valuable insights into the relationship between training, competition, recovery, and performance sustainability.

While there are several varying definitions of what athlete workload represents, the consensus is that a workload, in training or in competition, expresses some means of describing an athlete's effort in their training or competition performance (Gabbett & Domrow, 2007; Gabbett et al., 2017; Staunton et al., 2022). Generally, workload in the context of sport is accepted to include "both 'external' and 'internal' domains" (Staunton et al., 2022). These domains may work in combination or individually to represent elements reflective of an intensity and a volume or duration (Gabbett & Domrow, 2007).

Internal factors of load occur "inside" the body. These internal factors may represent physiological changes in heart rate, blood lactate, rating of perceived exertion, etc., psychological perceptions of performance, and even biomechanical responses to training or competition stress (Gabbett et al., 2017; McLaren et al., 2018; Impellizzeri et al. 2019).

External factors of load occur “outside” the body and are often considered to be reflective of the physical performance of the sport training or competition (Gabbett et al., 2017; Impellizzeri et al., 2019). Examples include speeds of movement, distances travelled, event duration, or even sport-specific skills like number of pitches, balls bowled, physical contact, etc. (Bradbury & Forman, 2012; Gabbett 2015; Hulin et al., 2014; Gabbett, 2016; Impellizzeri et al., 2019).

1.4 Measures of load

Given the wide range of both internal and external metrics that may describe change with a sport performance, it is not surprising that there exist numerous methods of quantifying the training and competition loads experienced by athletes.

1.4.1 Internal workload measures

Popular athlete workload measures include elements reflecting both an intensity and volume or duration (Gabbett & Domrow, 2007; Gabbett et al., 2017; Staunton et al., 2022). In the case of internal workload measures these include session rating of perceived exertion (sRPE-TL), and training impulse (TRIMP) (Halson, 2014; Haddad et al., 2017).

sRPE-TL or sRPE-CL (see Section 1.2) is the product of activity duration, in minutes, and an intensity scale. In most cases intensity comes from athlete self-reported perception of exertion using a scale like the Borg, or modified Borg CR-10 (Foster et al. 2001). Widely used, sRPE-TL is popular due to its accessibility and ability to be validly applied in environments where resources are limited or in those featuring a variety of training stimuli like team sports where the physical performance is not necessarily steady-state (Foster et al., 2001; Windt et al., 2020; Benson et al., 2021). While sRPE-TL or sRPE-CL is recommended to be calculated using sRPE values collected thirty minutes following the session, it has been found to be reasonably temporally valid using values collected days after the session in conditions where other similar sessions have not occurred (Foster et al., 2021). This measure is widely recognized as a holistic measure, rooted in the athlete’s subjective interpretation of their own attempt relative to their maximal potential performance (Haddad et al., 2017). sRPE-TL has been shown to have statistically significant associations with numerous external variables including total distance covered, high speed running volume and frequency, and accelerations (Haddad et al., 2017). Further, sRPE is associated with internal

measures of athlete load like heart rate (average, percentage above 80% and 90% of maximum), VO_2, and TRIMP (Banister's, Edwards', Lucia's, and Individualized TRIMP - Foster et al., 2001; Haddad et al., 2017).

TRaining IMPulse (TRIMP), first developed by Banister, quantifies athlete loads as the product of activity duration, heart rate, and an exponential weighting factor based on exercise intensity, referred to as Banister's TRIMP (Halson, 2014). Modified versions include Lucia's TRIMP and, Edwards' TRIMP, both of which assign discrete weightings to heart rate zones, and Individualized TRIMP (iTRIMP), which uses percent heart rate reserve and duration and an individualized exponential weighting factor (Edwards, 1993; Lucía et al., 2000; Manzi et al., 2009; Akubat et al., 2012). TRIMP models are used most commonly in endurance sports, in which athlete's physical efforts are more sustained and reflect physiological fitness and tools like heart rate monitors are regularly deployed in the sport environment (Halson, 2014). Team TRIMP values have been associated with aerobic performance (Akubat & Ali, 2011; Fox et al., 2018). iTRIMP has also shown to be significant correlation with 20m sprint time in soccer (Manzi et al., 2013; Fox et al., 2018). The flexibility of TRIMP measures make them applicable both to individuals as well as group training environments.

Associations between internal and external workloads stems from the relative stresses placed on the body (Halson, 2014). These stresses present a fatiguing element and are mediated by a variety of factors, often recognized as a fitness-fatigue model (Halson, 2014; Imbach et al., 2022). Generally, the strain from physical training can be mitigated by improved aerobic and anaerobic fitness, while psychological stressors are mediated by the concept of psychological resilience which includes mental processes and behaviours that protect an individual from negative effects of stressors (Busso et al., 1994; Sarkar, et al., 2014). Physiologic stressors affecting workload stem from changes, usually increases, in physical demands (Busso et al., 1994; Bossi et al., 2024). Psychological stressors affecting workload include perceptions around importance of the competition, level of opposition, as well as demands from media, etc. (Schultz de Arruda et al., 2019; Ferioli et al., 2021; Barrett et al., 2018; Mellalieu et al., 2009). Further to these stressors, athletes' abilities to tolerate workloads of varying intensities may be mediated by external factors including weather, like the temperature and humidity, and field-of-play conditions, like surface type, and demands surrounding competition, like travel and available resources to recovery, etc.

(Benjamin et al., 2020; Ponce-Bordón et al., 2024 ; Waterhouse & Edwards, 2004; Mellalieu et al., 2009). Considerations for drivers of internal workload measures, or drivers of physiological and psychological stressors, warrant recognition as they may explain variance in inter-individual responses to load (Bossi et al., 2024; Imbach et al., 2022).

1.4.2 Internal metrics used to inform workload

There are many internal metrics that both measure and assess the effects of athlete load, including the body's response to and perception of load. Popular metrics have been studied, including physiological measures, like heart rate and heart rate variability, biochemical measures, metabolite and enzymes measures in blood, and psychological metrics, such as subjective perceptions of wellness and effort, as well as measures that span many internal systems like sleep.

1.4.2.1 Heart rate to inform workload

The use of heart rate is popular in individual duration-based sports like running and cycling (Mujika, 2017). The use of zones of training based on the maximum heart rate achieved in a maximal exertion test is popular to inform training strategies for aerobic and anaerobic energy systems (Dellal et al., 2012; Scanlan et al., 2018). Further to this, with the widespread use of smartphone and smartwatch technology heart rate, and related measures such as heart rate variability (HRV), are easy to measure relatively accurately (Lundstrom et al., 2023). HRV represents the variation in duration between sinus beats and reflects the autonomic nervous system's balance between sympathetic and parasympathetic branches (Lundstrom et al., 2023). An increased HRV is generally reflective of parasympathetic activity, suggesting a more recovered state, whereas a lower HRV indicates more sympathetic nervous activity suggesting fatigue (Lundstrom et al., 2023). HRV has been associated with sRPE-TL in individual gymnasts, suggesting that HRV taken first thing in the morning reflects the previous day's training sRPE-TL (Sartor et al., 2013).

1.4.2.2 Biochemical measures informing workload

Biochemical measures are also indicative of stress, including blood lactate, lactate dehydrogenase, and creatine kinase (Goodwin et al., 2007; Baird et al., 2012; Coppalle et al., 2019). These

measures, while responding uniquely, will generally increase during physiologically fatiguing or physically fatiguing or traumatic events and may be associated with changes in sRPE or, in the case of creatine kinase, instances of high-intensity collisions, such as those in rugby league (McLellan et al., 2011; Goodwin et al., 2007; Coppalle et al., 2019).

1.4.2.3 Hooper-Mackinnon questionnaire to inform the effects of workload

In addition to physiological measures of load, psychological measures including subjective wellness questionnaires like the Hooper-Mackinnon questionnaire or ratings of perceived exertion are also widely used in sports to inform athlete load management strategies (Saw et al., 2016; Duignan et al., 2020). The Hooper-Mackinnon tracks self-reported questions about an athlete's fatigue, sleep quality, sleep hours, muscle soreness, stress, and mood (Gallo, 2016). While popular and widely used, questionnaires like the Hooper-Mackinnon have been shown to have limited relationships with load measures like sRPE-TL (Duignan et al., 2020). Duignan et al. (2020) demonstrated that only sleep quality, stress, and mood had limited association with sRPE-TL across a meta-analysis of the use of subjective measures in team sports, given that subjective questionnaires measure different constructs from athlete workload metrics. Duignan et al. (2020) also highlighted potential instances of multicollinearity such as mood which may be tied to stress and fatigue.

1.4.2.4 Rating of perceived exertion to inform workload

The rating of perceived exertion (RPE) is an athlete self-reported scale used to reflect how effortful an athlete felt their performance was between verbal anchors of no effort/low effort and maximal effort possible (Borg, 1998). This scale is a foundational measure used in the calculation of sRPE-TL and is usually reported using either a Borg scale (6-20), Modified Borg scale (Borg CR-10, 0-10 scale), or a pictural anchored version referred to as the OMNI scale (Borg, 1998; Foster et al., 2021). The sRPE data from this dissertation is based on the Borg CR-10 scale, first developed by Foster et al. (2001) which uses "American idiomatic English" terms such as "easy" and "hard" to provide verbal anchors along a 0 to 10 scale. This sRPE tool is useful both in steady state as well as intermittent- and high-intensity exercise spaces given the slightly skewed bias towards "hard" to "maximal" effort items whereby a 2 on the scale is "easy", a 3 is "moderate" a 4 is "somewhat hard" and a 5 is "hard", with 10 being "maximal" and is meant to capture the average RPE from

an entire “session” or activity (Foster et al., 2001). Like other subjective measures, the sRPE scale features limitations in application as it relies on the athlete’s perception of effort which may be influenced by factors like mood, performance outcome, or even weather conditions (Impellizzeri et al., 2004; Eston 2012; Viana et al., 2016; Schultz de Arruda et al., 2019). Further to this, the use of subjective scales are experience-dependent where less experienced athletes may struggle to accurately gauge their efforts, leading to inconsistent ratings (Gallo et al., 2014; Becirovic et al. 2024).

Another challenge with the use of self-reported measures, including but not limited to sRPE is the data collection process itself. Relying on athletes to provide information in a timely, accurate manner has improved with the use of athlete monitoring apps, telecommunications, and survey tools, however it remains the job of the athlete and practitioners in the environment to make sure the datasets are complete (Griffin et al., 2021). This presents challenges in competition settings where athletes face demands outside of their usual routine such as the presence of media, anti-doping requirements, medical treatments, spectators, etc. (Mellalieu et al., 2009). Further, in some professional sport settings, athletes’ and their unions have expressed concerns over the collection of athlete self-reported data due to concerns about athlete autonomy and potential misuse affecting contract negotiations or playing time decisions (McCall et al., 2023). This means that in many cases self-reported data is unavailable or not available at key times in competition for effective regular use.

1.4.2.5 Sleep to inform effects of workload

Finally, at the intersection of physiological and psychological performance factors, lies sleep. There are various ways that sleep is monitored; subjectively through questionnaires reporting on sleep time and quality, or objectively through REM periods requiring technology including wearables which may be invasive to sleep patterns (Clemente et al., 2021). Sleep has known associations with performance in that poor chronic sleep, less than seven hours a night for at least fourteen days, has some links to both poor performance, slower speeds in field sports and lower execution scores in judged sports, and a higher risk of injury (Dumortier et al., 2018; Lalor et al., 2019; Clemente et al., 2021; Huang & Ihm 2021; Fox et al. 2022;). However, acute disruptions to sleep quantity and quality may remain irresponsive to variations in load and vice versa (Knufinke et al., 2018). Ultimately, considerations for sleep as a factor in athlete load management often stem

from managing challenging travel and competition schedules as well as designing an optimal daily training environment including appropriate training times to enable optimal sleep quality and quantity for athletes (Lalor et al. 2019; Fox et al., 2022).

1.4.3 External workload measures and evaluation techniques of measures

Several popular methods exist for tracking external athlete load. These include acute-chronic ratios, exponentially weighted moving averages, and measures that reflect both intensity and duration and algorithms capturing athlete load and calculations of mechanical work (Hulin et al., 2015; Gabbett, 2016; Bourdon et al., 2017). These various approaches typically rely on data collected through GNSS units and/or inertial measurement units (IMUs) to provide comprehensive monitoring of athletic performance (Hulin et al., 2015; Gabbett, 2016; Bourdon et al., 2017).

One technique to evaluate measures of external athlete workloads is through the use of an acute:chronic workload ratio (ACWR) of measures like total distance and high-speed running distance are calculated in one of two ways (Hulin et al., 2015; Gabbett et al., 2016; Williams et al., 2017; Griffin et al., 2020a). The rolling average method uses measures from the day's training as compared to the previous seven days or from the seven-day training week compared to a 28-day training month (Hulin et al., 2015). The exponentially weighted moving average method assigns a decreasing weighting against loads over time, causing a time decay that more highly weights recent loads, usually over a 28-day period (Williams et al., 2017). The exponentially weighted moving average method is generally recognized as more sensitive to the nature of fitness and fatigue as the rolling average method is unable to account for variation within the set seven- or 28-day periods (Gabbett, 2016; Williams et al., 2017; Menaspà 2017) The mathematical basis of ACWRs, calculated either through the rolling average or the exponentially weighted moving average method, has been called into question as the formulas are mathematically coupled (Griffin et al., 2020a), whereby the inclusion of acute training loads in the chronic time period means that the chronic training value is directly influenced by the more recent acute training. A proposed solution to this issue is to uncouple the two by removing the acute load from the chronic. This approach has led to investigations comparing the coupled and uncoupled ACWR and findings suggest limited differences despite the theoretical concerns that the coupled inputs to ACWR may produce

spurious correlations (Lolli et al., 2017; Windt & Gabbett, 2017; Gabbett et al., 2019). Ultimately, further investigation as well as more detailed methodological descriptions explaining the specific ACWR calculations are necessary to determine the best method of calculating and applying ACWR moving forward (Windt & Gabbett, 2017; Griffin et al., 2020a).

To further improve efficiency in external load monitoring, sport-focused GNSS companies have produced algorithms meant to reflect athlete load. These algorithms include elements of speeds, distances, accelerations and decelerations, athlete mass, and tri-axial IMU measurements (Tierney et al., 2016; Nicolella et al., 2018; Bredt et al., 2020). These load measures have been used across a variety of team sports and are associated with sRPE-TL (Casamichana et al., 2013; Marynowicz et al., 2020). One popular algorithm called "Player Load" purports to use IMU data but returns a value of arbitrary units and is defined by a variety of descriptions (Equation 1.1 - da Glória Teles Bredt, et al., 2020).

$$Player\ Load = \sqrt{\frac{\left(a_{y(t)} - a_{y(t-1)}\right)^2 + \left(a_{x(t)} - a_{x(t-1)}\right)^2 + \left(a_{z(t)} - a_{z(t-1)}\right)^2}{100}}$$

Equation 1.1: Player load metric.

In the case of "Player Load", it is supposed to represent the rates of change in acceleration (jerk), however, the equation does not include a robust means to evaluate rate of change (Equation 1.1 - Bredt et al., 2020). While in theory, this measure is based on the rates of change in acceleration (see numerator of Equation 1.1), in practicality, the sum of changes in acceleration are computed which is different from the rate whereby the sum of change represents the total of all increases and decreases in acceleration and the rate of change illustrates more of an average in change and may be less sensitive to fluctuations in acceleration and deceleration (Bredt et al., 2020). This example illustrates the application of a "black box" metric. Unfortunately, this measure demonstrates differences from workloads calculated using IMU raw data based on the same purported formula (Equation. 1.1 - Malone et al., 2017; Nicolella et al., 2018).

As an alternative to "black box" algorithms, the use of mechanical work to describe external load has been suggested (Winter et al., 2016). Not surprisingly, there are notable connections between

mechanical work and energy expenditure (Cavagna & Kaneko, 1977; Zelik & Kuo, 2012; Winter et al., 2016). Additionally, Tuft and Kavaliauskas (2020) demonstrated a significant moderate association (r=0.329, p<0.01) between mechanical work and sRPE-TL in male field hockey athletes and Buchheit (2019) demonstrating the use mechanical work to monitor training loads in male professional soccer.

1.4.4 External metrics used to inform workload

External measures of physical performance are objectively quantified using tools that measure duration and kinematics related to skills or overall sport performance. Activity duration is a popular external measure and used across both individual and team-based sports, to quantity various efforts or in combination with other metrics to produce load measures like sRPE-TL (Haddad et al., 2017; Mujika, 2017). Time to complete an activity may be combined with distances or speeds to produce rate-based measures like distance per unit time (m/min, km/h, etc.) or time spent in a particular speed zone (Gabbett, 2015). Time is also used to calculate accelerations and decelerations, which represent yet another external measure of performance (Johnston et al., 2014).

1.4.4.1 Distance as a measure to inform workload

Total distance is a quantifier of training volume and is generally reported a single number, typically the mean absolute in team sports (Paquette et al., 2020). Distance may also be applied against thresholds of speed to measure distances covered in particular speed zones (Gabbett, 2015). High-speed running distances are a popular tool for monitoring as there are protective benefits to exposure to maximal speeds (Gabbett & Ullah, 2012; Edouard et al., 2019). Distance and speed-based measures are often collected using GPS/GNSS sensors (Johnston et al., 2013). These tools use standard algorithms as presets to define thresholds; in most cases 5.5 m/s (15 km/h) is used as a threshold for high-speed running (Clarke et al. 2014). Notably, this value was determined from male team sport athletes and recommendations to individualize measures based on athlete's maximal sprint physical efforts are advised (Clarke et al. 2014).

1.4.4.2 Accelerations and decelerations as measures to inform workload

Through the use of commercial GNSS units, it is possible to assess calculated metrics like accelerations (positive accelerations) and decelerations (negative accelerations) (Harper et al., 2019). Accelerations and decelerations are reported in a variety of ways; through counts (absolute or relative), absolute peak, and distances spent accelerating above or decelerating below a threshold (sometimes referred to as explosive distance in commercial GNSS programs) (Harper et al., 2019). The thresholds used to define high acceleration/deceleration vary between 2.0 m/s^2 to 3.5 m/s^2 among both commercial software and the literature, with lower ranges being advised for use with female athletes given sex-specific differences in absolute maximal sprint velocity characteristics (Mara et al., 2017; Harper et al., 2019; Bradley & Vescovi, 2015). However, challenges exist in the methodological treatment of acceleration data, whereby, deceleration data may be converted to positive to "normalize" the data which dramatically changes the interpretation (Harper et al., 2019). Critically, deceleration-specific monitoring may provide a protective effect against eccentric loading related injuries above and beyond the acceleration and sprinting demands (McBurnie et al., 2022). Notably, a count of decelerations coupled with total distance were identified as risk factors for injury in basketball (Caparrós et al., 2018).

1.4.4.3 Sport-specific measures to inform workload

Count-based measures beyond accelerations, decelerations, or number of sprints may come from sport-specific skill instances such as number of pitches thrown in baseball, balls bowled in cricket, or physical impacts in rugby (Bradbury & Forman, 2012; Hulin et al., 2014; Gabbett 2015; Gabbett, 2016). These measures may be quantified from calculated metrics collected using GNSS/IMU units or from visual measures coded by trained analysts and coaches producing counts of events (Gabbett, 2016). A popular quantified metric using both strategies reflects either counts of collisions or contacts in rugby (all codes) and Australian Rules Football or the intensity of these events (Gabbett, 2015; Gabbett, 2016). Collisions are considered as a rapid slowing of the body due to either the self-slowing of the centre of mass or due to the application of opposing force (from another person). Collisions are usually preceded by a 0.5-s increase in velocity where athletes adjust speeds to evade or match that of the opponent (Gray et al. 2018). Collision detection has been automated using GPS/GNSS/IMU sensors and machine learning strategies with some

success, but when compared to coded video by trained performance analysts there is a discrepancy in reporting (Gastin et al., 2014; Hulin et al., 2017). Further, the automated detection units were developed using male athletes and have been shown to inaccurately report collisions in female athletes (Clarke et al., 2017a).

1.4.4.4 Video data to inform measures of workload

The use of trained performance analysts to assess video data of sport performance is also common in the evaluation of tactical strategies deployed by teams to assess outcomes in training and competition (den Hollander et al., 2018). It is generally recommended to apply some sort of framework to any video analysis, with standardized definitions of events such as styles of play, notable skills, and errors (Hendricks et al., 2020). The use of video data further informs athlete performance by providing information about individual performance and successful, or unsuccessful, skills displayed in training and competition (Hendricks et al., 2020). Specifically, in competition, combining video analysis of tactical strategies and competition characteristics like the type of competition or the opposition further informs the experience of athletes (Hendricks et al., 2020).

1.5 Strategies to evaluate and impute load measures

Presently, no gold standard exists for the quantification of athlete load (Lambert & Borresen, 2010; Dawson et al., 2024). The selection and application of measures is mostly based on available resources and familiarity and knowledge of the methods and tools (Lambert & Borresen, 2010). To that end, measures are evaluated against one another to establish associations for interchangeable use such as the relationship between sRPE-TL and mechanical work (Buchheit 2019; Tufts & Kavaliauskas, 2020). Further, the use of known associations between measures allows for the potential to impute data where select measures are missing or of low quality (Griffin et al., 2021).

Given the challenges with collecting self-reported data, as well as challenges in collecting data from objective monitoring methods such as limitations with using GPS/GNSS indoors or with

using electronic equipment in water or extreme temperatures, athlete load data are often incomplete (Windt et al., 2020; Benson et al., 2021; Griffin et al., 2021). Cases of missed self-report data may be classified as data missing at random because athletes may omit a reported value due to fatigue or competition stress (Hong & Lynn, 2020). In cases of technology failure, the data are missing completely at random, due to an external malfunction and cannot be explained by other existing observed variables and not a systematic pattern, unlike data missing at random where differences between missing and observed data may be completely explained by other observed variables (Bhaskaran & Smeeth, 2014; Hong & Lynn, 2020). Fortunately, in both cases of missing data, it is possible to impute, or estimate, a replacement value to ensure continuity and completeness of the athlete load dataset (Hong & Lynn, 2020). Where imputation is necessary, common strategies include the application of statistical and machine learning techniques (Musil et al., 2002; Waljee et al., 2013; Yin et al., 2019).

1.5.1 Statistical techniques for imputation of load measures

One of the simplest, and accessible, techniques for the imputation of missing values that is especially popular in team sports is the use of *group mean substitution*. In this case, the alternative value used in place of the missing one is the average (mean) of the team (group) for that session's data (Yin et al., 2019; Benson et al., 2021; Griffin et al., 2021). This technique is especially popular with single measures of internal or external load such as sRPE, or distance, as it is fast and easy to apply (Benson et al., 2021; Griffin et al., 2021). However, other domains, including medical research, suggest that mean substitution, in any capacity is inferior to other statistical techniques including regression models (Musil et al., 2002; Celton et al., 2010; Waljee et al., 2013; Schmitt et al., 2015; Carey et al., 2016). Regression is a statistical technique that models the association between a dependent variable and one or more independent variables and, in the case of imputation can be used to predict the dependent value based on known values of independent variables (Shi & Conrad, 2009).

1.5.2 Machine learning techniques for athlete load measure imputation

Machine learning techniques have enhanced the use of regression models to describe relationships between variables. These techniques enable more flexible methods that can handle non-parametric data, or data agnostic of form, including *k*-nearest neighbours, neural networks, and support vector

machine to establish associations between variables and therefore, impute missing values from other variables (Alamoodi et al., 2021). The use of classification models is also applicable in the imputation of missing data, with random forest classification appearing as a popular choice across domains (Waljee et al., 2013, Hong & Lynn, 2020).

1.6 Athlete load in sport performance

Given the diverse nature of sport performance, unique methods of monitoring athlete load exist across individual and team sports. Individual sport performances generally fall into aerobic or endurance, continuous duration-based sports, such as track and distance running, triathlon, swimming, or skiing, or into a sport that rewards skills measured through distances or scores such as jumps and throws, gymnastics, archery, or artistic skating. Team sports are generally invasion-based sports, whereby one team attempts to outscore the other including soccer, rugby codes, and ice hockey. An underlying assumption of competition is that, regardless of sport, all athletes are trying their hardest to successfully meet, or exceed, the demands of their sport. Given the reactive, irregular nature of the physical efforts within a competitive match, several pacing strategies may exist for athletes in competition including an "all out" strategy in which physical effort is not limited or scaled and a "reserve" strategy in which physical effort may not be maximal or supramaximal (Waldron & Highton, 2014). The "reserve" strategy may be employed in cases where athletes know they will be performing repeated efforts, such as in games played back-to-back or played on consecutive days (Waldron & Highton, 2014). Athlete load may reflect the moderated physical efforts in competition as well as the intermittent nature of the sport, and as such may behave differently than in continuous, individual activities (Waldron & Highton, 2014).

1.6.1 Athlete load responses in individual sports

In an endurance, or distance-based, sport in which competitions challenge athletes to cover a set distance in as little time as possible, load responses increase over time (Halson, 2014). In the case of sRPE, in submaximal physical efforts done to exhaustion, such as those in a race (running, swimming, triathlon, road cycling), sRPE drifts in a linear increase as a function of the total exercise duration (Horstman et al., 1979; Noakes, 2004; Eston, 2012). This suggests that the athlete's perception of the intensity of effort may be moderated based on an anticipatory

understanding of the event duration (Noakes, 2004; Eston, 2012). sRPE-TL also increases over time in a similar manner, in both steady-state and interval running and cycling (Foster et al., 2021). Like sRPE-TL, TRIMP also increases with activity duration, however, TRIMP tends to increase exponentially when working above the lactate threshold due to the shift towards greater anaerobic energy use with the increase in lactate accumulation in the blood (Akubat & Abt, 2011).

1.6.2 Athlete load responses in team sports

Field- or invasion-based team sports are intermittent in nature, with athletes performing high-intensity movements and sport-specific skill execution in response to events in the field of play interspersed with periods of low-intensity activities like walking or jogging (Baker et al., 2015).

TRIMP generally follows a curvilinear response in individuals exercising over time, both in individual and in team sports, increasing over the length of a single session, and over the course of a week or training period it may increase in a non-linear way before decreasing as rest occurs between individual sessions (Akubat & Abt, 2011; Taylor et al., 2018). Further, significant, positive linear relationship between Banister's TRIMP and sRPE-TL and Edwards' TRIMP and sRPE-TL have been demonstrated in soccer and gridiron football (Akubat et al., 2012; Clarke et al., 2013). Interestingly, Taylor et al. (2018) identified TRIMP methods using exponential weightings such as Banister's TRIMP and iTRIMP are better suited for use than linearly weighted methods such as Edwards' TRIMP and Lucia's TRIMP to inform the dose-response relationship in intermittent sports like rugby union over several days of training.

The use of sRPE-TL is especially popular in team sports due to ease of application (Haddad et al., 2017; Benson et al., 2021). Similarly to individual sports, sRPE-TL increases with duration as well as with increasing intensity of sport performance (Askow et al., 2021). Interestingly, sRPE-TL is closely related to TRIMP (Banister's and Edwards') as well as player load algorithms across soccer, gridiron football, and various rugby codes (Clarke et al., 2013; Roe et al., 2017; Hulin et al., 2017; Barr et al., 2019; Askow et al., 2021). While sRPE-TL is easy to deploy, the challenges with using RPE, including the fact that RPE is sensitive to experience with past use and past activities as well as challenges with self-reported values, moderate the effectiveness of the calculation of sRPE-TL from session duration and RPE (Eston, 2012;

Gammon et al., 2016; Haddad et al., 2017). This suggests that the subjective nature of RPE values is sensitive to a variety of factors, making sRPE-TL a multi-faceted holistic measure (Haddad et al., 2017).

Objective, external, load monitoring strategies such as the calculation of mechanical work or the application of player load algorithms are popular in team sport environments as GPS/GNSS units are commercially available with software that supports efficient processing and reporting for applied practitioners (Malone et al., 2017). With strong relationships to sRPE-TL as well as TRIMP, the use of these algorithms demonstrates similar behaviour to sRPE-TL and TRIMP in team sports; increasing over time with potential moderating effects based on the intensity of play (Hulin et al., 2018).

1.6.3 Considerations for athlete load monitoring in women's rugby sevens

Rugby Sevens (sevens) was first introduced as an Olympic Sport at the Rio 2016 Olympics (International Olympic Committee, 2019). It is a fast-paced derivative of traditional Rugby Union, following virtually the same rules, but with seven-minute halves and seven players per team on-field (World Rugby, 2025a). The sevens tournament format is such that teams of twelve athletes will play five or six games in a two- or three- day period (Furlan et al. 2015; World Rugby, 2025a). Sevens features both high-speed running as well as a similar level of physicality as Rugby Union games through tackles or ball carries. Furthermore, this is the only women's team sport in the Olympic programme where physical contact is sanctioned (International Olympic Committee, 2019). With the short game times, highly concentrated tournament schedule understanding the athlete experience to mitigate injury, win games and tournaments, and prepare players for success in future elite competition is critical.

A key challenge in the rugby sevens environment are the regulations around equipment use. World Rugby mandates the use of only specific monitoring tools to be worn following extensive safety testing and, in competition, only authorizing the use of GPS/GNSS units (World Rugby, 2025b). Without the ability to reliably monitor heart rate in competition, it is not possible to deploy TRIMP-based measures and instead the focus is on the use of objective load measures or sRPE-CL.

The player load algorithms available with commercial GPS/GNSS units authorized by World Rugby have been developed using male rugby union or league players, who are, on average, capable of achieving higher accelerations due to higher body masses (Clarke et al., 2017b). Data from the 2015 fifteens world cup suggests that male players range from 179.2(±6.4)-188.5(±7.1) cm in height with body masses between 87.2(±7.8)-115.9(±8.6) kg based on position (smaller players are Backs, larger players, Forwards - Fuller et al., 2017). Women's sevens players are smaller both in height and mass, averaging 166.8(±5.4) cm tall and 65.4(±5.7) kg in mass (Fuller et al., 2017). This discrepancy in anthropometrics between male and female players supports the conclusions made by Clarke et al. (2017a) that contact detection using wearable sensors requires further sex-specific investigation to produce reliable and valid information. Further, Bradley and Vescovi (2015) support this conclusion, suggesting that sex-specific thresholds for locomotion are needed given differences in maximum sprint velocities between male and female soccer players.

sRPE-TL or sRPE-CL may be deployed in both a training and competition environment as it does not reply on equipment to be worn in-competition but will enable practitioners to align training and competition data to gain a comprehensive understanding of the athletes' load in a team. However, the use of sRPE-TL is called into question given the high sensitivity of sRPE, one of the contributing metrics, to a variety of conditions, including match-specific factors like opponent and game importance, activity duration, and athlete experience with the sRPE scale, the other contributing metric to the calculation of sRPE-TL (Rejeski & Ribisl, 1980; Eston, 2012; Barrett et al., 2018; Schultz de Arruda et al., 2019; Ferioli et al., 2021). sRPE has been reported to be highly correlated to duration, which is reasonable given that duration is one of the two measures that make up sRPE-TL (Marynowicz et al., 2020). The holistic nature of sRPE-TL a single measure, is highly sensitive to the sRPE measure that produces it. The subjective nature of the RPE measure makes it challenging to identify the contributions of the various elements of sport performance. Without knowledge of the contributing objective factors to an athlete's load, it is not feasible to train for, mitigate, or even identify key drivers of athlete load related to actual, physical effort.

As a means to illustrate the variety of factors which may potentially influence the elements of athlete load, intensity and duration, in women's rugby sevens, a selection of variables have been presented in a directed acyclic graph in Figure 1.1.

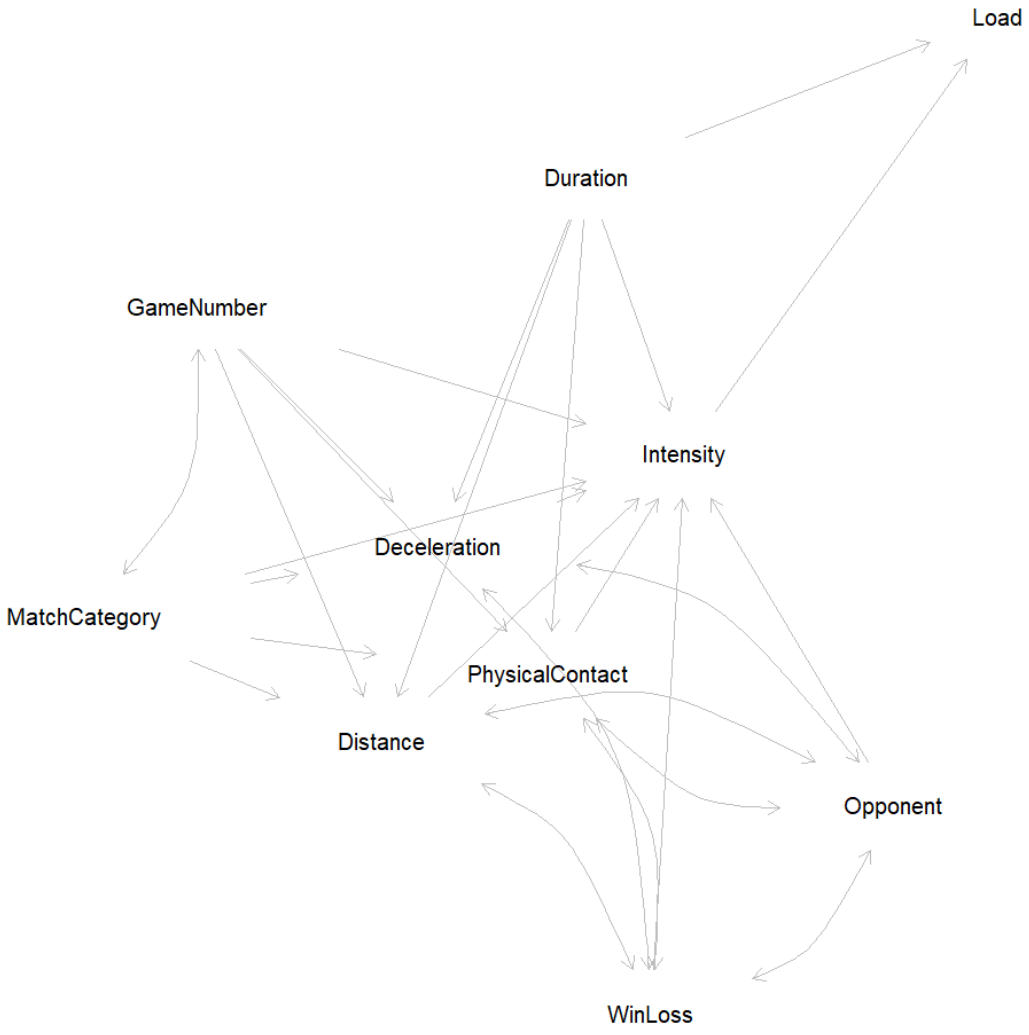

Figure 1.1: Directed acyclic graph of factors which may influence athlete workload.

In Figure 1.1, the inclusion of phrases unique to the rugby competition environment are included. GameNumber refers to the order in which the game, or match, appears in the tournament, with tournaments having between one and six total matches. MatchCategory refers to the type of match, or game, and may fit under the *pool play* where teams are ranked into a bracket based on results and move into *knockout* play where the team is ranked in the top eight for the tournament in order to play for medals or into *ranking* matches where teams play to determine tournament ranking in positions nine and below. WinLoss refers to the match outcome, a win or a loss – however, in *pool play*, ties may also occur. Opponent refers to the other team being played. PhysicalContact refers

to the inclusion of elements of contact that exists in rugby. Distance refers to the ground covered by athletes during play. Deceleration refers specifically to events of negative acceleration, or slowing, regardless of speed. All of these measures are explored in more detail in the subsequent chapters of this dissertation.

1.7 Outline and specific aims of this dissertation

The goal of this dissertation is to determine objective, salient factors that quantify physical athlete load in women's rugby sevens.

To address this goal, I hypothesized that objective metrics may be useful in imputing missing RPE data such that athlete workload could be continuously tracked without significant changes to currently monitoring methods available in rugby sevens (Chapter Two). Key objective factors were used to impute sRPE by a variety of machine learning strategies across various levels of missingness in the data. Unfortunately, given the poor performance of the models, with the best performing random forest classifier producing 26.5% accuracy, it was determined that the use of RPE and subsequently the use of sRPE-CL may not be dependable as a quantification of athlete workload and that an alternative measure may be more appropriate for use.

In parallel to this, I also investigated whether salient components can be identified from the existing, popular sRPE-CL workload metric using more objective measures. First, as there have been no investigations into the effect of contact in perceived experience of load we investigated what relationship measures of contact had with sRPE (Chapter Three). Data were collected from elite, international women's sevens matches and included measures of sRPE, athlete-worn GPS, coded video footage, and match characteristics. These measures were used to inform a linear regression investigating the influence of the total number of contacts experienced in a match by an athlete, their playing time, and their reported RPE. This model determined that 30.6% of global variance in RPE could be explained by the number of contacts and playing time.

The investigation into salient factors affecting an existing workload metric was further enhanced through analysis with a more detailed regression model that included athlete mass, number of contacts in a match, playing time, game number, speed, and acceleration data against sRPE-CL

(Chapter Four), informed by measures included in Figure 1.1. This model demonstrated improved explanatory power, whereby 79.3% of global sRPE-CL variance was accounted for through playing time, select game numbers in a tournament, and distances covered at low speeds and low accelerations as well as at moderate speeds and high decelerations.

Using objective models, such as mechanical work, and a speed-deceleration-contact (SDC) model, the imputation of sRPE is improved such that sRPE is able to be imputed by a variety of machine learning strategies (Chapter Five).

Further investigation into the influence of contextual, match-specific characteristics, identified in Figure 1.1, on sRPE-CL, the SDC model, and mechanical work highlighted key differences between the two athlete workload quantification strategies (Chapter Six). sRPE-CL appears to be highly subjectively variable as it is contextualized by numerous match characteristics including score differential, match outcome, match importance (pool, ranking playoff, medal final, etc.), opposition, and athlete expertise. Conversely, mechanical work and the SDC model appear to provide a less subjective understanding of workload, influenced by fewer contextual factors.

Portions of this dissertation have been published or are in the process of being published. Chapter two has been published in *Machine Learning & Knowledge Extraction* (Epp-Stobbe, Tsai, Klimstra, 2022a). Chapter three has been published in the *Journal of Strength and Conditioning Research* (Epp-Stobbe, Tsai, Morris, & Klimstra, 2022b). Chapter four is under preparation for publication. Chapter five has been published in *Applied Sciences* (Epp-Stobbe, Tsai, & Klimstra, 2025a). Chapter six has been published in *Applied Sciences* (Epp-Stobbe, Tsai, & Klimstra, 2025b).

2 Comparison of imputation methods for missing rate of perceived exertion data in rugby

2.1 Abstract

Session rate of perceived exertion (sRPE) is used to calculate athlete load. Incomplete load data, due to missing athlete-reported sRPE, can increase injury risk. The current standard for missing sRPE imputation is daily team mean substitution. However, sRPE reflects an individual's perceived, group mean substitution may be suboptimal. This investigation assessed an ideal method for imputing sRPE. 987 rows of data were collected from women's rugby sevens competitions. Daily team mean substitution, *k*-nearest neighbours, random forest, support vector machine, neural network, linear, stepwise, lasso, ridge, and elastic net regression models were assessed at different missingness levels. Statistical equivalence of true and imputed scores by model were evaluated. An ANOVA of accuracy by model and missingness was completed. While all models were equivalent to the true RPE, differences by model existed. Daily team mean substitution was the poorest performing model and random forest, the best. Accuracy was low in all models, affirming sRPE as multifaceted and requiring quantification of potentially overlapping factors. While group mean substitution is discouraged, practitioners are recommended to scrutinize any imputation method relating to athlete load.

2.2 Introduction

A standard, and widely accepted, sport metric calculated to determine an athlete's training and competition load is the session rating of perceived exertion-competition load (sRPE-CL (Haddad et al., 2017; Benson et al., 2021; Griffin et al., 2021). sRPE-CL is calculated by multiplying the athlete self-reported rate of perceived exertion (sRPE), on a 10 point scale, by the duration of the activity (Haddad et al., 2017). The sRPE scale represents ten, unique points, or classes, whereby each class is distinct, usually in increasing level of effort where low numbers represent low effort and the highest number, ten, represents an athlete's maximal possible level of effort with key phrases anchoring the scale and offering a frame of reference to the levels of effort (Haddad et al., 2017). For example, 0 may have the descriptor of "rest", 2 "easy", 7 " very hard", and 10 "maximal" (Haddad et al., 2017). For elite team sports, the sRPE data for each session is collected through athlete self report and the activity duration is

normally collected through an athlete-worn tracking device (ATD) which collects various kinematic metrics including distance, speed, acceleration and time (Cummins et al., 2013a). While the data from the ATD can be reliably collected in training and competition with the guidance of a sport science technician, there are difficulties in athlete adherence to self report sRPE data. Due to these difficulties, missing sRPE data is a common issue in sport training and competition environments and therefore sRPE-CL data cannot be dependably calculated (El-Masri & Fox-Wasylyshyn, 2005; Cummins et al., 2013a; Windt et al., 2018; Griffin et al., 2021). Considering the small sample size of elite athlete populations, this missing data limits the statistical assessment of training and competition load to support data-driven sport decision making (Benson et al., 2021; Griffin et al., 2021).

Incomplete and potentially inaccurate athlete load data can result in a number of deleterious outcomes and places athletes at risk of inappropriate training recommendations, potentially leading to physical unpreparedness, injury, or burnout (Gabbett, 2016). An important example is the female rugby sevens competition environment where teams play five or six games in a two or three day tournament, often with multiple tournaments happening in a few weeks. This high volume of competition requires a critical focus on athlete management. The use of ATDs, including GPS monitors, is an option for athlete load monitoring, as data collected from ATDs has been used to develop proprietary algorithms that model athlete load and are related to rate of perceived exertion (Cummins et al., 2013a). However, these proprietary algorithms include details that are not disclosed (Clarke et al., 2015; Clarke et al., 2017a). Further, it has been suggested that ATD load algorithms alone may not be optimized to accurately quantify the loads experienced by female rugby athletes, leading to an increased reliance on athlete-reported sRPE for the evaluation of training load (Clarke et al., 2015; Clarke et al., 2017a).

In order to potentially mitigate missing sRPE data in sport, mathematical techniques for the imputation, or prediction, of missing values present a unique solution (El-Masri & Fox-Wasylyshyn, 2005; Cummins et al., 2013b; Windt et al., 2018; Griffin et al., 2021). Traditionally, in sport research, missing value imputation (MVI) occurs via value substitution or through classification and regression models (El-Masri & Fox-Wasylyshyn, 2005). Substitution uses alternative values in place of the missing value (El-Masri & Fox-Wasylyshyn, 2005), while classification or regression models may use other known variables

to predict the missing one and therefore these methods may be better because of the inclusion of associated athlete-specific metrics and not just one variable (El-Masri & Fox-Wasylyshyn, 2005). In dealing with sRPE from athletes, Benson et al. (2021), and Griffin et al. (2021) advocate for the use of group mean substitution, referred to as daily team mean substitution, whereby the average of the known group sRPE data is used in place of a single missing athlete's RPE value for that same day without influence from any other variables (Benson et al., 2021; Griffin et al., 2021). Carey et al. (2016) used more detailed approaches including linear regression, multivariate adaptive regression splines, random forests, support vector machines, neural networks, naïve Bayes, C5.0 decision rules, and ordered logistic regression. Unfortunately, the literature is sparse in terms of comparing single imputation with machine learning methods. Benson et al. (2021) did compare single imputation methods against a least-squares boosted regression tree model, finding that this method was not as robust as daily team mean substitution. However, it must be noted that only one regression imputation strategy was used and therefore comparisons between alternative regression or classification strategies with group mean substitution relating to athlete load data is limited (Benson et al., 2021).

Arguments for group mean substitution focus on the ease of implementation as a team average is a simple calculation and preferable to a missing datapoint (Benson et al., 2021; Griffin et al., 2021). However, given that sRPE is a reflection of an individual athlete's subjectively percieved effort, group mean substitution may over- or under-estimate training load data (sRPE-TL). Further, there is evidence from other domains that mean substitution may be inferior to other more common statistical approaches such as linear regression, random forest classification, or neural networks (Musil et al. 2002; Celton et al., 2010; Waljee et al., 2013; Schmitt et al., 2015; Carey et al., 2016). Given the plethora of imputation methods available to account for missing data in many fields, from simple linear regression to alternative machine learning models, it is important to consider if an optimal method of imputation is available and how this method may compare to the current standard of group mean substitution (Fox-Wasylyshyn & El-Masri, 2005; Cummins et al., 2013b; Waljee et al., 2013; Schmitt et al., 2015). Therefore, the purpose of this investigation is to compare the current standard method of sRPE imputation (daily team mean substitution) to other methods for predicting sRPE data in elite women's rugby sevens competition.

2.3 Materials and methods

Through retrospective qualitative analysis, the effectiveness of sRPE missing value imputation was explored through statistical modeling of other objective metrics collected during games in a cohort, observational study. Twenty-one women's sevens players (25.5 ± 3.90 years old, 169.4 ± 5.89 cm tall, and 71.0 ± 5.64 kg) provided sRPE data for 101 international matches (2017-2020). The University of Victoria provided ethics approval for the use of voluntary data collection and the investigation complied with the principles outlined in the Declaration of Helsinki. Further, match date, match number within the tournament, and opponent, were provided for each match. All data were anonymized by team staff prior to analysis.

Subjective sRPE data were collected following the completion of the match using a 0-10 scale with athletes providing one sRPE rating for the whole match (Eston, 2012; Comyns & Flanagan, 2013; Mujika 2017). This sRPE data were collected as players returned to the team's designated rest area following the match, usually within 15 minutes post-match. Players self-reported by recording their value on a clipboard that displayed a CR-10 scale with verbal anchors as reference. Values outside of the bounds of the 0-10 range, as well as integer values were discarded from analysis. Additionally, objective variables from ATDs, worn between the shoulder blades in a custom harness of each athlete, collected athlete playing time and total distance covered in each match (Apex v2.50, StatSports, Newry: UK), were available for potential inputs into imputation models.

Footage of each match were evaluated to produce a count of all contacts (sum of tackles, carries, contested restarts of play following a try, and rucks) (Sportscode v 11, Hudl, Lincoln: USA). The operational definitions used to code the forms of contacts were developed by coaching and analysis staff, maintaining the team's current analysis practices and applied by one trained analyst (Gabbett & Kelly, 2007; Gabbet et al., 2007; King et al., 2010; Wheeler et al., 2011).

A six match subset of 65 complete player-match rows of data were coded twice by one trained analyst on two separate occasions. A two-way mixed effects, absolute agreement, single rater

intraclass correlation (ICC 3,1) determined the reliability was 0.99 (95% Confidence Intervals at 0.98-0.99), demonstrating excellent intra-rater reliability (Koo & Li, 2016).

As a means to model the sRPE relationship (dependent variable) from objective metrics available through athlete-worn ATD units, match video footage, and provided by team staff (independent variables), statistical models were used to classify and predict sRPE data, before comparisons of true sRPE data and model-predicted sRPE data were made (R version 3.4.4, Vienna, Austria). A total of 987 rows of data were used for analysis.

In all models except for daily team mean substitution, sRPE data were predicted using match number, player, opponent, total distance in meters, playing time in minutes, and contact count. Prior to modeling, the residual plots and normality plots of sRPE data were evaluated for normality. Further details on the explanatory variable data selected as objective variables to improve the imputation of subjective data are found in Table 2.1.

Table 2.1: Details on explanatory variable data used in models.

Variable	**Type**	**Method of Data Collection**
sRPE	Integer	Athlete self-report, measured in arbitrary units
Match Number	Integer	Integer reflecting match order in tournament (i.e.: first game played = 1, second game = 2, etc.)
Player	Integer	Integer used in place of name to anonymize athlete
Opponent	Integer	Integer used in place of name to anonymize opponent
Total Distance	Float	ATD, measured in meters,
Playing Time	Float	ATD, measured in minutes
Contact Count	Integer	Match footage, coded and evaluated by team analyst

The models used to classify and predict sRPE data were selected based on a combination of models used in current sport literature to impute missing sRPE data as well as models that can be completed using open-source software (Musil et al., 2002; Celton et al., 2010; Carey et al.,

2016; Benson et al., 2021; Griffin et al., 2021). In this investigation, sRPE values were classified and predicted using daily team mean substitution (Benson et al., 2021; Griffin et al., 2021), regression models (linear, (R *stats* package) stepwise (R *MASS*), lasso, ridge, elastic net (lasso, ridge, and elastic net using R *glmnet*)), k-nearest neighbours (R *FNN*), random forest (R *randomForest*), support vector machine (R *e1071*), and neural network models (R Core Team, n.d.; Riedmiller 1994; Tibshirani, 1996; Breiman, 2001; Zou & Hastie, 2005; Fritsch et al., 2019; Beygelzimer et al., 2022; Friedan et al., 2022; Liaw & Wiener, 2022; Meyer et al., 2022; Ripley et al., 2022; Hsu et al., 2003).

Data were divided into a training and test dataset, whereby 80% of the data were designated for training models and 20% for testing models to produce predicted sRPE scores, iterated 100 times with mean values used for downstream analysis. The same explanatory variables (Equation 2.1) were used for each model, with variance influence factors reported in Appendix 1:

$$\text{sRPE} = \text{Match Number} + \text{Player} + \text{Opponent} + \text{Total Distance} + \text{Playing Time} + \text{Contact Count}$$

Equation 2.1: Equation for imputation of sRPE.

Predicted sRPE scores were then compared to the true sRPE scores from the test dataset and the accuracy of each model were calculated. Accuracy, or the rate of correctly predicted sRPE scores, R^2, and root mean square error (RMSE) were identified as key metrics of interest in evaluating if the models were able to appropriately impute the sRPE value in comparison to the true sRPE.

The imputed values from the test dataset, at 20% missingness, were compared against the true sRPE values using a paired-samples equivalence test (paired TOST) to establish statistical equivalence, or more practically interchangeability of models (Lakens et al., 2017; Lakens, 2018). 20% missingness was determined to be a reasonable level of missingness as practically that was the equivalent to 2-3 missing sRPE values within a team of rugby sevens players which, on advice of team staff, represented the regular outcomes of data collection (Benson

et al., 2021; Griffin et al., 2021). The paired-samples equivalence test used bounds of *Cohen's d* × $\boldsymbol{\sigma}$, in this case using a *Cohen's d* of 0.2 to represent a small effect size (Lakens, 2017).

To explore the cases of divergence in accuracy, all imputation strategies were tested at different levels of missingness, in 5% increments, from 5% to 30% and iterated 100 times. A one-way ANOVA compared model accuracy by type, by missingness, and the interaction of model type and missingness in recognition that model accuracy may be dependent on level of data missingness. This investigation hypothesized that different types of models would improve accuracy over daily team mean substitution (Hawthorne & Elliot, 2005; Kang, 2013). Further it was hypothesized that as levels of missingness increase, accuracy would decrease, both in general across all models as well as by model type.

Finally, a supervised model based on a relationship between sRPE and total distance (Equation 2.2), was used to identify how accuracy changes by model type across levels of missingness (5% increments from 5% to 30% imputed data).

$$\text{sRPE} = \text{Total Distance}$$

Equation 2.2: Equation for supervised application of sRPE.

Regressions were generated from data imputed using the different models and across different levels of missingness. One-way ANOVAs assessed the influence of model type and missingness on the slope. It was hypothesized that model type may drive particular significant differences in the regressions, especially as 0% missing data, or data with no imputed values, was included for analysis. This analysis would highlight cases where model selection diverged from the true data across levels of missingness.

2.4 Results

2.4.1. Description of data

The frequency of athlete-reported sRPE values used in developing the models is shown in Figure 2.1 (mean sRPE = 7 ± 1.9 au). Investigation of the residual plot showed a random scatter of points and normality plot showed the residuals fall on a straight-line indicating the normality assumption was appropriate for sRPE (Appendix 2). On average, athletes covered

1082.86 m of total distance (± 439.78 m), played 11.04 minutes (± 4.67 minutes), and experienced 5 contacts (± 3 contacts) per match.

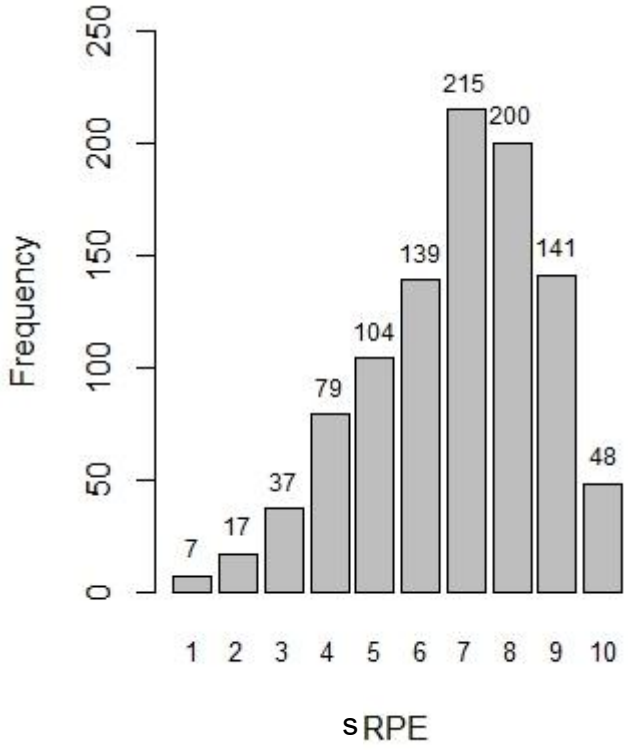

Figure 2.1: Frequency of athlete self-reported sRPE values.

2.4.2. Model performance

Imputation model accuracy, R^2, and RMSE values are reported in Table 2.2.

Table 2.2: Imputation model accuracy, R^2, and RMSE at 20% missingness

Model	Accuracy	R^2	RMSE
Daily Team Mean Substitution	0.216	0.009	1.832
Linear Regression	0.248	0.306	1.602
Stepwise Regression	0.247	0.305	1.603
Lasso Regression	0.227	0.264	1.651
Ridge Regression	0.233	0.274	1.650
Elastic Net Regression	0.227	0.265	1.651
k-Nearest Neighbours	0.239	0.268	1.653
Random Forest	0.265	0.407	1.480
Support Vector Machine	0.255	0.371	1.541
Neural Network	0.226	0.157	1.862

2.4.3. Comparison of models

Paired-samples equivalence tests of each imputed model against the true sRPE resulted in all tested models being deemed statistically equivalent to the true sRPE data ($p < 0.05$).

The one-way ANOVA of the data at 20% missingness found a statistically significant difference in model accuracy by imputation model type (F (9, 5940) = 86.83, $p < 0.05$), however did not report statistically significant differences by missingness (F (5, 5940) = 0.99, $p > 0.05$) or the interaction of missingness and model type (F (45, 5940) = 0.86, $p > 0.05$). A Bonferroni post hoc test determined that statistically significant differences in the mean differences existed between select models (Figure 2.2). Both daily team mean substitution and random forest differed from all other models ($p < 0.05$) (Figure 2.2).

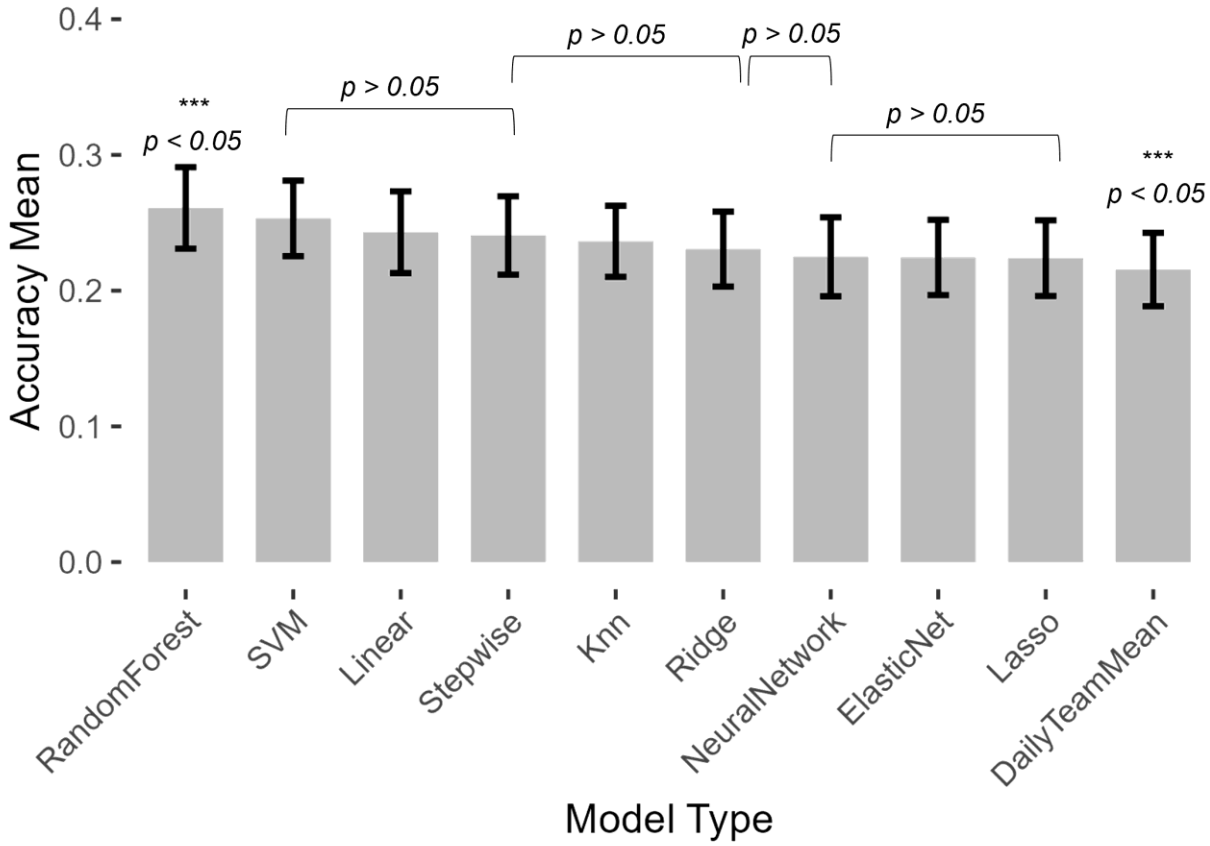

Figure 2.2: Model accuracy by type at 20% missingness, where *** indicates significance.

The accuracy of each model across cases of missingness is highlighted in Figure 2.3.

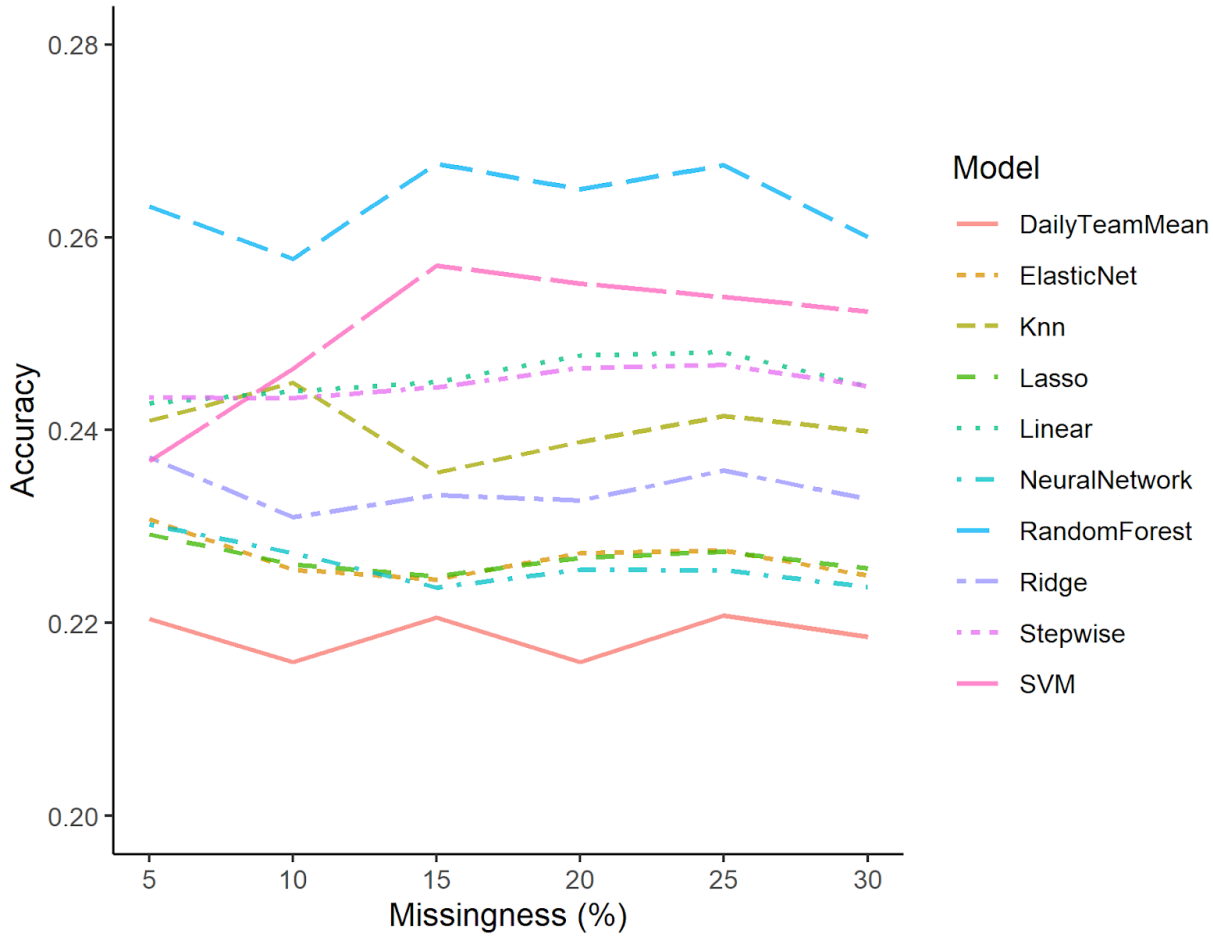

Figure 2.3: Model accuracy by data missingness from 5-30% missingness.

The one-way ANOVA of the slope of the supervised regression found statistically significant differences in slope by model type (F (9, 1454) = 1.93, $p < 0.05$), and by level of missingness (F (6, 1454) = 2.83, $p < 0.05$). A Tukey post hoc revealed that the daily team mean substitution and neural network model types were significantly different than all other models across all levels of missingness, including the complete dataset (0% imputed data).

2.5 Discussion

This study is the first to compare different methods of imputation across levels of missingness of sRPE data in women's rugby sevens. Overall, daily team mean substitution was outperformed by every other method in terms of accuracy, with the random forest model performing better than other models. Daily team mean substitution was not equivalent to any other model and the limited accuracy, R^2, and relatively high RMSE affirms that the team

average is not a suitable proxy for individual athlete data. Furthermore, all tested models performed poorly for accuracy and RMSE across multiple levels of missingness. Overall, our results suggest that the present substitution method as well as other common statistical models are not suitable imputation approaches and that the prediction of missing data may require more investigation and the use of more robust statistical approaches which consider the inclusion of the numerous factors affecting the individual's performance (Schmitt et al., 2015; Carey et al., 2016; Mujika et al., 2017; Benson et al., 2021; Griffin et al., 2021).

While the popularity of daily team mean substitution stems from the efficiency of substitution over other methods, the poor accuracy, R^2, and RMSE scores (Table 2.2) of the daily team mean substitution relative to other methods of imputation suggests that this is not the most robust option. The finding that mean substitution is a poor candidate for imputation is common in data with human subjects like those of the medical or athletic performance fields. Musil et al (2002) performed imputation using regression and substitution models and noted that while all methods have limitations, mean substitution was the least effective and linear regression was the most effective of the imputation models. Waljee et al. (2013) found that mean imputation produced the greatest error while random forest, the least. Further, in Australian football, an open-skill field sport with similar skill demands to rugby sevens, Carey et al. (2016) found success in imputing sRPE data using non-linear regression models, over neural networks. The very low accuracy across levels of missingness (Figure 2.4) suggests that daily team mean substitution consistently underperforms sRPE imputation compared to methods which rely on additional athlete information (Benson et al., 2021; Griffin et al., 2021). Daily team mean substitution may be a poor method for data from individual athletes in a team sport, such a rugby sevens, due to multiple varying levels of factors that could impact the athlete's perceived experience. For example, athletes may participate for different time periods (ie: starting player vs substitute player), be asked to perform specialized skills by position (ie: kicking) and experience different levels of sprint efforts or contact (Bartlett et al., 2017; Epp-Stobbe et al., 2022b). Additionally, given the global nature of the sRPE metric, tactical decision-making, mental fatigue, or other psychological states may influence perceived experience (Eston, 2012). This therefore may require multiple factors to be quantified and used to contribute to imputation methods and may account for the improved accuracy of other imputation models tested that use multiple variables for their calculation (Mujika, 2013;

Mujika 2017; Epp-Stobbe et al., 2022b). Conversely, it may be possible that daily team mean substitution may be a suitable model for sports where athletes are performing the same loads or skill demands with limited variable conditions and factors such as during race events (Faulkner et al., 2008; Bonacci et al., 2013).

The significant equivalence testing in this study further suggests that the imputed data are not different from true sRPE scores. This is true for all imputation methods and could suggest that any of the models tested are comparable for sRPE data imputation. However, this result needs to be considered alongside model accuracy (ANOVA) which suggests that there are significant differences between the models. These results demonstrated that daily team mean substitution and random forest were significantly different from all other imputation models with daily team mean substitution having the lowest accuracy and random forest having the highest. Further analysis from the supervised regression model demonstrated that model type and missingness did have a significant influence on the slope of the relationship between RPE and total distance with the daily team mean substitution and neural network model types being significantly different than all other imputation techniques, including the complete, non-imputed dataset, model. The supervised assessment demonstrates the relevance of model selection and level of data missingness on the relationship between valuable training load metrics and further highlights that daily team mean substitution and neural network are poor performing models. These results are contrary to current recommendations for sRPE imputation suggesting that daily team mean substitution or neural network models are viable imputation techniques (Carey et al., 2016; Benson et al., 2021; Griffin et al., 2021). Taken together, the results of the ANOVAs, as well as model performance data (Table 2.1, Figure 2.3, Figure 2.4) suggest that daily team mean substitution is the least robust imputation method and random forest, the most robust of the methods evaluated. In cases of relatively low to moderate missingness, support vector, linear, or stepwise regression techniques may also be applicable. In alignment with the results of this study, Waljee et al. (2013) identified the use of random forest classification as an imputation strategy with a high accuracy in medical data missing completely at random (MCAR). Hong and Lynn (2020), noted that random forest imputation yields high predictive accuracy in cases of data missing at random (MAR). It is reasonable to suggest that sRPE data falls within the case of MAR whereby an athlete's ability to report their sRPE value for a match may be affected by overall fatigue, mental stress, or

physical state. Since random forest models do not require data pre-processing and can handle a wide variety of datasets without relying on distributional assumptions, these classifiers present an appealing choice for imputation (Shah et al., 2014; Kokla et al., 2019). Nevertheless, while random forest models exhibit predictive accuracy, these models cannot estimate relationships involving imputed, or estimated, values (Hong & Lynn, 2020). This is because the imputed data are not real observtions, this may introduce false patterns or bias, which increase uncertainty in the results. Random forest models cannot account for the uncertainty that comes with using imputed data. Therefore, imputed values from different models may need to be tested in a more supervised manner to existing, explainable situations (Shataee, et al., 2012; Schmitt et al., 2015; Burkart & Huber, 2021). Additionally, alternative models to those not considered in this investigation including models using fuzzy clustering or Bayesian approaches may be explored (Schmitt et al., 2015). Fuzzy clustering may more appropriately describe outcomes given the overlap in ranges of input data and particularly small number of possible outcomes, further improving accuracy (Rahman & Islam, 2016).

Support vector machine regression and linear regression were secondary and tertiary top-performing models at 20% missingness. When comparing support vector machine and random forest imputation models, Shataee et al. (2012), noted that random forest models were somewhat superior to support vector machine models as random forest models did not necessarily require the reduction of predictors which is sometimes required in the use of support vector machine regression (Shen et al., 2016). Interestingly Musil et al. (2002) found that linear regression was the most optimal approach with their dataset, supporting the results of this study (Musil et al 2002). Simple linear regression and stepwise regression outperformed other more involved models like lasso and elastic net regression strategies, perhaps due to the nature of penalized regression present in elastic net and lasso strategies relative to simple linear regression (Waldmann et al., 2013). Further, the neural network model, a technique that is generally robust at imputation due to the presence of predictive functions within the layers enabling identification of combinations of properties, was not a top-performing model (Verpoort et al., 2018). This is most likely due to the constraints on the neural network model used in the present study of a sigmoid function with one hidden layer. Given the range of possible imputation models presented in the literature, it is possible that particular methods may in fact be equivalent or interchangeable. Therefore, future research

should seek to identify optimal imputation strategies in supervised settings that promote actionable outcomes.

Despite the presence of statistical differences by model type, accuracy was very low across all models (Yin et al., 2019). One potential reason for this low accuracy may be the nature of the dataset, as there was low variance of sRPE values (mean $sRPE_{true}$ = 7 ± 1.9 au) (Figure 2.1). Another potential reason for this stems from the use of accuracy as the means to evaluate model performance. Accuracy assesses agreement between true and imputed scores such that if a true score is 7 and an imputed score is 7 that represents an accurate imputation however if the imputed score is 8 the model performed inaccurately. This is a harsh threshold, which may in fact punish models that are predicting scores within 1 sRPE value thereby predisposing accuracy scores to be low through the inherent limitation of binary classification. Alternatives to accuracy may include graphical-based metrics like Receiver Operating Curves or Precision-Recall Curves (Rahman & Islam, 2016; Burkart & Huber, 2021). To counter this limitation, R^2 and RMSE scores are reported (Table 2.1) and a supervised regression was also completed. It remains important to recognize that sRPE has been found to potentially have scalar properties, in competition environments sRPE increases across competitive matches with maximal sRPEs most often reported in and around finals or standalone events (Faulkner et al., 2008; Bonacci et al., 2013). Therefore future analysis including a broader range of sRPE values, such as using the sRPE values generated across a season of diverse training periods, may enable improved accuracy of sRPE imputation. Including additional associated variables with known relationships to sRPE, sport- or individual-specific, may improve training load accuracy (Mujika 2013; Mujika 2017). To that end, the identification of potentially overlapping factors would further enable the development of optimal strategies for working with missing athlete data (Mujika 2013; Mujika 2017).

Efficiency of analysis, for faster dissemination of knowledge, has been identified as a key consideration for sport practitioners working in high-performance environments (Coutts, 2016; Bartlett & Drust, 2021; Brocherie & Beard, 2021). This means that while the results of this study suggest that several imputation models may be interchanged for the current standard of group mean substitution (daily team mean substitution) and still produce results that are not statistically different and statistically equivalent to true sRPE scores, some models may be

more applicable than others in the applied sport environment. This study highlights that random forest classification outperforms the existing group mean substitution standard as well as more complex machine learning models such as neural networks in cases of low to moderate missing data. This study also offers practitioners the possibility of leveraging methods like random forest, support vector machine, or even simple linear regression or stepwise regression to complete datasets allowing for further evaluation of training load monitoring. Practically, random forest classification, or even simple linear or stepwise regression offer reasonable options for prediction of missing sRPE values in comparison to group mean substitution approaches, such as daily team mean substitution. Low imputation method accuracy across all methods means that any attempt to predict missing data requires care, including scrutiny of methods and data used to develop the models. Finally, practitioners are advised to prioritize data collection from athletes directly above applying any imputation methods, daily team mean substitution, linear regression, or otherwise to predict missing data.

3 The influence of physical contact on athlete load in international female rugby sevens

3.1 Abstract

While self-reported rating of perceived exertion (sRPE) is a simple and popular metric for monitoring player loads, this holistic measure may not adequately represent the distinct contributing factors to athlete loading in team sports, such as contact load. The purpose of this investigation is to determine the relationship between the number of contacts experienced and playing time on sRPE in elite women's rugby sevens athletes during competition. The data collected included sRPE, playing time, and number of contacts from one team participating in 74 international women's sevens matches. The relationship was modelled using multiple linear regression. Results, including the coefficients for number of contacts and playing time, were significant (p<0.001) and $R^2_{adjusted}$ was 0.3063. Since contacts are accounted for within the measure of sRPE in the proposed model, this further supports the value of sRPE as a global measure of athlete experience. However, this study has found a different relationship between sRPE and playing time dependent on the number of contacts, such that the influence of playing time on sRPE decreases as the number of contacts increase. Ultimately, this may mean that the weighting of individual salient factors affecting player loads, such as number of contacts or playing time, depend on the levels of all known and potentially unknown factors experienced and may limit the use of sRPE when contextualizing player load across athletes. Taken together, the findings suggest that number of contacts, playing time, and sRPE should be considered when monitoring athlete loads while further substantiating the need for more, and higher resolution, measures to better quantify competition loads in contact team sports.

3.2 Introduction

Effective athlete monitoring is integral to optimizing training improvements and minimizing injury risk in support of exceptional sport performance (Bourdon, et al. 2017; Haddad et al., 2017; Williams et al., 2017). One of the most popular tools for monitoring training and competition load in team sports is the self-reported sRPE (Bourdon et al., 2017; Haddad et al., 2017). sRPE represents a holistic measure of an athlete's workload and may include contributions from stressors

internal (e.g. perception of cardiovascular effort) and external (e.g. perception of physical contact) to an athlete's psychological state (Morgan, 1994). sRPE has further been combined with training time to produce a sRPE-TL or sRPE-CL measure, which has been validated for use as a load monitoring metric across team and contact sports (Gabbett, 2004; Gabbett et al., 2010; Clarke et al., 2013; Lovell et al., 2013; Mujika, 2017; Fuller & Taylor, 2018; Barr et al., 2019). While the use of a simple measure like sRPE is appealing, Mujika (2017) argues that accurate athlete load monitoring requires effective quantification of each distinct contributing factor, like duration, speed, or power. The precise and reliable quantification of these independent load metrics, which may account for the multi-faceted nature of team sports, should improve the accuracy of athlete load monitoring, inclusive or exclusive of subjective measures like sRPE (Casamichana et al., 2013; Mujika 2013; Bourdon et al., 2017; Barr et al., 2019).

Rugby sevens (sevens) is a variation of the traditional fifteen-a-side rugby union whereby seven players per team play on a full-size regulation rugby pitch, for two short, seven minute, halves (Worldrugby.org, n.d.; Goodale et al., 2017). Sevens is based on similar rules as tradition rugby union matches, however, since the matches are short, lasting only about 14 minutes in length, multiple matches are played in tournaments (Worldrugby.org, n.d.; Goodale et al., 2017). At international tournaments, teams generally play four to six matches over two or three days, with a maximum of three matches daily (Goodale et al., 2017). With fewer players on the field of play, the players must all demonstrate proficiency in similar match demands and technical rugby skill (Goodale et al., 2017). Beyond the technical skills, sevens players require strength, speed, power, anaerobic and aerobic training stimulus to meet the demands of competitive play which include running, sprinting, repeated physical contact, and repeated bouts of high intensity efforts (Goodale et al., 2017). Similar to the presence of high speed running as a metric for monitoring athlete load and injury risk, the number and magnitude of contacts present in rugby are a relevant objective training load metric (Gaudino et al., 2015; Rossi et al., 2018). Presently, there have been very few studies linking contacts and sRPE. Investigation into number of contacts in male rugby league by McLellan et al. (2011) found that repeated contacts sustained in a match led to lasting skeletal muscle damage up to 72 hours after a match. Further, in men's rugby union it has been shown that physical contacts, such as carries or tackles, are linked to increases in mean sessional heart rate, neuromuscular fatigue (upper and lower body), creatine kinase levels as well as poorer perceived well-being and higher sRPE-TL or sRPE-CL (Fuller et al., 2008; Roe et al., 2017). In men's rugby

union and league matches lasting 80 minutes, the number of contacts for teams, per match, are between 293-1274 tackles, roughly 19.5-84.9 contacts per player per match (Quarrie & Hopkins, 2008; Cunniffe et al., 2009; McLellan et al., 2011).

Given the high rate of contact in all forms of rugby, and the associated physiological effects from competition play, it is important to consider contact metrics in any competition load evaluation (McLellan et al., 2011; Cunniffe et al., 2009; Quarrie & Hopkins, 2008). To that end, GPS units equipped with inertial measurement units (IMUs) or video analysis tools have been used to quantify contacts and assess the kinetic and kinematic loads athletes experience (Clarke et al., 2015; Hulin et al., 2017; Barr et al., 2019). Kinematic and kinetic information from contacts has been described using video or GPS data for several areas of men's rugby including rugby union, rugby league, as well as men's sevens however, limited information is available for women's rugby (Cummins et al., 2013a; Clarke et al., 2015; Gabbett, 2015; Clarke et al., 2017a; Clarke et al., 2017b). It is especially important to determine the contribution of contact to training and competition loads in women's sevens since 89.3% of match injuries are contact-related and an international team of 12 athletes experiences about 44 contacts per match (Foster et al., 2001; Clarke et al., 2015; Fuller & Taylor, 2018). Considering the known fatiguing effects from contact in male rugby players in combination with the tremendous potential for injury from contact that women seven's athletes experience, it is critical to understand how contact is being accounted for and influences a holistic competition load measure like sRPE for female sevens athletes (Morgan 1994; Mujika 2013; Bourdon et al., 2017; Mujika 2017; Haddad et al., 2017; Fuller & Taylor, 2018). Therefore, as a first step in improving the understanding of the role of contacts in competition load management, the purpose of this investigation is to determine the relationship between the number of contacts experienced and sRPE in elite women sevens athletes during competition.

3.3 Methods

3.3.1 Approach to the problem

A cohort observational study was carried out through the retrospective analysis of data to evaluate the relationship between athlete self-reported sRPE and physical contacts experienced in international women's sevens matches. The data collected included athlete self-reported sRPE, GPS coded measure of playing time, and match footage from one team participating in 74

international women's sevens matches. As both sRPE and number of contacts may be closely linked to playing time, playing time was also included as a potential contributor to the model (Fuller et al., 2007a; Eston, 2012). As a means to model the relationship between the independent variables (number of contacts, playing time) and their interaction, and the dependent variable, sRPE, multiple linear regression was used.

3.3.2 Subjects

Data from 20 athletes, all international-level women's sevens players (25.5 ± 3.90 years old, 169.4 ± 5.89 cm tall, and 71.0 ± 5.64 kg) were collected. All athletes were in a full-time rugby sevens-specific training environment while competing. Ethics approval was obtained from the University of Victoria for the use of anonymized data from competitions that players had volunteered to take part in and complied with the principles outlined in the Declaration of Helsinki. Further, all data were provided to researchers by team staff in an anonymized form prior to analysis.

3.3.3 Procedures

Data from 74 international women's sevens matches (2017-2019) were analyzed. Rosters of 12 players, seven on field and five substitutes attended, whereby participants played in an average of 40 matches ± 23 matches, where the fewest matches were 6 and the most were all 74 matches. Match footage were evaluated using Sportscode (v11, Hudl, Lincoln: USA) to produce a summed count of contacts (tackles, tackle-assists, carries, contested restarts, and rucks). Coded, operational definitions of the forms of contacts are shown in Table 3.1. These operational definitions were developed by expert coaching and analysis staff and are aligned with the team's current analysis practices and training priorities, similar to the work of Gabbett & Kelly (2007) and others (Gabbett & Kelly, 2007; Gabbett et al., 2007; Wheeler et al., 2011; Kelly et al., 2012). Consistency in the application of coded definitions of the contacts during match footage evaluation was achieved through the use of one trained analyst. A subset of the data (six matches, 65 complete player-match rows of data were coded twice by the trained analyst on two separate occasions and a two-way mixed effects, absolute agreement, single rater intraclass correlation (ICC 3,1) found reliability to be 0.99 (95% Confidence Intervals at 0.98-0.99), indicating excellent intra-rater reliability (Koo & Li, 2016).

Table 3.1: Operational definitions used in coding contact events in rugby sevens competition.

Contact Event	Coded Definition
Tackle	Defensive event; defender connects with, holds, and may bring ball carrier to the ground.
Tackle-Assist	Defensive event; assist is awarded to secondary tackler who aids in bringing ball carrier to ground but was not the first to contact the ball carrier.
Carry	Offensive event; player holding the ball moves into defender with the goal of advancing the ball by breaking the tackle or maintaining possession during the tackle.
Contested Restart	To start play, or following a try, offensive team kicks from centre to begin play. The contest occurs when players from either team may bid, for the ball once it is kicked, this usually involves aerial contact between players or contact upon landing with the ball.
Ruck	Occurring between offensive players looking to retain possession and defending player, specifically during the time between a tackle going to ground and the next ball play. The ruck may include contact between defenders looking to cause a turnover and offensive players working to secure possession on the ground.

Athlete-worn GPS monitors collected coded playing time from matches, expressed in minutes (Apex v2.50, StatSports, Newry: UK). sRPE, was collected using the Borg CR-10, 0-10, scale with verbal anchors and pictural cues, used regularly by participants in their training and competition environments (Muyor, 2013). RPE data were provided by athletes following the completion of the match, cool-down, and debrief and re-hydration/snack, about 30 minutes after the completion of the match (Foster et al., 2001; Comyns et al., 2013). Athletes provided one rating for the whole match as the sRPE value. With all data incorporated from matches and players, a total of 1019 complete rows of data were available for analysis, with no missing data used.

3.3.4 Statistical analysis

A multiple regression model was used to determine the influence of number of contacts and playing time on sRPE, as well as any interaction effects between number of contacts and playing time (R version 3.4.4, Vienna, Austria). Data was tested for sphericity and normality prior to modelling (Appendix 3).

3.4 Results

Athletes experienced, on average, 4.7 ± 2.9 contacts, played 11.11 ± 4.66 minutes, and reported sRPE values of 6.8 ± 1.9 arbitrary units (au) per match. To mitigate outlying cases of low playing time and high contacts and ease interpretation, number of contacts experienced were grouped by count; 1-3, 4-6, 7-9, 10-12, and 13-15 (Figure 1). The minimum playing time of all subjects in each group was set as the minimum for the modelled result with minimum and maximum playing times for all subjects in each group as follows: 1-3 contacts: 1.22-18.18 mins, 4-6: 1.82-19.65 mins, 7-9: 5.33-19.83 mins, 10-12: 9.63-20.17 mins, 13-15: 14.05-19.23 mins. The results from the multiple regression model showed the model was significant ($F(3, 1015) = 150.9$ $p<0.00001$). All model coefficients were significant with the interaction effect of number of contacts and playing time ($\beta_3 = -0.026$, $p<0.001$, 95%CI [-0.03, -0.02]), number of contacts ($\beta_2 = 0.459$, $p<0.001$, 95%CI [0.35, 0.57]), and playing time ($\beta_1 = 0.260$, $p<0.001$, 95%CI [0.22, 0.30]) with an intercept of 3.28 (95%CI [2.86, 3.71]). $R^2_{adjusted}$ is 0.3063, suggesting that a portion of approximately 30% of the variance of sRPE is accounted for by the variables in the model. The model output is defined as sRPE = 0.260*Playing Time + 0.459*Number of Contacts – 0.026*(Playing Time X Number of Contacts Interaction) + 3.28 (Figure 3.1). Investigation of the residual plot showed a random scatter of points, indicating constant variance, and the normality plot showed the residuals fall on a straight-line indicating the normality assumption was appropriate for the multiple regression model sRPE data (Appendix 3).

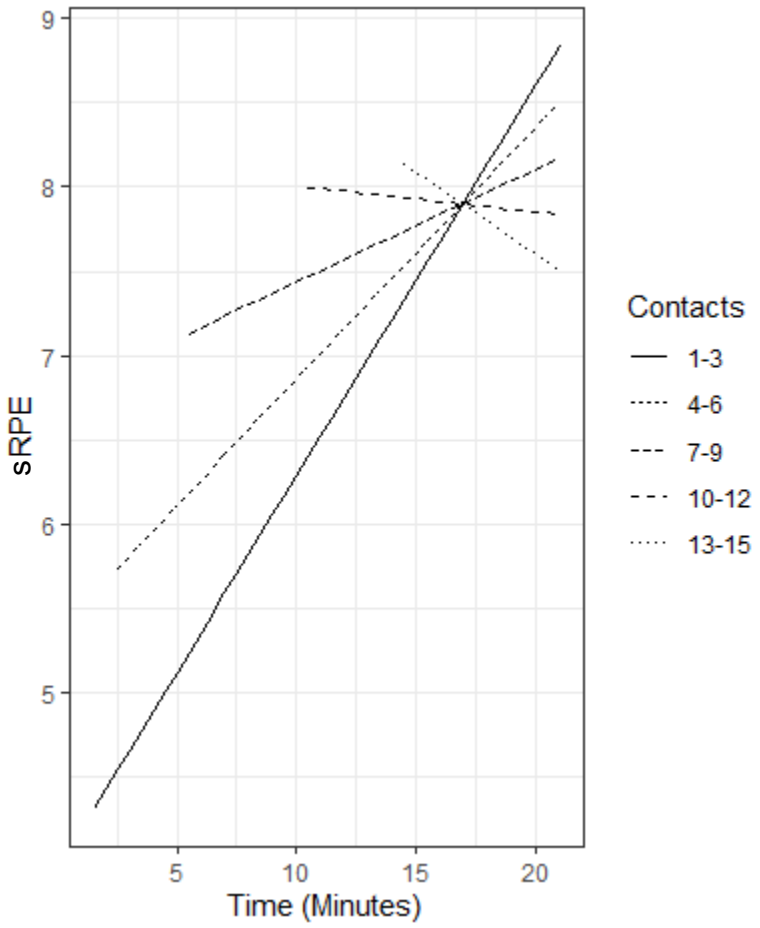

Figure 3.1: Athlete sRPE as compared to playing time, in minutes, for contact groupings by number of contacts.

3.5 Discussion

This study is one of the first exploring the relationship between number of contacts and sRPE in international level women's rugby sevens. Overall, the results suggest that number of contacts, playing time, and their interaction significantly account for 30.63% of the variability in sRPE. This finding supports that sRPE does account for number of contacts and affirms the role of number of contacts in an athlete's perceived competition load. Additionally, the modelled results found that playing time significantly contributes to sRPE. Further, this study has found that there is a different relationship between sRPE and playing time dependent on the number of contacts, such that the influence of playing time, on an athlete's sRPE, decreases as the number of contacts increase. This suggests that the weighting of individual salient competition load factors on an athlete's perceived workload may depend on the levels of all factors experienced by the athlete, thereby, this may limit the use of sRPE when comparing different training or competition sessions. Coyne et al.

(2018) also recognize the unique variations of decay of performance and fatigue in athletes whereby sports may require custom training load monitoring strategies and models. Taken together, the findings highlight the importance of number of contacts and playing time as salient factors which influence athlete competition load. These factors, as well as others like high speed running, may need to be collected to more accurately quantify athlete loads.

The significant relationship between number of contacts and sRPE in women rugby athletes found in the present study suggests that contacts are associated with an athlete's perceived workload and supports the use of sRPE as a measure of competition load in contact sports. However, an athlete's perceived training or competition load is complex and the subjective, holistic measure of sRPE should be considered alongside other quantifiable, objective metrics such as number of contacts, playing time, or distance measures like total distance and distance covered at high speed (Mujika, 2013). In the present study, number of contacts and playing time were only able to account for 30.63% of the variability in sRPE. The low coefficient of determination infers that global workload, as denoted by sRPE, is not solely driven by the experiences of contact and playing time. sRPE may be influenced by internal reference points, such as known endpoints to exercise, or that sRPE is sensitive to sensations like "feelings from working muscles" (Pandolf, 1982), heart rate, or ventilation rate (Pandolf, 1982; Morgan, 1994; Baden et al., 2005; Eston 2012). This further supports precisely and reliably quantifying all salient factors which contribute to measures of athlete load in order to evaluate effects on performance (Mujika, 2013; Mujika, 2017).

The interesting result of the significance of playing time in the model brings into question the contribution of playing time to sRPE-TL. The use of sRPE-TL is an appealing metric because it is the product of two easily obtained measures; sRPE (a measure of perceived intensity) and activity duration (time as a surrogate of volume) (Fuller et al., 2007b). While these two component measures are theoretically valid representations of training it is important to evaluate the training or competition load component measures and ensure that the product or summation adequately reflects the athlete's physical experience as well as accounting for unique and non-overlapping aspects of training. In the present study we found a small, yet significant relationship between playing time and sRPE, suggesting that playing time may already by accounted for in sRPE. Therefore, it may be possible that sRPE-CL accounts for playing time as the product of the contributing metrics: sRPE, and activity duration. As the model has shown, playing time is already

accounted for in sRPE, therefore the incorporation of activity duration as an "external" source into training load measures that already use sRPE may require further consideration. In parallel to the assessment of the role of playing time in sRPE and sRPE-CL, this evaluation reaffirms the need to statistically assess all potential contributing factors influencing competition load such as number of contacts, playing time, or even high-speed running or skill-based efforts. This is similar to the work of Mujika (2017) who argues that evaluating the links between external and internal loads and sport performance are key to supporting meaningful training interventions and competition outcomes (Lovell et al., 2013; Hulin et al., 2014; Bartlet et al., 2017; Mujika 2017).

The present study demonstrates a significant interaction between number of contacts and playing time on sRPE suggesting that the number of contacts may have a different effect on perceived workload based on playing time (Figure 3.1). From the modeled results, there is a positive relationship between playing time and sRPE for athletes with a low number of contacts (<10) and a negative relationship between playing time and sRPE for athletes with a high number of contacts (>10). In instances of low contacts, 1-9 per match, playing time has a larger effect on sRPE, with the shorter time frame for exposure to match demands, including contact, aerobic, and anaerobic efforts, driving greater perception of exhaustion. At longer playing times (>10 min), a player may be exposed to a higher number of contacts and this higher number of contacts may moderate the playing time-sRPE relationship, potentially decreasing the contribution of playing time on sRPE. The observed interaction effect, and subsequent different relationships by number of contacts, suggests that further exploration of contributing factors to sRPE is necessary. This raises the importance of evaluation of the use of sRPE as the level of one variable (number of contacts) influences the relationship of the other variable (playing time) to sRPE. It is possible to infer that the level of different variables (e.g. distance) could also affect the relationship of other factors and therefore result in an entirely different subjective measure of perceived workload for athletes in different situations. In agreement with this assertion, it has been shown that a combination of sRPE-TL and GPS-based measures of running distance and speeds appropriately model athlete training load in soccer and Australian Rules Football, improving on sRPE-TL quantification alone (Casamichana, et al., 2013; Bartlett et al., 2017). In fact, Bartlett et al. (2017), found that machine-learning generalized estimating equations better predicted training response from a combination of sRPE and GPS data than traditional monitoring measures, like high-speed running distance, alone (Bartlett et al., 2017). Lovell et al. (2013) found that 62.4% of the adjusted variance in sRPE-

TL was explained by total distance, high speed running distance, number of impacts, body load, and training impulse (as determined using Banister's model), suggesting that other factors, such as the potential inclusion of high-speed running distance, may be worth exploration. Therefore, critical loading factors, such as contacts, high speed running, and playing time, need to have clearly established relationships with holistic load measures like sRPE to ensure that each factor is appropriately accounted for (Mujika, 2013; Mujika, 2017).

3.6 Practical applications

This study offers insight into considerations for training load monitoring in physical contact sports, particularly for women's sevens athletes.

First, this study has highlighted that sRPE includes contributions from number of contacts and playing time, which substantiates the value of sRPE as a subjective competition load measure. Further this presents the possibility that number of contacts and playing time may be valuable inputs into an objective competition load calculation. However, as these measures were only able to predict 30.63% of the variability in sRPE, other salient variables may be needed to create a comprehensive training or competition load measure. Using objective measures like number of contacts and playing time may be simple to administer for sport analysis practitioners in competition settings given that it is very common to use video analysis software to track these metrics. Gathering this objective information from a single source, like a coded video stream, in competition, may speed up the collection, analysis, and dissemination of results with team leaders. In a sevens environment, where several matches are played over the span of two or three days, efficiency in carrying out data-driven decisions is a critical factor for success.

Second, this analysis has shown that the influence of playing time on sRPE decreases as number of contacts increases. In this case, as an athlete sustains more contacts, the effect of playing time on sRPE lessens, inferring that the numbers of contacts occupy a larger influence on the sRPE value. This may mean that an athlete's perceived competition load, which may include measures of sRPE, may depend on multiple variables and the interactions between these variables. Therefore, practitioners are advised to take care when comparing sRPE-based metrics across different sessions due to the sensitive nature of sRPE. For example, the sRPE of an athlete who

played fewer minutes and sustained fewer contacts (lower overall load) may not be comparable to the sRPE of an athlete who played more minutes and sustained more contacts (higher overall load).

Finally, sRPE includes numerous components which should be recognized as contributors to athlete load. In women's sevens, both number of contacts and playing time influence how athletes perceive their exertion in competition. The finding of relatively low proportional variance of playing time and number of contacts as factors of sRPE infers that some factors of the athlete's experience remain unexplained. Consideration for the appropriate measurement of the many facets of a sport's demands may assist strength and conditioning coaches and sport scientists in optimizing athlete training for performance.

4 Predicting athlete workload in women's rugby sevens using GNSS sensor data, contact count and mass

4.1 Abstract

The use of session-rating of perceived exertion (sRPE-TL) as a measure of workload is a popular athlete load monitoring tool. However, the nature of sRPE-TL means the contribution of salient, sport-specific factors to athlete load in field sports is challenging to isolate and quantify. In rugby sevens, drivers of load include high-speed running and physical contact. In soccer and men's rugby union acceleration/deceleration also influence load. These metrics are evaluated using data from global navigation satellite system (GNSS) sensors worn by athletes. Research suggests that sensor data methods for identifying load in men's rugby do not accurately quantify female athlete loads. This investigation examined how mass, contact, and accelerations and decelerations at different speeds contribute to load in women's rugby sevens. The study evaluated 99 international matches, using data from 19 full-time athletes. GNSS-measures, sRPE, athlete mass, and contact count were evaluated using a linear mixed model regression. The model demonstrated significant effects for playing time, distance covered in low acceleration at low speeds, distance covered at high deceleration in moderate speeds, and select game numbers in a tournament explaining 79.3% of the global variance of sRPE-CL. The use of acceleration/deceleration and speed from GNSS sensors alongside playing time and game number presents a novel approach to quantifying load.

4.2 Introduction

4.2.1 Monitoring athlete workload in sport

Understanding the loads that athletes experience in competition may provide important benchmarks for optimizing preparedness (Hopkins, 1991; Mujika, 2017). Monitoring athlete load may be accomplished using various tools including athlete-worn sensors, video-based event tagging and ratings of subjective effort. While there is no gold-standard method applicable across sports, a common current approach is to use session-rating of perceived exertion (sRPE-TL) for both male and female team sports (Haddad et al., 2017). sRPE-TL is the product of an athlete's subjective self-reported session rating of perceived exertion (sRPE) and the objective duration of the session in minutes (Foster et al., 2001; Haddad et al., 2017). sRPE is considered a reflection of

an “internal” training load, as it reflects the athlete’s own perceived effort, rooted in a subjective interpretation of their performance relative to their maximal effort (Haddad et al., 2017). It has been found that while sRPE alone does not entirely reflect the load experienced by an athlete, multiplying this subjective measure by distance or duration, as is done when calculating sRPE-TL, has shown to provide a valuable metric for athlete load determination (Gabbett et al., 2017; McLaren et al., 2018; Impellizzeri et al., 2019). sRPE-TL has been shown to have relationships to other internal training load measures, like training impulse (TRIMP), including Banister’s, Lucia’s, Edwards’, and external load measures like total and high-speed distance (Haddad et al., 2017).

While sRPE-TL is popular across sports, due to its ease of collection and simplicity in analysis, its subjective component (sRPE) means that the influence of individual factors and their specific impact on athlete load are not clearly identified when using this metric alone (Haddad et al., 2017; Epp-Stobbe et al., 2022a). The calculation of sRPE-TL relies on self-reporting from athletes, unfortunately this subjective information is often missing at random and without a consistent determination of load it is difficult to properly monitor athlete performance (Mujika, 2013; Gabbett, 2016; Hong & Lynn, 2020). Therefore, there is a need to identify specific objective factors that contribute to athlete load and there exists great opportunity to use specialized kinematic analysis of athlete worn sensor data to provide suitable metrics. Identifying and quantifying salient, objective, sport-specific drivers of athlete load should enable a more targeted approach to load monitoring to optimize performance and minimize injury risk (Mujika, 2013; Gabbett, 2016; Mujika, 2017).

This information will also be critical for sport coaches and practitioners to apply load metrics to objectively evaluate technical and tactical interventions (Mujika, 2013; Gabbett 2016; Mujika, 2017). Fortunately, the use of global navigation satellite system (GNSS) data collected from athlete worn sensors during training and competition, in team sports, is of great value to objectively quantify an athlete’s performance.

4.2.2 Athlete workload in rugby sevens – GNSS metrics

As an international governing body, World Rugby, mandates the use of specific monitoring tools, such as GNSS units, to be worn in competition (World Rugby, 2025b). GNSS sensors provide

speed and distance data at 10-20Hz and are often used to provide average or cumulative kinematic variables, such as providing and overall distance or duration of activity for use in sRPE calculations. While these variables are important performance summaries, this simple approach to GNSS sensor data analysis does not provide adequate detail of the unique demands experienced by the athlete such as during bouts of high intensity effort. This is of particular importance to note as high speed running events and physical contact have been identified as two of the many objective drivers of load in rugby (Schuster et al., 2018; Epp-Stobbe et al., 2022b). Therefore, there have been attempts to further identify more detailed objective kinematic measures available through GNSS, such as the quantification of bouts of high-speed running, and compare these metrics to the established athlete load measure (sRPE-TL) (Furlan et al., 2015; Delaney et al., 2016; Gabbett et al., 2017; Haddad et al., 2017; Delaney et al., 2018;). From these investigations it has been shown in men's rugby research that there are associations between high-speed running, distances covered and athlete load (sRPE-TL) (Gabbett et al., 2017; Haddad et al., 2017). Delaney et al. (2018) demonstrated that accelerations and decelerations are a driver of load in men's rugby league as well as show associations to metabolic power and perceived muscle soreness. Other research in men's rugby and soccer has focused on using acceleration and deceleration zones from GNSS data as a unique approach to characterizing objective load metrics (Haddad et al., 2017; Schuster et al., 2018; Delaney et al., 2018; Epp-Stobbe et al., 2022b).

4.2.3 Athlete workload in rugby sevens – considerations for acceleration as workload

Despite showing great potential as an approach to characterizing athlete load, existing methods for evaluating the impact of acceleration/decelerations on athlete load vary substantially and often lack precision. While either representing acceleration as all positive values (Delaney et al., 2016), quotients (Delaney et al., 2018), or fixed thresholds (Furlan et al., 2015), all approaches potentially oversimplify the distinct effects or acceleration and deceleration which may limit the-unique and important impact of acceleration and deceleration on athlete load (Furlan et al., 2015; Delaney et al., 2016; Clarke et al., 2017a). Further, while this novel process for evaluating workload may be applied similarly across both male and female athletes, work by Clarke et al., (2017a) and Mara et al., (2017) suggest that the use of zone-based data must be appropriately adjusted to the team as the proprietary zones used in commercial GNSS units, or recommended in the literature are not suitable for use in female team sport populations.

Before considering the use of acceleration and deceleration derived from GNSS sensor data to characterize athlete load in team sports it is important to appreciate how they are distinctly developed. Acceleration is normally associated with volitional increase in speed to overcome an opponent or gain a positional advantage (Bureau international des poids et mesures, 2019). Deceleration may result from volitional slowing to evade contact or may be necessary during a defensive event. Deceleration may also result from contact with an opponent. Given that changes of direction are common in invasion sports like rugby, there is great potential for acceleration and deceleration events to separately and uniquely contribute to athlete load (Harper et al., 2019). Therefore, acceleration and deceleration events may need to be quantified separately as they may be related to different aspects of physical effort as well as related to different physical mechanisms (Harper et al., 2019).

4.2.4 Athlete workload in women's rugby sevens – combining multiple objective factors

While there have been important research findings in men's rugby concerning the characterization of objective load metrics through GNSS sensor data, more recent work suggests that ways of identifying load in the men's game do not appropriately quantify the experience of female players, with speculation pointing to the generally lower masses of female players as a contributing factor (Delves et al., 2021). With the growth of women's rugby internationally following the inclusion of women's sevens in the Olympic programme, it is essential that salient factors of competition load relevant for women's sevens athletes are quantified (World Rugby, 2025c). As there is substantial evidence that multiple interacting factors are at play to contribute to athlete load in team sports there is great need to develop statistical approaches that can fuse data from multiple sources and domains. Further, the specific inclusion of acceleration and deceleration zones alongside other salient characteristics, such as contacts, have not yet been investigated as an approach to quantify drivers of athlete load in women's rugby sevens**.** To support this a linear mixed effect model (LMM) presents a powerful, repeatable, and reliable statistical approach that can help to fuse GNSS sensor data with other categorical variables, such as athlete- or match-specific characteristics to improve the prediction of load as successfully demonstrated by Iannaccone et al. (2021). By including multiple factors, LMMs can better model the relationship between training load and performance than traditional linear regression models as they account for individual variability across athletes. This leads to more accurate insights into their training adaptations and injury risk. The identification of the main factors of load relevant for women's

sevens athletes, may enable a simple, targeted, and actionable approach to load monitoring and management compared to the conventional strategies of a sRPE-TL measure or the use of proprietary approaches by commercial software.

Therefore this investigation's purpose is to leverage a novel approach to categorize GNSS sensor data based on accelerations across speed zones, coupled with athlete mass, contact count, and game number using LMM to determine the contribution of these variables to athlete load in women's rugby sevens. Ultimately, the aim of the investigation is to provide sport coaches and practitioners with key objective metrics that can be developed and applied both in the monitoring of athlete performance in competition as well as in training to inform decision making related to tactical and technical parameters that may be impacted by athlete load and better prepare athletes to meet the demands of their sport.

4.3 Materials and methods

4.3.1 Study design

This cohort observational study analyzed the contributions of objective measures of physical performance from GNSS sensor kinematics, contact count and athlete mass to athlete workload, sRPE-CL, through retrospective analysis of data from international women's rugby sevens competition. The data collected included GNSS-coded measures, coded match footage, athlete self-reported sRPE, athlete mass, and match-related data for one team, participating in 99 international women's sevens matches (2017-2020). Matches were played in outdoor, uncovered facilities. The relationship between independent variables (athlete mass, movement distance, movement zone, acceleration type by movement zone, and contact count) and the dependent variable (sRPE-CL, as determined by multiplying sRPE and playing time), were assessed using a linear mixed model regression.

4.3.2 Subject information

Nineteen international women's sevens players in a full-time rugby sevens training program participated (26.5 ± 4.20 years, 169.5 ± 5.90 cm, 69.7 ± 6.24 kg). Participants volunteered for the study and gave their written informed consent to participate. Ethical approval for the study was obtained from the University's Human Research Ethics Board and followed the principles described in the Declaration of Helsinki. All data used by researchers were anonymized by team

staff prior to analysis. One team of players was included in the study due to limitations in access to additional teams. While this restricted the overall sample size, it enabled a focused analysis within the available population. Consequently, findings should be interpreted with consideration for the specific characteristics of the team studied.

4.3.3 Methodology

During all matches athletes wore GNSS monitors (Apex v2.50, StatSports, Newry: UK). The monitors were installed between the shoulder blades of the athlete, fixed under the uniform, in a custom vest (Figure 4.1).

Figure 4.1: (a) GNSS monitor used in data collection and (b) GNSS monitor in vest worn by athlete.

These sensors measured distance in meters, speed in meters per second and playing time in minutes at 10 Hz. Athlete acceleration was calculated as the first derivative of the GNSS speed measurement. The use of GNSS measures of distances and speeds have been demonstrated to be appropriately accurate at resolutions of 5-10 Hz and more recent work supports that the application of filters to the data before deriving acceleration further improves accuracy (Malone et al., 2017; Roell et al., 2018; Delves, 2023; Ellens et al., 2024,). Following each match, data from the GNSS monitors were downloaded from the unit and exported. Measures of distances travelled at certain speeds, and accelerations/decelerations were organized into twelve zones (three speed zones by four acceleration/deceleration zones) to enable the differential contribution of movement speed and acceleration/deceleration zones within the statistical analysis (Figure 4.2). Preliminary data

processing were completed using Python (Python Language Reference, version 2.7, Amsterdam, Netherlands).

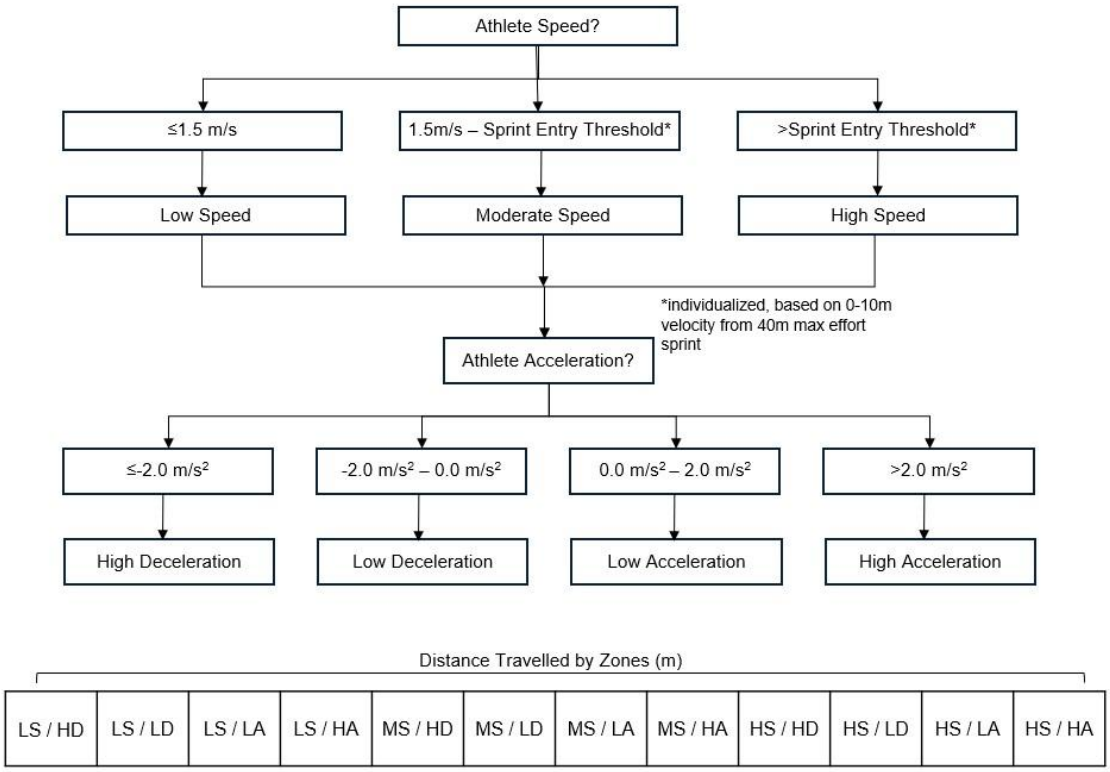

Figure 4.2: Visual depiction of organization of movement categories based on low, moderate, and high athlete speeds and low or high accelerations or decelerations, resulting in 12 zones.

The use of these zones is consistent with previous studies in team sports linking movement speed and acceleration/deceleration zones to performance demands from soccer, field hockey, and men's rugby union fifteens and sevens (Furlan et al., 2015; Sweeting et al., 2017; Ihsan et al., 2021). Distance travelled, in metres, was divided into three speed zones: low speeds (walking - 0m/s-1.5m/s), moderate speeds (running - 1.5m/s to each athlete's entry to sprinting threshold), and high speeds (sprinting - entry to sprinting and above), expressed relative to each athlete to ensure appropriate reflection of the individual physical efforts of each athlete (Cunniffe et al., 2009; Gabbett, 2015). The use of three zones represents an efficient combination of the traditional five zone division available in team sport GNSS systems, intended to offer an applied workflow where movements are easy to identify and communication of results to coaches is efficient in plain language (Lockie et al., 2016; Bartlett & Drust, 2021; Torres-Ronda et al., 2022). Further, each athlete's entry to sprinting threshold was individualized based on the athlete's velocity achieved

from 0-10m of a maximal physical effort 40m sprint performed in training and updated regularly through-out the season to ensure accuracy over periods of physical development (Gabbett, 2015; Núñez et al., 2018).

Once categorized in speed zones, an acceleration threshold of $2m/s^2$ was used to further categorize distances covered at low ($<2.0m/s^2$) and high ($>2.0m/s^2$) accelerations as well as low ($<-2.0m/s^2$) and high ($>-2.0m/s^2$) decelerations (Mara et al., 2016). While various means of quantifying acceleration have been presented in the literature, it remains paramount that the use of a threshold is appropriately reflective of the population and as such, $2m/s^2$ was used to reflect the physical effort of female team sport athletes as identified by Mara et al. (2017) (Delaney et al., 2016; Mara et al., 2017; Delaney et al., 2018; Delves et al., 2021).

The 10Hz data were smoothed using a low, dual-pass second order Butterworth filter. Outliers were identified using the Tukey rule whereby values lying beyond the upper and lower bounds of 1.5 times the interquartile range were removed from analysis.

A summed count of contacts per player was produced from match footage, evaluated by a trained analyst working with the team using Sportscode (v11, Hudl, Lincoln: USA).

Mass, in kilograms, was collected pre-match using a portable weigh scale (ES-310, Anyload, Burnaby: CAN).

One sRPE value per match was self-reported by athletes using a 0-10 scale, familiar to participants from regular use in training and competition, following each match (roughly 30 minutes after each match) (Foster et al., 2001). Athletes recorded their own values on a clipboard with a reference scale on the same page. Values above 10 and integer values were not included for analysis. One sRPE-CL value per match per athlete was calculated by multiplying sRPE by the athlete's playing time.

With all data incorporated, and cases of missingness dropped ahead of analysis, a total of 1009 complete rows of data were available for analysis.

4.3.4 Statistical analysis

A LMM was used to determine the influence of both fixed and random effects on sRPE-CL by athlete. Fixed effects included continuous kinematic variables derived from GPS data, specifically select movement categories, athlete mass, playing time, contact count, and game number were also included by athlete, where athlete represented the random effect.. This modeling approach allowed the integration of continuous and categorical data to predict sRPE-CL, accommodating both overall patterns and individual-specific responses. The LMM was fitted using R statistical software package "lme4" with *p*-values manually calculated from the *z*-distribution (version 4.2.1, Vienna, Austria).

The movement categories included in the model were determined by completing a Pearson correlation coefficient matrix of all possible variables. In cases where correlations between two variables were equal or greater than 0.90, the variable representing the lower acceleration or deceleration were removed given known effects of relatively higher accelerations and decelerations on athlete load, with decelerations given preference of accelerations given the known effects of deceleration relative to acceleration on athlete load (Table 4.1 – McBurnie et al., 2022; Hills et al., 2024).

Table 4.1: Correlation between variables where correlation is equal to or greater than 0.9, including resulting action for analysis.

Variable 1	Variable 2	Correlation	Action
Low Acceleration-Moderate Speed	Low Deceleration-Moderate Speed	0.980	Remove Low Acceleration-Moderate Speed
High Acceleration-Moderate Speed	High Deceleration-Moderate Speed	0.968	Remove High Acceleration-Moderate Speed
Low Deceleration-Moderate Speed	High Deceleration-Moderate Speed	0.960	Remove Low Deceleration-Moderate Speed
High Acceleration-High Speed	High Deceleration-High Speed	0.954	Remove High Acceleration-High Speed
High Acceleration-High Speed	Low Deceleration-High Speed	0.944	No action, High Acceleration-High Speed already removed
High Acceleration-Moderate Speed	Low Deceleration-Moderate Speed	0.941	No action, High Acceleration-Moderate Speed already removed
Low Acceleration-High Speed	Low Deceleration-High Speed	0.936	Remove Low Acceleration-High Speed
Low Deceleration-High Speed	High Deceleration-High Speed	0.926	Remove Low Deceleration-High Speed

Following the LMM, collinearity was reviewed by calculating the variance inflation factor and a leave one athlete out cross validation was performed to provide further insights into the model's RMSE and generalizability.

4.4 Results

4.4.1 General summary of data

On average, athletes experienced 4 ± 2.5 contacts, played for 11.8 ± 4.53 total minutes, experienced an sRPE of 7 ± 1.8 au, and subsequently an sRPE-CL of 99.2 ± 44.09 au. The distances covered, with standard deviations, in each of the speed zones are detailed in Table 4.2. Sample

sensor data providing insight into the speeds and accelerations of players participating in one match are detailed in Figure 4.3.

Table 5.2: Distances covered by category of movement speed and acceleration type, including standard deviations.

Movement Speed	**Acceleration Type**	**Distance Covered (m)**
Low Speed (LS)	Low Acceleration (LA)	11.2 ± 5.66
	High Acceleration (HA)	2.0 ± 1.30
	Low Deceleration (LD)	259.4 ± 130.09
	High Deceleration (HD)	2.9 ± 9.40
Moderate Speed (MS)	Low Acceleration (LA)	21.7 ± 9.62
	High Acceleration (HA)	16.8 ± 7.39
	Low Deceleration (LD)	703.6 ± 298.06
	High Deceleration (HD)	16.4 ± 7.00
High Speed (HS)	Low Acceleration (LA)	2.4 ± 1.60
	High Acceleration (HA)	2.6 ± 1.64
	Low Deceleration (LD)	82.5 ± 52.29
	High Deceleration (HD)	1.8 ± 1.31

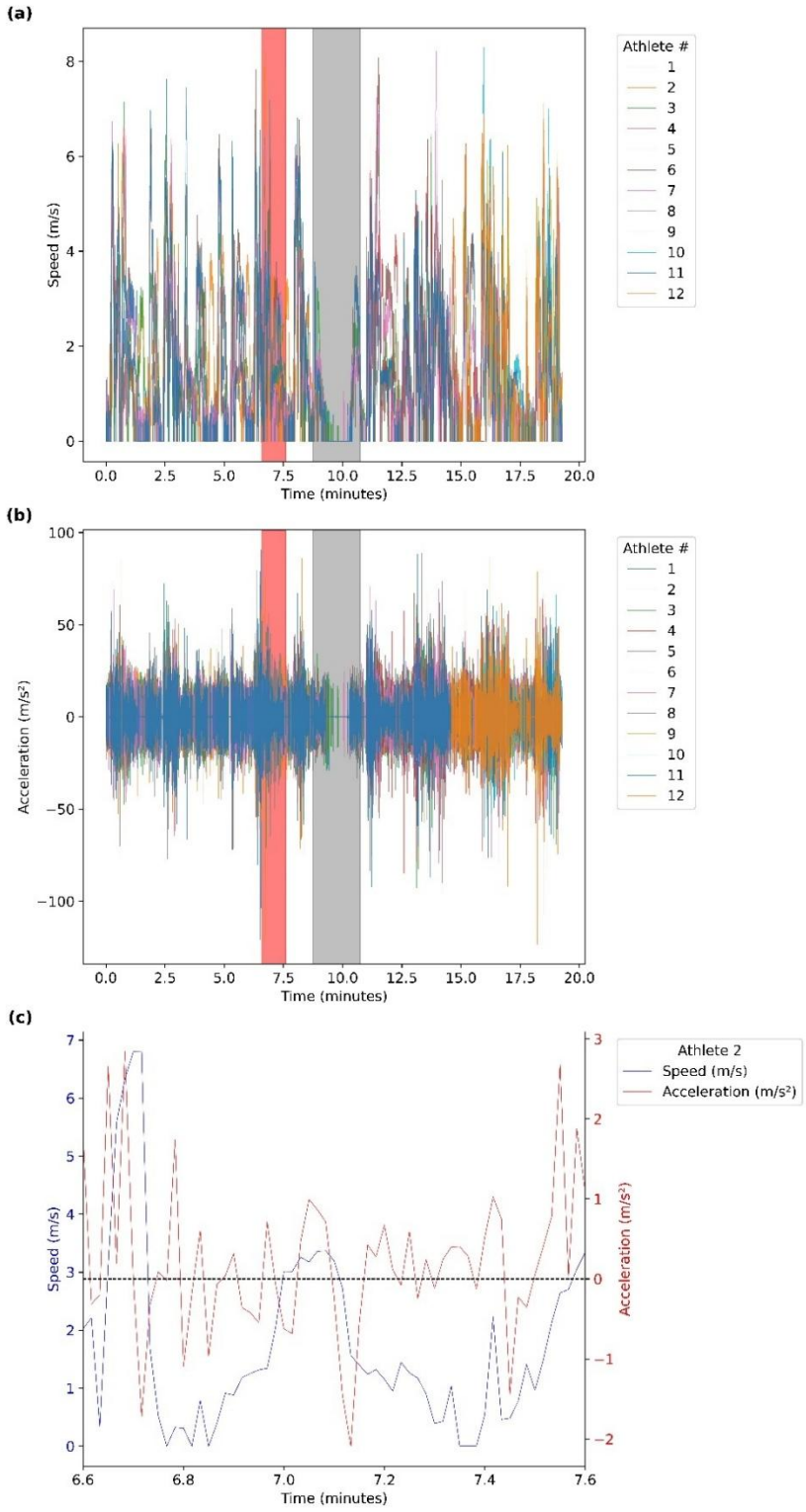

Figure 4.3: Sample sensor data from one match showing (a) all athlete speeds, (b) all athlete accelerations and decelerations, and (c) a single athlete subset for a one minute period of play. The grey shaded box in (a) and (b) denotes the two-minute halftime period of the match, where players remain on field. The red shaded box in (a) and (b) denotes the subset time period in (c).

Further information on the general dataset used for analysis in this investigation. Global summary data for variables are reported in Figures 4.4-4.13 (inclusive). A correlation matrix for variables is reported in Figure 4.14.

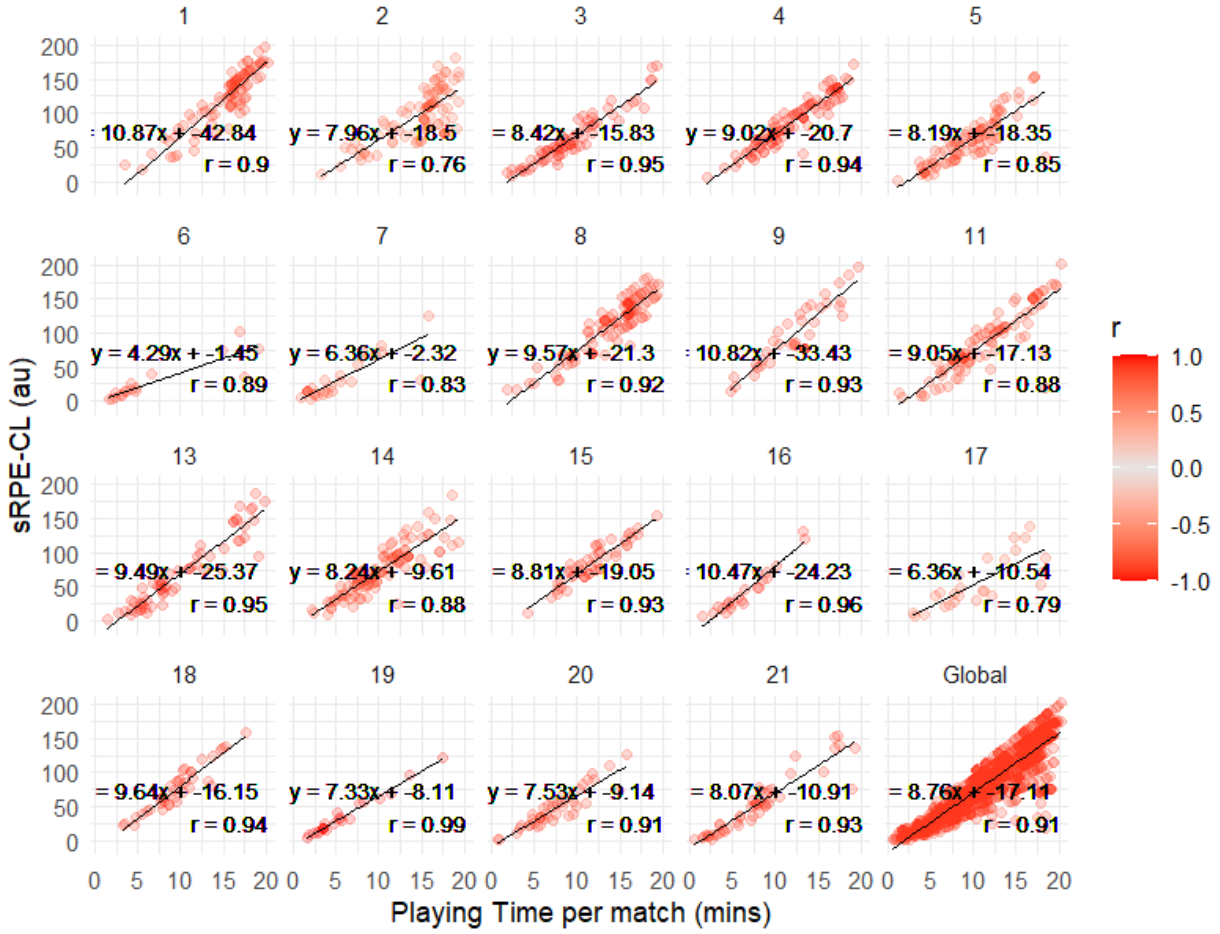

Figure 4.4: Playing time (mins) and sRPE-CL (au) per match, by athlete number with global data including regression line, equation, and r values, as colour coded by r value and opacity indicating frequency of reported value.

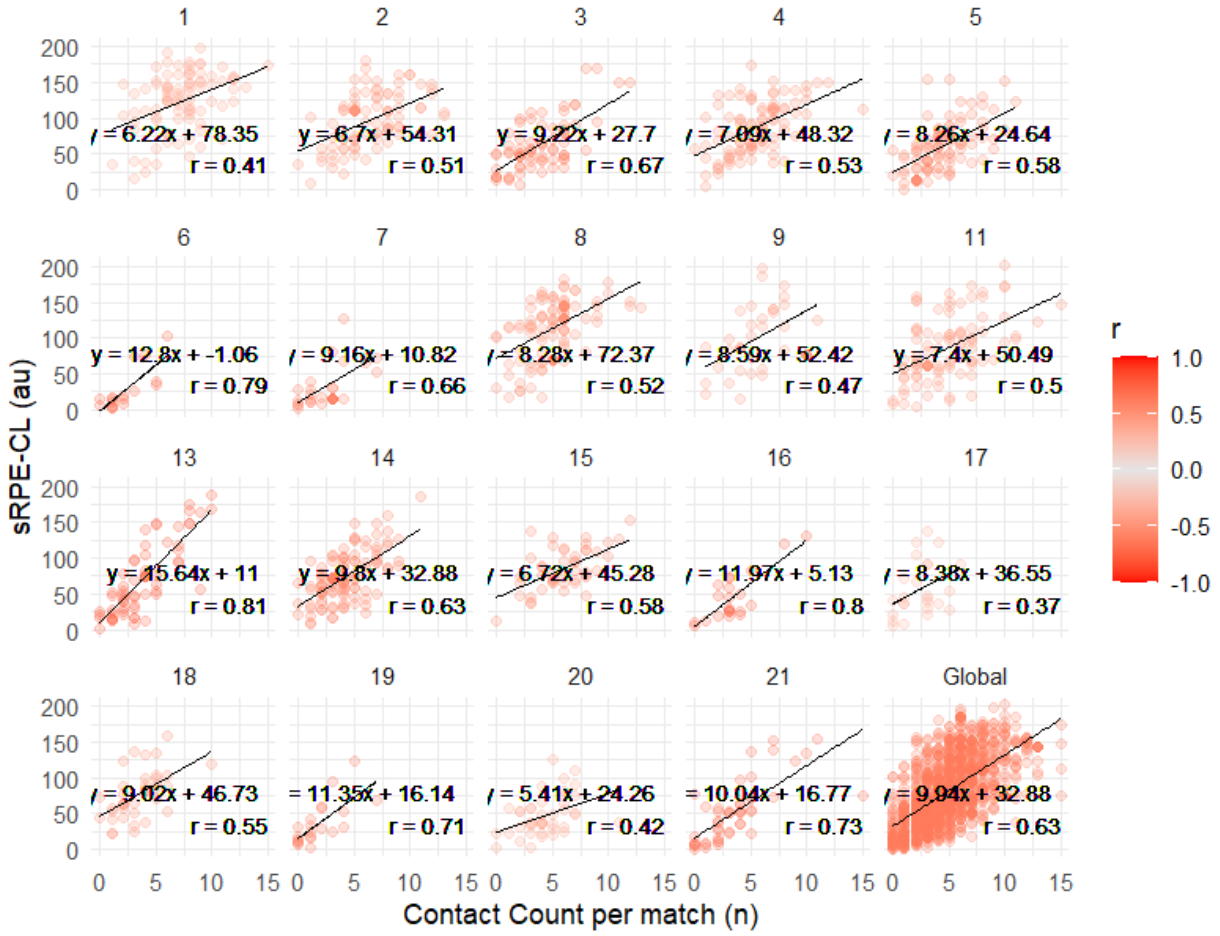

Figure 4.5: Contact count (n) and sRPE-CL (au) per match, by athlete number with global data including regression line, equation, and r values, as colour coded by r value and opacity indicating frequency of reported value.

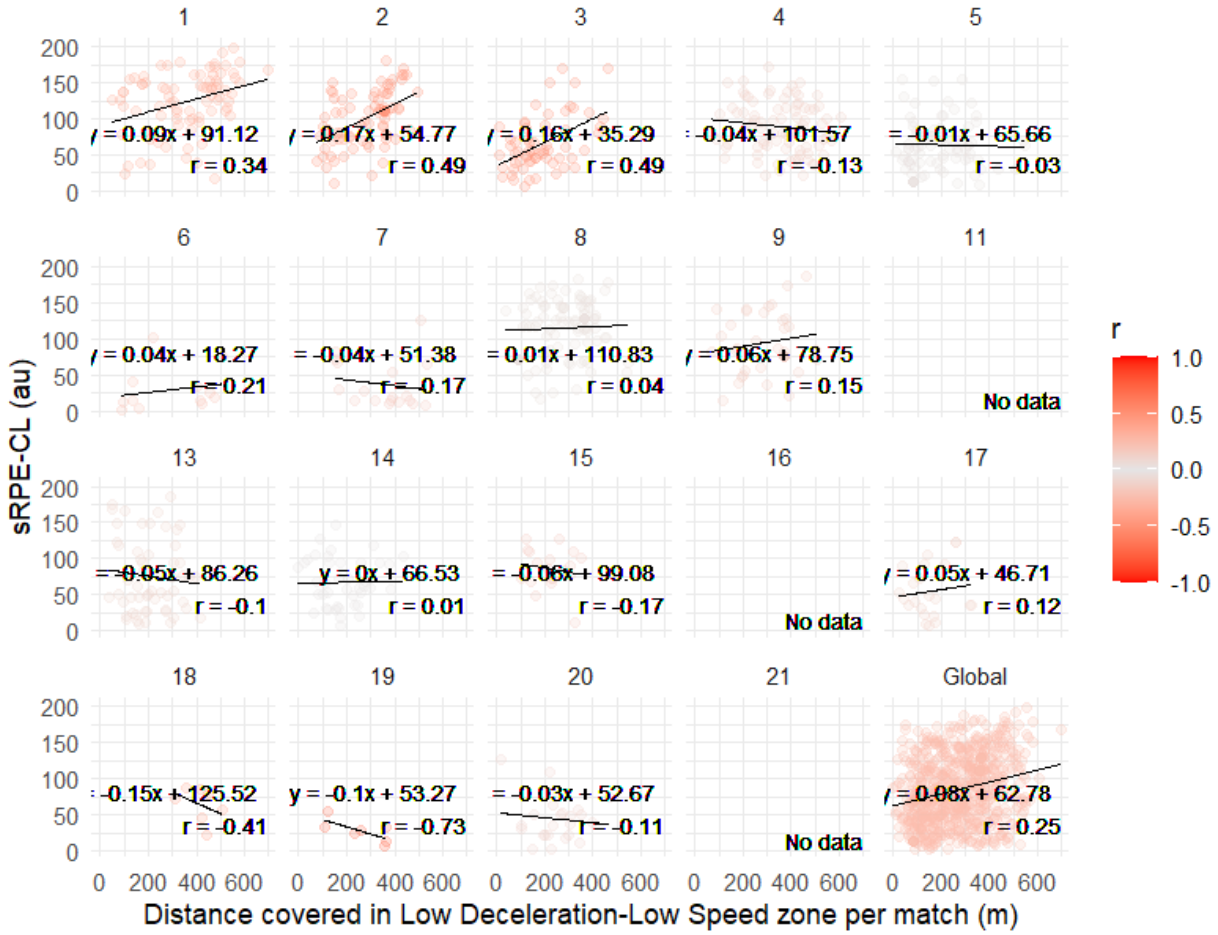

Figure 4.6: Distance covered in low deceleration-low speed zone (m) and sRPE-CL (au) per match, by athlete number with global data including regression line, equation, and r values, as colour coded by r value and opacity indicating frequency of reported value.

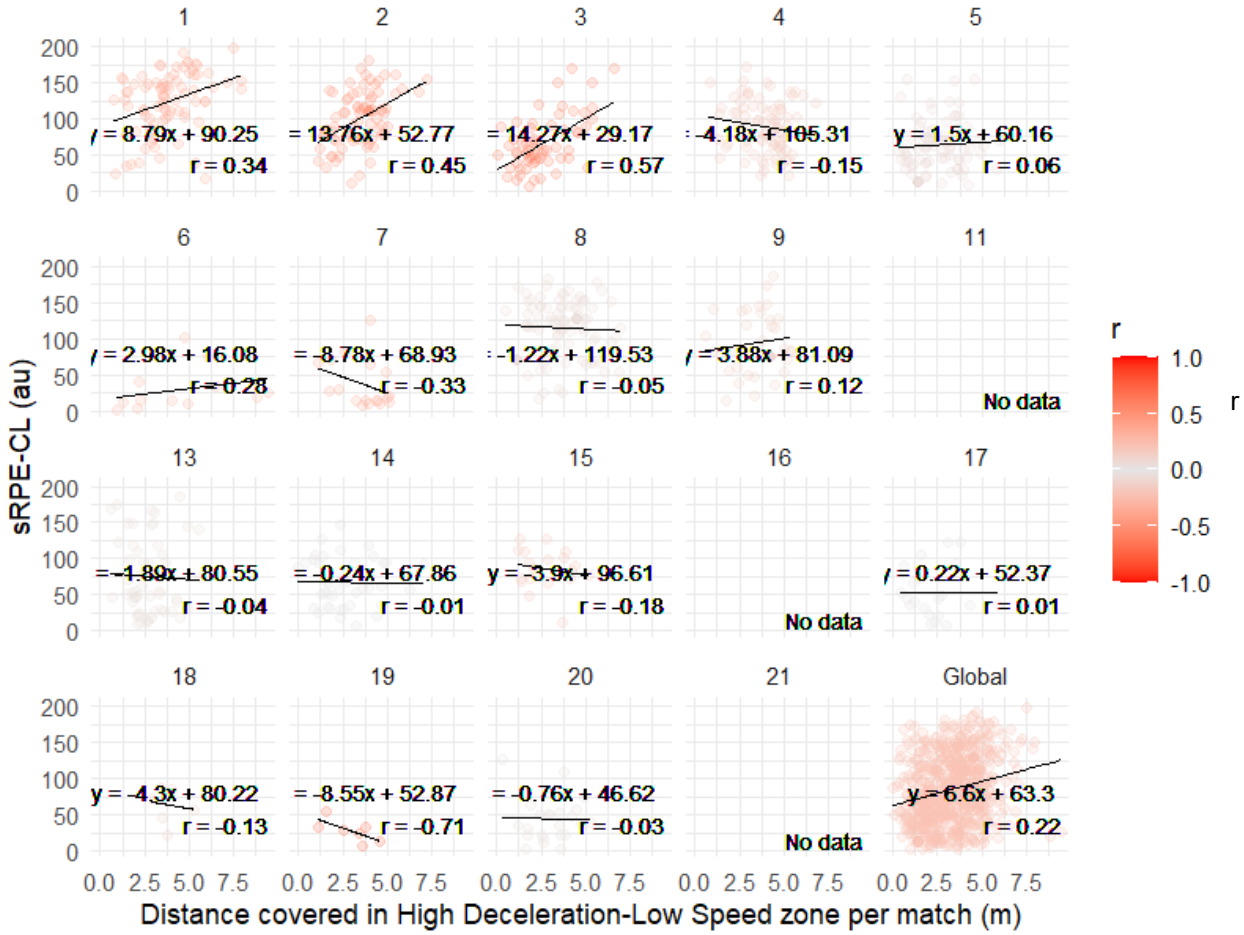

Figure 4.7: Distance covered in high deceleration-low speed zone (m) and sRPE-CL (au) per match, by athlete number with global data including regression line, equation, and r values, as colour coded by r value and opacity indicating frequency of reported value.

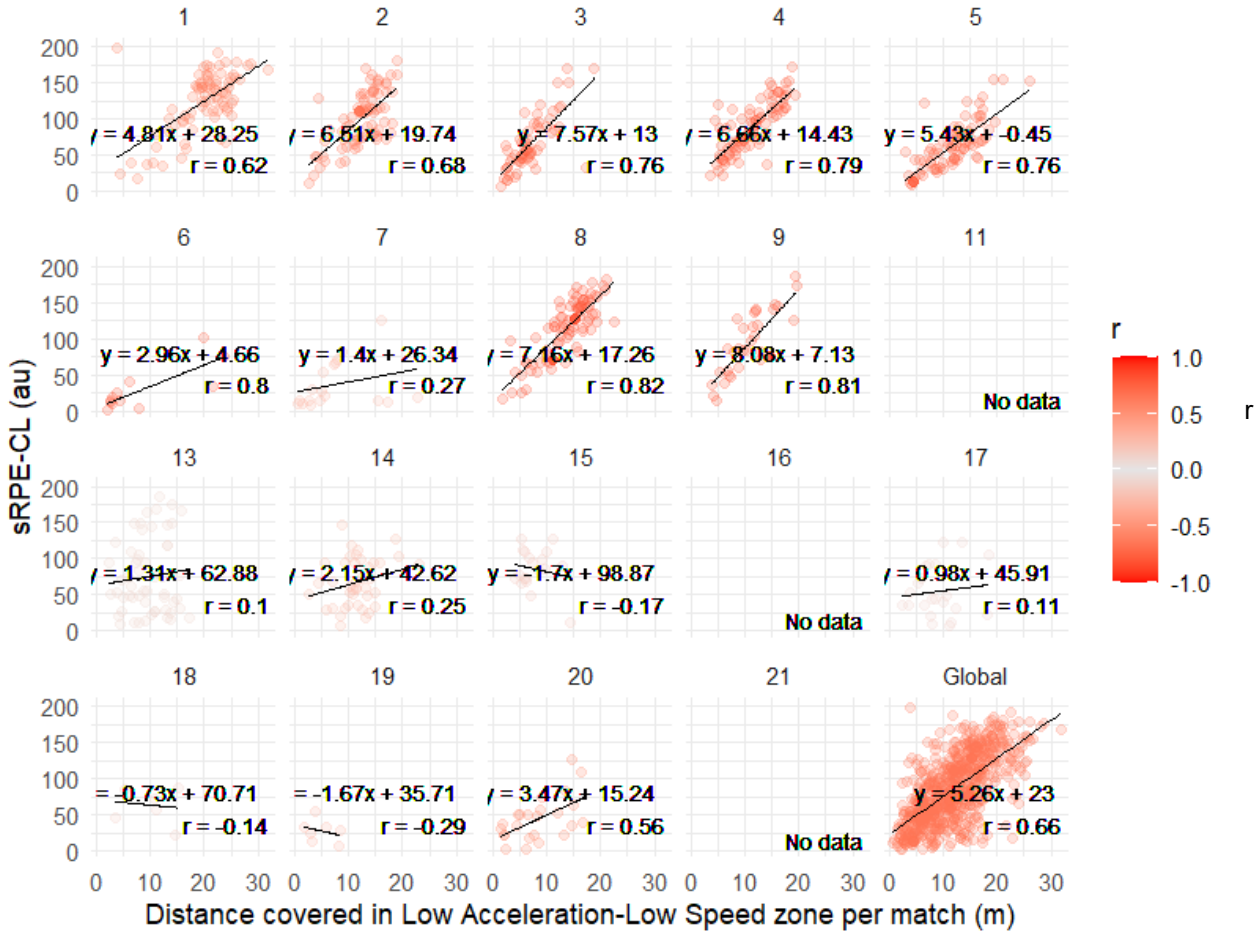

Figure 4.8: Distance covered in low acceleration-low speed zone (m) and sRPE-CL (au) per match, by athlete number with global data including regression line, equation, and r values, as colour coded by r value and opacity indicating frequency of reported value.

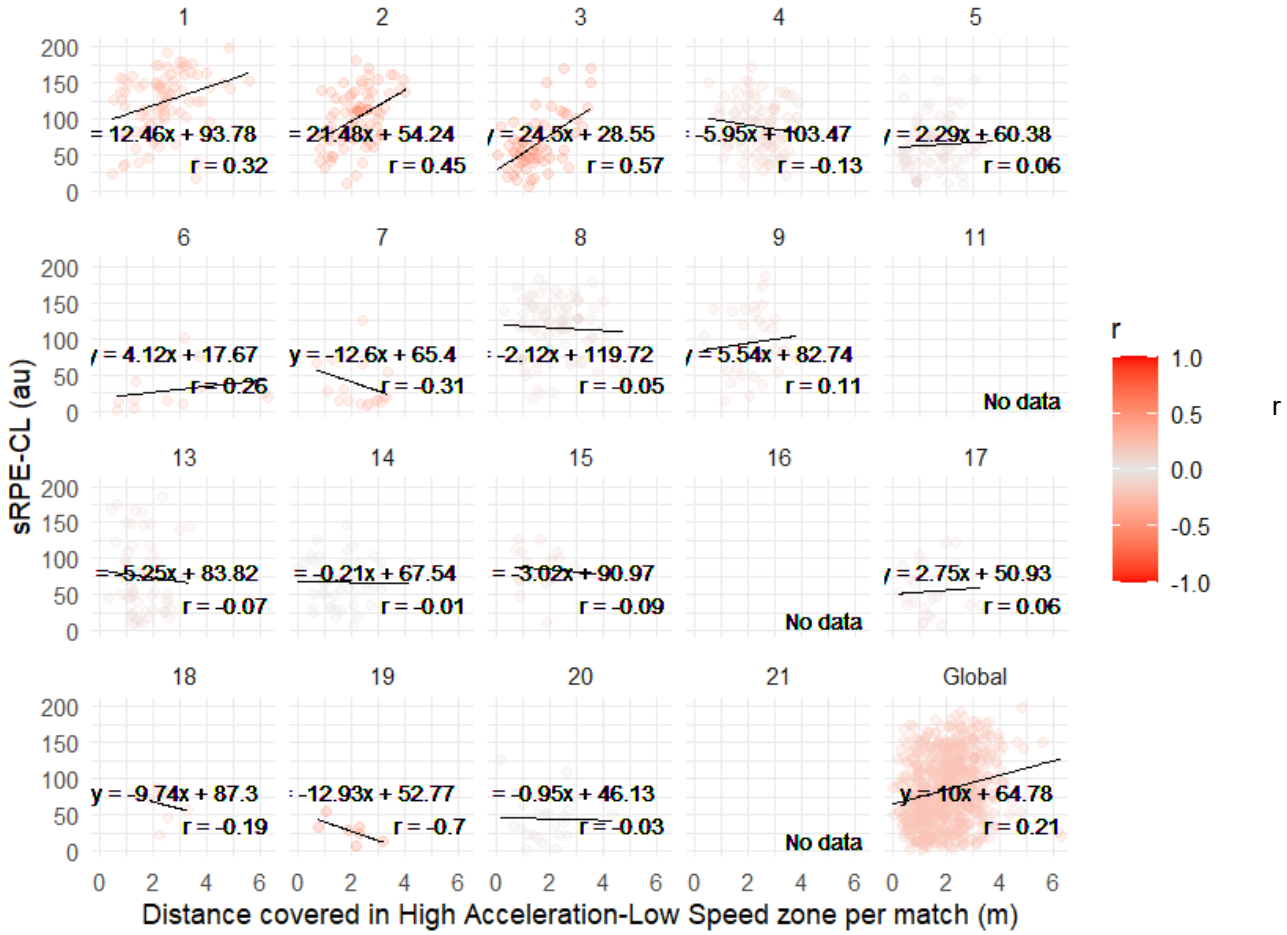

Figure 4.9: Distance covered in high acceleration-low speed zone (m) and sRPE-CL (au) per match, by athlete number with global data including regression line, equation, and r values, as colour coded by r value and opacity indicating frequency of reported value.

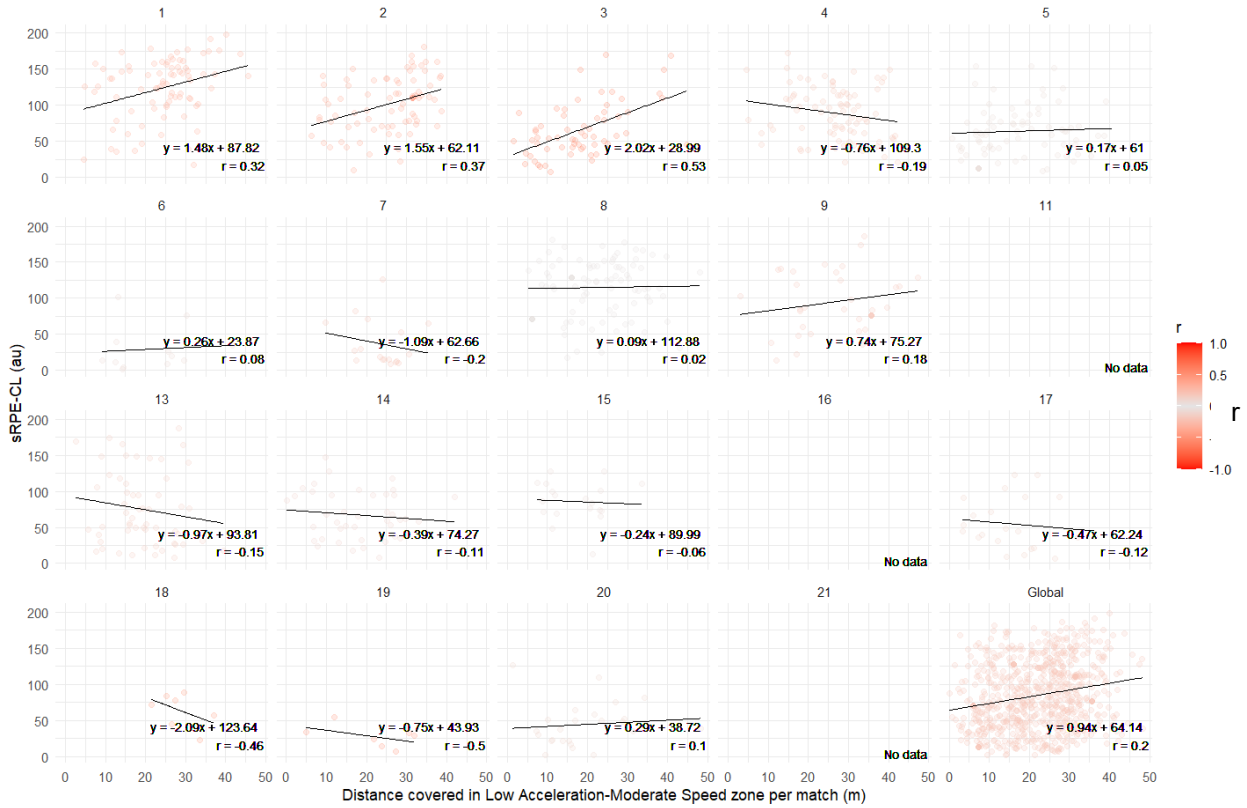

Figure 4.10: Distance covered in low acceleration-moderate speed zone (m) and sRPE-CL (au) per match, by athlete number with global data including regression line, equation, and r values, as colour coded by r value and opacity indicating frequency of reported value.

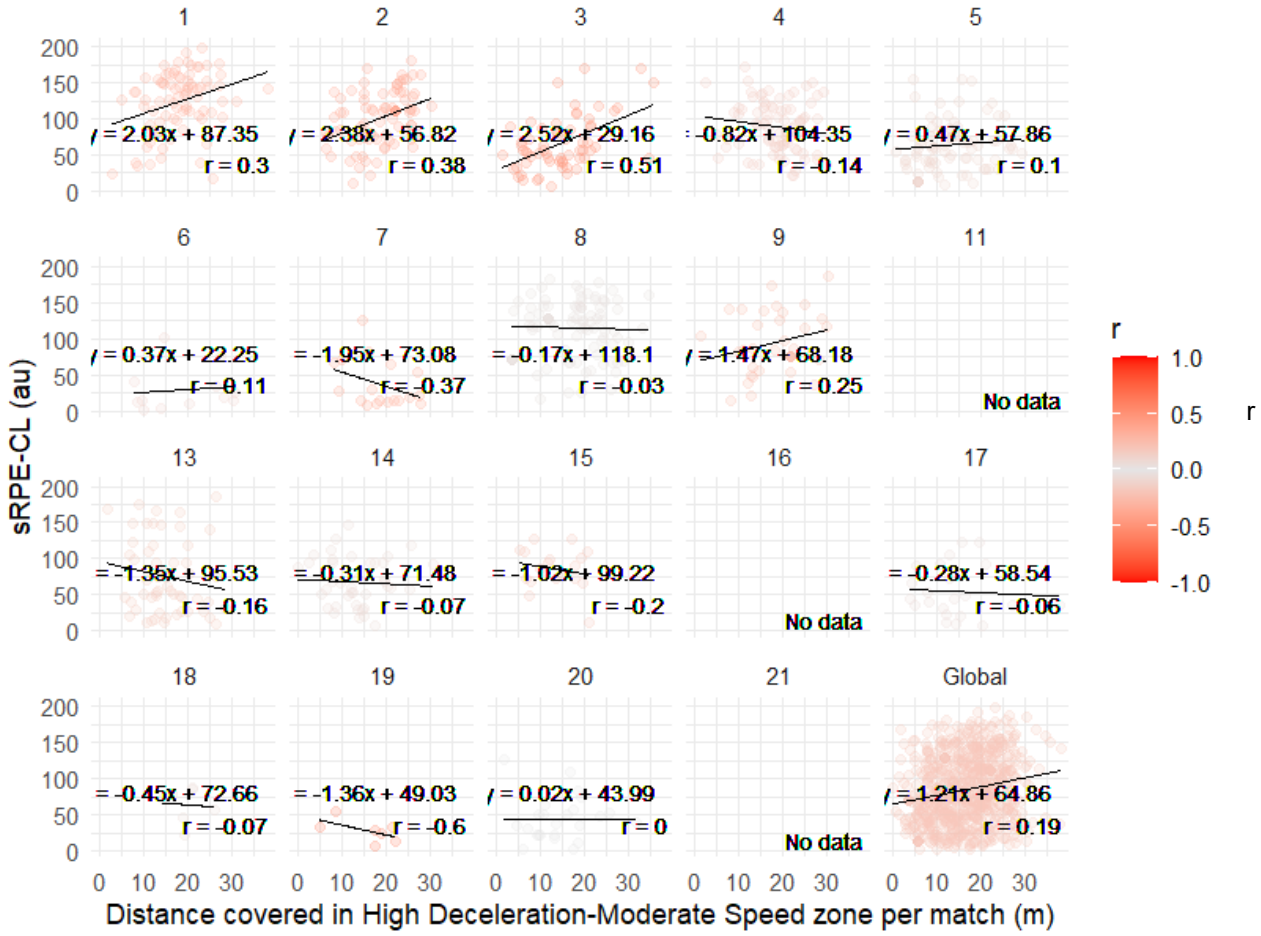

Figure 4.11: Distance covered in high deceleration-moderate speed zone (m) and sRPE-CL (au) per match, by athlete number with global data including regression line, equation, and r values, as colour coded by r value and opacity indicating frequency of reported value.

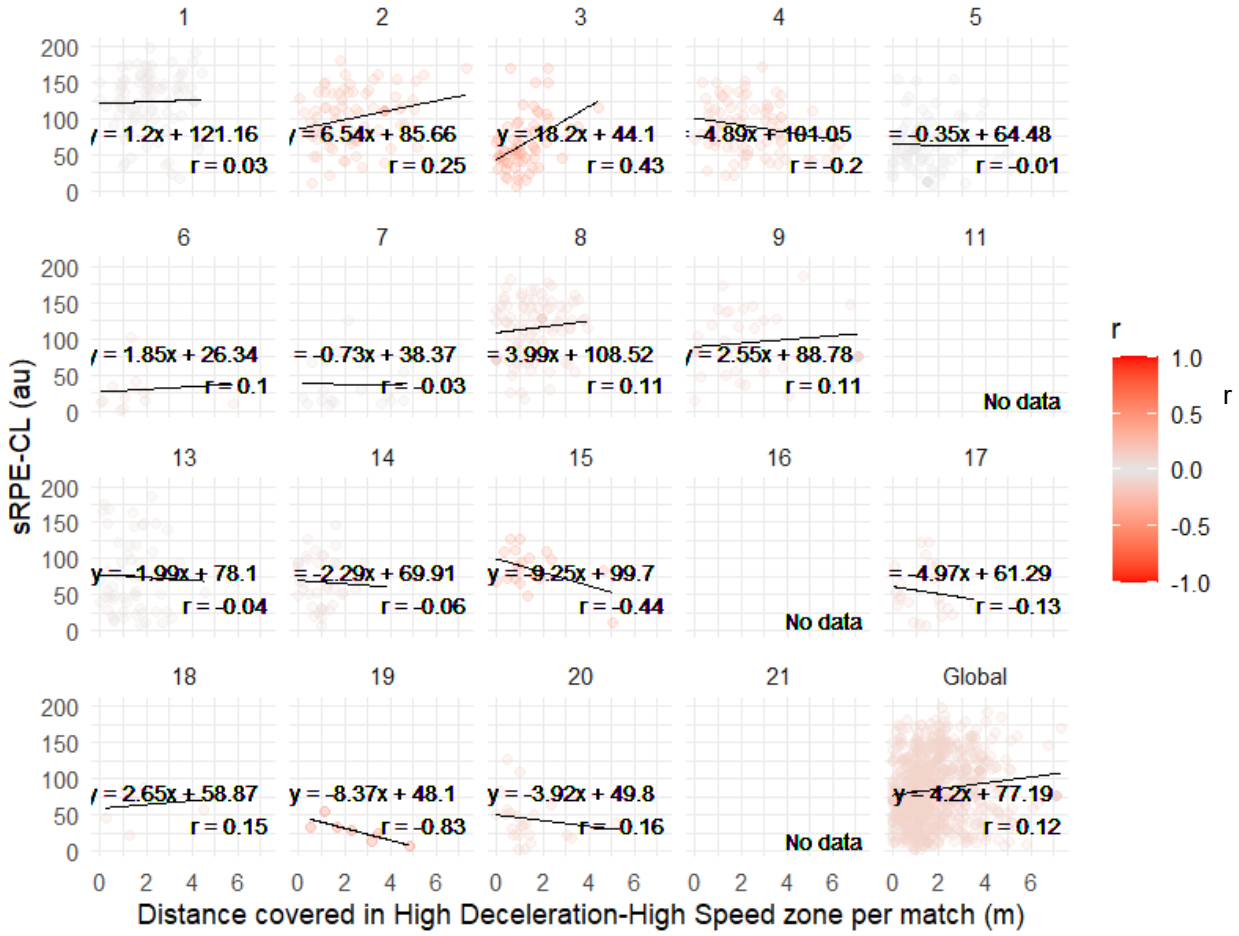

Figure 4.12: Distance covered in high deceleration-high speed zone (m) and sRPE-CL (au) per match, by athlete number with global data including regression line, equation, and r values, as colour coded by r value and opacity indicating frequency of reported value.

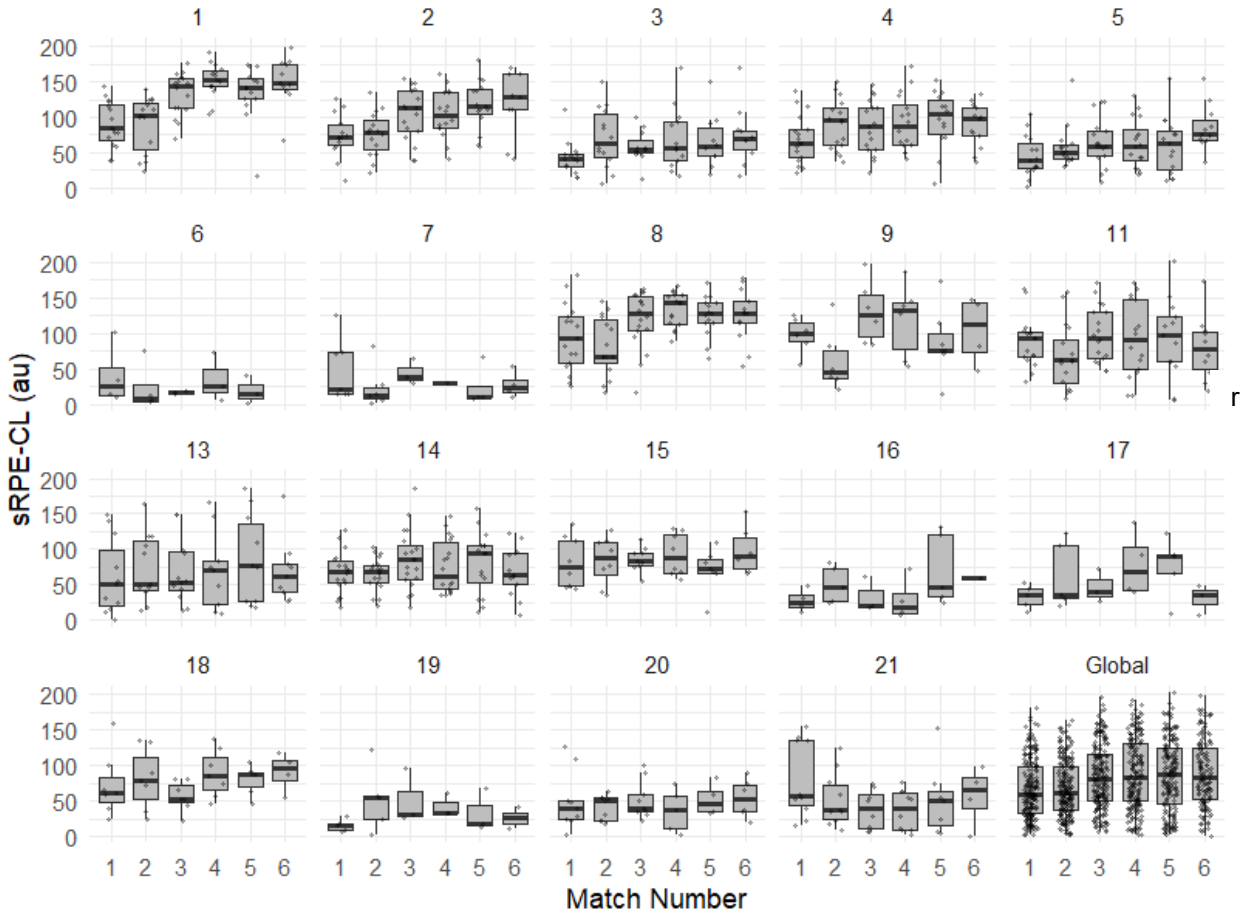

Figure 4.13: Match number and sRPE-CL (au) per match, by athlete number with global data.

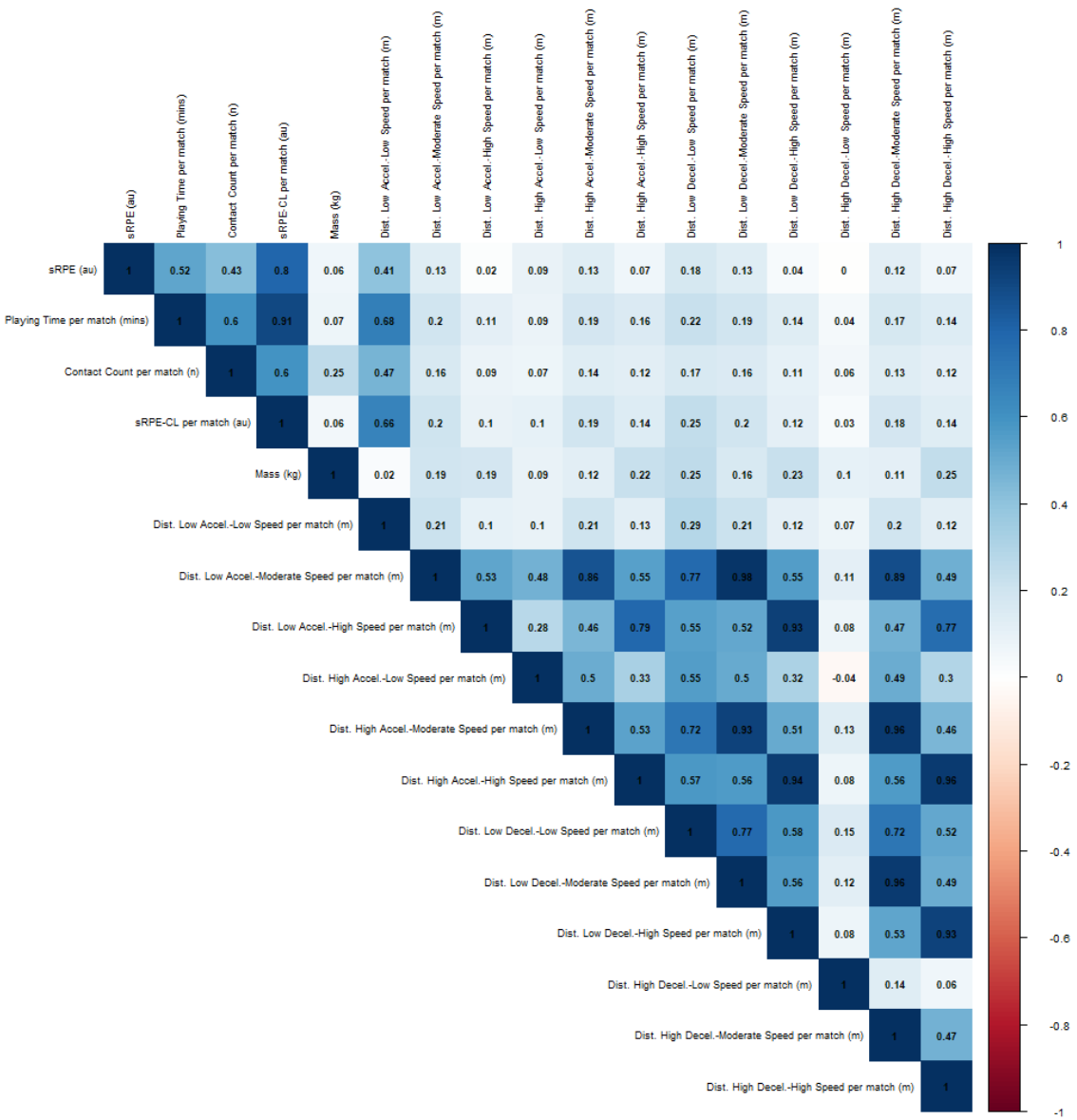

Figure 4.14: Correlation graph of key exploratory variables, colour coded by level of correlation.

4.4.2 Results of linear mixed model

The correlation matrix indicated high correlation between several variables, enabling removal of six of the twelve categories (Table 4.1; Figure 4.14). The final model included the following variables: athlete mass, contact count, game number, playing time, and distances travelled at low speed-low acceleration, low speed-low deceleration, low speed-high acceleration, low speed-high deceleration, moderate speed-low acceleration, moderate speed-high deceleration, and high speed-high deceleration.

The model results showed a significant main effect of playing time, $t(18) = 31.70$, $p<0.01$, distance travelled at low speed-low acceleration, $t(18) = 3.11$, $p<0.01$, and distance travelled at moderate speed-high deceleration $t(18) = 3.30$, $p<0.01$, as well as game 3, $t(18) = 3.33$, $p<0.01$, game 4, $t(18) = 3.37$, $p<0.01$, and game 5, $t(18) = 4.37$, $p<0.01$.There were, however, no significant main effects for mass, contact count, game numbers 1, 2, or 6, or the other movement categories present, $p>0.01$.

Overall, 79.3% of the variability of sRPE- CL was accounted for by variables in the model, as indicated by the $R^2_{adjusted} = 0.793$. The relationship between sRPE-CL and the model variables with parameter estimates by zone further illustrated in Figure 4.15.

Visual inspection of the Q-Q plot (Figure 4.16a) indicated that residuals were approximately normally distributed, though mild deviation was observed at the tails. The residuals vs. fitted values plot (Figure 4.16b) showed some heteroscedasticity, with greater variability in residuals at mid-range fitted values, suggesting potential non-constant variance. The leverage plot (Figure 4.16c) revealed no major influence concerns, with only two observations exhibiting high leverage. Variance inflation factors (Table 4.3) were within acceptable limits, with all adjusted VIFs below 2.5, indicating no serious multicollinearity concerns among predictors. Overall, model diagnostics support the adequacy of the linear mixed model for inference.

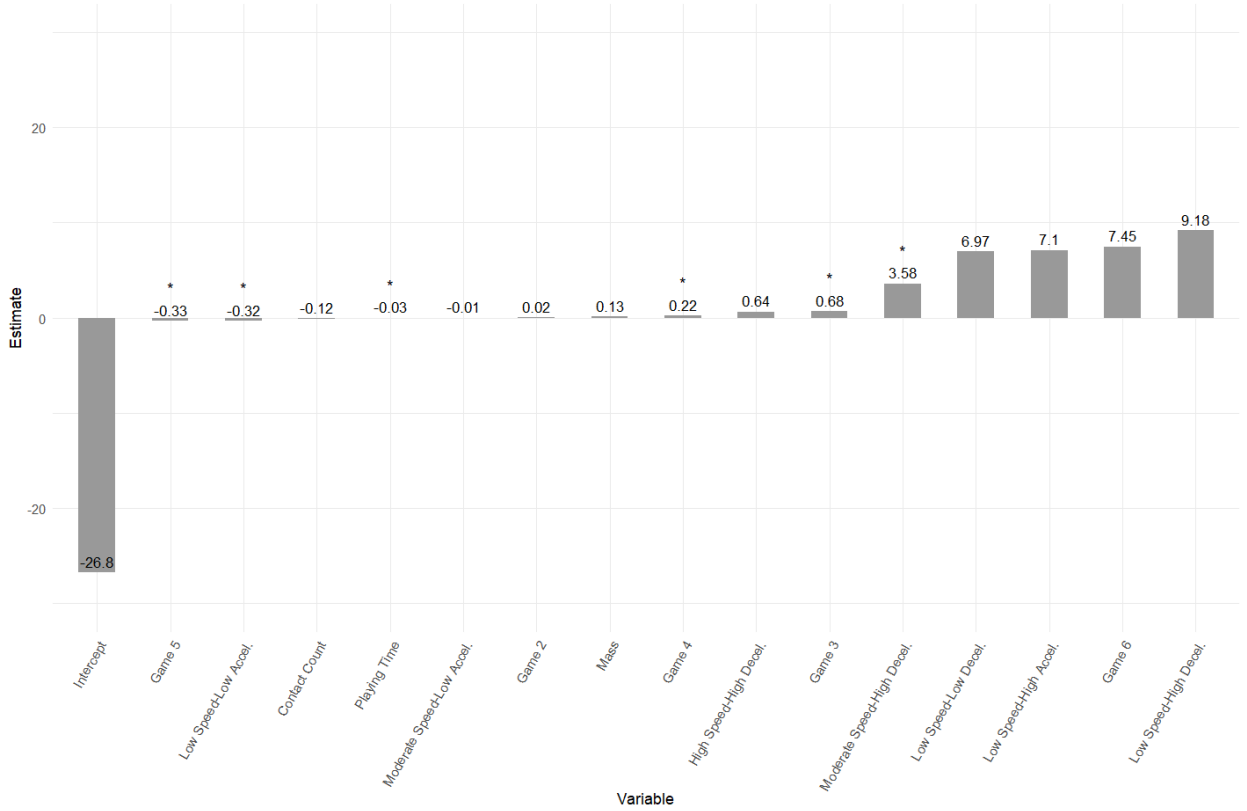

Figure 4.15: Parameter estimates by variable, including zones, * denotes statistical significance ($p<0.01$).

Model diagnostics are presented in Figure 4.16. The variance inflation factor is presented in Table 4.3.

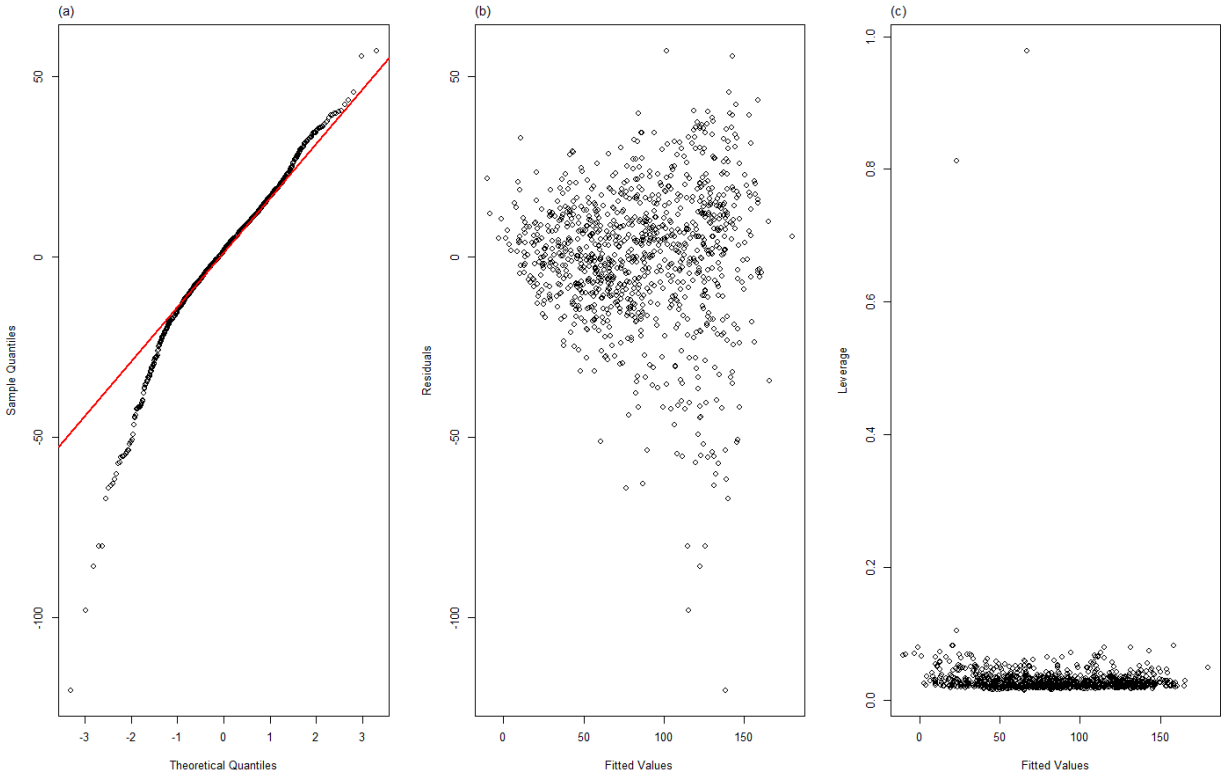

Figure 4.16: (a) Q-Q, (b) residual, and (c) leverage plots for linear mixed model.

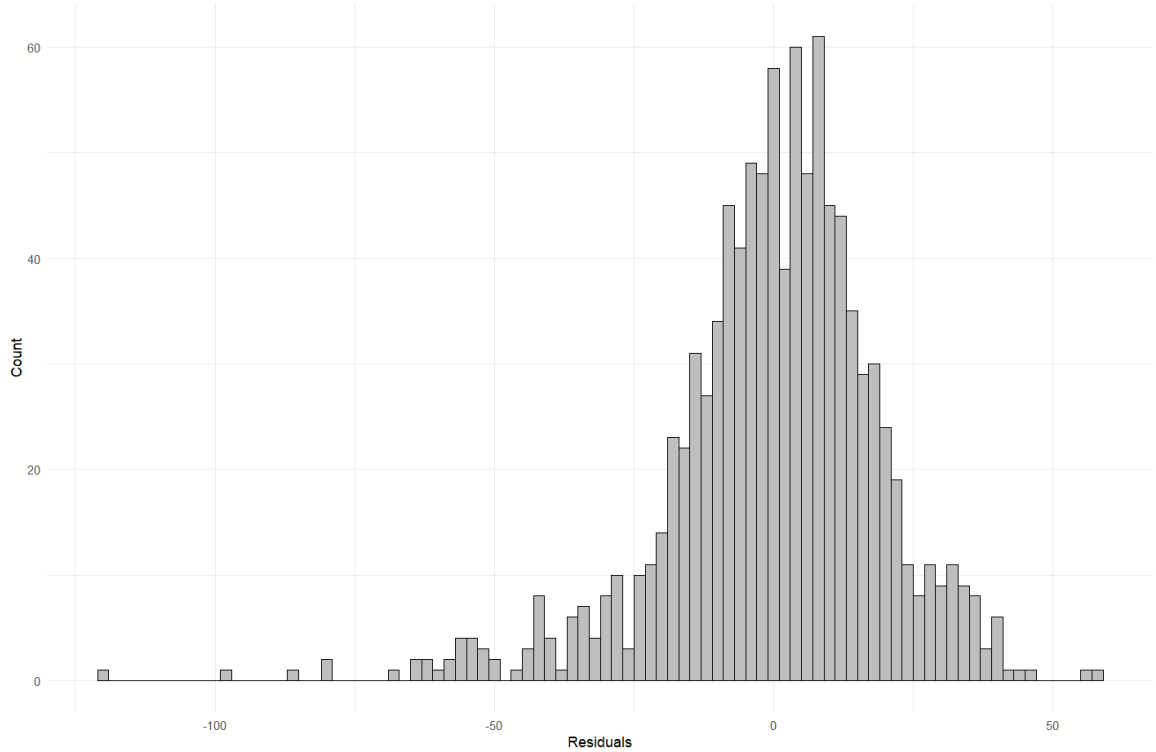

Figure 4.17: Histogram of residuals for linear mixed model.

Table 4.3: Variance inflation factors for linear mixed model.

Term	VIF [95% CI]	Adjusted VIF	Tolerance [95% CI]	Correlation Level	Action (James et al., 2017)
Mass	1.01 [1.00, 2.79]	1.01	0.99 [0.36, 1.00]	Low	Retain
Contact Count	1.01 [1.00, 17.71]	1.00	0.99 [0.06, 1.00]	Low	Retain
Game	1.10 [1.05, 1.21]	1.05	0.91 [0.83, 0.95]	Low	Retain
Playing Time	2.31 [2.11, 2.55]	1.52	0.43 [0.39, 0.47]	Low	Retain
Low Speed-Low Acceleration	2.33 [2.13, 2.58]	1.52	0.43 [0.39, 0.47]	Low	Retain
Low Speed-Low Deceleration	3.10 [2.81, 3.44]	1.76	0.32 [0.29, 0.36]	Low	Retain
Low Speed-High Acceleration	1.60 [1.48, 1.75]	1.26	0.63 [0.57, 0.69]	Low	Retain
Low Speed-High Deceleration	1.05 [1.02, 1.19]	1.03	0.95 [0.84, 0.98]	Low	Retain
Moderate Speed-Low Acceleration	5.61 [5.03, 6.27]	2.37	0.18 [0.16, 0.20]	Moderate	Retain
Moderate Speed-High Deceleration	4.99 [4.48, 5.57]	2.23	0.20 [0.18, 0.22]	Low	Retain
High Speed-High Deceleration	1.28 [1.20, 1.39]	1.13	0.78 [0.72, 0.83]	Low	Retain

As a means to further illustrate model performance, the actual sRPE-CL values are visualized against the model-predicted sRPE-CL values in Figure 4.17.

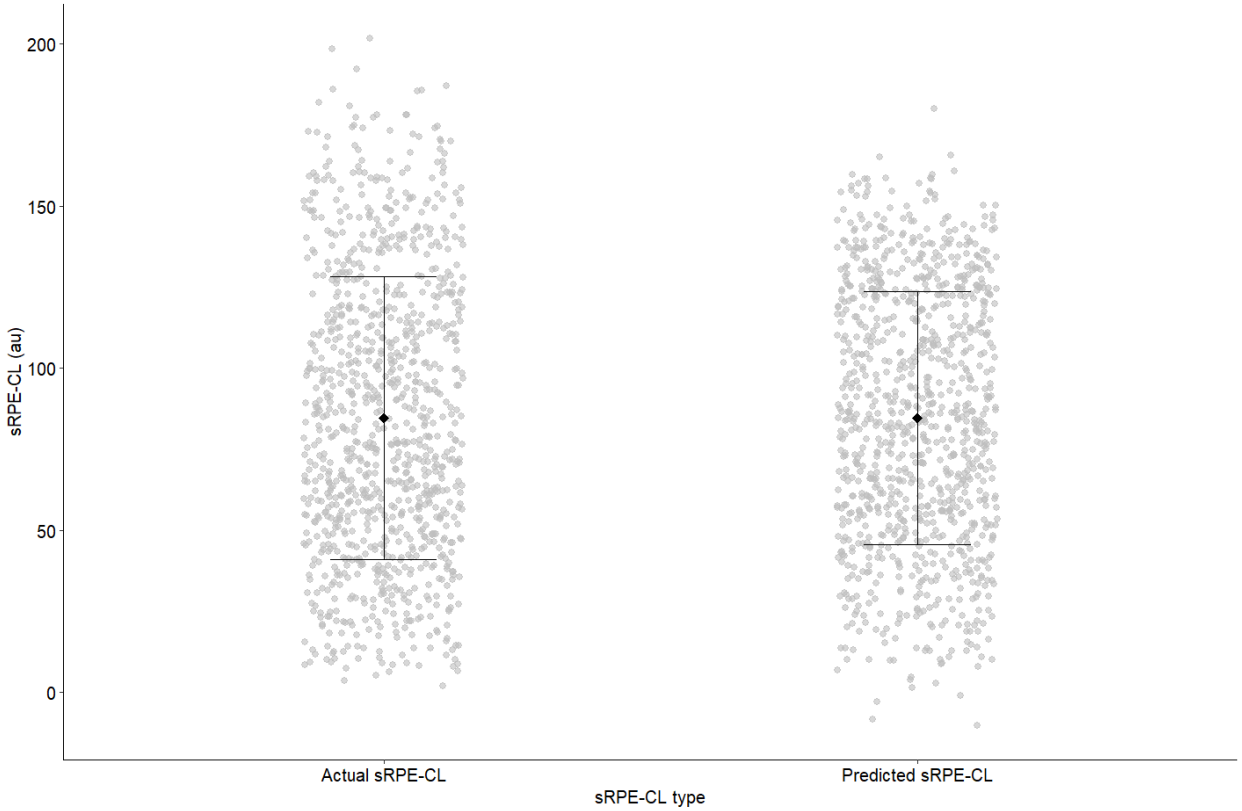

Figure 4.18: sRPE-CL type (actual or predicted) by sRPE-CL value from linear mixed model output with mean and standard deviations shown.

4.4.3 Results of Model Cross Validation

Leave-one-athlete-out cross-validation revealed an overall RMSE of 21.4 across all athletes, indicating acceptable predictive performance of the linear mixed model. However, substantial variability in RMSE was observed between athletes, ranging from 10.3 to 38.3 (Figure 4.18). Notably, prediction error was highest for athletes with fewer games played, suggesting limited training data may have reduced prediction accuracy for those individuals (Table 4.4). Conversely, athletes with larger sample sizes tended to show lower RMSEs. These results support the model's generalizability while highlighting individual-level differences in prediction error.

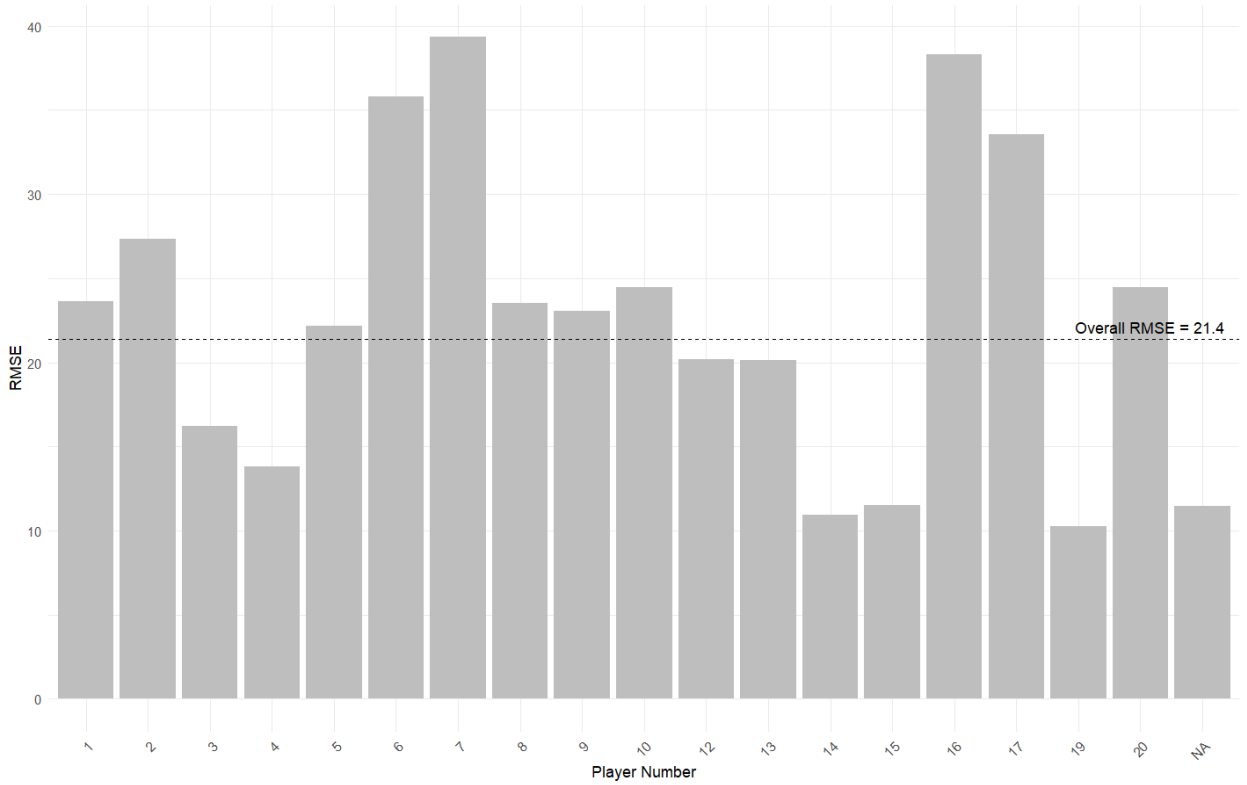

Figure 4.19: RMSE by player number, including overall RMSE, for cross validation.

Table 4.4: RMSE by player number, including number of games played (observations).

Number	RMSE	Number of Games Played
1	23.64	87
2	27.34	82
3	16.20	65
4	13.82	88
5	22.17	83
6	35.79	12
7	23.52	91
8	23.52	91
9	23.05	35
10	24.45	85
12	20.19	54
13	20.12	99
14	10.94	42
15	11.54	20
16	38.30	21
17	33.56	35
19	10.26	36
20	24.49	41
NA	11.45	14

4.5 Discussion

4.5.1 General findings

This investigation has demonstrated that objective factors categorized by accelerations and decelerations derived from athlete worn GNSS sensor data, across different speed zones, as well as a quantification of playing time and game number can provide a reasonable prediction of athlete perceived workload in women's rugby sevens. The use of different categories of

acceleration/deceleration and speed zones presents a novel approach to quantify the contribution of athlete movement to athlete load using GNSS sensors. Additionally, the significant contribution of distance traveled during high decelerations at moderate speeds as well as during low accelerations and low speeds to athlete load in this cohort provides impetus for future research. The significant effect of particular game numbers affirms the importance of considering sport-specific contextual factors in female rugby athlete load monitoring (Epp-Stobbe et al., 2025b). Further, the significant effect of playing time supports the important consideration of the role time plays in influence load.

4.5.2 Playing time as a driver of athlete load

The significant main effect of playing time is not surprising given the role of duration as an element of athlete load. Indeed, sRPE-CL is the product of sRPE and duration (Haddad et al., 2017). There are two key considerations as a result of this finding. The first is the methodological implications of the existing relationship between the two values, this mathematical coupling does create a fundamental concern in that the playing time of an athlete is the most highly influential factor on that athlete's workload in this particular model (Lolli et al., 2019). However, if an athlete's sRPE value were to remain completely independent of duration there should be a perfect linear relationship present, however, that is not the case, suggesting that the relationship is moderated by other factors. The second consideration is of the moderating factors of sRPE-CL. In this case, it is anticipated that other factors outside of time are also driving athlete load, which this investigation shows to be the case. Ultimately, this result infers that further investigation into the influence of playing time as well as care in the deployment of time-related metrics for athlete monitoring are necessary.

4.5.3 High deceleration at moderate speed as a driver of athlete load

The significant main effect of high decelerations occurring at moderate speeds invites consideration for both methodological and sport-specific considerations. From the standpoint of methodological considerations, the treatment of deceleration data in this investigation may contribute to the observed effect. For example, Delaney et al. (2016), chose to express all acceleration and deceleration values as positive values to give an indication of total accelerations. Alternatively, Delaney et al. (2018) divided accelerations by decelerations, resulting in a quotient value for analysis. Similarly, Furlan et al. (2015) used thresholds of acceleration and deceleration to explore metabolic demands of men's sevens. The present study categorized acceleration and

deceleration using one threshold applied both as a positive and negative value to create four possible outcomes which meant there were more opportunities for some element of deceleration (high or low) to influence load in the model. The considerations for acceleration information are further summarized in Table 4.5.

Table 4.5: Methodological approaches to processing acceleration and deceleration data.

Authors	**Population**	**Strategy for Data Processing**	**Data Processing Output**
Delaney et al. (2016) [16]	Men's rugby league	Absolute value (all values expressed as positive acceleration)	Total acceleration distance
Delaney et al. (2018) [15]	Men's rugby league	Accelerations/Decelerations	Acceleration ratio
Furlan et al. (2015) [17]	Men's rugby sevens	Thresholds of acceleration and deceleration	Four bins of acceleration and deceleration distances
Current study	Women's rugby sevens	Thresholds of speed and acceleration and deceleration	Twelve bins of acceleration and deceleration distances

The current model demonstrates that decelerations do have a significant role in athlete load in women's sevens, however, the variety of methodological approaches to handling acceleration and deceleration creates discrepancies in reporting (Schuster et al., 2018). However, the role of deceleration remains unclear in the literature as high deceleration cases may be included within a low acceleration threshold as deceleration is a negative value or may be processed to become a positive value (Delaney et al., 2016; Delaney et al., 2018). Nevertheless, deceleration is present in field sports, and as such, is warranted being included in training load models (Delaney et al., 2016; Delaney et al., 2018; Schuster et al., 2018; Torres-Ronda et al., 2022). Ultimately, clarity in methodological constructs will support further development of training load models as our study suggests that high deceleration cases are particularly impactful on sRPE-CL.

The sport-specific consideration for the presence of high-decelerations may also explain the observed effect. In rugby, as opposed to other field sports, deceleration may come from the athlete volitionally slowing to step around another player, or it may come from the athlete encountering another body in physical contact, from a tackle, carry, or fend (Hendricks et al., 2014). However, in this model, the unique causes of deceleration (negative acceleration) were not isolated and as such, physical contacts may be captured in the deceleration categories as well as in the contact count. Further investigation into what contributes to the nature of deceleration in each category is essential to appropriately quantify the potential role of contact on athlete deceleration. For example, as a potential interpretation of the significance of high deceleration at moderate speeds it is possible that when scrummaging or rucking substantive force to cause a high deceleration occurs as bodies come in contact at close distances. Alternatively, players on the edges of the field, may experience a high deceleration as they react to play, cutting to evade an opponent or sharply changing direction in response to adjustments of the defensive line. As cases of high deceleration across different movement speeds remain a strong influencing factor of sRPE-CL in women's rugby sevens, more research investigating the exact nature of these relationships is needed.

4.5.4 Low acceleration at low speed as a driver of athlete load

The significant main effect of low acceleration at low speed is an interesting finding given that, compared to other movement categories, there is a relatively lower amount of distance accrued in this space. When considering the sport demands unique to rugby, this movement category suggests that events occurring at low speeds, from standing or walking, and with very limited acceleration, influence athlete load. Considering rugby, there is a wide range of tactical and technically skilled movements that are required to be executed at either low speed or with limited space to accelerate, accruing limited distance. The presence of rucks, scrums, tackles, and ensuring grappling or wrestling movements may all occur at these speeds and accelerations (Hendricks et al., 2020). In these movements, players may generate high forces but without significant physical displacement (Hendricks et al., 2012; Hendricks et al., 2014). These activities may be physically exhausting, resulting in an influence on athlete workload (McLellan et al., 2011). Future investigations may consider evaluating the presence of time, instead of distance in this space.

Another tactical consideration is the efficiency of the player to return to position after participating in the play of the game (Deutsch et al., 2007). Players who spend more time lying on the ground

after a tackle or walking back to position after a break in play would accrue more distance in this movement category, potentially highlighting a greater relative distance in this category compared to the opportunity to move at other speeds and accelerations/decelerations (Deutsch et al., 2007). The consideration for efficient return-to-position ensures team strategies may be appropriately executed; however, the significant main effect of this category suggests that these seemingly low-intensity movements carry influence over athlete perceived workload that other movement categories fail to capture.

4.5.5 Game number as a driver of athlete load

Interestingly, specific game numbers produced significant main effects in the model, suggesting that the game number within a tournament influences sRPE-CL in this cohort. Generally, in rugby sevens tournaments, five or six games are played over two or three days, this means that in a six-game two-day tournament three games are played on day one and three on day two, and in a five-game three-day tournament two games are played day one, and either a third and fourth on day two or just a third on day two with the remaining game(s) played on day three. This suggests that fatigue accumulates over the course of the tournament, whereby higher games appear to influence athlete load. In the case of the third game exhibiting a significant main effect, it may be driven by the two-day tournament format in which this game comes at the end of a long day of rugby. In addition, in the pool/knockout structure of sevens tournaments, pool-play consists of three games and therefore the third game may come with added psychological pressure when players feel the game must be won in order to advance to the knockout stages to play for a medal (REF). The format may also explain why games four and five are significant drivers of athlete load in this model: they represent knockout matches in which a win ensures the team moves forward with the goal of winning a medal, whereas a loss may mean that the team does not achieve a top finish in the tournament (REF). This insight, coupled with the accumulation of physical exertion in a short time frame, provides a plausible basis for why these games produce significant main effects (REF; McClellan et al., 2011).

Interestingly, however, the final game of a tournament, game six, was not identified as a significant main effect and in this case the possible explanation stems from the tournament format once more whereby this game either represents playing for a gold medal, a bronze medal, or simply for a ranking. When the value of the match is high (i.e. playing for a medal) players may identify a

different level of effort (i.e. the effort was worth the outcome so on the whole it was not as fatiguing) and when the value of the match is low (i.e. no chance for a medal, playing for a ranking) players may not feel a high level of effort is warranted (Schultz de Arruda et al., 2019; Ferioli et al., 2021). No matter the case, the sport-specific considerations for the accumulation of load throughout a competition period are highlighted by the influence of particular game numbers on athlete load in this investigation.

4.5.6 Future considerations

The use of GNSS sensor data to obtain distances, speed, and calculate accelerations and decelerations presents a simple, efficient strategy for data collection (Malone et al., 2017). It is worth noting that multisensory approaches leveraging inertial measurement units (IMUs) for the collection of speed and acceleration data are also popular strategies in athlete monitoring (Roell et al., 2018). However, the processing strategies to fuse data between GNSS and IMU sensors in commercially available products is not well known (Malone et al., 2017). Time intervals, smoothing, and filtering strategies are not always publicly available to practitioners and as such use of GNSS data offers data that is sufficiently robust and transparent (Malone et al., 2017; Bartlett & Drust, 2021). Further investigation into sensor fusion methodologies present in commercially available products, as well as the validity across sporting populations is encouraged.

4.5.7 Practical applications for coaches and practitioners

This investigation identified the contribution of several meaningful variables to athlete load in women's rugby sevens, with implications for both the competition and training environments.

High deceleration events at moderate speeds, equivalent to running but not sprinting, are particularly impactful to athlete load in women's sevens rugby. In competition, monitoring these events through post-game analysis of GNSS data may help coaches gain insight into player experiences from the previous match, and encourage adjustments to line-ups and substitution plans for subsequent matches. In training, practitioners should focus on training for optimal movement strategies, whether through self-slowing and cutting drills or in rugby contact drills around tackle or carry technique, to ensure athletes are able to safely experience the demands of high deceleration (Caparrós et al., 2018; McBurnie et al., 2022).

Practitioners are encouraged to consider the stage of competition of their athletes. In sevens, the game number influences athlete load across a tournament. This, coupled with tracking playing time which also influences athlete load may support team strategies around managing athlete load across a tournament, such as when to rest certain players. The sevens tournament format is such that teams of twelve athletes will play five or six games in a two- or three- day period, making roster management a critical element of team success (Furlan et al., 2015; World Rugby, 2025a).

Finally, practitioners are strongly encouraged to determine relevant drivers of athlete load for their sport environment and clearly document how their data is measured to ensure high quality, reproducible cases.

4.6 Conclusion

This investigation presents a novel approach to the quantification of athlete movement as a contributor to athlete load in competition. The model explained 79.3%of global variance of sRPE-CL, with playing time, select game numbers, low accelerations at low speeds, and high decelerations at moderate speeds proving significant factors. The significant contribution of distance in high decelerations at moderate speeds invite practitioners to consider the role of deceleration in women's rugby sevens as an important load metric. Further the significant contribution distances covered at low accelerations at low speeds highlights considerations for in-game actions by athletes including a consideration for workrate. The significant contribution of select game numbers in a tournament continues to attest to the importance of considering sport-specific factors in athlete load monitoring. Finally, the significant contribution of playing time reinforces the importance of time as a primary driver of athlete loadTaken together, the results infer that objective metrics like movement category, game number, and playing time may be useful in the quantification of athlete load in women's rugby sevens.

5 Rugby sevens sRPE workload imputation using objective models of measurement

5.1 Abstract

While accurate athlete load monitoring is crucial for preventing injury and optimizing performance, the commonly used session rating of perceived exertion (sRPE-CL) method faces limitations due to compliance issues related to missing subjective athlete self-reported data and subsequent challenges in imputing the rating of perceived exertion (RPE) component. This study investigated imputing missing sRPE from mechanical work and from a Speed-Deceleration-Contact (SDC) model. 1002 rows of data were collected from women's rugby sevens competitions. Using each of mechanical work or SDC, linear regression, and random forest imputation models were assessed at different missingness levels and compared to a common method of daily team mean substitution, using an ANOVA of accuracy by model type and missingness. Statistical equivalences were evaluated from true and imputed sRPE scores by model and strategy. Significant interactions between model type and missingness were found, with all imputed scores being deemed statistically equivalent. From the ANOVA, daily team mean substitution was the poorest performing model and random forest was the best. However, the best performing model was not superior to previously reported imputation approaches which confirms the difficulty in using subjective measures of load when missing data is a prevalent issue in team sports. Practitioners are encouraged to critically evaluate any method of imputation for athlete load.

5.2 Introduction

In the women's rugby sevens competition environment, teams play five or six games over two to three days, with multiple tournaments happening in the span of a few weeks. This high competition volume means that managing athlete loads to ensure safe and successful participation is of critical importance. The use of sensor technology including athlete tracking devices (ATDs) like GNSS monitors, is a popular option for athlete load monitoring in team sports (Cummins et al., 2013a, 2013b). However, proprietary algorithms may include metrics that are either not defined or not optimized to accurately quantify loads experienced by female athletes (Clarke et al., 2015; Clarke et al., 2017).

With concerns over the suitability of proprietary ATD metrics, many teams continue to rely on subjective workload measures of sRPE-TL (Haddad et al., 2017; Benson et al., 2021; Griffin et al., 2021). sRPE-TL is the product of an athlete's playing time and their self-reported sRPE which is used to assign a value to a level of effort the athlete felt they worked at during the session or match (Haddad et al., 2017). While the use of sRPE-TL is a fast way to glean insight to an athlete's workload, it relies on regular reporting by athletes and is known to be sensitive to factors like experience with the RPE scale (Eston 2012). Further, in competition settings the demands on athletes between matches to recover and prepare for the upcoming match mean that sRPE values may not be reported, with a lower compliance and missing data, making the calculation of sRPE-CL not feasible (Saw et al., 2015). Missing or incomplete load information may create unnecessary risk for athletes through an inability to consistently quantify athlete performance, either over or underestimating the load experience (Gabbett, 2016). Therefore, there is a need to find ways to mitigate missing data, such as using imputation techniques, as well as potentially identifying other load monitoring approaches that may rely less on subjective input and compliance.

The use of mathematical techniques for the imputation of missing values presents a unique solution to incomplete datasets (El-Masri & Fox-Wasylyshyn, 2005; Cummins et al., 2013b; Windt et al., 2018; Griffin et al., 2021). Missing value imputation (MVI) commonly occurs in sport research through value substitution, classification, or regression models (El-Masri & Fox-Wasylyshyn, 2005). In sport data, the use of group mean substitution is a popular strategy for its ease of use (Benson et al., 2021; Griffin et al., 2021). While group mean substitution is a common approach, this technique was outperformed by multiple different models for the imputation sRPE in a cohort of rugby sevens athletes included linear regression, random forests, support vector machines, and neural networks (Carey et al., 2016). This evaluation relied upon available sRPE data for group mean substitution compared to a simple multicomponent model (sRPE = Match Number + Player + Opponent + Total Distance + Playing Time + Contact Count) augmented with machine learning techniques (Epp-Stobbe et al., 2022a). It is important to note that the best performing model in this study was a random forest classifier. While this approach resulted in improved imputation performance compared to the standard daily team mean substitution, the model accuracy was still only 26.5% with and R^2 of 0.407 as compared to daily team mean which had a model accuracy of 20.6% and a R^2 of 0.09 (Epp-Stobbe et al., 2022a). This suggests that the simple model employed may not be a suitable objective approximation of athlete load or sRPE. This supports an

opportunity for other, statistically or theoretically driven, objective models to improve the ability to impute missing sRPE data.

An alternative workload metric, which objectively quantifies athlete load using data gathered by ATDs, is mechanical work (Delaney et al., 2018; Buchheit 2019; Tuft & Kavaliauskas, 2021; Epp-Stobbe et al., 2024b). The use of mechanical work is a popular strategy as it is minimally invasive to the athlete and represents a measure of both intensity and duration (Delaney et al., 2018; Buchheit 2019; Tuft & Kavaliauskas, 2021; Epp-Stobbe et al., 2024b). Mechanical work is the product of force and distance, and can be calculated using data from an ATD, including speed and session duration, alongside athlete mass data (Delaney et al., 2018). As an alternative to mechanical work, other objective statistical models have been suggested to quantify athlete load. (Carey et al., 2016; Bartlett et al., 2017). One such example is the Speed-Deceleration-Contact (SDC) model proposed by Epp-Stobbe et al. (2024a) which uses distances travelled by the athlete in distinct speed and deceleration zones as well as the number and frequency of contacts experienced by the athlete. This model was found to have reasonable explanatory power of sRPE-CL ($R^2_{adjusted} = 0.487$) suggesting it may be used in place of sRPE-CL as a load monitoring tool. Considering the relationships between objective load metrics, such as mechanical work and SDC, and sRPE-CL and the need for imputation of sRPE, it is possible that current objective load metrics, augmented with different statistical approaches, could improve the ability to impute sRPE over current methods.

The use of objective mathematical models like mechanical work and the SDC are appealing alternatives to athlete self-reported data and may further be useful metrics to support MVI in the case of missing sRPE data when sRPE-CL is used as an athlete load metric. Therefore, the purpose of this investigation is to assess the accuracy of alternative load metrics of mechanical work and SDC, augmented with linear regression and random forest classification to impute sRPE as compared to daily team mean substitution. Outcomes of this investigation could provide important alternatives for the imputation of sRPE which could better support the use of sRPE-CL to calculate athlete loads when missing data is experienced.

5.3 Materials and methods

5.3.1 General methods

Twenty-one women's rugby sevens players (25.5 ± 3.90 years old, 169.4 ± 5.89 cm tall, and 71.0 ± 5.64 kg) in a full-time training program provided data for 101 international matches over several years in a retrospective qualitative analysis. All data were anonymized by team staff prior to analysis. Ethics approval was provided by the University of Victoria for the use of voluntary data collection with the investigation complying with the principles outlined in the Declaration of Helsinki.

Date, match number within the tournament, and opponent were provided for each match in the dataset. Objective variables from ATDs, worn between the shoulder blades in a custom harness of each athlete, collected athlete playing time and total distance covered in each match (Apex v2.50, StatSports, Newry: UK), were available for potential inputs into imputation models. Athlete masses were collected before each match using a portable weigh scale (ES-310, Anyload, Burnaby:CAN). Athlete self-reported RPE was collected post-match using a modified Borg scale familiar to athletes (Eston 2012). Further, footage of each match was evaluated by one trained analyst to determine a summed count of all contacts experienced in the match (tackles, carries, contested restarts, and rucks) as developed by staff maintaining the team's current match analysis practices (Sportscode v 11, Hudl, Lincoln: USA) (Gabbett & Kelly, 2007; King et al., 2010; Wheeler et al., 2011).

Overall, absolute game mechanical work (W) was calculated as the cumulative sum of the product of instantaneous absolute power (P) and time (t) (Equation 5.2). Instantaneous absolute power was calculated as the product of athlete mass, acceleration, and velocity. Overall, absolute game mechanical work (W) was calculated as the cumulative sum of the product of instantaneous absolute power (P) and time (t) (Equation 5.2).

$$W = \sum_{i=1}^{n} (P_i \cdot \Delta t_i)$$

Equation 5.1: Overall absolute game mechanical work.

5.3.2 Imputation of sRPE

To model the relationship between sRPE (dependent variable) and imputation strategy (independent variable), statistical models were used to classify and predict sRPE based on daily team mean, mechanical work or the SDC model. Following this, comparisons of the true RPE calculated data and the imputed model data were made (R version 3.4.4, Vienna, Austria). A total of 1002 complete rows of data were available for analysis.

Daily team mean substitution was one imputation strategy which relied on the sRPE values of other teammates from the match and day. Mechanical work data were used to impute sRPE by player and match number. SDC model data were used to imputed sRPE data by player, opponent, match number, playing time, contact count, and total distance covered in zones of high deceleration and high speed, and high deceleration and low speed (Epp-Stobbe et al., 2024a).

This investigation used daily team mean substitution, linear regression models (linear, (R stats package) and random forest (R randomForest) to classify and predict sRPE (R Core Team n.d.; Riedmiller 1994; Tibshirani 1996; Brieman 2001; Hsu et al., 2003; Zou & Hastie, 2005; Fritsch et al., 2019; Benson et al., 2021; Griffin et al., 2021; Beygelzimer et al., 2022; Friedman et al., 2022; Liaw & Wiener, 2022; Meyer et al., 2022; Ripley et al., 2022). Linear regression was selected as it is a common and easy to administer approach currently used in sport research. Random forest classification was chosen as it has been shown to be a superior method for RPE imputation that can be executed with open-source software (Musil et al., 2002; Celton et al., 2010; Carey et al., 2016; Benson et al., 2021; Griffin et al., 2021). Random forest classification is an adaptation of the decision tree process where data is divided according to parameters over numerous decision trees (Brieman, 2001). In imputation, random forests leverage the idea that a group of models will generally outperform a single one, thereby multiple decision trees are used to determine final predicted class based on the majority class that is predicted, producing a more robust result (Brieman, 2001). While the use of random forest was previously shown to be a viable option in imputation of sport data, it is important to recognize that the inputs for this investigation differ from the case demonstrated by Epp-Stobbe et al. (2022a) (Chapter 2). However, given the presence of multiple decision trees as well as the relatively limited size of the feature sets suggests that random forest may be a useful and relevant tool for the data in this investigation (Brieman, 2001).

This investigation used daily team mean substitution, linear regression models (linear, (R stats package) and random forest (R randomForest) to classify and predict sRPE (R Core Team n.d.; Riedmiller 1994; Tibshirani 1996; Brieman 2001; Hsu et al., 2003; Zou & Hastie, 2005; Fritsch et al., 2019; Benson et al., 2021; Griffin et al., 2021; Beygelzimer et al., 2022; Friedman et al., 2022; Liaw & Wiener, 2022; Meyer et al., 2022; Ripley et al., 2022).

The data were split into training and test datasets, with 80% allocated for training the models and 20% for testing them. This process was repeated 100 times, and the mean values from these iterations were used for further analysis to generate predicted mechanical work scores. Following the process for the imputation of sRPE from mechanical work, the same models were used to impute sRPE from the Speed-Deceleration-Contact (SDC) model using the same datasets.

The predicted sRPE values were then compared to the true, calculated values from the test dataset. Accuracy, R^2, and root mean square error (RMSE) were used to evaluate the models' ability to impute sRPE in comparison to the true athlete-reported sRPE scores.

To establish statistical equivalence and, in turn consider model interchangeability, a paired-samples equivalent test (paired TOST) and the imputed values from the test dataset at 20% missingness, were compared against the true sRPE values (Lakens et al., 2017; Lakens et al., 2018). Practically, 20% missingness was determined as reasonable as that was equivalent to two to three missing sRPE values within a team, which represented regular outcomes of data collection based on advice of team staff (Benson et al., 2021; Griffin et al., 2021). The bounds of Cohen's *d* $\times \boldsymbol{\sigma}$, were used in the paired-samples equivalence test, using a Cohen's *d* of 0.2 to represent a small effect size (Lakens et al., 2017).

To explore the cases of divergence in accuracy, all imputation strategies were tested at different levels of missingness, in 5% increments, from 5% to 30% and iterated 100 times. A one-way ANOVA compared model accuracy, by imputation strategy-model type, by missingness, and the interaction of the strategy-model type and missingness. A Bonferroni planned comparison test was performed to identify differences in imputation strategy-model types at 20% missingness. This investigation hypothesized that different objective models, using linear regression or random

forest, would exhibit improved accuracy over daily team mean substitution (Hawthorne & Elliot 2005; Kang, 2013).

5.4 Results

5.4.1 Description of data

On average, athletes covered 1082.86 m of total distance (± 439.78 m), played 11.04 minutes (± 4.67 minutes), and experienced 5 contacts (± 3 contacts) per match, reporting a sRPE of 7 au (± 1.8 au) and a sRPE-CL of 84.80 au (± 43.67 au), a mechanical work demand of 56236.38 Joules (± 21413.36 Joules), and an SDC model workload of 7.77 au (± 1.596 au).

5.4.2 Model performance for imputation of sRPE from mechanical work

Imputation model accuracy, R^2, and RMSE values are reported in Table 5.1.

Table 5.1: Imputation model accuracy, R^2, and RMSE at 20% missingness.

Strategy	Model	Accuracy	R^2	RMSE
Daily Team Mean Substitution		0.0000	0.0377	1.80
Mechanical Work	Linear Regression	0.1841	0.0854	1.78
	Random Forest	0.1891	0.1590	1.71
SDC Model	Linear Regression	0.2200	0.2287	1.61
	Random Forest	0.2724	0.3383	1.51

Paired-samples equivalence tests of the mechanical work-imputed sRPE and SDC-imputed sRPE against the true sRPE resulted in all tested models being deemed statistically equivalent to the true sRPE data ($p < 0.05$).

The one-way ANOVA of the data at 20% missingness found a statistically significant difference in accuracy by imputation strategy-model type (e.g. Mechanical Work-Linear Regression, SDC-Random Forest) ($F(4, 2994) = 24.78$, $p < 0.05$), by missingness ($F(5, 2994) = 0.39$, $p < 0.05$), and their interaction ($F(20, 2994) = 0.89$, $p < 0.05$). A Bonferroni planned comparison of the strategy-model type accuracy at 20% missingness found that all strategy-model types were significantly different from one another ($p < 0.05$) (Figure 5.1).

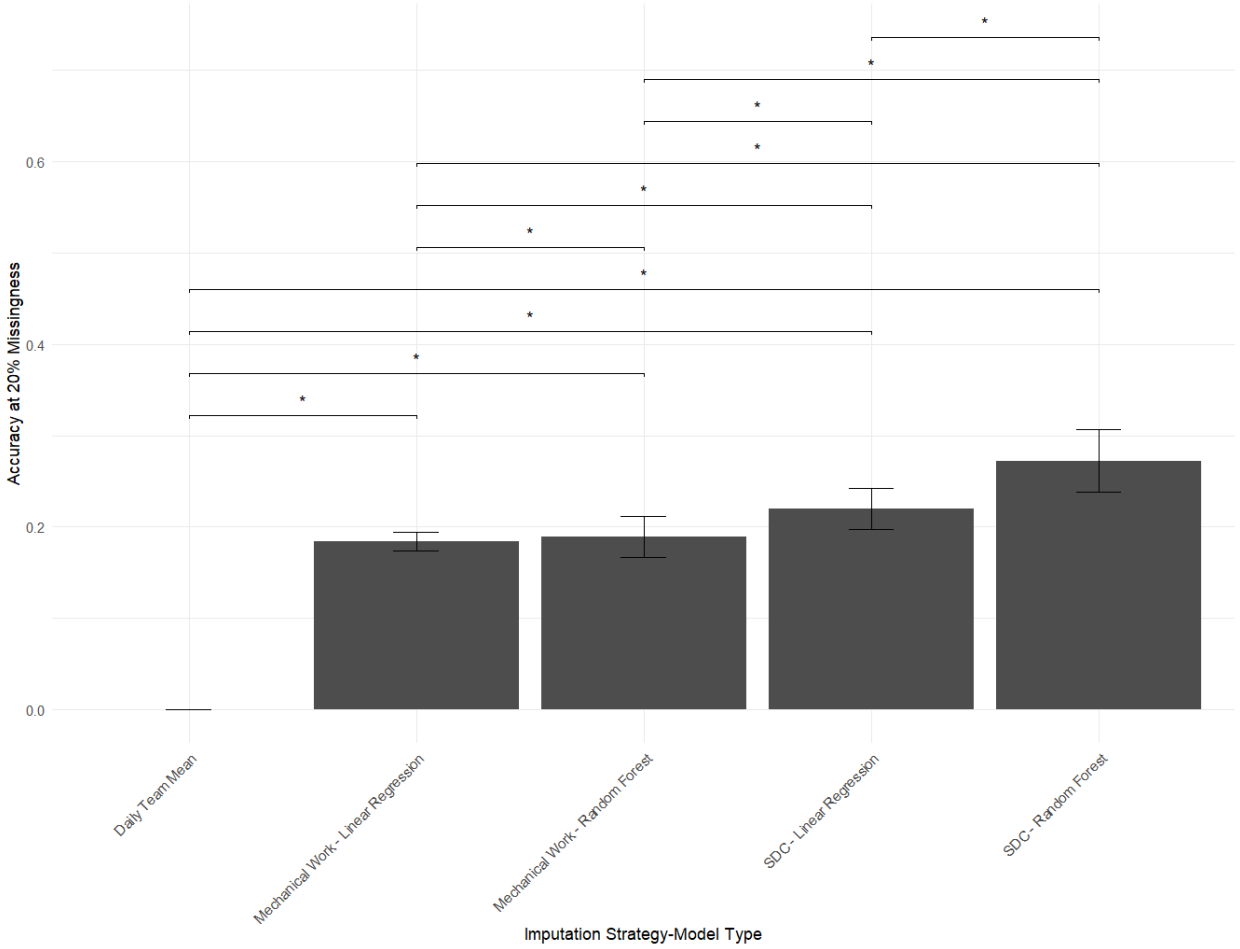

Figure 5.1: Average accuracy by missingness for all imputation strategy-model types, * denotes statistical significance

5.4.4 Comparison of all models for sRPE imputation explanatory power by missingness

The accuracy of all models, across all levels of missingness for sRPE imputation strategies, daily team mean, mechanical work and the SDC model respectively, is shown in Figure 5.2.

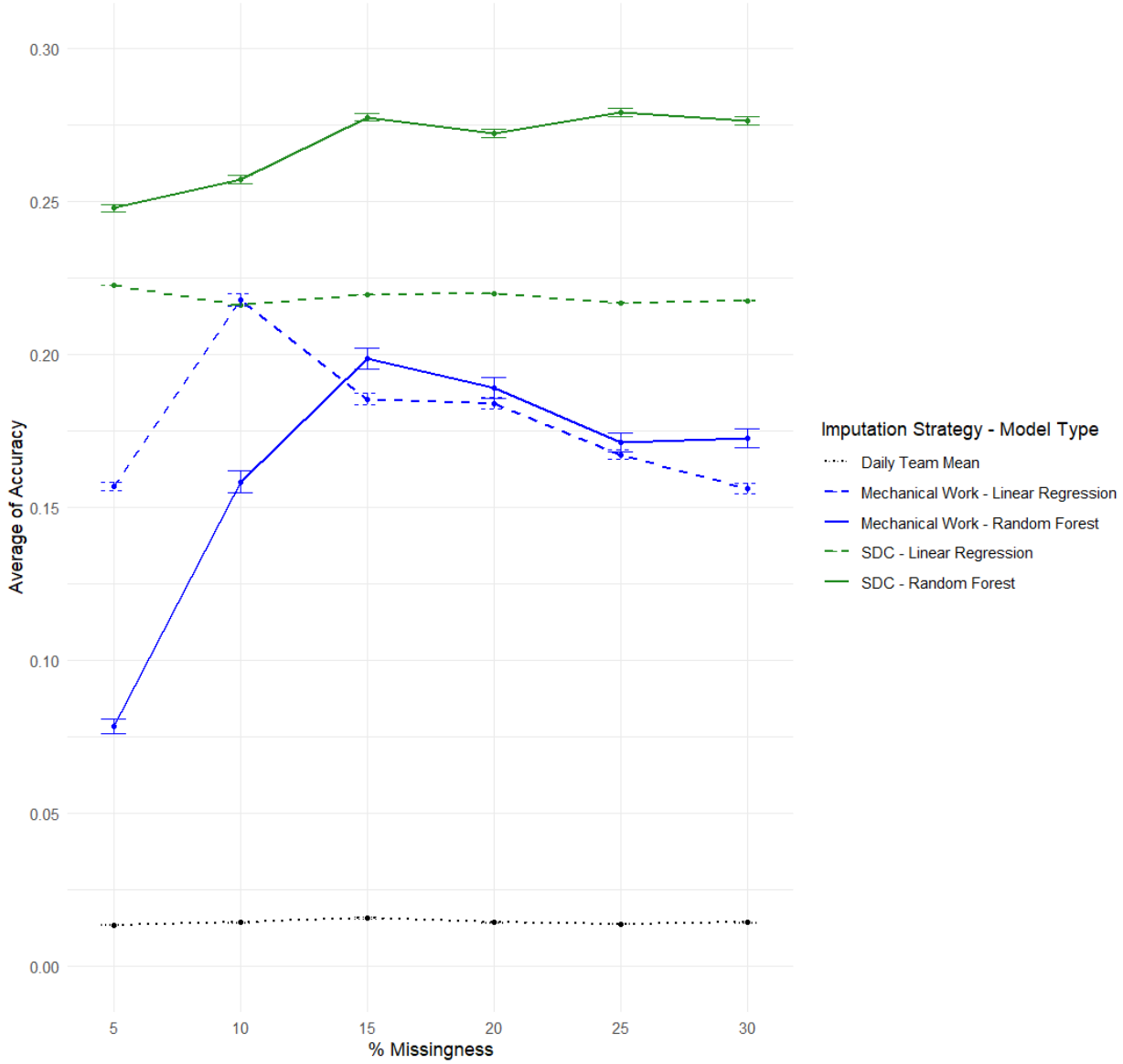

Figure 5.2: Average accuracy by missingness for all model types and both objective imputation strategies.

5.5 Discussion

To our knowledge, this is the first study to examine the imputation of sRPE using known objective load models, and not just a collection of metrics, to support the calculation of sRPE-CL when athlete subjective sRPE data is missing. First, it was found that both objective workload measures,

mechanical work and the SDC model, outperformed daily team mean substitution irrespective of the statistical approach used (Table 5.1, Figures 5.1 & 5.2). Further, in terms of statistical approaches, for each objective model compared, random forest performed the best in terms of accuracy and explanatory power in imputing sRPE. Finally, the SDC model using random forest regression resulted in the best accuracy and explanatory power of all strategies and models. However, regardless of the models used, the model accuracy and goodness of fit statistics would be considered poor. This finding further substantiates the difficulty in using subjective measures for athlete load calculation when adherence may limit reporting and data missingness is possible. Overall, these results suggest that while objective models can be used to impute missing RPE data for use in the calculation of sRPE-CL the true athlete reported value is far superior and it is recommended that either strategies are developed to minimize missing data or the use of objective load metrics, in place of subjective metrics, would be advised (Schmitt et al., 2015; Carey et al., 2016; Benson et al., 2021; Griffin et al., 2021).

The very poor performance of daily team mean substitution is consistent with previous imputation research in this women's rugby cohort. This finding aligns with work in medical and athletic fields suggesting that mean substitution is less effective, with greater error and variability, compared to regression or classification strategies (Musil et al., 2002; Waljee et al., 2013; Carey et al., 2016; Epp-Stobbe et al., 2022a). While mean substitution is an efficient strategy, the low accuracy suggests that it does not account for variance of individual performances within a team sport whereby an athlete may be required to participate for different times (starter, substitute), perform specialized tactical skills (kicking, scrummaging), or experience different physical demands (sprints, tackles, carries). The consistent, poor performance of mean substitution strategies in imputing missing athlete workload data emphasizes the limitations of this strategy; the inability to accurately reflect individual physical demands and individual athlete's perceptions of the competition environment (Carey et al., 2016; Epp-Stobbe et al., 2022a).

The results of the equivalence testing suggests that the imputed sRPE data are not different from true sRPE scores no matter if the imputed data is produced using mechanical work or the SDC model. However, these results must be considered in parallel with the results of the ANOVA test for model accuracy which suggest that there are significant differences between the models across levels of missingness and imputation strategy-model type in Figures 5.1 and 5.2. Taken together,

the SDC model may be the better strategy for the imputation of missing sRPE values, and imputation using random forest or linear regression may be preferred over that of the daily team mean. Specifically, random forest explained the most global variability in the SDC model (highest R^2) while having the highest accuracy. In other fields, Waljee et al. (2013) identified random forest classification as a highly accurate imputation strategy for medical data missing completely at random (MCAR). Similarly, Hong and Lynn (2020) noted that random forest imputation provides high predictive accuracy for data missing at random (MAR). Random forest models are an appealing option for imputation as the models can handle a variety of data types without relying on distributional assumptions (Shah et al., 2014; Kokla et al., 2019). However, despite their predictive accuracy, random forest models cannot estimate relationships among imputed values (Hong & Lynn, 2020). This suggests that other models, not considered in this study, may be beneficial, in order to better identify the relationships between variables (Schmitt et al., 2015). For example, fuzzy clustering or Bayesian approaches may be able to more accurately describe outcomes for sport datasets as these approaches work with an overlap in input data ranges and limited possible outcomes (Schmitt et al., 2015; Rahman & Islam, 2016). Conversely, linear regression models provide a simple, yet viable option for imputation that may be more easily employed than random forest models by practitioners in an applied sport environment where computational resources and time may be limited (Musil et al., 2002; Waldmann et al., 2013).

The potential to use objective workload measures in team environments is appealing and should be considered in light the consistency of data collection from ATD for objective workload calculations and the great potential for missing subjective data in sRPE-CL collection. Objective measures can also provide a great opportunity to interpret changes in workload connected to discrete factors where subjective measures may not allow such determination. The SDC model, especially using random forest, was shown to be the best model for the imputation of missing sRPE data. This model was shown to have reasonable prediction of sRPE-CL (Epp-Stobbe et al., 2024a) and have components such as contact which has been shown to be related to sRPE in this cohort (Epp-Stobbe et al., 2022b). The combination of both physical and tactical components within the SDC model, may more broadly encompass the athletic experience as it relates to workload (Epp-Stobbe et al., 2024a) and support this model as a useful strategy for the imputation of missing RPE values as well as a standalone workload metric. Mechanical work was chosen as another objective load metric for imputation of RPE in this study because it is a purely objective representation of

the cumulative athlete kinematics and kinetics collected using ATDs. It has been previously suggested that mechanical work does not necessarily account for psychological factors relating to competition, leading to different workload values than sRPE-CL, as demonstrated by Epp-Stobbe et al (2024a). Therefore, it is reasonable that it did not perform as well as the SDC model which combines both physical and tactical information about the athlete (Epp-Stobbe et al., 2024a). Further, in cases of few missing values, the limited variability in sRPE values may not necessarily be reflected in mechanical work, that is to say athletes may still demonstrate variable physical outputs, making it difficult to appropriately generalize mechanical work output to sRPE. However, this strategy may be effective in cases where ATD data is readily available relative to tactical data which may not always be available through film or broadcast footage for particular sports.

It has been demonstrated here and previously that imputation of sRPE is difficult and that the true collected value is vastly superior to the imputed value (Epp-Stobbe et al., 2022a). This strongly supports the need to develop strategies to minimize missing data and maximize athlete reporting adherence (Griffin et al., 2020). In their development and evaluation of a training monitoring system in rugby union athletes Griffin et al. (2020) found that there are many reasons for data missingness and perhaps the most common is the athletes perceived usefulness of the training load measure. Further, obtaining initial and continued compliance is paramount to sustained data collection. They suggest that developing appropriate educational resources for athletes as well as an effective feedback loop between the coach, and or, sport scientist is critical to create increased value in both adherence and athlete buy-in into the use of any training load measure (Bourdon et al., 2017). Nevertheless, ensuring a complete dataset in an efficient and reliable manner is key for athlete load management (Saw et al., 2015). While the results of this study have identified imputation strategies that may be applied instead of the popular daily team mean substitution method, the application of the strategy may depend on the nature of the high-performance sport environment including access to appropriate computing hardware and software, time demands, and practitioner knowledge (Coutts 2016; Yin et al., 2019; Bartlett et al., 2021; Brocherie & Beard, 2021).

5.6 Conclusion

Practically, the ability to use mechanical work or the SDC model to impute missing sRPE values for the calculation of sRPE-CL means that practitioners are not completely reliant on athlete self-

reported data to understand elements of an athlete's performance. This investigation suggests that mechanical work or the SDC model may be reasonable alternatives for the imputation of missing RPE data. However, further investigation into the ability of these models, derived from objective physical data, to appropriately reflect athletic performance as a whole, beyond the physical output is required. Ultimately, practitioners are encouraged to collect and clean data from ATDs directly wherever possible before applying any imputation methods to workload models that include meaningful and relevant factors to their sport.

6 A Comparison of the application of load monitoring metrics for key match characteristics in women's rugby sevens

6.1 Abstract

In rugby sevens, multiple high-speed matches in quick succession make effective workload monitoring essential to support decision-making around athlete preparedness and competition strategy. Match characteristics like score differential, player's competition experience, match type, and opponent may influence workload. The purpose of this investigation was to examine the relationships between match and player characteristics and three workload measures, sRPE-CL mechanical work, and an alternative Speed-Deceleration-Contact (SDC) model. Twenty-two female rugby sevens athletes were monitored across 103 international matches. Data from GNSS-derived playing times, speeds, accelerations, as well as athlete mass, and self-reported ratings of perceived exertion were collected. sRPE-CL and mechanical work were computed, and the SDC model produced predicted values. Associations between player experience, game category, opponent rank, and score differential with each workload measure were tested using ANOVAs with Tukey post-hoc. Player experience and match category were significant for all three workload measures. Opponent was significant associated with sRPE-CL and the SDC model and match outcome was only associated with sRPE-CL. All three workload measures, sRPE-CL, mechanical work, and the SDC model, are valuable but differ in response to contextual and experiential factors.

6.2 Introduction

There are numerous factors influencing a team sport athlete's workload in competition, including competitive experience, systems of play, opposition, and physical preparedness (Folgado et al., 2014; Ariol-Serrano, et al., 2021). Oliviera et al. (2021) identified higher workloads in an elite men's soccer team following losses to top-ranked teams and Gallo et al. (2014) have demonstrated that more experienced AFL players have higher training workloads than less experienced players for the same sessions. These workload differences impact player preparation and performance, making workload monitoring valuable for assessing athlete's physical preparedness to meet, or exceed, the competitive demands of the sport (Gabbett, 2016; Bourdon et al., 2017).

Common examples of workload measurement in team sports include the use of sRPE-TLand mechanical work (Haddad et al., 2017; Delaney et al., 2018). sRPE-TL, is the product of an athlete's subjective self-report sRPE and the objective measurement of duration of the activity, normally obtained from an athlete worn sensor (Haddad et al., 2017). This metric is considered a workload measure as it includes both an indicator of intensity (sRPE) and duration (time) and has relationships to other relevant measures like total and high-speed distance, and training load including training impulse (TRIMP) (Haddad et al., 2017). Mechanical work is the product of force and distance produced by athletes in their sport performance (Haddad et al., 2017; Mara et al., 2017; Delaney et al., 2018; Staunton et al., 2022). This can be further calculated by using an athlete's mass and acceleration to calculate force, and the same athlete's velocity and activity time to calculate distance. Mechanical work has been demonstrated to be associated with sRPE-CL in athletes (Epp-Stobbe et al., 2024b). Statistical and machine learning workload models have also been developed using training and competition data (Bartlett et al., 2017; Epp-Stobbe et al., 2024a). All models offer unique value in monitoring athletes in the training and competition environments yet may result in different outcomes.

Rugby sevens, like other team field sports, includes many unique factors that could contribute to an experience of workload such as elements of high-speed running, changes in speed through accelerations and decelerations, and physical contact (Schuster et al., 2018). Both high-speed running and distances covered are known drivers of sRPE (Schuster et al., 2018). Since sRPE-CL includes time, more time spent in a match can lead to greater acceleration and deceleration opportunities at various speeds to meet the demands of play. Further, accelerations and decelerations have been shown to be associated with metabolic power in men's rugby league (Delaney et al., 2016; Delaney et al., 2018). Therefore, the use of high-speed accelerations and decelerations as specific objective components may enhance the understanding of drivers of load enabling coaches and staff to appropriately manage player exposure in a tournament. Additionally, aspects of physical contact, especially in sports like rugby can potentially impact athlete load. Epp-Stobbe et al. (2024a) found that even within sRPE-CL, contact load was observed to be accounted for. This suggests that there is value in including multiple objective factors in load monitoring models.

As an alternative workload model, Epp-Stobbe et al. (2022a) used factors of acceleration/deceleration and contact as a proxy to predict sRPE-CL with a Speed-Deceleration-Contact (SDC) model (Epp-Stobbe et al., 2024a). Identifying that distances covered at low deceleration (between –2.0 and 0.0m/s^2) while at low and high speeds, as well as mass and counts -of contact explained almost half of the global variance of sRPE-CL ($R^2_{adjusted}$ = 0.487 - Epp-Stobbe et al., 2022a). This model demonstrated strong agreement with sRPE-CL using objective, external measures and presented a potential new model for athlete workload measurement in women's rugby sevens. While this model shows promise, it has not been compared to sRPE-CL or mechanical load using applied, match-specific comparisons.

In rugby sevens, workload monitoring preferences may vary based on data collection methods and emphasis on subjective or objective factors. As such, workload models in a rugby sevens tournament setting may be uniquely impacted by factors such as the opposition, the type of match (pool play vs knockout), athlete's competitive experience, and match outcome (win or loss) (Folgado et al., 2014; Ariol-Serrano et al., 2021). The effect of match-specific characteristics on athlete workload is an important consideration in workload monitoring and the choice of model. For example, the potential for the type of game to alter perceived workload (sRPE-CL) has been identified in basketball, where regular season game workloads were perceived differently than playoff game workloads (Schultz de Arruda et al., 2018; Ferioli et al., 2021). Further, the training and competition experience of athletes alters their perception of the events and execution of skills in training and competition (Gallo et al., 2014; Schultz de Arruda et al., 2018).

While there are commonalities between sRPE-CL, mechanical work and the SDC model, some differences between all models remain (Epp-Stobbe et al., 2024a; Epp-Stobbe et al., 2024b). The associations between sRPE-CL and mechanical work, as well as between sRPE-CL and the SDC model suggest that athlete load measures yield comparable interpretations regarding relative inter-game differences in athlete load. Further investigation may be warranted to examine potential nuances between load modeling approaches, particularly concerning systematic bias and effect magnitude (Epp-Stobbe et al., 2024a). Despite this preliminary investigation, there remains a lack of direct comparison between load models such as mechanical work, sRPE-CL, and the new athlete load model in their ability to distinguish athlete load variations across different match conditions. Therefore, the purpose of this investigation is to provide a comparison between three workload

models and their relationships to match-specific and player-specific characteristics in women's rugby sevens.

6.3 Materials and methods

6.3.1 Participant information

A total of twenty-two female athletes participated in this investigation (26.5 ± 4.20 years, 169.5 ± 5.90 cm, 70.5 ± 6.43 kg). All participants were members of a full-time national elite amateur centralized rugby sevens training program, providing voluntary, written, informed consent to participate. The University of Victoria Human Research Ethics Board provided ethical approval for this investigation which followed the principles described in the Declaration of Helsinki (approval code 19-0546). Furthermore, all data used by researchers were anonymized by team staff prior to analysis.

6.3.2 Procedures

103 international elite women's sevens matches were analyzed retrospectively for collected data including GNSS-coded measures, athlete mass and self-reported sRPE data, and match characteristics.

Distance, in metres, speed, in metres per second, and playing time, in minutes, were collected from GNSS monitors worn by the athletes between the shoulder blades in custom vests sampling at 10 Hz (Apex v2.50, StatSports, Newry:UK). Acceleration, in metres per second squared, were calculated based on GNSS speeds using a low, dual pass second order Butterworth filter to smooth the data before calculating the rate of change in the smoothed speeds over time (Python, v3.9.8, python.org).

Athlete mass, in kilograms, was collected using a portable weight scale before each match (ES-310, Anyload, Burnaby:CAN). One sRPE value was self-reported by each athlete per match, using a 0-10 scale, in arbitrary units, familiar to the participants from regular use in training and competition, collected within thirty minutes following each match as players returned to the team's designated area following post-match cool-down and any media obligations. sRPE-CL, in arbitrary units, values were calculated as the product of the sRPE value and playing time using Python (Python, v3.9.8, python.org).

Mechanical work, in joules, was calculated as the product of mass, acceleration, velocity, and time increments of 0.1 seconds. The overall absolute mechanical work, in joules, (W) was obtained by summing the instantaneous mechanical work across the entire game, using 0.1-second intervals (Equation 6.1). In this context, instantaneous mechanical work was derived from the product of instantaneous absolute power (P) and time (t), where instantaneous absolute power was defined as the product of the athlete's mass, acceleration, and velocity.

$$\boldsymbol{W} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} (\boldsymbol{P_i} \cdot \Delta \boldsymbol{t_i})$$

Equation 6.1: Overall absolute game mechanical work.

The SDC model, first proposed by Epp-Stobbe et al. (2024a) was used to provide predicted workload values, in arbitrary units. This model was previously found to have reasonable explanatory power of sRPE-CL and therefore provides an alternative means of evaluating athlete load. The SDC model is a regression that uses similar metrics as those necessary for the calculation of mechanical work but uses thresholds to define twelves speed and acceleration zones that are used to organize the distanced travelled by the athlete as well as the mass and count of physical contacts experienced by the athlete. The SDC model is represented in Equation 6.2 where $u_{athlete}$ denotes the random error of each athlete.

$$\begin{aligned} sRPE = {} & -0.852 + 53.87(Total\ High\ Deceleration\ Distance) + 0.159(Contact\ Count) \\ & - 53.46(High\ Speed \times High\ Deceleration\ Distance) \\ & - 26.59(Low\ Speed \times High\ Deceleration\ Distance) + u_{athlete} \pm 10.989 \end{aligned}$$

Equation 6.2: SDC model equation.

The distance travelled, in metres, was binned into three speed zones: low speeds (walking - 0m/s-1.5m/s), moderate speeds (running - 1.5m/s to each athlete's entry to sprinting threshold), and high speeds (sprinting - entry to sprinting and above) (Epp-Stobbe et al., 2024a). The use of athlete-specific entry to sprint thresholds enabled the individual physical efforts to be appropriately scaled to reflect individual physical efforts (Epp-Stobbe et al., 2024a; Gabbett et al., 2025; Cunniffe et al., 2009). After being categorized into speed zones, an acceleration threshold of 2m/s^2 was applied to further organize the distances into four zones of low (<2.0m/s^2) and high (>2.0m/s^2)

accelerations and low (<-2.0m/s^2) and high (>-2.0m/s^2) decelerations (Epp-Stobbe et al., 2024a; Mara et al., 2017; Vescovi, 2015). The significant zones for the regression were: high speed/high deceleration and low speed/high deceleration (Equation 6.2).

A summed count of contacts for each athlete was produced from match footage, evaluated using Sportscode (v11, Hudl, Lincoln: USA).

The model used the distances by zone, mass, and count of contacts per athlete to generate a workload value by a linear mixed model regression using a custom R script (R version 4.2.1, Vienna, Austria).

Key characteristics of match category, opponent, match outcome, and player experience were collected for each game. Player experience, in years playing at the senior international level, was defined as first year of play (rookie season), two-four years of play, or five or more years of play and was determined for each match based on the date of play relative to the player's debut, making all matches available for inclusion in analysis. Match category was defined by the type of match, with categories outlined in Table 6.1. In tournaments, matches where the team-of-study did not finish in the top half of their respective pool were excluded from analysis. The opponent rank groupings were determined as the top four and bottom, or lowest, four ranked teams that the team-of-study played, and the rank of match's opponent as defined by World Rugby at the start of each tournament. Matches against teams falling outside of the top or bottom ranking, within central rankings, were excluded from analysis. Match outcome was identified as either a win or loss, with ties excluded as ties were possible under the code during the period of collection.

Table 6.1: Operational definition of match categories.

Match Category	Operational Definition
Pool	Denotes games happening in the first phase of play (pool play) to determine a ranking within a small group of teams determined ahead of the tournament.
Cup Quarterfinal	First knockout stage of the tournament for teams, after pool play, whereby team finished in top half of their respective pool and advance in the tournament.
Cup Semifinal	Second knockout stage of the tournament, after quarterfinal, whereby team is guaranteed a top four finish in the tournament.
Cup Final	Final knockout stage of the tournament, after Medal Semifinal whereby team will play for a final placing, with a win meaning either a bronze medal (third place) or gold medal (first place) in the tournament.

6.3.3 Statistical analysis

A total of 1002 complete rows of data were available for analysis. Four one-way ANOVA tests were used for each of the three load measures (mechanical work, sRPE-CL, and the SDC model), one for each match characteristic (match category, opponent, match outcome, and player experience) using subsets of the complete dataset in alignment with each match characteristic. (R version 4.2.1., Vienna, Austria). Tukey's Honest Significant Difference (HSD) tests were applied post-hoc to significant effects (R version 4.2.1., Vienna, Austria). ANOVA tests were selected over a mixed model approach for the efficiency of group comparison and the focus was on the fixed effects of the workload and match- and player-specific characteristics. A post-hoc power analysis was conducted based on our smallest match characteristic sample size (N=345) and on power calculations for the main effects. Assuming a large effect size of f=0.40 for main effects resulted in a power level of 0.99.

6.4 Results

On average, athletes played for 11.8 ± 4.53 minutes, experienced an sRPE of 7 ± 1.9au, and subsequently a calculated sRPE-CL of 79.6 ± 45.59 au, 56.25 ± 21.40 kJ of mechanical work per match played.

6.4.1 Athlete workload measures and match category

Of the matches analyzed (N=890), 60.90% (N=542) were pool play, 12.58% (N=112) were cup quarterfinals, 13.71% (N=122) were cup semifinals, and 12.81% were cup finals. The greater incidence of cup semifinals to cup quarterfinals stem from differences in tournament structures by hosted event where some tournaments moved from pool play to quarterfinals and others moved from pool play directly to semifinals. There was a main effect for match category for mechanical work ($F(3, 886) = 10.12$, $p < 0.05$), sRPE-CL ($F(3, 886) = 10.12$, $p < 0.05$), and the SDC model ($F(3, 886) = 8.79$, $p < 0.05$). Post-hoc Tukey tests identified significant differences in workload from pool play and cup quarterfinals and cup finals for mechanical work (pool play-cup quarterfinals mean difference = -6512.51, 95% CI [-12168.26, -856.76], p <0.05, d=-0.31 indicating a small-to-moderate effect; pool play-cup finals mean difference = -6627.18, 95% CI[-12222.40, -1033.96], p <0.05, d=-0.31 indicating a small-to-moderate effect), sRPE-CL (pool play-cup quarterfinals mean difference = -15.31, 95% CI [-26.87, -3.75], p <0.05, d=-0.36 indicating a small-to-moderate effect; pool play-cup finals mean difference = -21.09, 95% CI[-32.52, -9.66], p <0.05, d=-0.50, indicating a moderate effect), and the SDC model (pool play-cup quarterfinals mean difference = -22665.50, 95% CI [-40647.23, -4683.77], p <0.05, d=-0.06 indicating a trivial effect; pool play-cup finals mean difference = -25629.96, 95% CI[-39629.52, -11830.39], p <0.05, d=-0.45 indicating a moderate effect)(Figure 6.1). Further differences between pool play and cup semifinals were present in sRPE-CL (mean difference = -10.22, 95% CI[-21.34, 0.90], $p < 0.05$, d=-0.24 indicating a small effect), and between cup quarterfinals and cup semifinals in the SDC model (mean difference = -30380.84, 95% CI[-47950.51, -12811.18], $p <$ 0.05, d=-0.21 indicating a small effect) were identified (Figure 6.1).

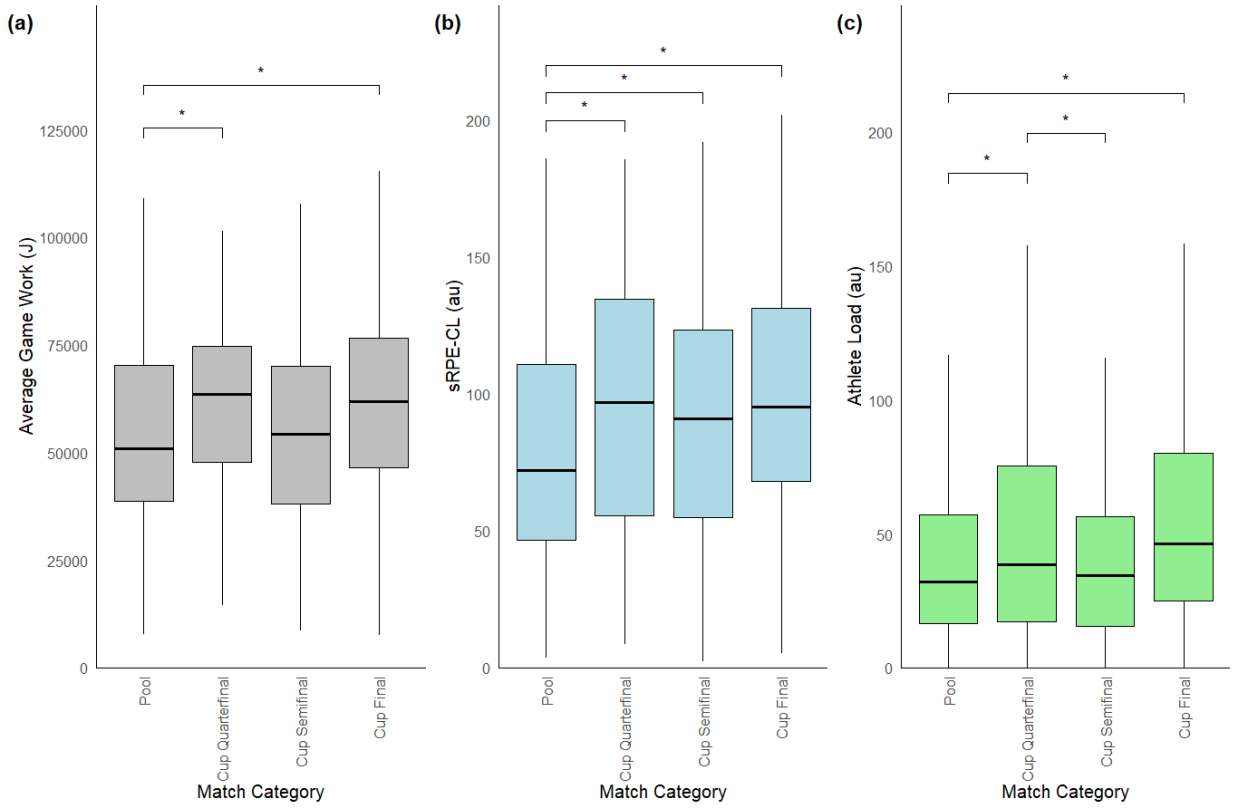

Figure 6.1: (a) Average game work, (b) sRPE-CL, (c) SDC model, workload measures by match category, * denotes significant comparison ($p < 0.05$).

6.4.2 Athlete workload measures and opponent

Of the matches analyzed (N=345), 69.57% (N=240) were against a top four ranked team and 30.43% (N=105) were against a bottom four ranked team. Team rankings were determined before each tournament based on World Rugby standings. There was a main effect for opponent for sRPE-CL ($F(1, 343) = 12.66$, $p < 0.05$) and the SDC model ($F(1, 343) = 4.58$, $p < 0.05$). There was not a significant main effect of opponent for mechanical work ($F(1, 343) = 0.04$, $p > 0.05$) (Figure 6.2). Post-hoc Tukey tests identified significant differences in workload from matches against top four ranked teams and bottom four ranked teams for both sRPE-CL (mean difference = 18.31, 95% CI[8.41, 28,20], $p < 0.05$, d=0.43 indicating a moderate effect) and the SDC model (mean difference = 11215.46, 95% CI[1490.05, 20940.87], $p < 0.05$, d=0.27 indicating a moderate effect) (Figure 6.2).

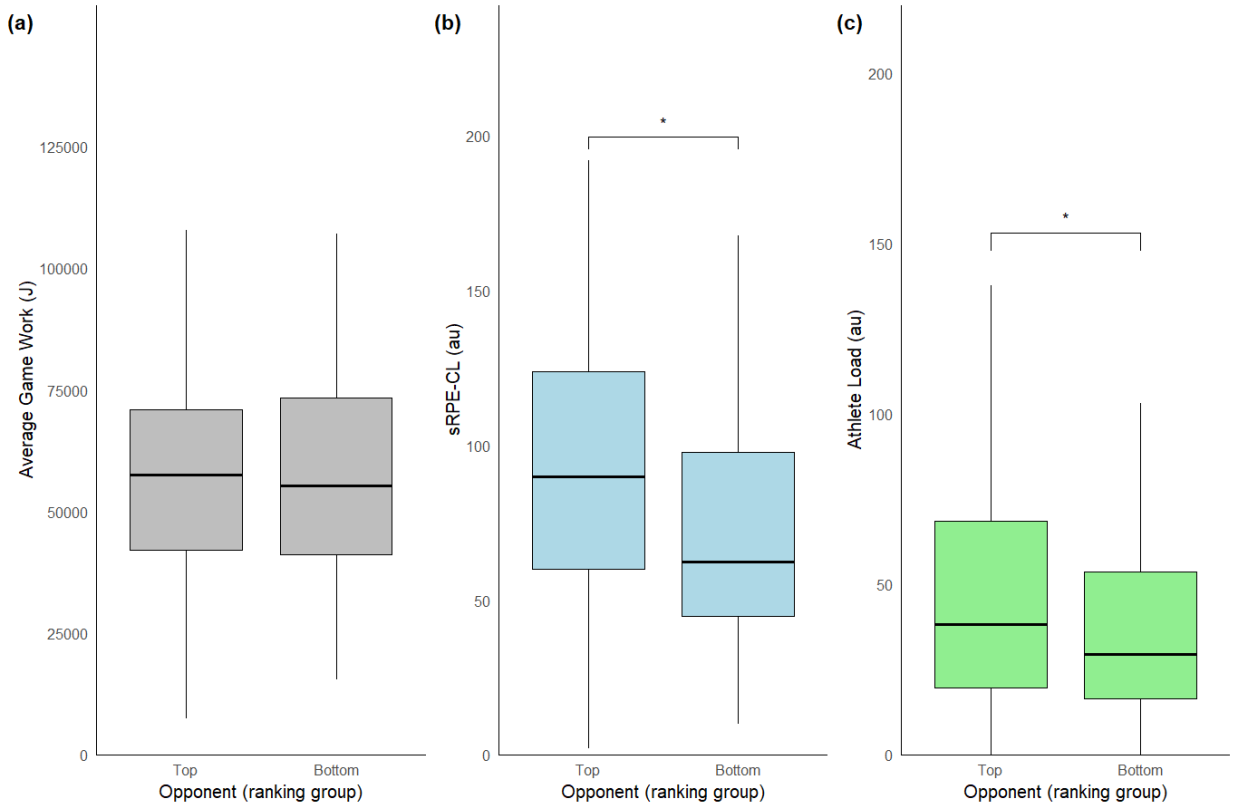

Figure 6.2: (a) Average game work, (b) sRPE-CL, (c) SDC model, workload measures by opponent, * denotes significant comparison ($p < 0.05$).

6.4.3 Athlete workload measures and match outcome

Of the matches analyzed (N=992), 67.94% (N=674) were wins and 32.06% (N=318) losses. There was a main effect for match outcome for sRPE-CL ($F(1, 990) = 6.90, p < 0.05$) and the SDC model ($F(1, 343) = 4.58, p < 0.05$). There was not a significant main effect of match outcome for mechanical work ($F(1, 990) = 1.25, p > 0.05$) or the SDC model ($F(1, 990) = 0.85, p > 0.05$) (Figure 6.3). A post-hoc Tukey test identified that workloads from wins were significantly different from losses (mean difference = -7.94, 95% CI[-13.73, -2.14], $p < 0.05$, d=-0.18 indicating a small effect) (Figure 6.3).

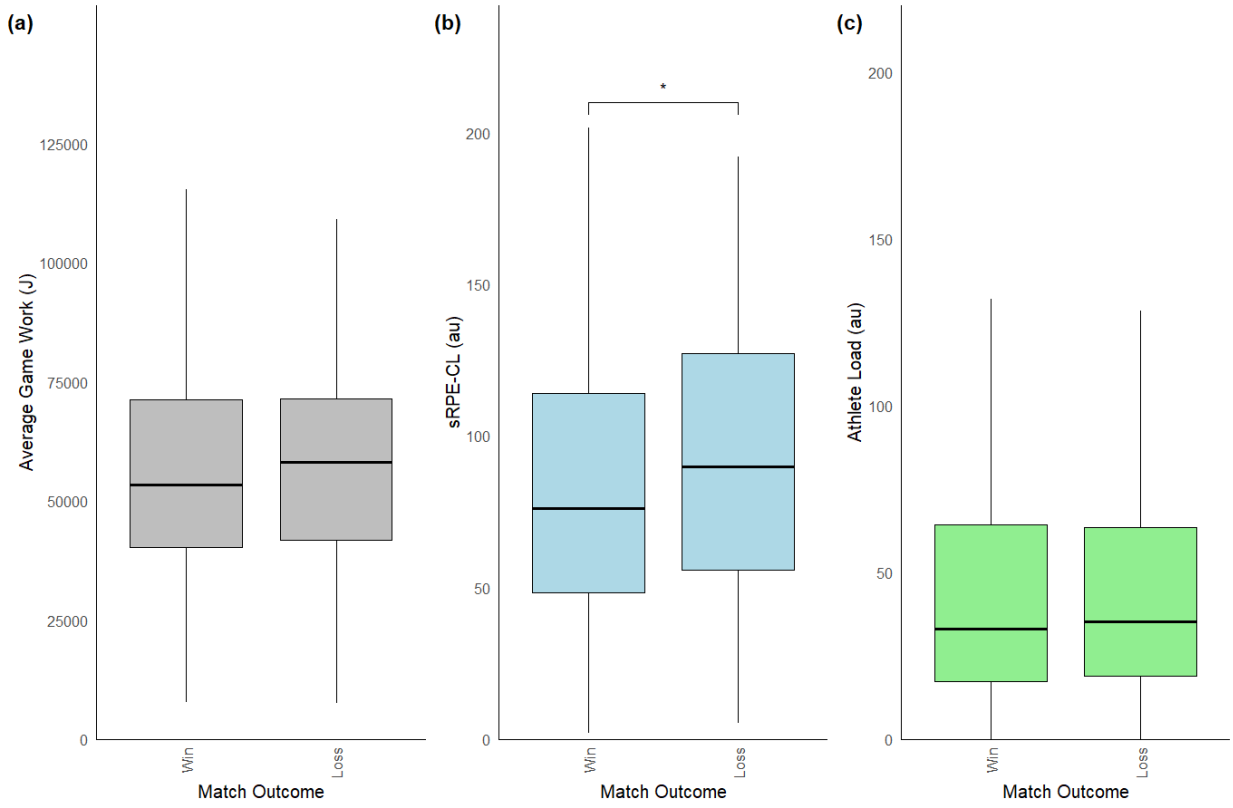

Figure 6.3: (a) Average game work, (b) sRPE-CL, (c) SDC model, workload measures by match outcome, * denotes significant comparison ($p < 0.05$).

6.4.4 Athlete workload measures and player experience

Of the matches analyzed (N=1002), 3.59% (N=36) were played with athletes in their first year on the field, 45.51% (N=456) with athletes in their second to fourth years, and 50.90% (N=510) with athletes in their fifth year or greater of international competition experience. There was a main effect for player experience for mechanical work ($F(2, 999) = 23.81$, $p < 0.05$), sRPE-CL ($F(2, 999) = 28.48$, $p < 0.05$) and the SDC model ($F(2, 999) = 41.10$, $p < 0.05$). Post-hoc Tukey tests identified significant differences in workload from all three player experience groups for mechanical work (Year 1-Year 2-4 mean difference = 16121.72, 95% CI[7615.11, 24628.33], $p < 0.05$, d=0.78 indicating a large effect; Year 1-Year 5+ mean difference = 21991.25, 95% CI[13519.01, 30463.49], $p < 0.05$, d=1.05 indicating a very large effect; Year 2-4-Year 5+ mean difference = 5869.53, 95% CI[2713.09, 9025.97], $p < 0.05$, d=0.28 indicating a small effect) and sRPE-CL (Year 1-Year 2-4 mean difference = 18.14, 95% CI[0.90, 35.38], $p < 0.05$, d=0.45 indicating a moderate effect; Year 1-Year 5+ mean difference = 36.19, 95% CI[19.02, 53.36], $p < 0.05$, d=0.83 indicating a large effect; Year 2-4-Year 5+ mean difference = 18.04, 95% CI[11.65,

24.44], $p < 0.05$, d=0.52 indicating a moderate effect) (Figure 6.4). In the SDC model significant differences were present between the Year 1 and Year 5+ player experience groups (mean difference = 46026.03, 95% CI[25395.47, 66656.58], $p < 0.05$, d=-0.73 indicating a moderate-to-large effect) and the Year 2-4 and Year 5+ player experience groups (mean difference = 26967.84, 95% CI[19281.67, 34654.01], $p < 0.05$, d=0.52 indicating a moderate effect) (Figure 6.4).

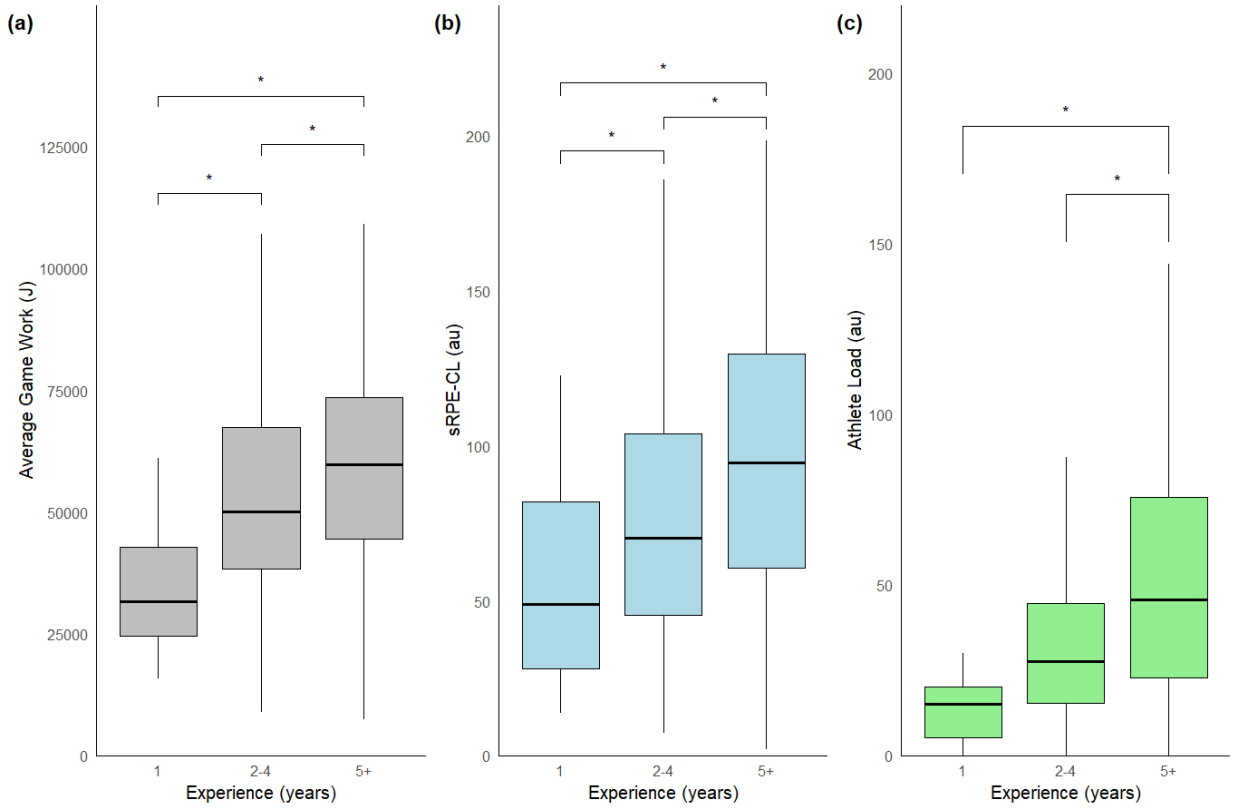

Figure 6.4: (a) Average game work, (b) sRPE-CL, (c) SDC model, workload measures by player experience, * denotes significant comparison ($p < 0.05$).

6.5 Discussion

This investigation was the first to compare associations between three different workload measures for differences in match characteristics in a national women's rugby sevens cohort. sRPE-CL detected differences for all levels examined across all match characteristics. Mechanical work showed limited differences for match category, detected differences for player experience, like sRPE-CL, and showed no differences for opponent or match outcome. The SDC model detected differences for opponent, like sRPE-CL, showed no differences for match outcome, like mechanical work, and demonstrated unique differences across match category. This demonstrates that the model chosen will influence the determination of athlete workload. While mechanical

work objectively quantifies the athlete's external workload it may not account for the internal effect of play. Alternatively, while sRPE-CL may account for perception of exertion, it may not objectively represent the true internal and external workload. Further, the SDC model benefits from comparable associations to sRPE-CL yet includes unique objective measures, positioning this model as valuable for consideration.

6.5.1 Athlete workload measures and match category

Match category exhibited significant associations with all three workload measures. Interestingly the post-hoc analysis highlighted slight nuances between the measures where pool play differed from cup quarterfinals and cup finals across average game work, sRPE-CL, and the SDC model, however, pool play also differed from cup semifinal workload using sRPE-CL and cup semifinal and cup quarterfinal workloads differed using the SDC model. Vescovi (2015) suggested that female soccer players experienced higher speeds and covered greater distances in playoff games compared to regular season games. This suggests that games of critical importance, such as playoffs, which would be similar to cup quarterfinal, cup semifinal, and cup final matches in this dataset, would potentially see players experiencing greater speeds, accelerations, and subsequently greater mechanical work measures. Research in men's basketball has shown increased sRPE-CL in playoffs compared to regular (pool) play and finals compared to semifinals, suggesting that the athletes' perceptions of the value or importance of the match drives increases in sRPE, and subsequently sRPE-CL (Schultz de Arruda et al., 2019; Ferioli et al., 2021). Further, Schultz de Arruda et al., (2019) concluded that the observed increases in workload may be explained by the psychological aspects associated with the competition environment, which is to say that the athletes' identified value of playing in a medal final elicits the perception of increased effort (Schultz de Arruda et al., 2019). Given the differences in psychological perception of the importance of the match, or game category, it is possible to infer that the physical output changes with the category of the game (Schultz de Arruda et al., 2019). For example, it is possible that in the medal knockout rounds, the quarter, semi, and final, players may have a higher physical output as they recognize the psychological importance of the match (Schultz de Arruda et al., 2019; Ferioli et al., 2021). This is consistent with the observed results where sRPE-CL shows most differences between match category. It is possible that any psychological perception of load is best expressed using this measure. Additionally, when considering both mechanical work and the SDC model, this psychological effect is not overtly accounted for and this shows the unique value of all

athlete workload measures, to either objectively evaluate the external measured work with (sRPE) or without (mechanical work, SDC model) psychological perception of effort.

6.5.2 Athlete workload measures and opponent

Opponent rank group demonstrated significant associations with sRPE and SDC model workloads, however, not with mechanical work. When considering opposition, Oliviera et al. (2021), noted lower sRPE-CL values in elite male soccer players when the team-of-study won against a higher ranked opponent (Oliviera et al., 2021). The work of Barrett et al., (2018) supports this assertion having identified higher sRPE scores when elite male soccer players faced top-ranked teams, ranked above the team-of-study, suggesting that the opponent rank may influence perceived effort and required further investigation. Recalling that sRPE-CL is the product of playing time and an athlete's self-reported sRPE, the significance of opposition may suggest some level of subjectively perceived difference in effort identified by the athlete (Haddad et al., 2017). However, this association may not necessarily be driven by the match characteristic measures, rather by the sRPE component of sRPE-CL. Nevertheless, the significant association present with the SDC model in this case suggests that there is some difference in physical output when facing top-ranked teams, however, not one necessarily driven by differences in speeds or accelerations which would be reflected in mechanical work. Instead, tactical differences in the style of play required to face highly skilled teams may result in changes in workload. For example, since the SDC model includes a count of physical contacts, it may suggest that this factor is a key driver in the difference in workloads from facing top versus bottom ranked teams. Physical contact has known associations with sRPE as well as sRPE-CL in women's sevens (Epp-Stobbe et al., 2022b; Epp-Stobbe et al., 2024a).

6.5.3 Athlete workload measures and match outcome

Match outcome only demonstrated significant associations with sRPE-CL. The work of Oliviera et al., (2021) and Fessi and Moalla (2018) support this, having identified higher RPE-CL values reported after losses compared to wins or draws in elite male soccer players. A lower sRPE-CL after a win compared to a loss suggests that athletes identify losses as being more effortful and would report a higher sRPE. However, the more objective workload measures of mechanical work and the SDC model were not significant, suggesting that physical outputs were not different in each case. This is potentially different from the work of Gabbett (2013) who identified greater distances covered when rugby league matches were won, which would potentially drive higher

workloads. In contrast, Hills et al. (2024) identified differences in locomotor demands where wins resulted higher in periods of physical output for both men's and women's rugby sevens. Ultimately, both the research of Gabbett (2013) and Hills et al. (2024) recognize that there are a variety of factors that affect physical activity within a match, therefore match workloads, regardless of outcome.

6.5.4 Athlete workload measures and player experience

Player experience demonstrated significant associations with all three workload measures. Again, the post-hoc analysis emphasized slight differences between the measures where significant differences existed between all three experience groups (Year 1, Years 2-4, Years 5+) for mechanical work and sRPE-CL but only between Year 1 and 5+ and Years 2-4 and 5+ in the SDC model. In all cases, the most experienced group of players demonstrated the highest workload, an important consideration for team staff as they plan substitutions and strategies across a tournament. The higher sRPE-CL experienced by more experienced players may point to increased playing time but also to the increased reliance on these players relative to rookies in their first year, inferring that players perceive an increased load due to their increased presence in multiple matches across a tournament. Gallo et al. (2014), found similar results in men's AFL training, hypothesizing that less experienced players may be less involved in executing skills or drills due to unnecessary and inefficient movements. Applying this logic to a game scenario, it would make sense that a less experienced player has a lower game acumen to participate as effectively as their more experienced counterparts, leading them to perceive they participated less in the match and report a lower RPE, and subsequently a lower sRPE-CL. In the same case, less effective involvement in matches may see less experienced athletes cover less ground and at different speeds and accelerations than seasoned veterans, leading to lower mechanical work and lower workload values from the SDC model. Again, the work of Gallo et al. (2014), in men's AFL training, may apply whereby inexperienced players may struggle to make the appropriate movements to complete a skill, drill, or in-game, making them less involved in plays and therefore experiencing less mechanical work. Conversely, players with high levels of experience have an improved understanding of the game and can participate more effectively in active play, increasing their mechanical work (Becirovic et al., 2024).

6.5.5 Considerations for use of workload measures

When considering the statistical differences between workloads across match characteristics it is apparent that sRPE-CL has more differences compared to the SDC model and mechanical work. Observed differences for all four key match characteristics investigated are associated with sRPE-CL, three of the four are associated with the SDC model, and only two are associated with mechanical work.

The investigation has demonstrated that an athlete's perception of their performance effort, through sRPE-CL, is moderated by match characteristics. This suggests that the SDC model may represent an objective alternative to relying on athlete-reported values for the calculation of sRPE-CL alone (Epp-Stobbe et al., 2024a). The value of objective data in place of athlete-reported values means that the usual problem of missing data when using survey-based measures is avoided, ensuring continuity and completeness of data over time (Griffin et al., 2021).

Despite this study being the first to investigate the associations between mechanical work and key match characteristics in women's rugby sevens, mechanical work was only influenced by the key match characteristics of player experience and game category. This may be due to an athlete's participation in the match stemming from the skills, knowledge, and effortful actions (Gallo et al., 2014; Becirovic et al., 2024). This is a reasonable conclusion as no matter the opponent or match outcome, an athlete is still exerting themselves physically on the field to make the most optimal plays for the team in the moment. However, it is also entirely possible that since mechanical work is a biomechanical expression of the physical output of a player, representative of external load, determined by inferring internal forces that cannot be measured in situ, it is not fully representative of all stresses experienced by players, accounting for some differences in its relationship with match characteristics. When considering rugby-specific stresses the physical contacts experienced in matches, already known to influence sRPE-CL, may also influence general physical output (Epp-Stobbe et al., 2022a; Epp-Stobbe et al., 2022b; Epp-Stobbe et al., 2024a). The SDC model identified both contact and deceleration as variables that influence workload and as such it is possible that contacts and decelerations are drivers of the differences in workload model by match characteristic. For instance, Hills et al., (2024) found periods of higher physical output by winning teams, suggesting that teams may be moving more and as such exposed to a wider variety of physical movements including contact in tackles and carries, as well as changes of direction

requiring deceleration. Irrespective of outcome, mechanical work, as a workload measure, quantifies the physical effort of the athlete more similarly to the SDC model than sRPE-CL.

Interestingly, the type of game being played may influence workload irrespective of opponent and opponent rank differential. This suggests that the level of game, quarterfinal, medal vs ranking final, etc. is more impactful on athlete workload than the opponent. This points to some interaction between the match category, outcome, and opponent. While the current investigation leveraged ANOVA tests to evaluate the presence of any significant interactions between these categorical variables, future research should consider means of quantifying the effects of these interactions. This is an important consideration in team sports given the bracketed nature of knockout or playoff rounds whereby teams often want to achieve a certain standing in pool play to avoid facing certain highly ranked or similarly ranked opponents in early playoff rounds (Guyon, 2022).

6.6 Limitations

This study only includes participants from women's rugby sevens therefore direct application of the findings may be limited. Further, However, the elite nature of international competition may mean that match characteristics may not necessarily be applied in the same way to different competition levels. Practitioners are encouraged to consider how these findings may or may not be applied in their respective environments.

As a future direction, considerations between the player experience and match-specific characteristics should be considered. It is entirely possible that experienced players may express different physical output, as well as subjectively perceive differences based on opponent and match category which would impact any of the proposed workload measures from this investigation.

6.7 Practical implications

These findings underscore the complexity of interpreting workload data in competitive sports settings, where broader interaction effects may not always translate into statistically significant differences in measures. The influence of match-specific and player-specific characteristics should be considered when monitoring athletic performances in competition as subjective perception may not match physical performance (Saw et al., 2016; Coyne et al., 2018). Practically, the findings support the application of more than one strategy when monitoring athletes (Bourdon et al., 2017; Gabbett et al., 2017). A summary for the application of the workload models for practitioners

appears in Table 6.2. For example, the use of mechanical work may be coupled with either sRPE-CL or the SDC model as a means of gathering insight into the physical output of the athlete as well as match-specific experiences such as those from contacts or perceived exertion. Alternatively, where sRPE-CL is already in use, the inclusion of mechanical work is complimentary as it provides additional context to athletic performance through objective measures physical output. While it is acknowledged that athlete performance in team sports is multifaceted, a significant gap exists between perceived and objective physical performance which may potentially be bridged through the consideration of how match and player characteristics relate to existing workload monitoring strategies (West et al., 2021). The value of a variety of measures to accurately quantify an athlete's sporting experience and physical effort is strongly encouraged (Bourdon et al., 2017; Gabbett et al., 2017). The combination of objective measures, such as mechanical work or the SDC model, with measures that encompass psychological elements, like sRPE-CL, provides a well-rounded understanding of the athletic experience.

Table 6.2: Recommendations for athlete workload model use.

Athlete Workload Model	Application
Mechanical Work	An objective model; useful for reflecting actual physical output, may not reflect perceptions of effort or stressors in the competition environment. Consider using in combination with sRPE-CL where athlete perceptions are useful to inform the experience but may be used as a standalone. Does not need to be combined with SDC model.
SDC	An objective model; useful for reflecting actual physical output, may not reflect perceptions of effort or stressors in the competition environment. Consider using in combination with sRPE-CL where athlete perceptions are useful to inform the experience but may be used as a standalone. Does not need to be combined with mechanical work model.
sRPE-CL	A subjective model; useful for reflecting perceptions of effort or stressors in the competition environment, may not reflect actual physical effort. Consider using in combination with either the mechanical work or SDC models for well-rounded understanding of experience. Not recommended for standalone use.

6.8 Conclusions

All three workload measures, sRPE-CL, mechanical work, and the SDC model, remain viable workload monitoring tools. The use of sRPE-CL or the SDC model to monitor workloads may be more sensitive to key match characteristics than mechanical work. Given, the potential interaction between match characteristics such as opponent, score difference, and type of match, and their influence on workload, practitioners are encouraged to consider the implications of the workload tools in their respective applied environments.

7 Conclusion

7.1 Summary of dissertation

Many strategies exist to monitor athlete workload, as a means to objectively quantify athletic workload. In the challenging team sport environment, constraints on monitoring tools and the feasibility of techniques must be balanced with the quality of the measure. This dissertation enforces the consensus that athlete monitoring cannot, and should not, be compressed to a single metric. Rather, this dissertation encourages the inclusion of objective factors that are clearly relevant to the sport demands in monitoring athletic workload.

Chapter two highlights the challenges in imputing sRPE as a standalone measure from objective data available in GNSS units, calling into question the utility of sRPE as a contributing factor of the popular sRPE-TL workload monitoring metric. This chapter also challenges the current approach of using daily team mean as an imputation strategy. Chapter three identifies the influence of contact, uniquely present in rugby in women's elite amateur sport, on sRPE as well as the interplay of playing time and contact on sRPE, suggesting a moderating relationship between the two; and that the weighting of individual components may depend on the levels of other factors experienced by the athlete. Chapter four produces a novel model, as an alternative monitoring strategy based on objective measures relevant to rugby sevens. Chapter five provides a comparison of imputation methods for sRPE using objective strategies: the SDC model and mechanical work, as a potential approach to improve sRPE imputation. The outcome of this chapter further demonstrates that sRPE is difficulty to impute, even from theoretical and experimentally determined objective workload measure. This chapter also further challenges the utility of daily team mean substitution for imputation of missing sRPE values given consistently poor accuracy. Finally, chapter six systematically evaluates sRPE-CL, the SDC model, and mechanical work under novel match- and player-specific conditions to enhance the understanding of the behaviour of each model. Each model behaves slightly differently; therefore, chapter six demonstrates the importance of model selection and encourages the use of an objective measure as well as the addition of another strategy; enforcing that monitoring athlete workload is not limited to single

measures. Further, this outcome supports the use of the SDC model or mechanical work as a robust, objective alternative workload strategies.

Ultimately, this dissertation determines relevant objective factors that could be used to appropriately quantify physical athlete workload in women's rugby sevens. This dissertation adds to the existing body of knowledge in this space by providing alternative, object options for monitoring that are both robust and feasible for use in women's rugby sevens and the broader team-sport environment.

7.2 Practical applications

In alignment with the consensus statement from Bourdon et al., (2017), this dissertation encourages the use of multiple methods of athlete monitoring. This is because, as shown in chapter two and five, sRPE cannot be reliably imputed from objective data in cases where athlete self-reported values are unavailable. Therefore, practitioners are recommended to consider the use of one objective method, such as mechanical work or the SDC model, that compliments the use of sRPE-CL, as shown in chapters and six. Chapters three, four, five, and six demonstrate this need by identifying other unique drivers of physical athlete workload such as contact, speeds, decelerations, distance travelled, mass, and playing time. Further, chapter six highlights that there are match-specific and player-specific characteristics which affect athlete workload such as playing experience, match category, match outcome, and opponent rank. Taken together it may appear there are a large number of variables to consider, however, practitioners are encouraged to consider key measures that can be influenced through training and preparation. In this way, the speed, and subsequent acceleration and/or deceleration, present in the SDC model, and in mechanical workpresent "trainable" metrics. Exposing athletes to high speeds as well as decelerations has been shown to offer a protective effect against injury and may also enhance resilience to high workloads in competitive settings (Edouard et al., 2019; Gabbett & Ullah, 2012).

In the competition setting, efficient, feasible strategies to gather information about the athlete's experiences remain of the utmost importance – both from an intra-tournament strategic perspective as well as from an inter-tournament recovery and preparedness perspective. The ability to reliably impute RPE, and subsequently calculate sRPE-CL, by a variety of objective models, as shown in

chapter five, highlights the high value of the incorporation of relevant objective metrics in a workload monitoring strategy. While it is possible to impute sRPE such as to mitigate the presence of missing data, (Chapter five), the athlete's perspective, and perhaps sRPE, may shift with increased playing experience, by match outcome, category, and opponent (Chapter six). The imputation of sRPE can take a variety of forms, although practitioners are encouraged to use the SDC model for the imputation of sRPE to be used in the calculation of sRPE-CL, as identified in chapter five.

7.3 Limitations

There are several limitations in the interpretation and application of these studies. From the overall dataset, some key considerations include the fact that the data was limited to data from one team of athletes. As such, the athlete characteristics and training background were shaped by the demands of the sport and the elite amateur, full-time daily training environment, as such, the results may not be generalizable to other sports, across sexes, across playing levels (collegiate, elite amateur, professional, recreational, etc.) or even across teams. A larger dataset of players across multiple teams and/or multiple playing levels would allow for a more enhanced assessment of objective metrics that influence physical athlete workload.

While the limitations of GNSS devices in team sport environments are known, there are some considerations to the processing of this particular dataset which must be acknowledged (Carey et al., 2016; Jennings et al., 2010; Rampinini et al., 2015). Using the GNSS data to obtain speed, the speed data was filtered using a dual pass, low frequency second order Butterworth filter with a 2Hz cutoff. This is a relatively mild low pass filter, however the 2Hz cutoff applied to a 10Hz signal may risk over-smoothing and blurring events which may lead to diminishing speeds and subsequent accelerations, decelerations, and distances travelled used in analysis. In chapter four, speed data in the highest speed zone and high acceleration and deceleration zones are most at risk for over-smoothing. However, considering the results of chapter four, the overfitting risks may not have blunted the speed zone response to a significant effect. Additionally, the frequency of high speed, high acceleration, or high deceleration events during rugby sevens matches are relatively infrequent compared to movement at lower speeds, accelerations, or decelerations.

Chapter two uses sRPE values from competition, which are by nature, a limited range of possible values. This limited range, with the bulk of values across sRPE scores 6, 7, and 8, may potentially inflate the accuracy of all models with fewer possible outcomes available. However, the limited variability may also inhibit some models, by making it harder to discern patterns or clusters and identify natural variability across the dataset. This reduction in discriminant power may explain why the neural network model behaved relatively poorly to other models – the concentrated range of sRPE values may have produced smaller or lower weight magnitudes to features, leading to slower overall learning as the model cannot easily differentiate between outputs. Carey et al., (2016) used a neural network to impute sRPE data in professional Australian football players whereby the neural network was one of the top performing models used, along with random forest. It is possible that since the work of Carey et al., (2016) used data from training sessions, a greater variance in the data enabled improved model outcomes.

Chapter three identifies the potential for variables with overlapping influence within sRPE: playing time and contact. This instance of multicollinearity, where both variables are featured within sRPE and where one variable (contact) influences the relationship of another (playing time) to sRPE, suggests that it may be difficult to clearly identify the magnitude of contributing variables.

Chapter four highlights potential overlap in variables between contact and deceleration distance in the model as well as considerations for mathematical coupling of playing time and sRPE-CP. By accounting for contacts within the deceleration distance zones, the role of physical contact may be artificially inflated. This same phenomenon may also occur given the presence of playing time.

Chapter five follows a similar methodology as chapter two, however, in this case only three models were used – the top performing random forest, linear regression, and the daily team mean substitution. Accurate imputation of sRPE may similarly be hampered by limited variation from either the SDC model predicted workload, or mechanical work given the data comes from competition settings. This may lead to cases of overfitting, particularly where playing time and random forest models are considered. Since random forest classifiers work to reduce variance, in cases of low variance they may over-fit leading to poorer generalization across populations. In this case it may be worthwhile to consider applying regularization penalties such as through lasso,

ridge, or elastic net regression to constraint the complexity, ensuring appropriate variable selection and coefficient shrinkage.

Chapter six uses ANOVA tests, however, in two cases the tests behave identically to t-tests given the inclusion of only two groups (match outcome – win or loss, and opponent ranking – top or bottom ranked teams). Further, this chapter does not explore interactions between match- and player-specific characteristics. Meant as a first exploration of these associations, the interactions that may be present are certainly plausible. Anecdotally, given the tournament structure, it is possible that only top-ranked teams face each other at particular match categories (such as cup quarterfinals, cup semifinals, or cup finals, not in pool play) or that certain match outcomes exist based on certain opponents (wins are more likely against teams ranked below the team of study or ranked at the bottom of the group compared to teams ranked above or in the top of the group). In this way, interactions would enhance the interpretation of the association of effects whereby the impact of the main effect may be moderated by other variables. This would further enhance the potential for predictive accuracy. The use of linear regression in place of ANOVAs would further enhance predictive accuracy as ANOVAs are only able to identify differences in group means, not relationships between variables, including interaction cases. Further, additional investigation into the minimum data required to implement the proposed objective models, both mechanical work and the SDC are needed to ensure that team sports practitioners are able to seamlessly adopt and integrate these strategies into their sport environments.

7.4 Future directions

Identifying and understanding the behaviour of unique, objective factors that influence athlete workload is a popular area of sport research. However, in team sports, and in particular contact sports and women's team sport, the field remains relatively untouched compared to the depth of knowledge from endurance, time-based, individual sport. A key area of investigation remains generalizability across sporting groups; these studies looked exclusively at women's rugby sevens athletes however inclusion of data from other team sports and those athletes may yield enhanced variability in the data, leading to more robust models and an improved understanding of factors affecting athlete workload in team sport environments. Further, the inclusion of a more varied

rugby sevens dataset to encompass both competition and training would support more robust model development and metric identification, countering the challenges identified in chapters two and six.

In the investigation into and identification of relevant objective metrics and subsequent development of objective models, additional considerations for model development including fuzzy clustering and Bayesian approaches may be worthwhile. Since fuzzy clustering allows for data points to belong to multiple clusters, this may help adjust for overlap that exists in the interpretation of sRPE (Chapters two and five) and subsequent calculated sRPE-TL or sRPE-CL values (Schmitt et al., 2015; Rahman & Islam, 2016). Bayesian approaches, while certainly computationally powerful, may not be the most feasible in the applied environment, however, the incorporation of prior information used to updated the probability of an event offer significantly improved predictive capabilities with respect to attempts to predict future performance capacity of athletes, such as coaches looking to determine tactical strategies for knockout rounds of a rugby sevens tournament (Schmitt et al., 2015; Rahman & Islam, 2016). In fact, Bayesian fuzzy clustering combines approaches to offer a probabilistic interpretation of cluster membership of data, providing an improved understanding of the nuances of the data and interactions (Glenn et al., 2014).

Chapter five investigated the imputation of sRPE through objective models used to calculate sRPE. While imputed sRPE data demonstrated poor accuracy, similarly to the outcome of chapter two; similarly suggesting that future investigations should proceed with the knowledge that sRPE may be biased by a variety of diverse skills, tactical strategies, and team plays used in competition, such is the case in team sports and shown in chapter six (Baker et al., 2015).

To that end, it is of the utmost importance that future work includes a focus on identifying and describing the interaction between objective measures, seeking ways to mitigate cases of multicollinearity wherever possible. Cases of both physical and match- or player-specific characteristics should be included for investigation and research be included, as identified in chapters three, four, and five, should extend to psychological drivers of load. This is especially important given that the athlete's psychological perception of their effort may not necessarily be

in alignment with their physical output. By seeking to understand all facets of the athlete's experience it is possible to support the athlete in the pursuit of improvement and success in sport.

7.4.1 Considerations for future sRPE data collection

While the utility of sRPE for quantifying athlete load is well-established, collection protocols can significantly influence data quality and subsequent interpretations. The following recommendations address key methodological considerations that warrant attention in future research and applied practice relative to the collection strategy used in this dissertation.

7.4.1.1 Recommendations for sRPE data collection

The sRPE data collection from this dissertation were collected as athletes returned to the team's designated space within the competition grounds, within approximately 30 minutes post-match using the Borg CR-10 scale. This scale, with verbal, and in 2017 additional visual, anchors was displayed on a clipboard with a data table of athlete names and match numbers. Upon returning to the team space, athletes were asked by a staff to report their sRPE either verbally, by pointing to the correct value on the scale, or by writing it down themselves in the table. The verbal prompt used was "how did you find that game". In most cases the staff member collecting the data was either the sport scientist or the team manager, based on staff travel assignments, and the information for collection was included on the reporting page (including the question for prompting). Coaches did not collect the information from players.

Nevertheless, the sRPE collection protocol presents some methodological concerns that may compromise data validity and reliability. Cognitive interference from post-competition activities, including conversations with teammates on the walk back to the team space may contribute to drift in sRPE scoring (Pageaux, 2016). Additionally, the use of multiple data collectors without introduces inter-rater variability that can systematically bias results. The shared clipboard approach, despite attempts at confidentiality and ensuring responses were athlete-initiated and not prompted by teammates, etc., may create a social desirability bias where athletes may anchor their responses to previously observed ratings or modify their reports based on peer influence (Eston, 2012). This is particularly problematic when collecting sRPE data, as the construct should reflect individual internal load perception rather than socially influenced responses.

7.4.1.2 sRPE construct validity and measurement consistency

To ensure that sRPE data accurately reflects the underlying construct of training intensity and internal load, several modifications are recommended for future investigation in this space. The specific wording used when collecting sRPE data is critical for ensuring valid and reliable measurements, as subtle variations in question phrasing can significantly alter athlete responses and compromise data integrity. For sRPE collection, the recommended standardized question is: "How was your workout?" followed by a 15-30 second pause to allow cognitive processing, then "On a scale of 0 to 10, how hard was your entire session?" (Foster et al., 2001). It may be possible to adjust the phrasing of “workout” to “game”, however, it may be that differences in phrasing may limit comparisons between training and competition sRPE-based information. In any case, phrasing emphasizing the global, integrated nature of the exercise experience rather than focusing on peak intensity moments or specific exercise components is key to ensure consistency. Data collectors should avoid leading questions or comparative language such as "Was that harder than the last game?" or "That looked tough, what would you rate it?" as these introduce external reference points that can bias responses away from the athlete's internal perception (Eston, 2012).

The scale should be administered individually in a private setting to eliminate social influence, with athletes providing responses either digitally or on individual forms. Reducing the collection window to within 10-15 minutes post-competition when feasible may better capture the integrated exercise intensity while minimizing recall decay or interfering factors, potentially asking athletes to report as they move from the field-of-play to the cool-down space ahead of returning to the team room (Impellizzeri et al., 2004). Furthermore, establish clear operational definitions for what constitutes the "session" being rated - whether it includes warm-up, competition, and cool-down phases - as inconsistent interpretation can lead to systematic measurement error. Consistent use of neutral, open-ended phrasing with standardized verbal anchors for scale extremes (e.g., "0 means rest, 10 means maximal effort") helps ensure that the RPE reflects the intended construct of subjective exercise intensity rather than external influences or misinterpretation of the measurement scale. Regular “calibration” sessions with athletes to explain or reinforce scale understanding and periodic inter-rater reliability assessments among data collectors will help maintain measurement consistency across time and personnel changes (Haddad et al., 2017).

7.4.1.3 Implications for future sRPE research and practice

The methodological considerations outlined above represent valuable steps toward improving the precision and interpretability of sRPE-based load monitoring strategies, like sRPE-TL or sRPE-CL. Future research should prioritize the development of protocols that can be consistently implemented across sports contexts while maintaining ecological validity. Additionally, investigations into the optimal timing of sRPE collection within the sevens competition space may further enhance measurement quality. By addressing these methodological considerations, practitioners can ensure that sRPE data more accurately reflects the underlying constructs of training intensity and internal load, ultimately leading to more effective load monitoring and training prescription strategies.

7.4.2: Comprehensive model approach for predicting sRPE-CL

Building upon the findings from the individual chapters, a logical next step would be to integrate all explored variables into a unified model, which may enhance prediction of sRPE-CL. A comprehensive approach that incorporates all available GNSS-derived kinematic measures, match-specific variables, and player-specific characteristics, including those referenced across Chapters 2-7 and in Figure 1.1, may benefit prediction of athlete load as determined by sRPE-CL. Identifying the optimal combination of objective factors would streamline data collection processes, allowing practitioners to focus monitoring resources on the most informative variables rather than collecting potentially redundant data. This could lead to more efficient and targeted load monitoring strategies that balance comprehensive assessment with practical constraints.

The implementation of this approach would require a model capable of handling the exploratory variables both as main effects, as well as interactions between multiple variable types while addressing potential concerns around multicollinearity. Initial data preprocessing should include correlation matrix analysis and variance inflation factor (VIF) assessment to identify highly correlated predictors, with variables exceeding VIF thresholds of 5-10 requiring careful consideration for removal or transformation. A crossed-factor linear mixed-effects regression LMER model would be a solid next step in modelling, treating athletes as random effects while systematically examining all possible interactions between GNSS-derived metrics, match-specific, and player-specific characteristics.

To ensure optimal model selection with minimal collinearity, traditional stepwise regression procedures (forward, backward, or bidirectional) could serve as complementary selection methods, particularly useful for their interpretability and established performance in identifying parsimonious models from large predictor sets. Additionally, a multi-stage approach combining several regularization techniques may be another option: elastic net regression could initially reduce the variable space by balancing ridge and lasso penalties for refined variable selection. Chapter two indicates that stepwise regression performed more accurately at imputing sRPE than lasso, ridge, and elastic net regression and so presents a logical step forward for this model.

A substantive model of this nature would be expected to increase R^2 values beyond those achieved in individual chapter analyses for several key reasons. While the inclusion of additional predictor variables naturally tends to increase explained variance, the primary gains would likely stem from capturing previously unaccounted-for interactions between variable types. By including these interactions, the model would better reflect the multifaceted nature of how athletes experience competition. The systematic variable selection process would also ensure that only the most informative predictors and interactions are retained, mitigating the inclusion of redundant variables that could otherwise dilute predictive power. Collectively, this approach should provide a more comprehensive representation of the factors influencing sRPE-CL, resulting in substantially improved sRPE-CL predictivity as measured by $R^2_{adjusted}$.

7.5 Reflection

This dissertation was informed by my personal work in the field, completing data collection, analysis, and reporting on the objective, physical, performance data and athlete-reported sRPE and sRPE-TL/sRPE-CL data from training and competition. Early on in working in the women's rugby sevens environment and after a particularly frustrating tournament result, in the debrief the coaching staff mused that the data from the GNSS units and what the athletes reported "looked the same whether we won or lost". The "why" behind this was a key driver in this doctoral work.

Unpacking the homogeneity of competition data led to some interesting considerations. While I believe that it is critical to recognize the distinct and unique experiences of each athlete in order to ensure that a team's performance is the best it can be, from an objective monitoring perspective of

physical readiness to meet the demands of the competitive environment, I realized early on that sRPE-TL monitoring may not be the be all and end all to which it is held in some training circles. Ultimately, this dissertation confirms what others have, one all-encompassing tool to monitor athlete workload is simply not sufficient.

Chapter six highlights different athlete workload tool responses in match scenarios. Chapters two and five identify the challenges in imputing athlete-reported data, while chapter six specifically speaks to the relative ease of imputing objective date (compared to sRPE). Taken together, coupled with my experience in the training and competition environment, I strongly believe that practitioners need to recognize that a major gap exists in the deployment of sRPE, and downstream sRPE-TL.

Foster et al., (2001) first identified sRPE as representing the "average" RPE of an entire session, but how does an athlete operationalize average? In endurance sports some of the strategies lie with central versus local RPE to differentiate between fatigue of the core processes like breathing (i.e. "lungs burning") and local musculoskeletal process around movement generation (i.e. "legs burning") (Pandolf et al., 1984). Following personal conversation with Carl Foster, I realized that his body of work, while incredibly powerful and relevant to the field, could not be extended to team sports even though categorically similar "intermittent" demands could be placed on team sport athletes as individual endurance sport athletes in training. In short, sRPE cannot be freely applied without serious consideration for the environment in which is it to be used.

While I do not believe that sRPE and the use of sRPE-TL are irrelevant, it is clear that this issue of "average" needs to be addressed. This dissertation challenges the impression that sRPE is solely an intensity-based metric, in particular chapter three demonstrates connections between sRPE and time. This supports the need to help athletes understand and define effort appropriately. This could be done by clarifying "average intensity", through training sessions where both instantaneous and end-of-session RPE (sRPE) are collected, athletes are debriefed and informed about key differences. Athletes would need to be exposed to a variety of intensities in a variety of settings – where their mental/psychological focus is challenged, where their physical output is challenged, and where their technical and tactical acumen is challenged.

Where athletes have improved understanding of the specific information they are contributing through the use of sRPE, the metric becomes more viable for use as an intensity measure of training load. However, that is not to say that objective measures supporting "invisible monitoring" of athletes are not also useful. In fact, I would argue in the competition and training space, the ability to gather data via unobtrusive means is critical to supporting athlete health and development. For example, understanding that an athlete in an end-stage return to play may self-report low sRPE scores as the session may not have been particularly taxing relative to pre-injury training sessions they remember, objectively gathered physical data may show the athlete working above their prescribed capacity in distances, decelerations, or physical contact exposure. As chapter four demonstrates, low speeds and deceleration events are associated with athlete workload. This was a common scenario in my working environment where athletes would accrue large amounts of distance at low speeds, still exposed to decelerations but because they were not running "all out" they felt the session was relatively easier than pre-injury training.

Finally, to affect growth and development in performance, both the physical and the perceptual responses need to be considered. I encourage practitioners to consider the factors which are associated with the athlete workload models of choice, these factors really provide the key to affecting positive change in sport performance. When I started working in this particular team space, there was only sRPE-CL data and tactical performance data available. The inclusion of GNSS data provided a bridge between the two, where tactical markers of game activity were replaced by GNSS-driven "workrate" (meters covered per minute). The inclusion of objective data about physical performance provided another layer of understanding, helping coaches to consider tactical skill execution from performance analysis data along with physical outputs and the athlete's perception of the event.

In closing, do things look the same when a team wins and loses? At the bird's eye view that summary data offers, this can be true. Nevertheless, the key is to recognize the different elements that contribute to the win or loss in different amounts – from physical output, measured through objectively gathered information to perceptual responses to the environment gathered through self-report to tactical data on the execution of sport skills. Taken together, these data more richly inform our understanding of the athlete's experience, giving practitioners more tools to support athletes, coaches, and teams in enhancing sport performance.

7.6 Conclusion

Monitoring physical athlete workload is a core tenet of ensuring athlete safety and competition preparedness. The work presented here demonstrates the importance of considering sport-specific drivers of athlete workload (Figure 1.1) to optimize understanding of the athlete's physical effort in sport performance. However, further analysis into the interaction of the many factors that influence athlete workload is necessary to ensure the application of workload monitoring strategies and training interventions centred around these contributing factors are both feasible and effective in the applied high-performance environment, and beyond.

Methodologically, this dissertation advances the field through the development of an alternative workload monitoring model and demonstrated connections between objective monitoring strategies like the SDC model and mechanical work and sRPE-CL, as well as the identification of salient factors in women's rugby sevens including contact, mass, speed, and deceleration that warrant consideration in athlete monitoring (Figure 1.1). The application of these models, and the inclusion of relevant features, provides an important context for future research into athlete monitoring strategies.

Bibliography

Akubat, I., & Abt, G. (2011). Intermittent exercise alters the heart rate-blood lactate relationship used for calculating the training impulse (TRIMP) in team sport players. Journal of Science and Medicine in Sport, 14(3), 249–253. https://doi.org/10.1016/j.jsams.2010.12.003.

Akubat, I., Patel, E., Barrett, S., & Abt, G. (2012). Methods of monitoring the training and match load and their relationship to changes in fitness in professional youth soccer players. Journal of Sports Sciences, 30(14), 1473–1480. https://doi.org/10.1080/02640414.2012.712711.

Alamoodi, A. H., Zaidan, B. B., Zaidan, A. A., Albahri, O. S., Chen, J., Chyad, M. A., Garfan, S., & Aleesa, A. M. (2021). Machine learning-based imputation soft computing approach for large missing scale and non-reference data imputation. Chaos, Solitons & Fractals, 151, 111236. https://doi.org/10.1016/j.chaos.2021.111236.

Arjol-Serrano, J. L., Lampre, M., Díez, A., Castillo, D., Sanz-López, F., Lozano, D. (2021). The influence of playing formation on physical demands and technical-tactical actions according to playing positions in an elite soccer team. *Int. J. Environ. Res. Public Health, 18*, 4148. https://doi.org/10.3390/ijerph18084148.

Askow, A. T., Lobato, A. L., Arndts, D. J., Jennings, W., Kreutzer, A., Erickson, J. L., Esposito, P. E., Oliver, J. M., Foster, C., & Jagim, A. R. (2021). Session rating of perceived exertion (sRPE) load and training impulse are strongly correlated to GPS-derived measures of external load in NCAA Division I women's soccer athletes. Journal of Functional Morphology and Kinesiology, 6(4), 90. https://doi.org/10.3390/jfmk6040090.

Babyak, M. A. (2004). What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models. Psychosomatic Medicine, 66(3), 411–421. https://doi.org/10.1097/01.psy.0000127692.23278.a9.

Baden, D. A., McLean, T. L., Tucker, R., Noakes, T. D., & St. Clair Gibson, A. (2005). Effect of anticipation during unknown or unexpected exercise duration on rating of perceived exertion, affect, and physiological function. *British Journal of Sports Medicine, 39*, 742-746.

Baird, M. F., Graham, S. M., Baker, J. S., & Bickerstaff, G. F. (2012). Creatine-kinase- and exercise-related muscle damage implications for muscle performance and recovery. Journal of Nutrition and Metabolism, 2012, 960363. https://doi.org/10.1155/2012/960363.

Baker, L. B., Rollo, I., Stein, K. W., & Jeukendrup, A. E. (2015). Acute Effects of Carbohydrate Supplementation on Intermittent Sports Performance. Nutrients, 7(7), 5733-5763. https://doi.org/10.3390/nu7075249.

Barr, M., Beaver, T., Turczyn, D., & Cornish, S. (2019). The validity and reliability of 15 hz global positioning system units for assessing the activity profiles of university football players. *Journal of Strength and Conditioning Research, 33*, 1371-1379.

Bartlett, J. D., O'Connor, F., Pitchford, N., Torres-Ronda, L., & Robertson, S. J. (2017). Relationships between internal and external training load in team-sport athletes: Evidence for an individualized approach. Int J Sports Physiol Perform, 12, 230-234. https://doi.org/10.1123/ijspp.2015-0791.

Barrett, S., McLaren, S., Spears, I., Ward, P., Weston, M. (2018). The influence of playing position and contextual factors on soccer players' match differential ratings of perceived exertion: A preliminary investigation. *Sports, 6*(13). https://doi.org/10.3390/sports6010013.

Bartlett, J. D., Drust, B. (2021). A framework for effective knowledge translation and performance delivery of Sport Scientists in professional sport. *European Journal of Sport Science, 21*(11), 1579–1587. https://doi.org/10.1080/17461391.2020.1842511.

Becirovic, M., Bischofberger, J., Kornfeind, P., Stifter, S., Baca, A. (2024). Analysis of sprint-related performance indicators in professional youth soccer across age groups & positions. ISBS Proceedings Archive, 42, Article 218. Available at: https://commons.nmu.edu/isbs/vol42/iss1/218.

Benjamin, C. L., Hosokawa, Y., Curtis, R. M., Schaefer, D. A., Bergin, R. T., Abegg, M. R., & Casa, D. J. (2020). Environmental Conditions, Preseason Fitness Levels, and Game Workload: Analysis of a Female NCAA DI National Championship Soccer Season. *Journal of strength and conditioning research*, *34*(4), 988–994. https://doi.org/10.1519/JSC.0000000000003535.

Benson, L.C., Stilling, C., Owoeye, O.B.A., & Emery, C.A. (2021). Evaluating methods for imputing missing data from longitudinal monitoring of athlete workload. *J Sports Sci Med, 20*, 188-196. https://doi.org/10.52082/jssm.2021.188.

Beygelzimer, A., Kakadet, S., Langford, J., Arya, S., Mount, D., & Li, S. (2022). Package "FNN". Available online: https://cran.r-project.org/web/packages/FNN/FNN.pdf (accessed 25 August 2022).

Bhaskaran, K., & Smeeth, L. (2014). What is the difference between missing completely at random and missing at random?. *International journal of epidemiology*, *43*(4), 1336–1339. https://doi.org/10.1093/ije/dyu080.

Bolker, B. M. (2015). Linear and generalized linear mixed models. In G. A. Fox, S. Negrete-Yankelevich, & V. J. Sosa (Eds.), *Ecological Statistics: Contemporary Theory and Application* (pp. 309–333). Oxford University Press.

Bonacci, J., Vleck, V., Saunders, P. U., Blanch, P., & Vicenzino, B. (2013). Rating of perceived exertion during cycling is associated with subsequent running economy in

triathletes. Journal of Science and Medicine in Sport, 16(1), 49-53. https://doi.org/10.1016/j.jsams.2012.04.002.

Borg, G. (1998). *Borg's perceived exertion and pain scales*. Human Kinetics.

Bossi, A. H., Naumann, U., Passfield, L., & Hopker, J. (2024). Modelling inter-individual variability in acute and adaptive responses to interval training: insights into exercise intensity normalisation. *European journal of applied physiology*, *124*(4), 1201–1216. https://doi.org/10.1007/s00421-023-05340-y.

Boullosa, D., Casado, A., Claudino, J. G., Jiménez-Reyes, P., Ravé, G., Castaño-Zambudio, A., Lima-Alves, A., de Oliveira Jr, S. A., Dupont, G., Granacher, U., & Zouhal, H. (2020). Do you play or do you train? Insights from individual sports for training load and injury risk management in team sports based on individualization. Frontiers in Physiology, 11, 995. https://doi.org/10.3389/fphys.2020.00995.

Bourdon, P. C., Cardinale, M., Murray, A., Gastin, P., Kellman, M., Varley, M. C., Gabbett, T.J., Coutts, A.J., Burgess, D. J., Gregson, W., & Cable, N. T. (2017). Monitoring athlete training loads: Consensus statement. *International Journal of Sport Physiology, 12*, 161–170.

Bradbury, J. C., & Forman, S. L. (2012). The impact of pitch counts and days of rest on performance among major-league baseball pitchers. Journal of Strength and Conditioning Research, 26(5), 1181–1187. https://doi.org/10.1519/JSC.0b013e31824e16fe.

Bradley, P. S., & Vescovi, J. D. (2015). Velocity thresholds for women's soccer matches: Sex specificity dictates high-speed running and sprinting thresholds—Female Athletes in Motion (FAiM). International Journal of Sports Physiology and Performance, 10(1), 112–116. https://doi.org/10.1123/ijspp.2014-0212.

Bredt, S. G. T., Chagas, M. H., Peixoto, G. H., Menzel, H. J., & de Andrade, A. G. P. (2020). Understanding player load: Meanings and limitations. Journal of Human Kinetics, 71, 5–9. https://doi.org/10.2478/hukin-2019-0072.

Breiman, L. (2001). Random forests. Machine Learning, 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Brocherie, F., & Beard, A. (2021). All alone we go faster, together we go further: The necessary evolution of professional and elite sporting environments to bridge the gap between research and practice. Frontiers in Sports Act Living, 2, 631147. https://doi.org/10.3389/fspor.2020.631147.

Buchheit, M. (2019). Programming high-speed running and mechanical work in relation to technical contents and match schedule in professional soccer. *Sport Performance & Science Reports, 69*, 1–3.

Buchheit, M., Samozino, P., Glynn, J. A., Michael, B. S., Al Haddad, H., Mendez-Villanueva, A., & Morin, J. B. (2014). Mechanical determinants of acceleration and maximal

sprinting speed in highly trained young soccer players. *Journal of sports sciences*, *32*(20), 1906-1913. https://doi.org/10.1080/02640414.2014.965191.

Bureau international des poids et mesures. (2019). Le Système international d'unités / The International System of Units ('The SI Brochure'), ninth edition. URL: http://www.bipm.org/en/publications/si-brochure/.

Burkart, N., & Huber, M. F. (2021). A survey on the explainability of supervised machine learning. Journal of Artificial Intelligence Research, 70, 245-317. https://doi.org/10.1613/jair.1.12228.

Busso, T., Candau, R., & Lacour, J. R. (1994). Fatigue and fitness modelled from the effects of training on performance. *European journal of applied physiology and occupational physiology*, *69*(1), 50–54. https://doi.org/10.1007/BF00867927.

Caparrós, T., Casals, M., Solana, Á., Peña, J. (2018). Low External Workloads Are Related to Higher Injury Risk in Professional Male Basketball Games. *Journal of Sports Science & Medicine, 17*(2), 289-297. PMCID: PMC5950746.

Carey, D., Ong, K., Morris, M., Crow, J., & Crossley, K. (2016). Predicting ratings of perceived exertion in Australian football players: Methods for live estimation. International Journal of Computer Science in Sport, 15, 64-77. https://doi.org/10.1515/ijcss-2016-0005.

Casamichana, D., Castellano, J., Calleja-Gonzalez, J., San Román, J., & Castagna, C. (2013). Relationship Between Indicators of Training Load in Soccer Players. Journal of Strength and Conditioning Research, 27(2), 369-374. https://doi.org/10.1519/JSC.0b013e3182548af1.

Cavagna, G. A., & Kaneko, M. (1977). Mechanical work and efficiency in level walking and running. The Journal of Physiology, 268(2), 467–481. https://doi.org/10.1113/jphysiol.1977.sp011866.

Celton, M., Malpertuy, A., Lelandais, G., & de Brevern, A. G. (2010). Comparative analysis of missing value imputation methods to improve clustering and interpretation of microarray experiments. BMC Genomics, 11, 15. https://doi.org/10.1186/1471-2164-11-15.

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. Neurocomputing, 408, 189–215. https://doi.org/10.1016/j.neucom.2019.10.118.

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Computer Science, 7, e623. https://doi.org/10.7717/peerj-cs.623.

Clarke, A. C., Anson, J. M., & Pyne, D. B. (2015). Physiologically based GPS speed zones for evaluating running demands in women's rugby sevens. *Journal of Sports Science, 33*, 1101–1108.

Clarke, A. C., Anson, J. M., & Pyne, D. B. (2017a). Proof of concept of automated collision detection technology in rugby sevens. *Journal of Strength and Conditioning Research, 31*, 1116–1120.

Clarke, A. C., Anson, J. M., & Pyne, D. B. (2017b). Game movement demands and physical profiles of junior, senior and elite male and female rugby sevens players. *Journal of Sports Science, 35*, 727–733.

Clarke, N., Farthing, J. P., Norris, S. R., Arnold, B. E., & Lanovaz, J. L. (2013). Quantification of training load in Canadian football: Application of session-RPE in collision-based team sports. *Journal of Strength and Conditioning Research, 27*, 2198–2205.

Clemente, F. M., Afonso, J., Costa, J., Oliveira, R., Pino-Ortega, J., & Rico-González, M. (2021). Relationships between sleep, athletic and match performance, training load, and injuries: A systematic review of soccer players. Healthcare, 9(7), 808. https://doi.org/10.3390/healthcare9070808.

Comyns, T., & Flanagan, E. P. (2013). Applications of the session rating of perceived exertion system in professional rugby union. *Strength and Conditioning Journal, 35*, 78-85.

Coppalle, S., Rave, G., Ben Abderrahman, A., Ali, A., Salhi, I., Zouita, S., Zouita, A., Brughelli, M., Granacher, U., & Zouhal, H. (2019). Relationship of pre-season training load with in-season biochemical markers, injuries and performance in professional soccer players. Frontiers in Physiology, 10, 409. https://doi.org/10.3389/fphys.2019.00409.

Coutts, A. J. (2016). Working fast and working slow: The benefits of embedding research in high performance sport. International Journal of Sports Physiology and Performance, 11, 1-2. https://doi.org/10.1123/IJSPP.2015-0781

Coyne, J. O. C., Haff, G. G., Coutts, A. J., Newton, R. U., & Nimphius, S. (2018). The current state of subjective training load monitoring – a practical perspective and call to action. Sport Medicine – Open, 4(58). https://doi.org/10.1186/s40798-018-0172-x.

Cummins, C., & Orr, R. (2013). Body loads and collisions in professional Australian Rugby League players. *Journal of Science and Medicine in Sport, 16S*, e17.

Cunniffe, B., Proctor, W., Baker, J. S., & Davies, B. (2009). An evaluation of the physiological demands of elite rugby union using global positioning system tracking software. Journal of Strength and Conditioning Research, 23(4), 1195-1203. https://doi.org/10.1519/JSC.0b013e3181a3928b.

Dawson, L., McErlain-Naylor, S. A., Devereux, G., & Beato, M. (2024). Practitioner usage, applications, and understanding of wearable GPS and accelerometer technology in

team sports. Journal of Strength and Conditioning Research, 38(7), e373–e382. https://doi.org/10.1519/JSC.0000000000004781.

Delaney, J. A., Duthie, G. M., Thornton, H. R., Scott, T. J., Gay, D., Dascombe, B. J. (2016). Acceleration-Based Running Intensities of Professional Rugby League Match Play. International Journal of Sports Physiology and Performance, 11(6), 802-809. https://doi.org/10.1123/ijspp.2015-0424.

Delaney, J. A., Cummins, C. J., Thornton, H. R., & Duthie, G. M. (2018). Importance, reliability, and usefulness of acceleration measures in team sports. Journal of Strength and Conditioning Research, 32(12), 3485-3493. https://doi.org/10.1519/JSC.0000000000001849.

Dellal, A., Varliette, C., Owen, A., Chirico, E. N., & Pialoux, V. (2012). Small-sided games versus interval training in amateur soccer players: Effects on the aerobic capacity and the ability to perform intermittent exercises with changes of direction. Journal of Strength and Conditioning Research, 26(10), 2712–2720. https://doi.org/10.1519/JSC.0b013e31824294c4.

Delves, R. I. M., Aughey, R. J., Ball, K., & Duthie, G. M. (2021). The quantification of acceleration events in elite team sport: A systematic review. *Sports Medicine – Open, 7*(1), 45. https://doi.org/10.1186/s40798-021-00332-8.

Delves, R. (2023). *Development and use of a consistent approach for processing GNSS-derived acceleration data in team sports* [Doctoral dissertation, Victoria University]. VU Research Repository. https://vuir.vu.edu.au/47136/.

den Hollander, S., Jones, B., Lambert, M., & Hendricks, S. (2018). The what and how of video analysis research in rugby union: A critical review. Sports Medicine - Open, 4(1), 27. https://doi.org/10.1186/s40798-018-0142-3.

Deutsch, M. U., Kearney, G. A., & Rehrer, N. J. (2007). Time – motion analysis of professional rugby union players during match-play. *Journal of Sports Sciences*, *25*(4), 461–472. https://doi.org/10.1080/02640410600631298

Duignan, C., Doherty, C., Caulfield, B., & Blake, C. (2020). Single-item self-report measures of team-sport athlete wellbeing and their relationship with training load: A systematic review. Journal of Athletic Training, 55(9), 944–953. https://doi.org/10.4085/1062-6050-0528.19.

Dumortier, J., Mariman, A., Boone, J., Delesie, L., Tobback, E., Vogelaers, D., & Bourgois, J. G. (2018). Sleep, training load and performance in elite female gymnasts. *European Journal of Sport Science, 18*(2), 151–161. https://doi.org/10.1080/17461391.2017.1389992.

Durand, R., Galli, M., Chenavard, M., Bandiera, D., Freund, H., & Messonnier, L. A. (2021). Modelling of blood lactate time-courses during exercise and/or the subsequent

recovery: Limitations and few perspectives. *Frontiers in Physiology, 12*, 702252. https://doi.org/10.3389/fphys.2021.702252.

Ebert, T. R., Martin, D. T., Stephens, B., & Withers, R. T. (2006). Power output during a professional men's road-cycling tour. *International Journal of Sports Physiology and Performance, 1*(4), 324–335. https://doi.org/10.1123/ijspp.1.4.324.

Edouard, P., Mendiguchia, J., Guex, K., Lahti, J., Samozino, P., & Morin, J. B. (2019). Sprinting: A potential vaccine for hamstring injury? *Sport Performance & Science Reports, 48*, 1–2.

Edwards, S. (1993). *The Heart Rate Monitor Book*. Polar Electro Inc.

El-Masri, M. M., & Fox-Wasylyshyn, S. M. (2005). Missing data: An introductory conceptual overview for the novice researcher. *Canadian Journal of Nursing Research, 37*(2), 156–171.

Ellens, S., Carey, D. L., Gastin, P. B., & Varley, M. C. (2024). Accuracy of GNSS-Derived Acceleration Data for Dynamic Team Sport Movements: A Comparative Study of Smoothing Techniques. *Applied Sciences*, *14*(22), 10573. https://doi.org/10.3390/app142210573.

Epp-Stobbe, A., Tsai, M.-C., Klimstra, M. (2022a). Comparison of imputation methods for missing rate of perceived exertion data in rugby. *Machine Learning and Knowledge Extraction, 4*, 827-838. https://doi.org/10.3390/make4040041.

Epp-Stobbe, A., Tsai, M., Morris, C., & Klimstra, M. (2022b). The influence of physical contact on athlete load in international female rugby sevens. Journal of Strength and Conditioning Research. https://doi.org/10.1519/JSC.0000000000004262.

Epp-Stobbe, A., Tsai, M.-C., Klimstra, M. D. (2024a). Predicting athlete workload in women's rugby sevens using GNSS sensor data, contact count and mass. *Sensors, 24*, 6699. https://doi.org/10.3390/s24206699.

Epp-Stobbe, A., Tsai, M. C., & Klimstra, M. (2024b). Work smarter not harder: Mechanical work as a measure of athlete workload. *ISBS Proceedings Archive, 42, Article 51*. Available at: https://commons.nmu.edu/isbs/vol42/iss1/51.

Epp-Stobbe, A., Tsai, M.-C., & Klimstra, M. (2025a). Rugby Sevens sRPE Workload Imputation Using Objective Models of Measurement. *Applied Sciences*, *15*(12), 6520. https://doi.org/10.3390/app15126520.

Epp-Stobbe, A., Tsai, M.-C., & Klimstra, M. (2025b). A Comparison of the Application of Load Monitoring Metrics for Key Match Characteristics in Women's Rugby Sevens. *Applied Sciences*, *15*(5), 2344. https://doi.org/10.3390/app15052344.

Eston, R. (2012). Use of ratings of perceived exertion in sports. *International Journal of Sports Physiology and Performance, 7*, 175-182.

Farhana, A., & Lappin, S. L. (2023). Biochemistry, Lactate Dehydrogenase. In *StatPearls*. StatPearls Publishing.

Faulkner, J., Parfitt, G., & Eston, R. (2008). The rating of perceived exertion during competitive running scales with time. Psychophysiology, 45, 977-985. https://doi.org/10.1111/j.1469-8986.2008.00712.x.

Ferioli, D., Scanlan, A. T., Conte, D., Tibilette, E., & Rampinini, E. (2021). The business end of the season: A comparison between playoff and regular-season workloads in professional basketball players. *International Journal of Sports Physiology and Performance, 16*(5), 655–662. https://doi.org/10.1123/ijspp.2020-0405.

Fessi, M.S., Moalla, W. (2018). Postmatch perceived exertion, feeling, and wellness in professional soccer players. *International Journal of Sports Physiology and Performance, 13*, 631-637. https://doi.org/10.1123/ijspp.2017-0725.

Folgado, H., Duarte, R., Fernandes, O., & Sampaio, J. (2014). Competing with lower level opponents decreases intra-team movement synchronization and time-motion demands during pre-season soccer matches. *PLoS ONE, 9*, e97145. https://doi.org/10.1371/journal.pone.0097145.

Foster, C., Florhaug, J. A., Franklin, J., Gottschall, L., Hrovatin, L. A., Parker, S., Doleshal, P., & Dodge, C. (2001). A new approach to monitoring exercise training. *Journal of Strength and Conditioning Research, 15*, 109–115.

Foster, C., Boullosa, D., McGuigan, M., Fusco, A., Cortis, C., Arney, B. E., Orton, B., Dodge, C., Jaime, S., Radtke, K., van Erp, T., de Koning, J. J., Bok, D., Rodriguez-Marroyo, J. A., & Porcari, J. P. (2021). 25 years of session rating of perceived exertion: Historical perspective and development. International Journal of Sports Physiology and Performance, 16(5), 612–621. https://doi.org/10.1123/ijspp.2020-0599.

Fox, J. L., Scanlan, A. T., Stanton, R., & Sargent, C. (2022). The effect of game-related contextual factors on sleep in basketball players. Journal of Strength and Conditioning Research, 36(10), 2942–2945. https://doi.org/10.1519/JSC.0000000000003943.

Fox-Wasylyshyn, S. M., & El-Masri, M. M. (2005). Handling missing data in self-report measures. Research in Nursing & Health, 28, 488–495. https://doi.org/10.1002/nur.20100.

Fox, J. L., Stanton, R., Sargent, C., Wintour, S. A., & Scanlan, A. T. (2018). The Association Between Training Load and Performance in Team Sports: A Systematic Review. *Sports medicine (Auckland, N.Z.)*, *48*(12), 2743–2774. https://doi.org/10.1007/s40279-018-0982-5.

Friedman, J., Hastie, T., Tibshirani, R., Narasimhan, B., Tay, K., Simon, N., Qian, J., & Yang, J. (2022). Package "glmnet". Available online: https://cran.r-project.org/web/packages/glmnet/glmnet.pdf (accessed 25 August 2022).

Fritsch, S., Guenther, F., Wright, M. N., Suling, M., & Mueller, S. M. (2019). Package "neuralnet". Available online: https://cran.r-project.org/web/packages/neuralnet/neuralnet.pdf (accessed 25 August 2022).

Fuller, C. W., Brooks, J. H. M., Cancea, R. J., Hall, J., & Kemp, S. P. T. (2007a). Contact events in rugby union and their propensity to cause injury. *British Journal of Sports Medicine, 41*, 862-867.

Fuller, C. W., Molloy, M. G., Bagate, C., et al. (2007b). Consensus statement on injury definitions and data collection procedures for studies of injuries in rugby union. *British Journal of Sports Medicine, 41*, 328-331.

Fuller, C. W., Ashton, T., Brooks, J. H. M., Cancea, R. J., Hall, J., & Kemp, S. P. T. (2008). Injury risks associated with tackling in rugby union. *British Journal of Sports Medicine, 44*, 159-167.

Fuller, C. W., Taylor, A., Kemp, S. P. T., & Raftery, M. (2017). Rugby World Cup 2015: World Rugby injury surveillance study. British Journal of Sports Medicine, 51(1), 51–57. https://doi.org/10.1136/bjsports-2016-096275.

Fuller, C. W., & Taylor, A. (2018). Women sevens World Series summary of results: 2011/12 to 2017/18. *World Rugby*.

Furlan, N., Waldron, M., Shorter, K., Gabbett, T.J., Mitchell, J., Fitzgerald, E., Osborne, M. A., & Gray, A. J. (2015). Running-intensity fluctuations in elite rugby sevens performance. *International Journal of Sports Physiology and Performance, 10*(6), 802-807. https://doi.org/10.1123/ijspp.2014-0315.

Gabbett, T. J. (2004). Reductions in pre-season training loads reduce training injury rates in rugby league players. *British Journal of Sports Medicine, 38*, 743-749. https://doi.org/10.1136/bjsm.2003.010229.

Gabbett, T. J., & Domrow, N. (2007). Relationships between training load, injury, and fitness in sub-elite collision sport athletes. Journal of Sports Sciences, 25(13), 1507–1519. https://doi.org/10.1080/02640410701215066.

Gabbett, T., & Kelly, J. (2007). Does fast defensive line speed influence tackling proficiency in collision sport athletes? *International Journal of Sports Science & Coaching, 2*, 467-472.

Gabbett, T., Kelly, J., & Pezet, T. (2007). Relationship between physical fitness and playing ability in rugby league players. *Journal of Strength and Conditioning Research, 21*, 1126-1133.

Gabbett, T. J. (2008). Influence of fatigue on tackling technique in rugby league players. *Journal of Strength and Conditioning Research, 22*, 625–632.

Gabbett, T., Jenkins, D., & Abernethy, B. (2010). Physical collisions and injury during professional rugby league skills training. *Journal of Science and Medicine in Sport*, *13*(6), 578-583.

Gabbett, T. J., Jenkins, D. G., Abernathy, B. (2011). Correlates of tackling ability in high-performance rugby league players. *Journal of Strength and Conditioning Research, 25*(5), 72–79. https://doi.org/10.1519/JSC.0b013e3181ff506f.

Gabbett, T. J., Jenkins, D. G. (2011). Relationship between training load and injury in professional rugby league players. *Journal of Science and Medicine in Sport, 14*, 204-209. https://doi.org/10.1016/j.jsams.2010.12.002.

Gabbett, T. J., & Ullah, S. (2012). Relationship between running loads and soft-tissue injury in elite team sport athletes. Journal of Strength and Conditioning Research, 26(4), 953–960. https://doi.org/10.1519/JSC.0b013e3182302023.

Gabbett, T. J. (2013). Influence of the opposing team on the physical demands of elite rugby league match play. *Journal of Strength and Conditioning Research, 27*, 1629-1635. https://doi.org/10.1519/JSC.0b013e318274f30e.

Gabbett, T. J. (2015). Relationship between accelerometer load, collisions, and repeated high-intensity effort activity in rugby league players. *Journal of Strength and Conditioning Research, 29*(12), 3424-3431. https://doi.org/10.1519/JSC.0000000000001017.

Gabbett, T. J. (2016). The training-injury prevention paradox: Should athletes be training smarter and harder? *British Journal of Sports Medicine, 50*(5), 273-280. https://doi.org/10.1136/bjsports-2015-095788.

Gabbett, T. J., Nassis, G. P., Oetter, E., Pretorius, J., Johnston, N., Medina, D., Rodas, G., Myslinski, T., Howells, D., Beard, A., & Ryan, A. (2017). The athlete monitoring cycle: A practical guide to interpreting and applying training monitoring data. *British Journal of Sports Medicine, 51*, 1451-1452. https://doi.org/10.1136/bjsports-2016-097298.

Gallo, T., Cormack, S., Gabbett, T., Williams, M., & Lorenzen, C. (2014). Characteristics impacting on session rating of perceived exertion training load in Australian footballers. *Journal of Sports Science, 33*, 467-475. https://doi.org/10.1080/02640414.2014.947311.

Gallo, T. F. (2016). Monitoring athlete preparedness in professional Australian football: Load, self-report measures and performance (Doctoral dissertation). Australian Catholic University.

Gammon, C., Pfeiffer, K. A., Pivarnik, J. M., Moore, R. W., Rice, K. R., & Trost, S. G. (2016). Age-related differences in OMNI-RPE scale validity in youth: A longitudinal analysis. Medicine and Science in Sports and Exercise, 48(8), 1590–1594. https://doi.org/10.1249/MSS.0000000000000918.

Gastin, P. B., McLean, O. C., Breed, R. V. P., & Spittle, M. (2014). Tackle and impact detection in elite Australian football using wearable microsensor technology. Journal of Sports Sciences, 32(10), 947–953. https://doi.org/10.1080/02640414.2013.868920.

Gaudino, P., Iaia, F. M., Strudwick, A. J., Hawkins, R. D., Alberti, G., Atkinson, G., & Gregson, W. (2015). Factors influencing perception of effort (session rating of perceived exertion) during elite soccer training. *International Journal of Sports Physiology and Performance, 10*, 860-864. https://doi.org/10.1123/ijspp.2015-0425.

Glenn, T. C., Zare, A., & Gader, P. D. (2015). Bayesian fuzzy clustering. IEEE Transactions on Fuzzy Systems, 23(5), 1545–1561. https://doi.org/10.1109/TFUZZ.2014.2370676.

Goodale, T. L., Gabbett, T. J., Tsai, M. C., Stellingwerff, T., & Sheppard, J. (2017). The effect of contextual factors on physiological and activity profiles in international women's rugby sevens. *International Journal of Sports Physiology and Performance, 12*, 370-376. https://doi.org/10.1123/ijspp.2016-0172.

Goodwin, M. L., Harris, J. E., Hernández, A., & Gladden, L. B. (2007). Blood lactate measurements and analysis during exercise: A guide for clinicians. Journal of Diabetes Science and Technology, 1(4), 558–569. https://doi.org/10.1177/193229680700100414.

Gray, A. J., Shorter, K., Cummins, C., Murphy, A., & Waldron, M. (2018). Modelling movement energetics using global positioning system devices in contact team sports: Limitations and solutions. Sports Medicine, 48(6), 1357–1368. https://doi.org/10.1007/s40279-018-0899-z.

Griffin, A., Kenny, I. C., Comyns, T. M., & Lyons, M. (2020). The association between the acute:chronic workload ratio and injury and its application in team sports: A systematic review. Sports Medicine, 50(3), 561–580. https://doi.org/10.1007/s40279-019-01218-2.

Griffin, A., Kenny, I. C., Comyns, T. M., Purtill, H., Tiernan, C., O'Shaughnessy, E., & Lyons, M. (2021). Training load monitoring in team sports: A practical approach to addressing missing data. Journal of Sports Science, 39, 2161-2171. https://doi.org/10.1080/02640414.2021.1923205.

Guyon, J. (2022). Choose your opponent: A new knockout design for hybrid tournaments. *Journal of Sports Analysis, 8*, 9-29. https://doi.org/10.3233/JSA-200527.

Haddad, M., Stylianides, G., Djaoui, L., Dellal, A., & Chamari, K. (2017). Session-RPE method for training load monitoring: Validity, ecological usefulness, and influencing factors. *Frontiers in Neuroscience, 11*, 612. https://doi.org/10.3389/fnins.2017.00612.

Halson, S. L. (2014). Monitoring training load to understand fatigue in athletes. Sports Medicine, 44(Suppl 2), S139–S147. https://doi.org/10.1007/s40279-014-0253-z.

Harper, D. J., Carling, C., Kiely, J. (2019). High-intensity acceleration and deceleration demands in elite team sports competitive match play: A systematic review and meta-

analysis of observational studies. *Sports Medicine, 49*(12), 1923-1947. https://doi.org/10.1007/s40279-019-01170-1.

Hawthorne, G., & Elliot, P. (2005). Imputing cross-sectional missing data: Comparison of common techniques. *Australian and New Zealand Journal of Psychiatry, 39*, 583-590.

Henderson, D. A., & Denison, D. R. (1989). Stepwise regression in social and psychological research. Psychological Reports, 64(1), 251–257. https://doi.org/10.2466/pr0.1989.64.1.251.

Hendricks, S., Karpul, D., Nicolls, F., & Lambert, M. (2012). Velocity and acceleration before contact in the tackle during rugby union matches. *Journal of Sports Science, 30*, 1215–1224. https://doi.org/10.1080/02640414.2012.684833.

Hendricks, S., Karpul, D., Lambert, M. (2014). Momentum and kinetic energy before the tackle in rugby union. *Journal of Sports Science and Medicine, 13*(3), 557-563.

Hendricks, S., Till, K., den Hollander, S., Savage, T. N., Roberts, S. P., Tierney, G., Burger, N., Kerr, H., Kemp, S., Cross, M., Patricios, J., McKune, A. J., Bennet, M., Rock, A., Stokes, K. A., Ross, A., Readhead, C., Quarrie, K. L., Tucker, R., & Jones, B. (2020). Consensus on a video analysis framework of descriptors and definitions by the Rugby Union Video Analysis Consensus group. British Journal of Sports Medicine, 54(10), 566–572. https://doi.org/10.1136/bjsports-2019-101293.

Hills, S. P., Izri, E., Howells, D., Lonergan, B., Kilduff, L. P., & Waldron, M. (2024). Assessing the locomotor demands of international men's and women's rugby sevens match-play according to passage of play. *PLoS ONE, 19*, e0304186. https://doi.org/10.1371/journal.pone.0304186.

Hong, S., & Lynn, H. S. (2020). Accuracy of random-forest-based imputation of missing data in the presence of non-normality, non-linearity, and interaction. BMC Medical Research Methodology, 20, 51. https://doi.org/10.1186/s12874-020-01080-1.

Hopkins, W. G. (1991). Quantification of training in competitive sports. Methods and applications. *Sports Medicine, 12*(3), 161-183. https://doi.org/10.2165/00007256-199112030-00003.

Horstman, D. H., Morgan, W. P., Cymerman, A., & Stokes, J. (1979). Perception of effort during constant work to self-imposed exhaustion. Perceptual and Motor Skills, 48(3 Pt 2), 1111–1126. https://doi.org/10.2466/pms.1979.48.3c.1111.

Hsu, C., Chang, C. C., & Lin, C. J. (2003). A practical guide to support vector classification. Proceedings of the International Conference on Machine Learning, 1396-1400. Available online: http://www.datascienceassn.org/sites/default/files/Practical%20Guide%20to%20Support%20Vector%20Classification.pdf.

Huang, K., & Ihm, J. (2021). Sleep and injury risk. Current Sports Medicine Reports, 20(6), 286–290. https://doi.org/10.1249/JSR.0000000000000849.

Hulin, B. T., Gabbett, T. J., Blanch, P., Chapman, P., Bailey, D., & Orchard, J.W. (2014). Spikes in acute workload are associated with increased injury risk in elite cricket fast bowlers. *British Journal of Sports Medicine, 48*, 708-712. https://doi.org/10.1136/bjsports-2013-092516.

Hulin, B. T., Gabbett, T. J., Kearney, S., & Corvo, A. (2015). Physical demands of match play in successful and less-successful elite rugby league teams. International Journal of Sports Physiology and Performance, 10(6), 703–710. https://doi.org/10.1123/ijspp.2014-0080.

Hulin, B. T., Gabbett, T. J., Johnston, R. D., & Jenkins, D. G. (2017). Wearable microtechnology can accurately identify collision events during professional rugby league match-play. *Journal of Science and Medicine in Sport, 20*, 638–642. https://doi.org/10.1016/j.jsams.2016.07.004.

Hulin, B. T., Gabbett, T. J., Johnston, R. D., & Jenkins, D. G. (2018). PlayerLoad variables: Sensitive to changes in direction and not related to collision workloads in rugby league match play. International Journal of Sports Physiology and Performance, 13(9), 1136–1142. https://doi.org/10.1123/ijspp.2017-0557.

Iannaccone, A., Conte, D., Cortis, C., & Fusco, A. (2021). Usefulness of linear mixed-effects models to assess the relationship between objective and subjective internal load in team sports. International Journal of Environmental Research and Public Health, 18, 392. https://doi.org/10.3390/ijerph18020392.

Ihsan, M., Yeo, V., Tan, F., Joseph, R., Lee, M., & Aziz, A. R. (2021). Running demands and activity profile of the new four-quarter match format in men's field hockey. *Journal of Strength and Conditioning Research, 35*(2), 512-518. https://doi.org/10.1519/JSC.0000000000002699.

Imbach, F., Sutton-Charani, N., Montmain, J., Candau, R., & Perrey, S. (2022). The Use of Fitness-Fatigue Models for Sport Performance Modelling: Conceptual Issues and Contributions from Machine-Learning. *Sports medicine - open*, *8*(1), 29. https://doi.org/10.1186/s40798-022-00426-x.

Impellizzeri, F. M., Rampinini, E., Coutts, A. J., Sassi, A., & Marcora, S. M. (2004). Use of RPE-based training load in soccer. Medicine and Science in Sports and Exercise, 36(6), 1042–1047. https://doi.org/10.1249/01.mss.0000128199.23901.2f.

Impellizzeri, F. M., Marcora, S. M., & Coutts, A. J. (2019). Internal and external training load: 15 years on. *International Journal of Sports Physiology and Performance, 14*(2), 270-273. https://doi.org/10.1123/ijspp.2018-0935.

International Olympic Committee. (2019). Rugby Sevens. Olympics.com. Retrieved February 13, 2025, from https://www.olympics.com/en/sports/rugby-sevens/

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). *An introduction to statistical learning: With applications in R*. Springer US Springer.

Jennings, D., Cormack, S., Coutts, A. J., Boyd, L. J., & Aughey, R. J. (2010). Variability of GPS units for measuring distance in team sport movements. International Journal of Sports Physiology and Performance, 5(4), 565–569. https://doi.org/10.1123/ijspp.5.4.565.

Johnston, R. J., Watsford, M. L., Pine, M. J., & Spurrs, R. W. (2013). Assessment of 5 Hz and 10 Hz GPS units for measuring athlete movement demands. International Journal of Performance Analysis in Sport, 13(1), 262–274. https://doi.org/10.1080/24748668.2013.11868646.

Johnston, R. D., Gabbett, T. J., & Jenkins, D. G. (2014). Applied sport science of rugby league. Sports Medicine, 44(8), 1087–1100. https://doi.org/10.1007/s40279-014-0190-x.

Jönsson, P., & Wohlin, C. (2004). An evaluation of k-nearest neighbour imputation using Likert data. Proceedings of the 10th International Symposium on Software Metrics (METRICS'04), 108–118. https://doi.org/10.1109/METRICS.2004.3.

Kang, H. (2013). The prevention and handling of the missing data. *Korean Journal of Anesthesiology, 64*, 402-406. https://doi.org/10.4097/kjae.2013.64.5.402.

Kelly, D., Coughlan, G. F., Green, B. S., & Caulfield, B. (2012). Automatic detection of collisions in elite level rugby union using a wearable sensing device. *Sports Engineering, 15*, 81–92. https://doi.org/10.1007/s12283-012-0090-0.

Kiely, M., & Warrington, G. (2019). Physiological and performance monitoring in competitive sporting environments: A review for elite individual sports. Strength & Conditioning Journal, 41(6), 61–63. https://doi.org/10.1519/SSC.0000000000000493.

King, D., Hume, P., & Clark, T. (2010). Video analysis of tackles in professional rugby league matches by player position, tackle height, and tackle location. *International Journal of Performance Analysis in Sport, 10*, 241–254.

Knufinke, M., Nieuwenhuys, A., Geurts, S. A. E., Coenen, A. M. L., & Kompier, M. A. J. (2018). Self-reported sleep quantity, quality, and sleep hygiene in elite athletes. Journal of Sleep Research, 27(1), 78–85. https://doi.org/10.1111/jsr.12509.

Kokla, M., Virtanen, J., Kolehmainen, M., Paananen, J., & Hanhineva, K. (2019). Random forest-based imputation outperforms other methods for imputing LC-MS metabolomics data: A comparative study. BMC Bioinformatics, 20, 124. https://doi.org/10.1186/s12859-019-3110-0.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*, 155-163. https://doi.org/10.1016/j.jcm.2016.02.012.

Lakens, D. (2017). Equivalence tests: A practical primer for t-tests, correlations, and meta-analyses. Social Psychological and Personality Science, 8, 355-362. https://doi.org/10.1177/1948550617697177.

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. Advances in Methods and Practices in Psychological Science, 1, 259–269. https://doi.org/10.1177/2515245918770963.

Lalor, B. J., Halson, S. L., Tran, J., Kemp, J. G., & Cormack, S. J. (2020). A complex relationship: Sleep, external training load, and well-being in elite Australian footballers. International Journal of Sports Physiology and Performance, 15(6), 777–787. https://doi.org/10.1123/ijspp.2019-0061.

Lambert, M. I., & Borresen, J. (2010). Measuring training load in sports. International Journal of Sports Physiology and Performance, 5(3), 406–411. https://doi.org/10.1123/ijspp.5.3.406.

Lambrick, D. M., Faulkner, J. A., Rowlands, A. V., & Eston, R. G. (2009). Prediction of maximal oxygen uptake from submaximal ratings of perceived exertion and heart rate during a continuous exercise test: The efficacy of RPE 13. European Journal of Applied Physiology, 107(1), 1–9. https://doi.org/10.1007/s00421-009-1093-7.

Liaw, A., & Wiener, M. (2022). Package "randomForest". Available online: https://cran.r-project.org/web/packages/randomForest/randomForest.pdf (accessed 25 August 2022).

Lockie, R. G., Schultz, A. B., Callaghan, S. J., Jeffriess, M. D. (2016). The relationship between dynamic stability and multidirectional speed. *Journal of Strength and Conditioning Research, 30*(11), 3033-3043. https://doi.org/10.1519/JSC.0b013e3182a744b6.

Lolli, L., Batterham, A. M., Hawkins, R., Kelly, D. M., Strudwick, A. J., Thorpe, R., Gregson, W., & Atkinson, G. (2019). Mathematical coupling causes spurious correlation within the conventional acute-to-chronic workload ratio calculations. British Journal of Sports Medicine, 53(15), 921–922. https://doi.org/10.1136/bjsports-2017-098110.

Lovell, T. W. J., Sirotic, A. C., Impellizzeri, F. M., & Coutts, A. J. (2013). Factors affecting perception of effort (session rating of perceived exertion) during rugby league training. *International Journal of Sports Physiology and Performance, 8*, 62–69.

Lucía, A., Hoyos, J., Pardo, J., & Chicharro, J. L. (2000). Metabolic and neuromuscular adaptations to endurance training in professional cyclists: A longitudinal study. Japanese Journal of Physiology, 50(3), 381–388. https://doi.org/10.2170/jjphysiol.50.381.

Lundstrom, C. J., Foreman, N. A., & Biltz, G. (2023). Practices and applications of heart rate variability monitoring in endurance athletes. International Journal of Sports Medicine, 44(1), 9–19. https://doi.org/10.1055/a-1864-9726.

Malone, J. J., Lovell, R., Varley, M. C., & Coutts, A. J. (2017). Unpacking the Black Box: Applications and considerations for using GPS devices in sport. International Journal of Sports Physiology and Performance, 12(Suppl 2), S218-S226. https://doi.org/10.1123/ijspp.2016-0236.

Manzi, V., Bovenzi, A., Franco Impellizzeri, M., Carminati, I., & Castagna, C. (2013). Individual training-load and aerobic-fitness variables in premiership soccer players during the precompetitive season. *Journal of strength and conditioning research*, *27*(3), 631–636. https://doi.org/10.1519/JSC.0b013e31825dbd81.

Manzi, V., Castagna, C., Padua, E., Lombardo, M., D'Ottavio, S., Massaro, M., Volterrani, M., & Iellamo, F. (2009). Dose-response relationship of autonomic nervous system responses to individualized training impulse in marathon runners. American Journal of Physiology-Heart and Circulatory Physiology, 296(6), H1733–H1740. https://doi.org/10.1152/ajpheart.00054.2009.

Mara, J. K., Thompson, K. G., Pumpa, K. L., & Morgan, S. (2017). The acceleration and deceleration profiles of elite female soccer players during competitive matches. Journal of Science and Medicine in Sport, 20(9), 867-872. https://doi.org/10.1016/j.jsams.2016.12.078.

Marynowicz, J., Kikut, K., Lango, M., Horna, D., & Andrzejewski, M. (2020). Relationship between the session-RPE and external measures of training load in youth soccer training. Journal of Strength and Conditioning Research, 34(10), 2800-2804. https://doi.org/10.1519/JSC.0000000000003785.

McBurnie, A. J., Harper, D. J., Jones, P. A., Dos Santos, T. (2022). Deceleration training in team sports: Another potential 'vaccine' for sports-related injury? *Sports Medicine, 52*(1), 1-12. https://doi.org/10.1007/s40279-021-01583-x.

McCall, A., Wolfberg, A., Ivarsson, A., Dupont, G., Larocque, A., & Bilsborough, J. (2023). A qualitative study of 11 world-class team-sport athletes' experiences answering subjective questionnaires: A key ingredient for 'visible' health and performance monitoring? *Sports Medicine, 53*, 1085-1100. https://doi.org/10.1007/s40279-023-01814-3.

McDonald, G. C. (2009). Ridge regression. WIREs Computational Statistics, 1(1), 93–100. https://doi.org/10.1002/wics.14.

McLaren, S. J., Macpherson, T. W., Coutts, A. J., Hurst, C., Spears, I. R., Weston, M. (2018). The relationships between internal and external measures of training load and intensity in team sports: A meta-analysis. *Sports Medicine, 48*(3), 641-658. https://doi.org/10.1007/s40279-017-0830-z.

McLellan, C. P., Lovell, D. I., & Gass, G. C. (2011). Biomechanical and endocrine responses to impact and collision during elite rugby league match play. *Journal of Strength and Conditioning Research, 25*(7), 1553–1562. https://doi.org/10.1519/JSC.0b013e3181dba0d9.

Mellalieu, S. D., Neil, R., Hanton, S., & Fletcher, D. (2009). Competition stress in sport performers: Stressors experienced in the competition environment. Journal of Sports Sciences, 27(7), 729–744. https://doi.org/10.1080/02640410902889834.

Menaspà, P. (2017). Are rolling averages a good way to assess training load for injury prevention? British Journal of Sports Medicine, 51(7), 618–619. https://doi.org/10.1136/bjsports-2016-096131.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C. C., & Lin, C. C. (2022). Package "e1071". Available online: https://cran.r-project.org/web/packages/e1071/e1071.pdf (accessed 25 August 2022).

Meylan, C. M., Cronin, J., Hopkins, W. G., & Oliver, J. (2014). Adjustment of measures of strength and power in youth male athletes differing in body mass and maturation. Pediatric Exercise Science, 26(1), 41–48. https://doi.org/10.1123/pes.2013-0029.

Miles, J. (2014). R squared, adjusted R squared. In Wiley StatsRef: Statistics Reference Online. Wiley. https://doi.org/10.1002/9781118445112.stat06627.

Mishra, P., Singh, U., Pandey, C. M., Mishra, P., & Pandey, G. (2019). Application of Student's t-test, analysis of variance, and covariance. Annals of Cardiac Anaesthesia, 22(4), 407–411. https://doi.org/10.4103/aca.ACA_94_19.

Misseldine, N. D., Blagrove, R. C., & Goodwin, J. E. (2018). Speed demands of women's rugby sevens match play. *Journal of Strength and Conditioning Research, 35*(1), 183-189. https://doi.org/10.1519/JSC.0000000000002638.

Morgan, W. P. (1994). Psychological components of effort sense. *Medicine and Science in Sports and Exercise, 9*, 1071–1077.

Mujika, I. (2013). The alphabet of sport science research starts with Q. *International Journal of Sports Physiology and Performance, 8*, 465-466.

Mujika, I. (2017). Quantification of training and competition loads in endurance sports: Methods and applications. *International Journal of Sport Physiology, 12*, S29–S217.

Musil, C. M., Warner, C. B., Yobas, P. K., & Jones, S. L. (2002). A comparison of imputation techniques for handling missing data. Western Journal of Nursing Research, 24(7), 815–829. https://doi.org/10.1177/019394502762477004.

Muyor, J. M. (2013). Exercise intensity and validity of the ratings of perceived exertion (Borg and OMNI scales) in an indoor cycling session. *Journal of Human Kinetics, 39*, 93–101. https://doi.org/10.2478/hukin-2013-0010.

Nicolella, D. P., Torres-Ronda, L., Saylor, K. J., & Schelling, X. (2018). Validity and reliability of an accelerometer-based player tracking device. PLOS ONE, 13(2), e0191823. https://doi.org/10.1371/journal.pone.0191823.

Noakes, T. D. (2004). Linear relationship between the perception of effort and the duration of constant load exercise that remains. Journal of Applied Physiology, 96(4), 1571–1572. https://doi.org/10.1152/japplphysiol.01124.2003.

Núñez, F. J., Toscano-Bendala, F. J., Suarez-Arrones, L., Martínez-Cabrera, F. I., De Hoyo, M. (2018). Individualized thresholds to analyze acceleration demands in soccer players using GPS. *Retos, 35*, 75-79. https://doi.org/10.47197/retos.v0i35.60402.

Oliveira, R., Brito, J. M., Loureiro, N., Padinha, V., Ferreira, B., Mendes, B. (2021). Effects of match location, match result and the quality of opposition in training load on a two-matches week in a top-class elite European soccer team. *Cuadernos de Psicología del Deporte, 21*, 183-197.

Pageaux B. (2016). Perception of effort in Exercise Science: Definition, measurement and perspectives. *European journal of sport science*, *16*(8), 885–894. https://doi.org/10.1080/17461391.2016.1188992.

Pandolf, K. B. (1982). Differentiated ratings of perceived exertion during physical exercise. *Medicine and Science in Sports and Exercise, 14*, 397–405.

Pandolf, K. B., Billings, D. S., Drolet, L. L., Pimental, N. A., & Sawka, M. N. (1984). Differentiated ratings of perceived exertion and various physiological responses during prolonged upper and lower body exercise. *European journal of applied physiology and occupational physiology*, *53*(1), 5-11.

Paquette, M. R., Napier, C., Willy, R. W., & Stellingwerff, T. (2020). Moving beyond weekly "distance": Optimizing quantification of training load in runners. Journal of Orthopaedic & Sports Physical Therapy, 50(10), 564–569. https://doi.org/10.2519/jospt.2020.9533.

Ponce-Bordón, J. C., Polo-Tejada, J., Sanabria-Pino, B., Rubio-Morales, A., García-Calvo, T., & Lobo-Triviño, D. (2024). The Influence of the Playing Surface on Workload Response in Spanish Professional Male Soccer Players. *Sensors*, *24*(14), 4506. https://doi.org/10.3390/s24144506.

Quarrie, K. L., & Hopkins, W. G. (2008). Tackle injuries in professional rugby union. *American Journal of Sports Medicine, 36*, 1705–1716. https://doi.org/10.1177/0363546508318912.

R Core Team. (n.d.). The R stats package. Available online: https://stat.ethz.ch/R-manual/R-devel/library/stats/html/stats-package.html (accessed 25 August 2022).

Rahman, M. G., & Islam, M. Z. (2016). Missing value imputation using a fuzzy clustering-based EM approach. Knowledge and Information Systems, 46, 389–422. https://doi.org/10.1007/s10115-015-0822-y.

Rampinini, E., Alberti, G., Fiorenza, M., Riggio, M., Sassi, R., Borges, T. O., & Coutts, A. J. (2015). Accuracy of GPS devices for measuring high-intensity running in field-based

team sports. International Journal of Sports Medicine, 36(1), 49–53. https://doi.org/10.1055/s-0034-1385866.

Rejeski, W. J., & Ribisl, P. M. (1980). Expected task duration and perceived effort: An attributional analysis. *Journal of Sport and Exercise Psychology, 2*(3), 227-236.

Riedmiller, M. (1994). Advanced supervised learning in multi-layer perceptrons — From backpropagation to adaptive learning algorithms. Computer Standards & Interfaces, 16, 265-278. https://doi.org/10.1016/0920-5489(94)90017-5.

Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., & Firth, D. (2022). Package "MASS". Available online: https://cran.r-project.org/web/packages/MASS/MASS.pdf (accessed 25 August 2022).

Roe, G., Darrall-Jones, J., Till, K., Phibbs, P., Read, D., Weakley, J., Rock, A., & Jones, B. (2017). The effect of physical contact on changes in fatigue markers following rugby union field-based training. *European Journal of Sport Science, 17*(6), 647-655. https://doi.org/10.1080/17461391.2017.1287960.

Roell, M., Roecker, K., Gehring, D., Mahler, H., & Gollhofer, A. (2018). Player monitoring in indoor team sports: Concurrent validity of inertial measurement units to quantify average and peak acceleration values. Frontiers in Physiology, 9, 141. https://doi.org/10.3389/fphys.2018.00141.

Ross, A., Gill, N., & Cronin, J. (2014). Match analysis and player characteristics in rugby sevens. *Sports Medicine, 44*(3), 357-367. https://doi.org/10.1007/s40279-013-0123-0.

Rossi, A., Pappalardo, L., Cintia, P., Iaia F. M., Fernàndez, J., & Medina, D. (2018). Effective injury forecasting in soccer with GPS training data and machine learning. *PLOS One, 13*, e0201264. https://doi.org/10.1371/journal.pone.0201264.

Sarkar, M., & Fletcher, D. (2014). Psychological resilience in sport performers: a review of stressors and protective factors. *Journal of sports sciences*, *32*(15), 1419–1434. https://doi.org/10.1080/02640414.2014.901551.

Sartor, F., Vailati, E., Valsecchi, V., Vailati, F., & La Torre, A. (2013). Heart rate variability reflects training load and psychophysiological status in young elite gymnasts. *The Journal of Strength & Conditioning Research, 27*(10), 2782-2790.

Saw, A. E., Main, L. C., & Gastin, P. B. (2016). Monitoring the athlete training response: Subjective self-reported measures trump commonly used objective measures: A systematic review. *British Journal of Sports Medicine, 50*, 281–291. https://doi.org/10.1136/bjsports-2015-094758.

Scanlan, A. T., Fox, J. L., Poole, J. L., Conte, D., Milanović, Z., Lastella, M., & Dalbo, V. J. (2018). A comparison of traditional and modified Summated-Heart-Rate-Zones models to measure internal training load in basketball players. *Measurement in Physical*

Education and Exercise Science, 22(4), 303-309. https://doi.org/10.1080/1091367X.2018.1457857.

Schmitt, P., Mandel, J., & Guedj, M. (2015). A comparison of six methods for missing data imputation. Journal of Biometric Biostatistics, 6, 1-10. https://doi.org/10.472/2155-6180.1000224.

Schultz de Arruda, A. F., Saldana Aoki, M., Drago, G., & Moreira, A. (2019). Salivary testosterone concentration, anxiety, perceived performance, and ratings of perceived exertion in basketball players during semi-final and final matches. *Physiology & Behavior, 198*, 102-107. https://doi.org/10.1016/j.physbeh.2018.10.008.

Schuster, J., Howells, D., Robineau, J., Couderc, A., Natera, A., Lumley, N., Gabbett, T. J., & Winkelman, N. (2018). Physical-preparation recommendations for elite rugby sevens performance. *International Journal of Sports Physiology and Performance, 13*(3), 255-267. https://doi.org/10.1123/ijspp.2016-0728.

Shah, A. D., Bartlett, J. W., Carpenter, J., Nicholas, O., & Hemingway, H. (2014). Comparison of random forest and parametric imputation models for imputing missing data using MICE: A CALIBER study. *American Journal of Epidemiology, 179*, 764-774. https://doi.org/10.1093/aje/kwt312.

Shataee, S., Kalbi, S., Fallah, A., & Pelz, D. (2012). Forest attribute imputation using machine-learning methods and ASTER data: Comparison of k-NN, SVR, and random forest regression algorithms. International Journal of Remote Sensing, 33, 6254-6280. https://doi.org/10.1080/01431161.2012.682661.

Shen, X., Mu, L., Li, Z., Wu, H., Gou, J., & Chen, X. (2016). Large-scale support vector machine classification with redundant data reduction. Neurocomputing, 172, 189-197. https://doi.org/10.1016/j.neucom.2014.10.102.

Shi, R., & Conrad, S. A. (2009). Correlation and regression analysis. *Annals of Allergy, Asthma & Immunology, 103*(4), S34-S41. https://doi.org/10.1016/j.anai.2009.05.005.

Smith, M. F. (2008). Assessment influence on peak power output and road cycling performance prediction. *International Journal of Sports Science & Coaching, 3*(2), 211-226. https://doi.org/10.1260/174795408785210465.

Staunton, C. A., Abt, G., Weaving, D., & Wundersitz, D. W. T. (2022). Misuse of the term 'load' in sport and exercise science. Journal of Science and Medicine in Sport, 25(5), 439-444. https://doi.org/10.1016/j.jsams.2021.08.013.

Steele, J. (2020). *What is (perception of) effort? Objective and subjective effort during attempted task performance* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/kbyhmResearchGate.

Sweeting, A. J., Cormack, S. J., Morgan, S., & Aughey, R. J. (2017). When is a sprint a sprint? A review of the analysis of team-sport athlete activity profile. *Frontiers in Physiology, 8*, 432. https://doi.org/10.3389/fphys.2017.00432.

Tanni, S. E., Patino, C. M., & Ferreira, J. C. (2020). Correlation vs. regression in association studies. Jornal Brasileiro de Pneumologia, 46(1), e20200030. https://doi.org/10.1590/1806-3713/e20200030.

Taylor, R. J., Sanders, D., Myers, T., Abt, G., Taylor, C. A., & Akubat, I. (2018). The dose-response relationship between training load and aerobic fitness in academy rugby union players. *International Journal of Sports Physiology and Performance, 13*(2), 163-169. https://doi.org/10.1123/ijspp.2017-0220.

Thomas, T., & Rajabi, E. (2021). Addressing missing data in a healthcare dataset using an improved k-Nn algorithm. In *International Conference on Computational Science* (pp. 223-230). Cham: Springer International Publishing.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological), 58*(1), 267-288.

Tierney, P. J., Young, A., Clarke, N. D., & Duncan, M. J. (2016). Match play demands of 11 versus 11 professional football using Global Positioning System tracking: Variations across common playing formations. *Human Movement Science, 49*, 1-8. https://doi.org/10.1016/j.humov.2016.02.004.

Tierney, P., Blake, C., & Delahunt, E. (2021). Physical characteristics of different professional rugby union competition levels. *Journal of Science and Medicine in Sport, 24*(12), 1267-1271. https://doi.org/10.1016/j.jsams.2021.05.009.

Torres-Ronda, L., Beanland, E., Whitehead, S., Sweeting, A., & Clubb, J. (2022). Tracking systems in team sports: A narrative review of applications of the data and sport-specific analysis. *Sports Medicine – Open, 8*(1), 15. https://doi.org/10.1186/s40798-022-00408-z.

Tuft, K., & Kavaliauskas, M. (2020). Relationship between internal and external training load in field hockey. International Journal of Strength and Conditioning, 1(1), 24. https://doi.org/10.47206/ijsc.v1i1.24.

Van Der Kruk, E., Van Der Helm, F. C. T., Veeger, H. E. J., & Schwab, A. L. (2018). Power in sports: A literature review on the application, assumptions, and terminology of mechanical power in sport research. *Journal of Biomechanics, 79*, 1-14. https://doi.org/10.1016/j.jbiomech.2018.07.013.

Verpoort, P. C., MacDonald, P., & Conduit, G. J. (2018). Materials data validation and imputation with an artificial neural network. Computational Materials Science, 145, 1-11. https://doi.org/10.1016/j.commatsci.2018.02.002.

Vescovi, J. D. (2015). Physical demands of regular season and playoff matches in professional women's soccer: A pilot from the Female Athletes in Motion (FAiM) study. In *International Research in Science and Soccer II* (pp. 12). Routledge.

Viana, B. F., Pires, F. O., Inoue, A., Micklewright, D., & Santos, T. M. (2016). Correlates of mood and RPE during multi-lap off-road cycling. Applied Psychophysiology and Biofeedback, 41(1), 1–7. https://doi.org/10.1007/s10484-015-9305-x.

Vigh-Larsen, J. F., Dalgas, U., & Andersen, T. B. (2018). Position-specific acceleration and deceleration profiles in elite youth and senior soccer players. *The Journal of Strength & Conditioning Research, 32*(4), 1114-1122. https://doi.org/10.1519/JSC.0000000000002014.

Vogt, S., Heinrich, L., Schumacher, Y. O., Blum, A., Roecker, K. A. I., Dickhuth, H. H., & Schmid, A. (2006). Power output during stage racing in professional road cycling. *Medicine & Science in Sports & Exercise, 38*(1), 147-151. https://doi.org/10.1249/01.mss.0000181667.68575.6f.

Waldmann, P., Mészáros, G., Gredler, B., Fürst, C., & Sölkner, J. (2013). Evaluation of the lasso and the elastic net in genome-wide association studies. Frontiers in Genetics, 4, 270. https://doi.org/10.3389/fgene.2013.00270.

Waldron, M., & Highton, J. (2014). Fatigue and pacing in high-intensity intermittent team sport: An update. *Sports Medicine, 44*(12), 1645-1658. https://doi.org/10.1007/s40279-014-0176-7.

Waljee, A. K., Mukherjee, A., Singal, A. G., Zhang, Y., Warren, J., Balis, U., Marrero, J., Zhu, J., Higgins, P. D. (2013). Comparison of imputation methods for missing laboratory data in medicine. BMJ Open, 3(8), e002847. https://doi.org/10.1136/bmjopen-2013-002847.

Waterhouse, J., Reilly, T., & Edwards, B. (2004). The stress of travel. *Journal of Sports Sciences*, *22*(10), 946–966. https://doi.org/10.1080/02640410400000264.

West, S. W., Clubb, J., Torres-Ronda, L., Howells, D., Leng, E., Vescovi, J. D., Carmody, S., Posthumus, M., Dalen-Lorentsen, T., Windt, J. (2021). More than a metric: How training load is used in elite sport for athlete management. *International Journal of Sports Medicine, 42*, 300–306. https://doi.org/10.1055/a-1268-8791.

Wheeler, W. K., Wiseman, R., & Lyons, K. (2011). Tactical and technical factors associated with effective ball offloading strategies during the tackle in rugby league. *International Journal of Performance Analysis in Sport, 11*, 392–409. https://doi.org/10.1080/24748668.2011.11868797.

Williams, S., Trewartha, G., Cross, M. J., Kemp, S. P. T., & Stokes, K. A. (2017). Monitoring what matters: A systematic process for selecting training-load measures. *International Journal of Sports Physiology and Performance, 12*(1), 101–106. https://doi.org/10.1123/ijspp.2016-0144.

Windt, J., & Gabbett, T. J. (2017). How do training and competition workloads relate to injury? The workload—injury aetiology model. *British Journal of Sports Medicine, 51*(5), 428-435. https://doi.org/10.1136/bjsports-2016-096547.

Windt, J., Ardern, C. L., Gabbett, T. J., Khan, K. M., Cook, C. E., Sporer, B. C., & Zumbo, B. D. (2018). Getting the most out of intensive longitudinal data: A methodological review of workload–injury studies. BMJ Open, 8, e022626. https://doi.org/10.1136/bmjopen-2018-022626.

Windt, J., MacDonald, K., Taylor, D., Zumbo, B. D., Sporer, B. C., & Martin, D. T. (2020). "To tech or not to tech?" A critical decision-making framework for implementing technology in sport. *Journal of Athletic Training, 55*(9), 902–910. https://doi.org/10.4085/1062-6050-0003.20.

Winter, E. M., Abt, G., Brookes, F. C., Challis, J. H., Fowler, N. E., Knudson, D. V., Knuttgen, H. G., Kraemer, W. J., Lane, A. M., van Mechelen, W., Morton, R. H., Newton, R. U., Williams, C., & Yeadon, M. R. (2016). Misuse of "power" and other mechanical terms in sport and exercise science research. *The Journal of Strength & Conditioning Research, 30*(1), 292-300. https://doi.org/10.1519/JSC.0b013e3182941e96.

World Rugby. (2025a). Rugby Sevens - Olympic dreams. World Rugby. https://www.world.rugby/the-game/beginners-guide/sevens. Accessed February 13, 2025.

World Rugby. (2025b). Approved Devices. World Rugby. https://www.world.rugby/the-game/facilities-equipment/equipment/devices/. Accessed February 13, 2025.

World Rugby. (2025c). Women's Rugby. World Rugby. https://www.world.rugby/organisation/about-us/womens. Accessed February 13, 2025.

Yin, M., Wortman Vaughan, J., & Wallach, H. (2019). Understanding the effect of accuracy on trust in machine learning models. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 9-29. https://doi.org/10.1145/3290605.3300509.

Zelik, K. E., & Kuo, A. D. (2012). Mechanical work as an indirect measure of subjective costs influencing human movement. *PloS One, 7*(2), e31143. https://doi.org/10.1371/journal.pone.0031143.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x.

Appendix 1

This appendix contains information relating to the general dataset used for analysis across the dissertation. Global summary data for variables are reported in Tables A.1.1 and A.1.2 as well as Figures A.1.1-A.1.22 (inclusive). A correlation matrix for variables is reported in Figure A.1.23.

Table A.1.1: Average values of key exploratory variables.

Metric	Average Value (± Standard Deviation)
Athlete Age	25.5 ± 3.90 years
Athlete Height	169.4 ± 5.89 m
Athlete Mass	71.0 ± 5.64 kg
Playing Time	11.04 ± 4.733 mins
Contact Count	4.69 ± 2.870 contacts
sRPE	7 ± 1.9 au
sRPE-CL	79.6 ± 45.59 au
Total Distance	1082.24 ± 447.996 m
Mechanical Work	56236.38 ± 21413.36 J
SDC Athlete Load	7.77 ± 1.596 au
Distance at Low Deceleration-Low Speed	259.4 ± 130.09 m
Distance at High Deceleration-Low Speed	2.9 ± 9.40 m
Distance at Low Acceleration-Low Speed	11.2 ± 5.66 m
Distance at High Acceleration-Low Speed	2.0 ± 1.30 m
Distance at Low Deceleration-Moderate Speed	703.6 ± 298.06 m
Distance at High Deceleration-Moderate Speed	16.4 ± 7.00 m
Distance at Low Acceleration-Moderate Speed	21.7 ± 9.62 m
Distance at High Acceleration-Moderate Speed	16.8 ± 7.39 m
Distance at Low Deceleration-High Speed	82.5 ± 52.29 m
Distance at High Deceleration-High Speed	1.8 ± 1.31 m
Distance at Low Acceleration-High Speed	2.4 ± 1.60 m
Distance at High Acceleration-High Speed	2.6 ± 1.64 m

Table A.1.2: Summary values of key exploratory variables.

Metric	Metric Level	Count
Number of Matches		103 matches
Match Category	*Pool Play*	13 matches
	Cup Quarterfinal	12 matches
	Cup Semifinal	12 matches
	Cup Final	12 matches
	All Other Match Categories	54 matches
Opponent Rank	*Top Four*	24 matches
	Bottom Four	10 matches
	All Other Opponents	28 matches
Match Outcome	*Wins*	33 matches
	Losses	2 matches
	Ties	68 matches
Player Experience	*Year 1 (Rookie)*	46 player-matches
	Years 2-4	514 player-matches
	Year 5+	459 player-matches

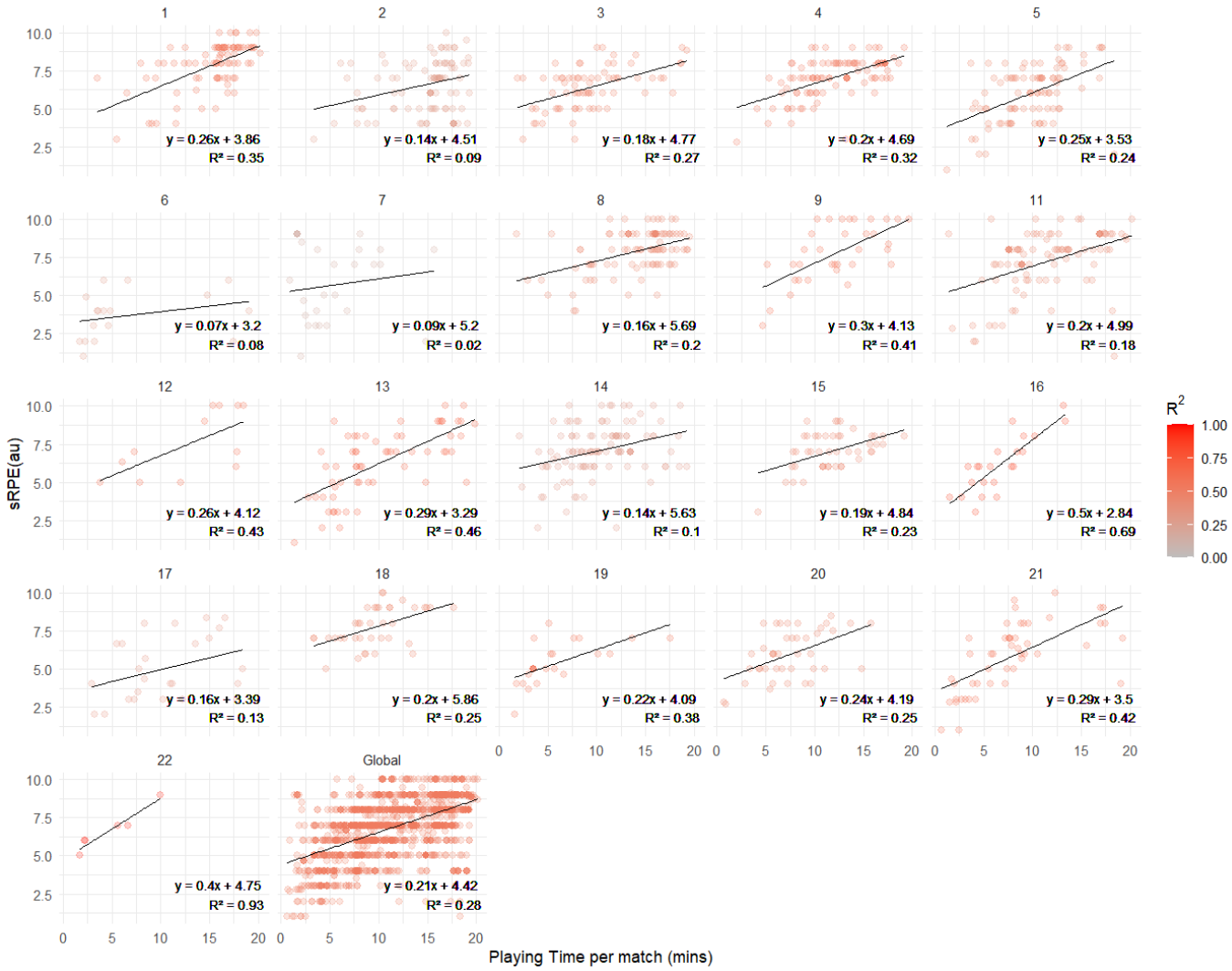

Figure A.1.1: Playing time (mins) and sRPE (au) per match, by athlete number with global data including regression line, equation, and R^2 values, as colour coded by R^2 value and opacity indicating frequency of reported value.

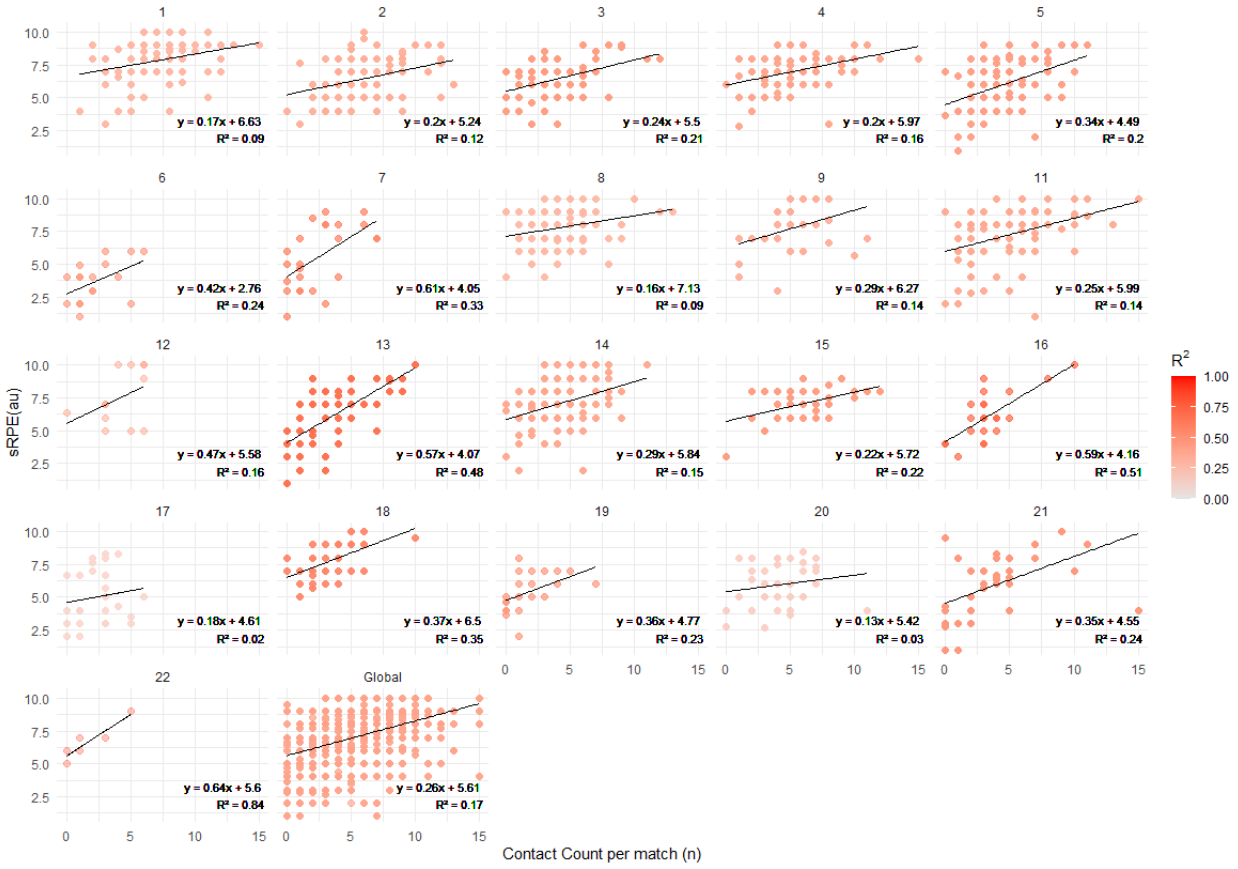

Figure A.2.2: Contact count (n) and sRPE (au) per match, by athlete number with global data including regression line, equation, and R^2 values, as colour coded by R^2 value and opacity indicating frequency of reported value.

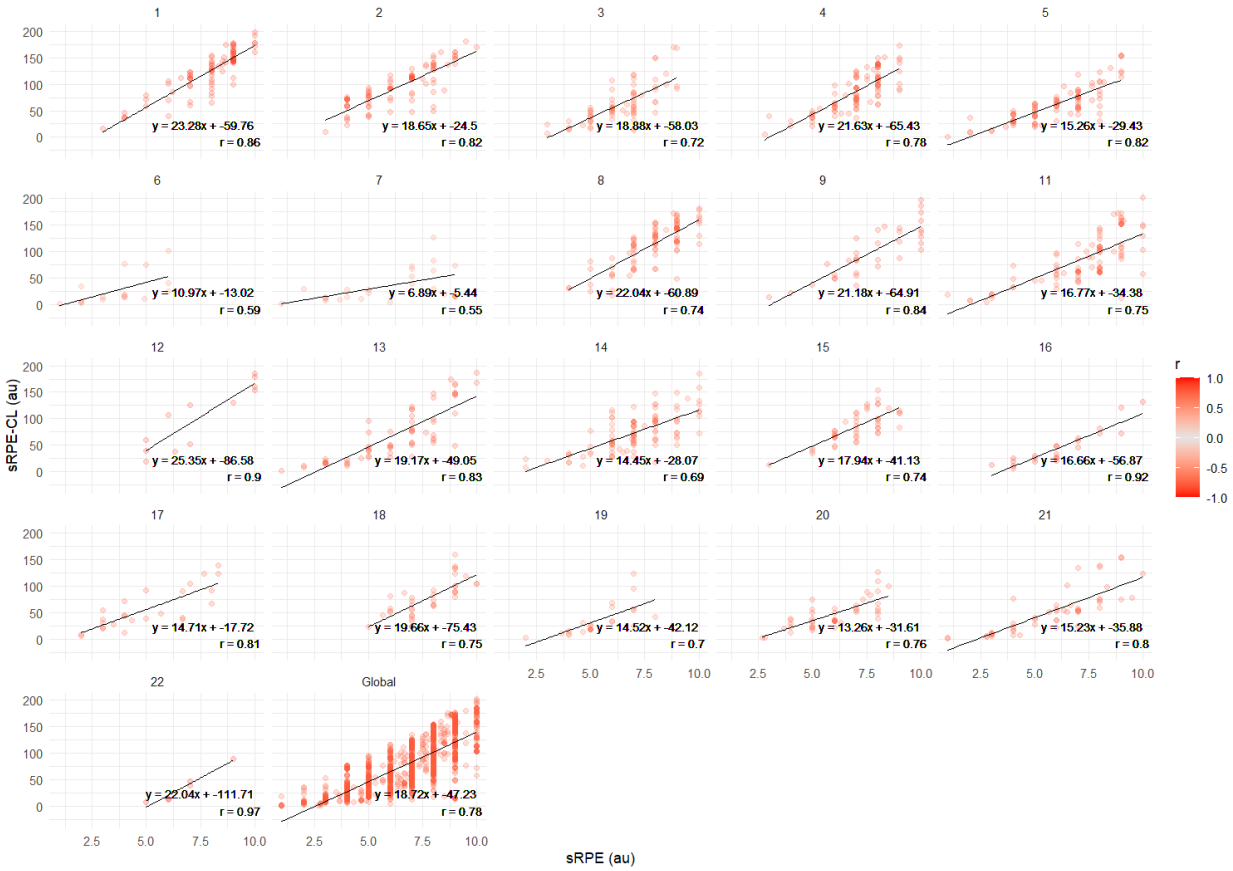

Figure A.2.3: sRPE-CL (au) and sRPE (au) per match, by athlete number with global data including regression line, equation, and r values, as colour coded by r value and opacity indicating frequency of reported value.

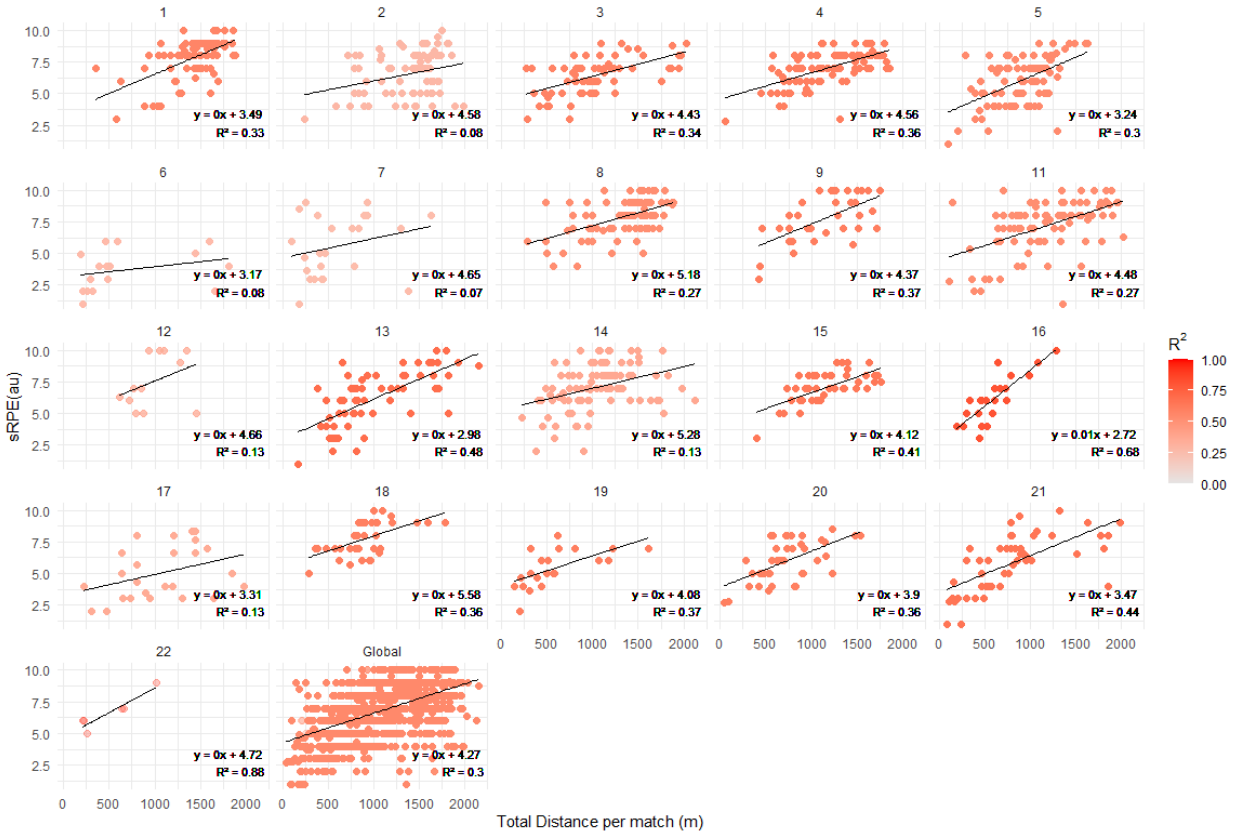

Figure A.1.4: Total distance (m) and sRPE (au) per match, by athlete number with global data including regression line, equation, and R^2 values, as colour coded by R^2 value and opacity indicating frequency of reported value.

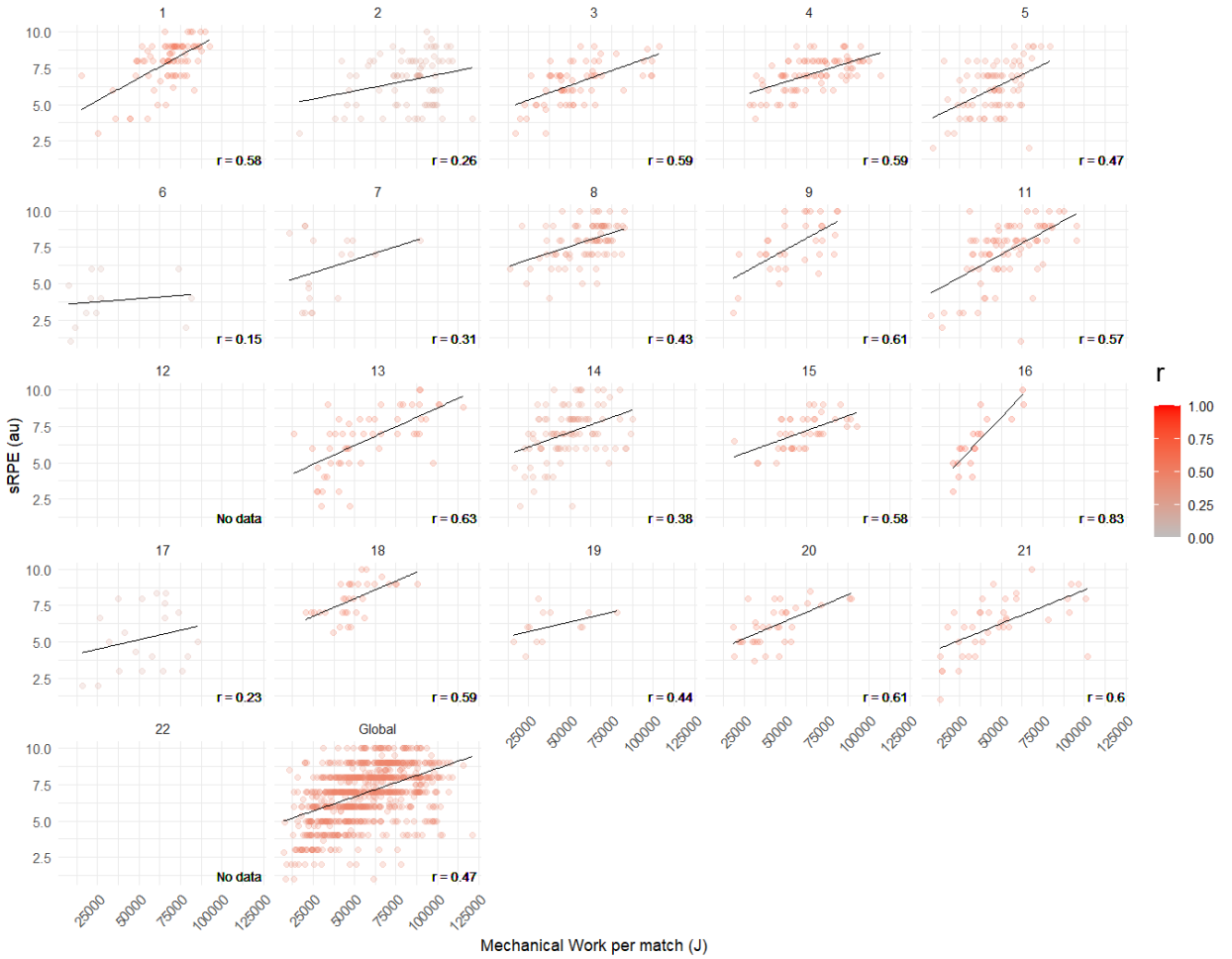

Figure A.1.5: Mechanical work (J) and sRPE (au) per match, by athlete number with global data including regression line, equation, and r values, as colour coded by r value and opacity indicating frequency of reported value.

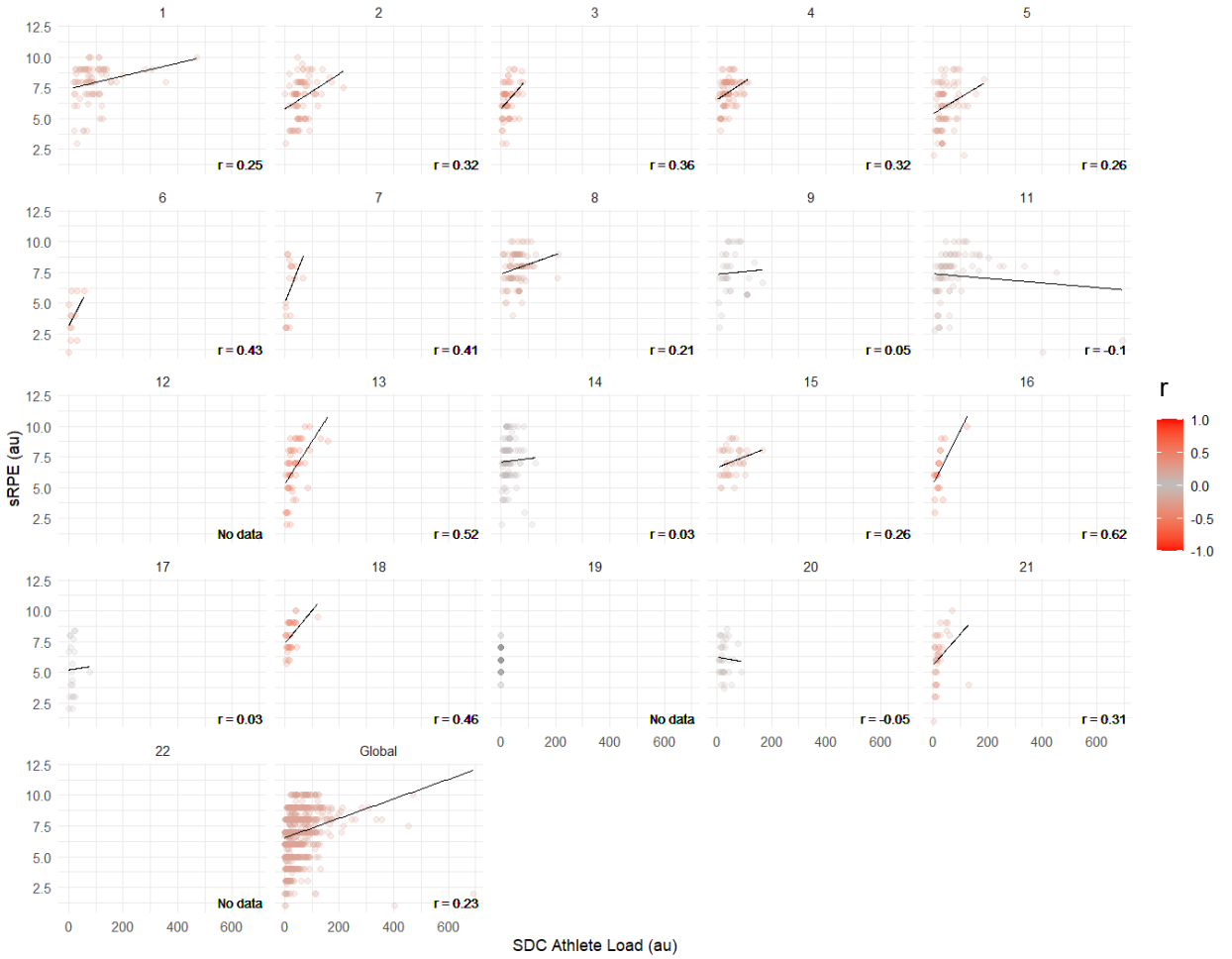

Figure A.1.6: SDC athlete load (au) and sRPE (au) per match, by athlete number with global data including regression line, equation, and r values, as colour coded by r value and opacity indicating frequency of reported value.

Figure A.1.7: Distance covered in low deceleration-low speed zone (m) and sRPE (au) per match, by athlete number with global data including regression line, equation, and r values, as colour coded by r value and opacity indicating frequency of reported value.

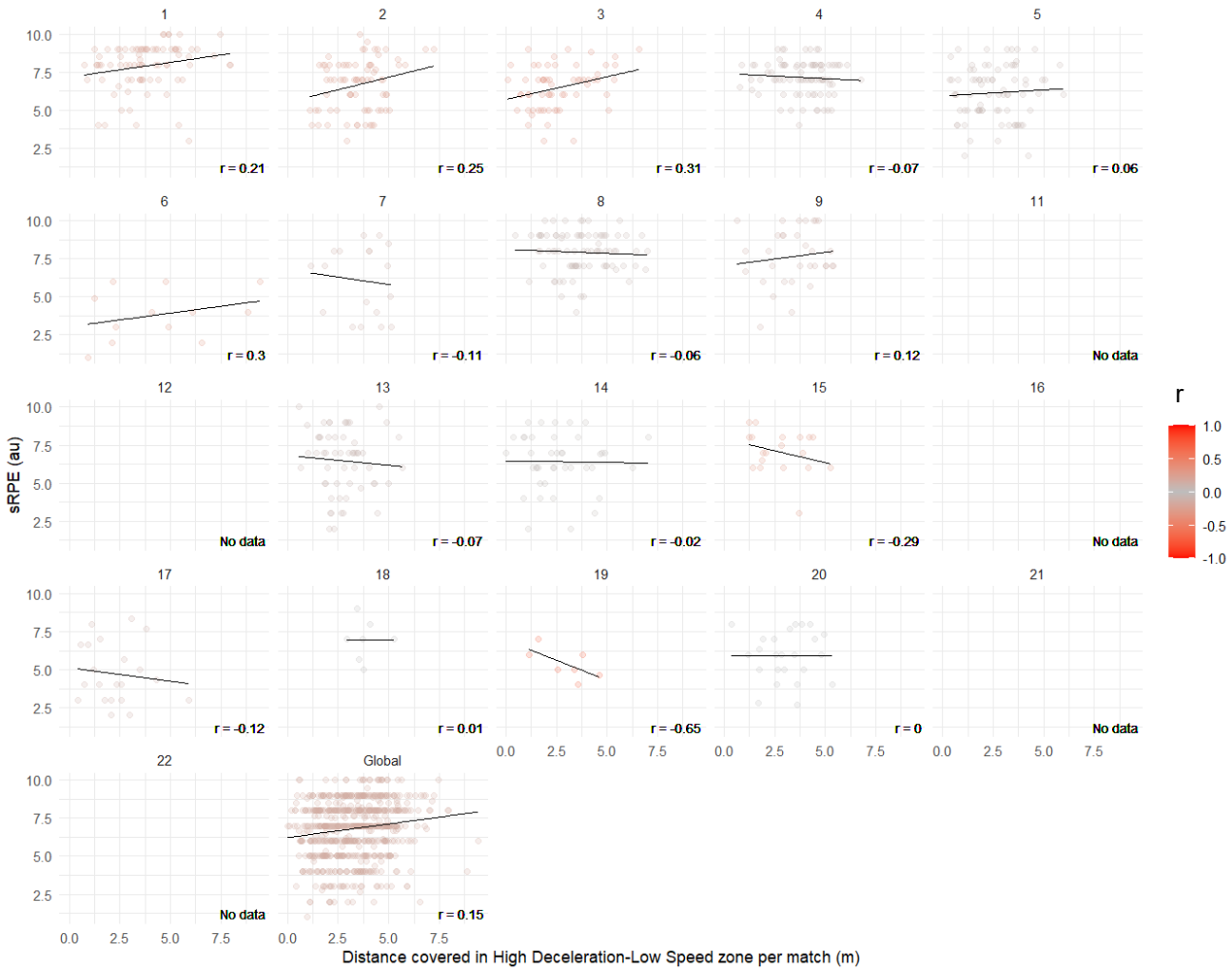

Figure A.1.8: Distance covered in high deceleration-low speed zone (m) and sRPE (au) per match, by athlete number with global data including regression line, equation, and r values, as colour coded by r value and opacity indicating frequency of reported value.

Figure A.1.9: Distance covered in low acceleration-low speed zone (m) and sRPE (au) per match, by athlete number with global data including regression line, equation, and r values, as colour coded by r value and opacity indicating frequency of reported value.

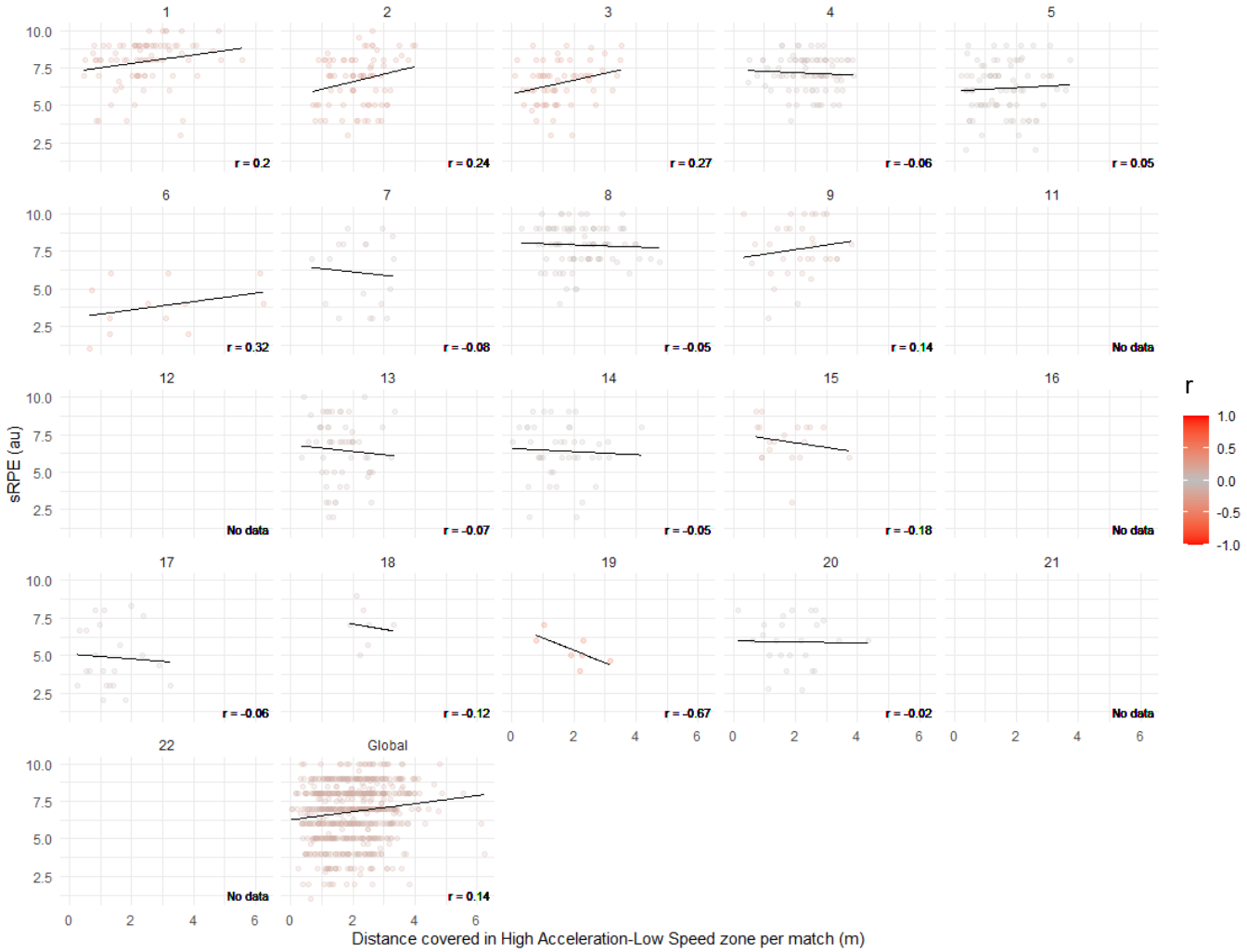

Figure A.1.10: Distance covered in high acceleration-low speed zone (m) and sRPE (au) per match, by athlete number with global data including regression line, equation, and r values, as colour coded by r value and opacity indicating frequency of reported value.

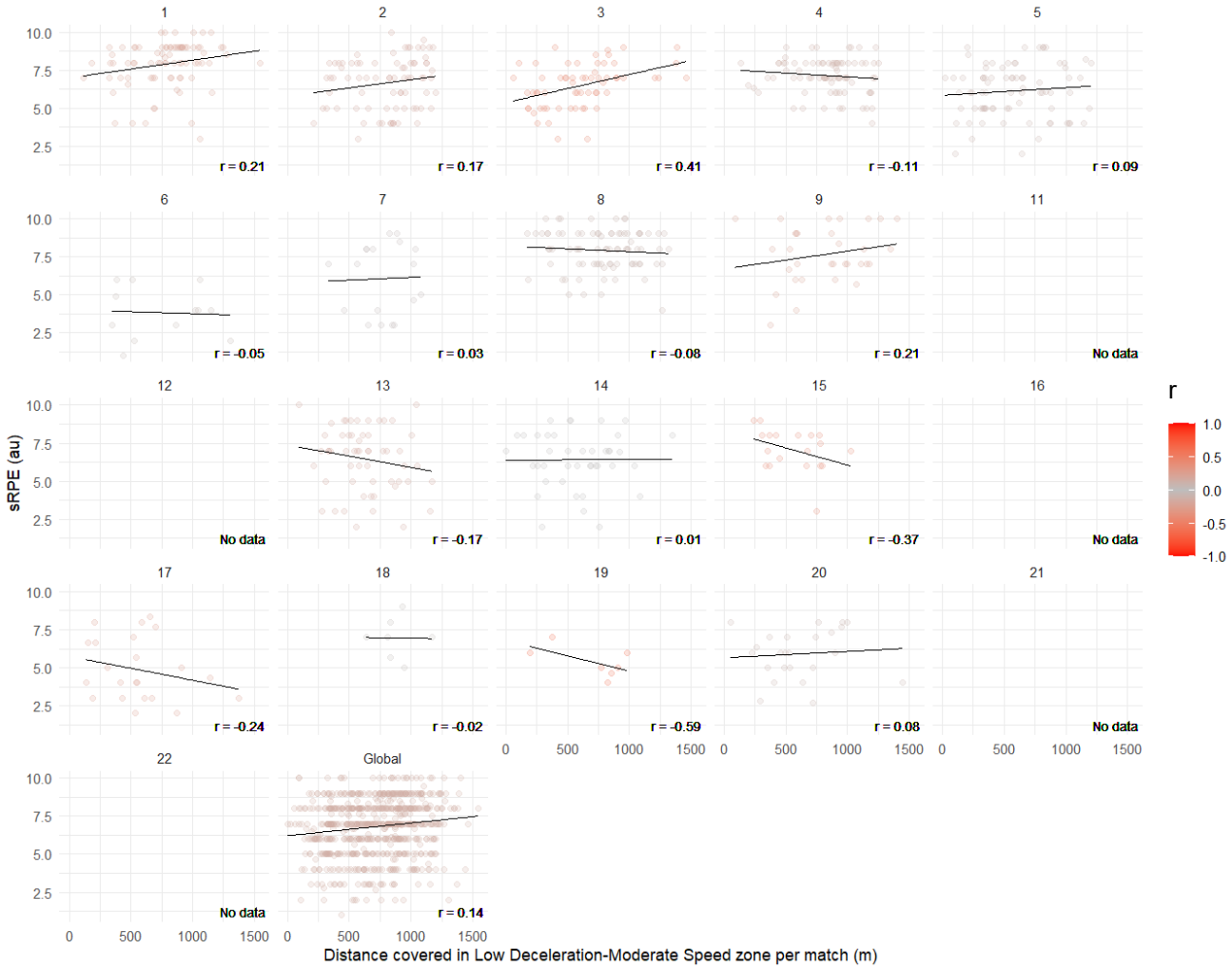

Figure A.1.11: Distance covered in low deceleration-moderate speed zone (m) and sRPE (au) per match, by athlete number with global data including regression line, equation, and r values, as colour coded by r value and opacity indicating frequency of reported value.

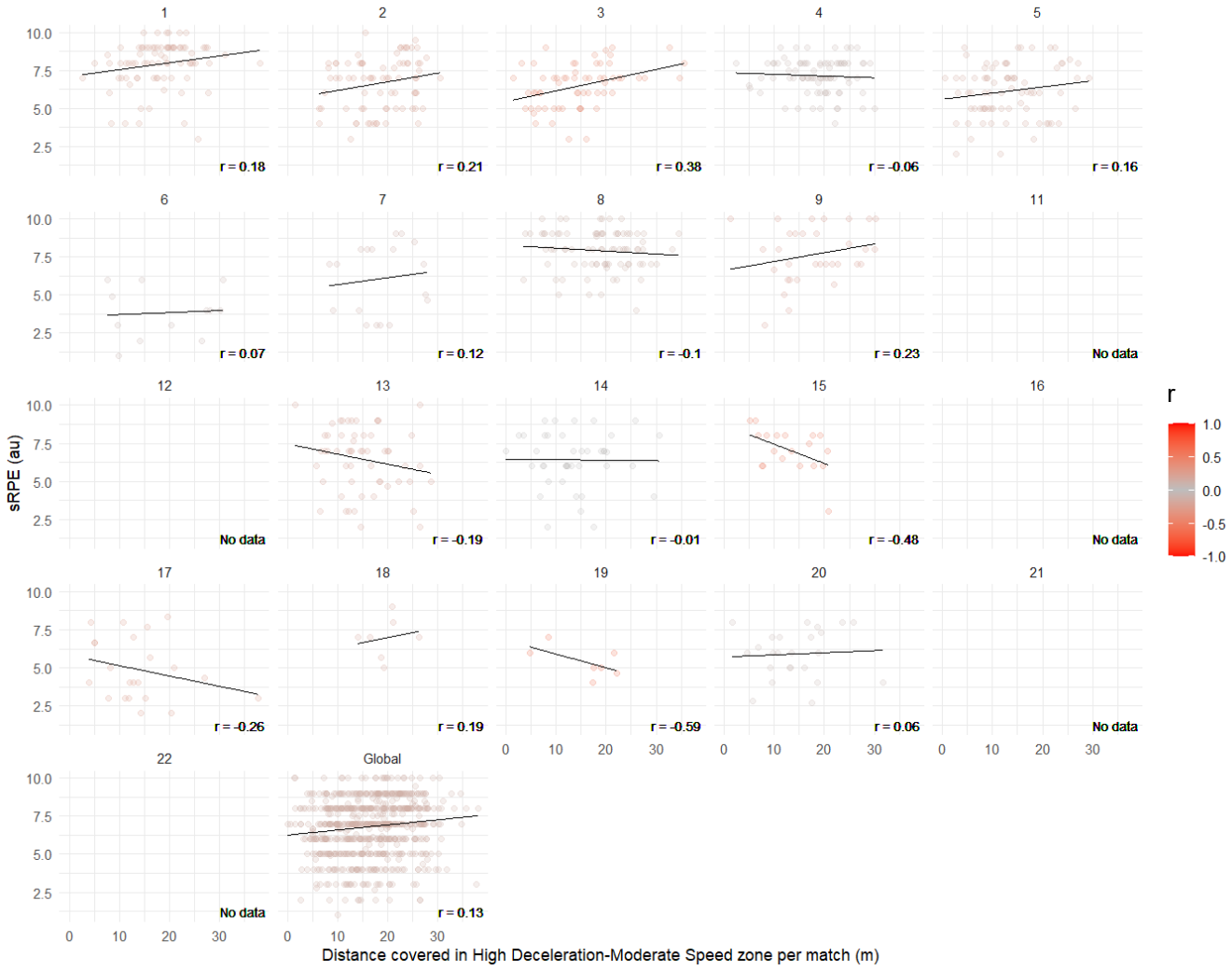

Figure A.1.12: Distance covered in high deceleration-moderate speed zone (m) and sRPE (au) per match, by athlete number with global data including regression line, equation, and r values, as colour coded by r value and opacity indicating frequency of reported value.

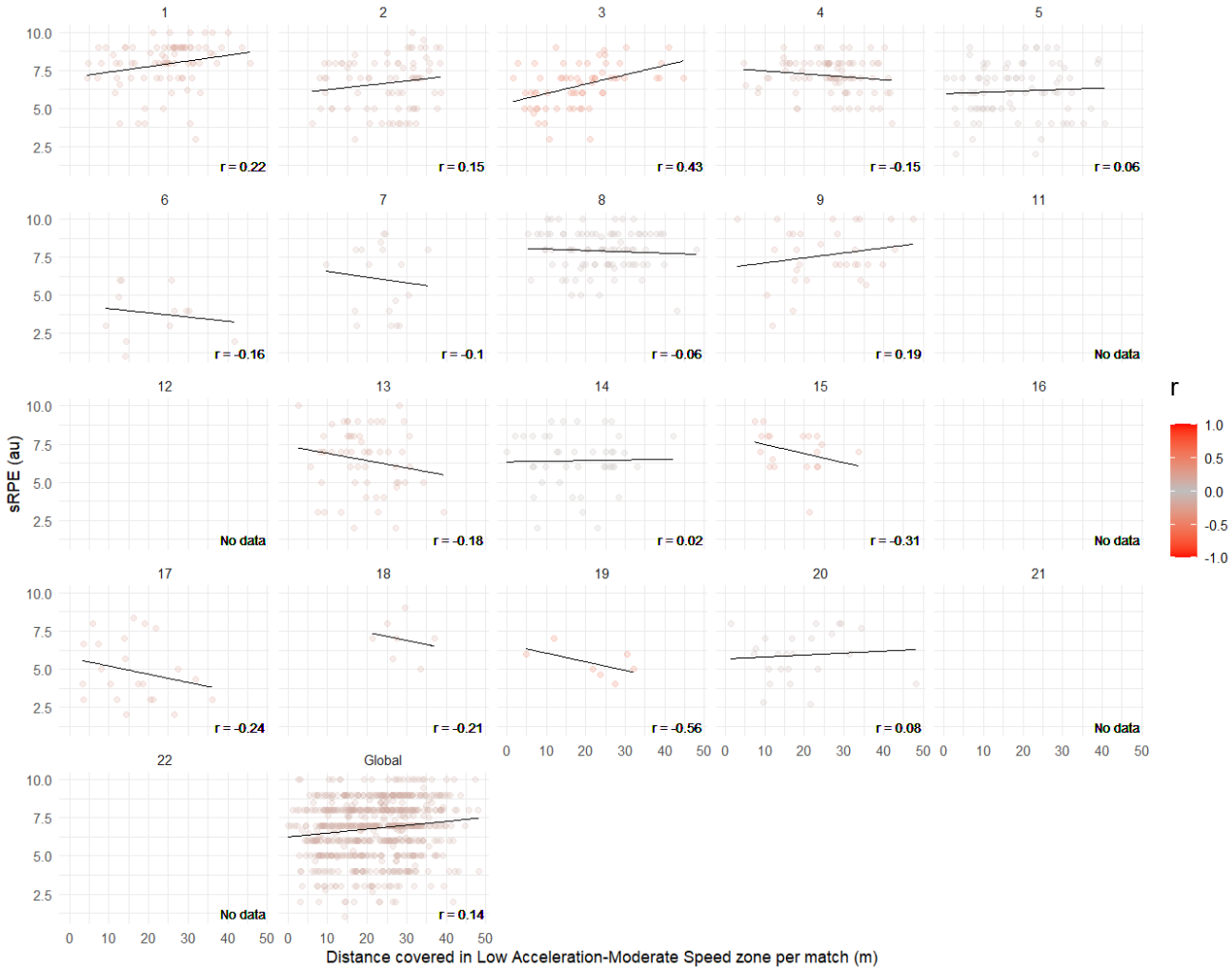

Figure A.1.13: Distance covered in low acceleration-moderate speed zone (m) and sRPE (au) per match, by athlete number with global data including regression line, equation, and r values, as colour coded by r value and opacity indicating frequency of reported value.

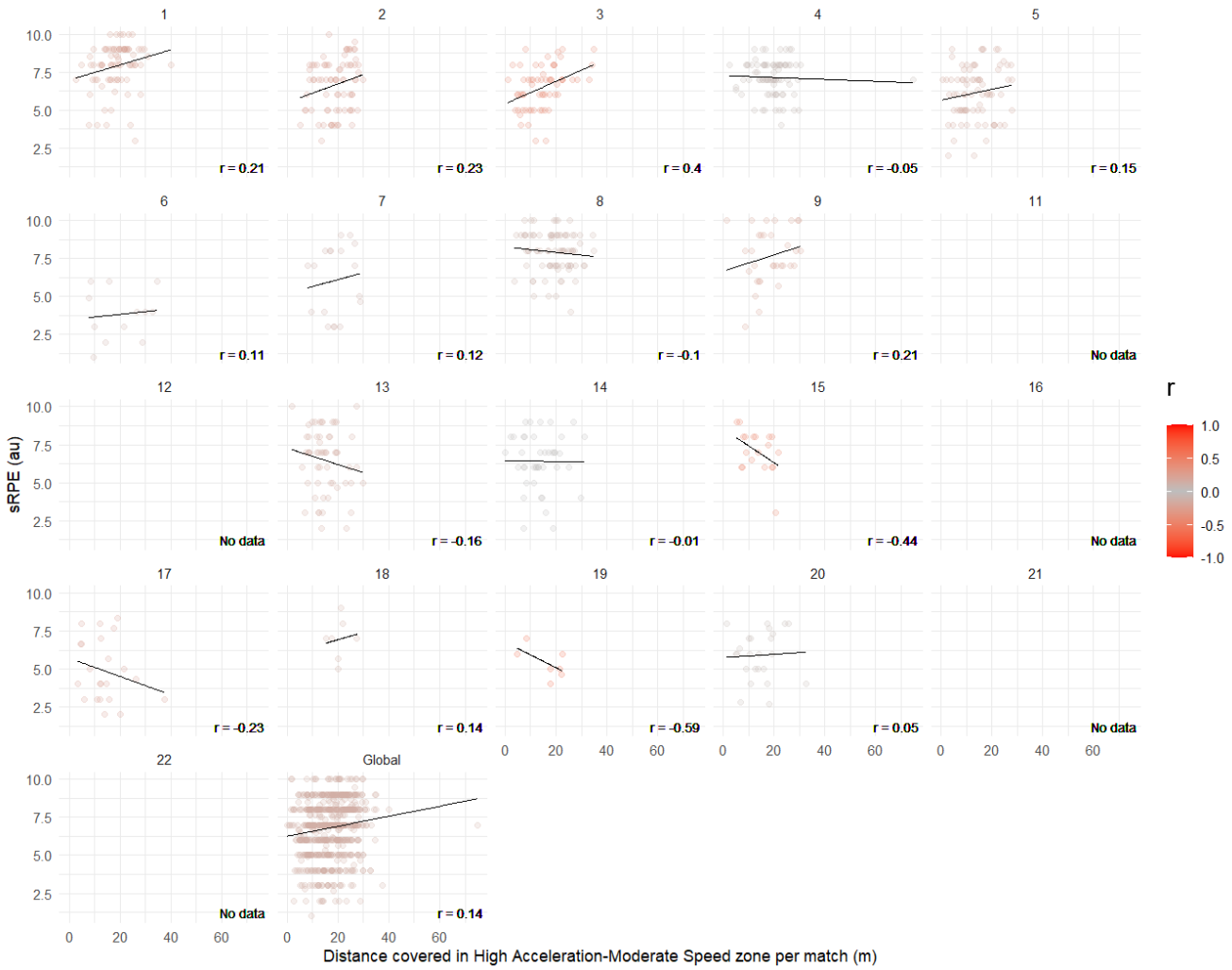

Figure A.1.14: Distance covered in high acceleration-moderate speed zone (m) and sRPE (au) per match, by athlete number with global data including regression line, equation, and r values, as colour coded by r value and opacity indicating frequency of reported value.

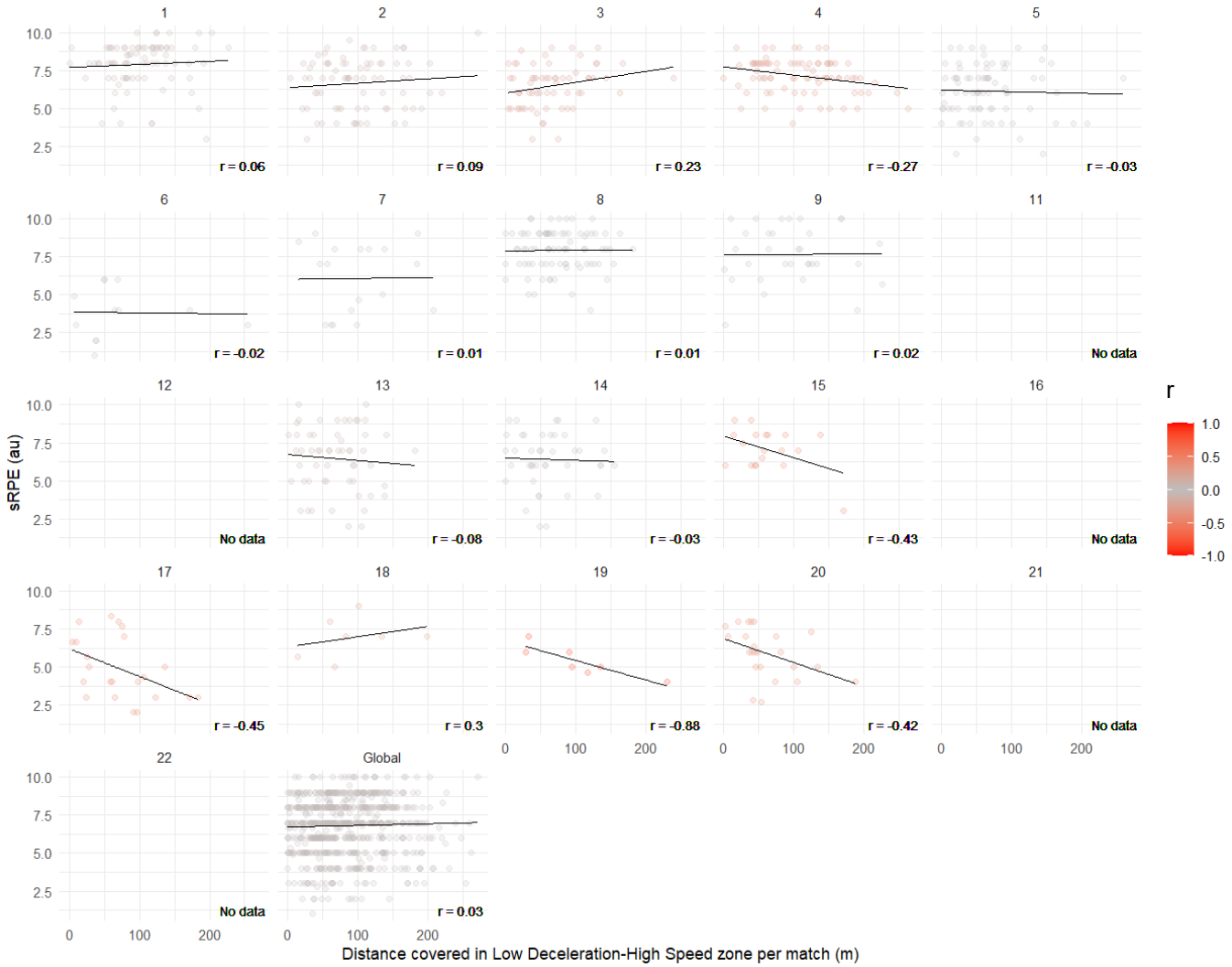

Figure A.1.15: Distance covered in low deceleration-high speed zone (m) and sRPE (au) per match, by athlete number with global data including regression line, equation, and r values, as colour coded by r value and opacity indicating frequency of reported value.

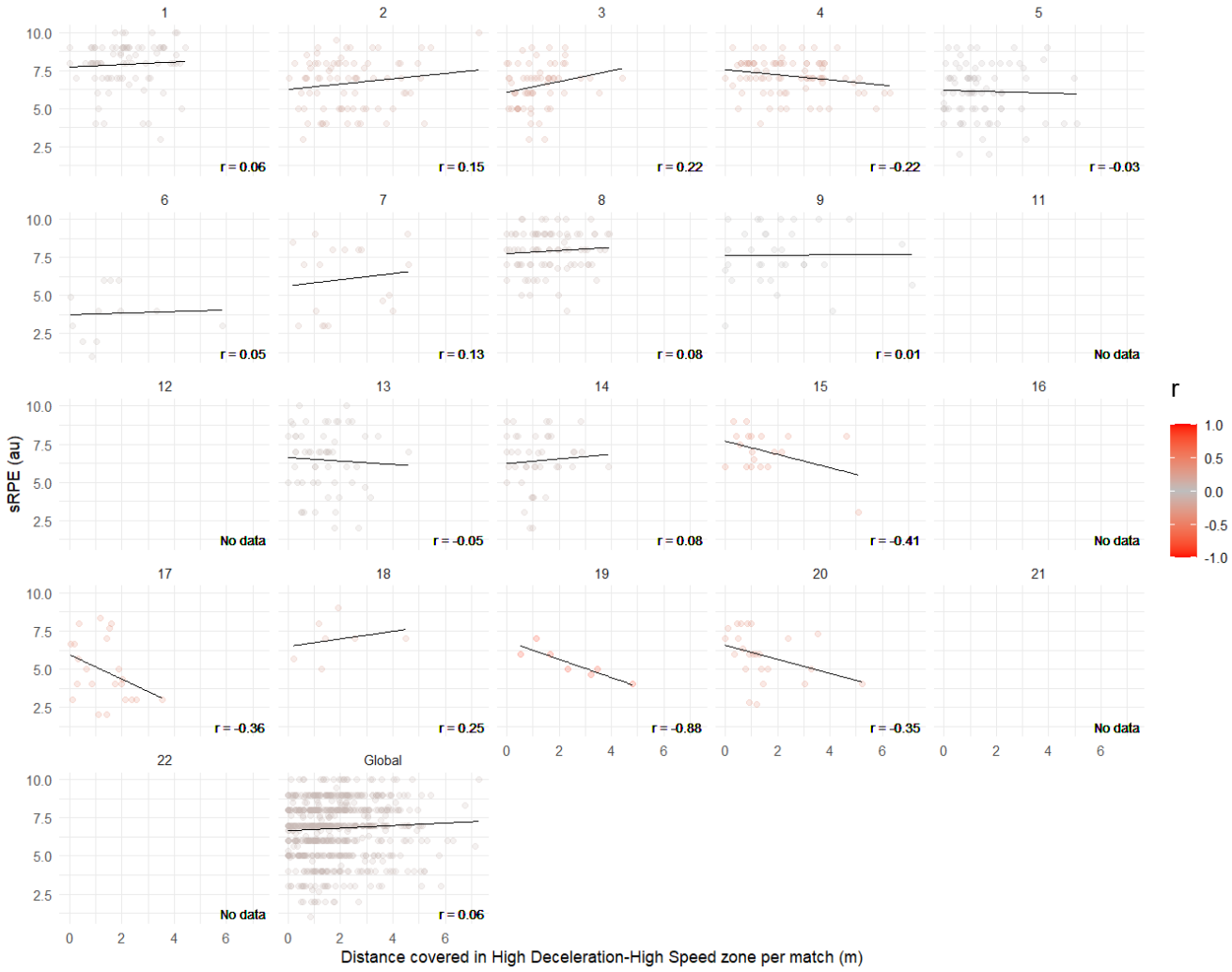

Figure A.1.16: Distance covered in high deceleration-high speed zone (m) and sRPE (au) per match, by athlete number with global data including regression line, equation, and r values, as colour coded by r value and opacity indicating frequency of reported value.

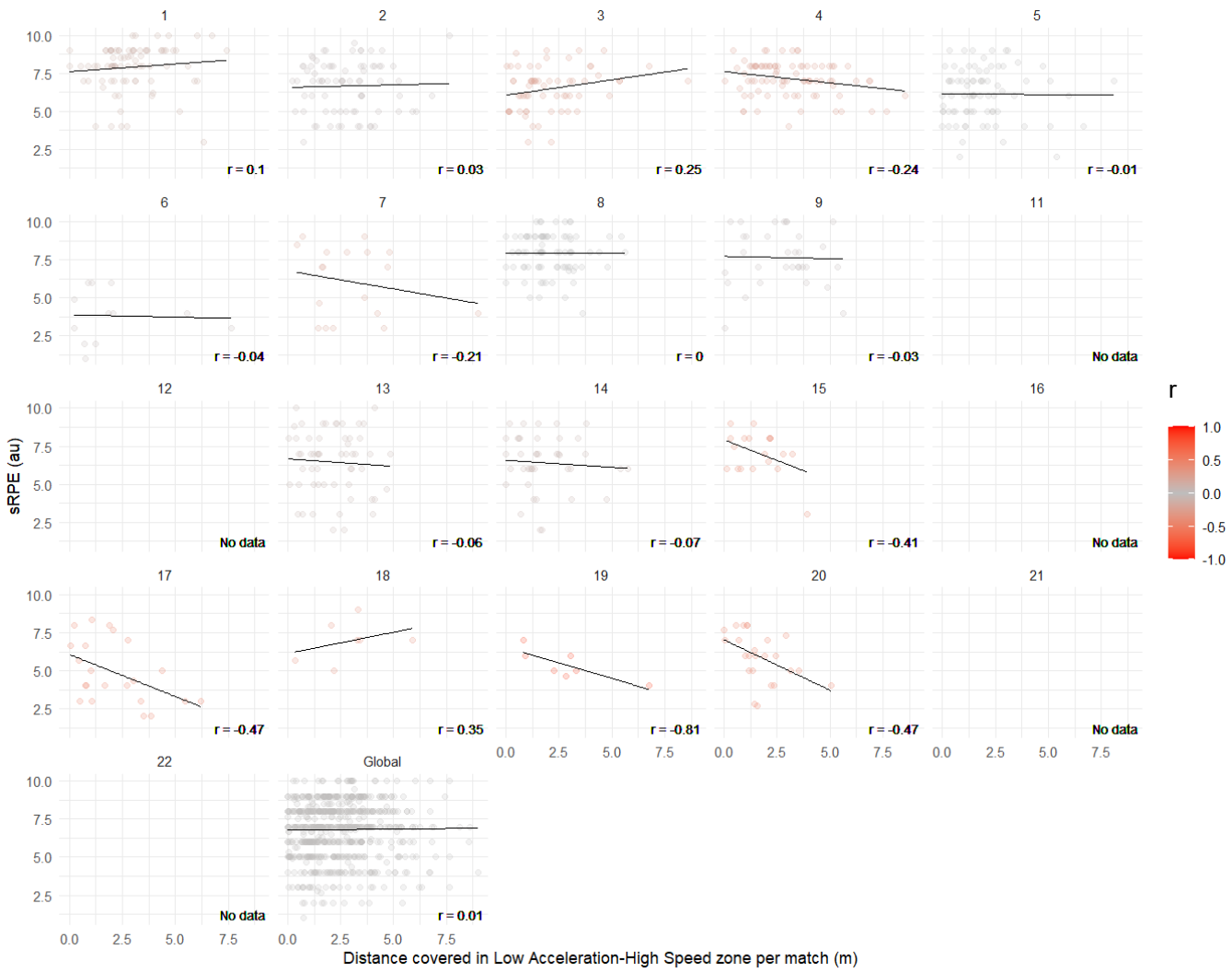

Figure A.1.17: Distance covered in low acceleration-high speed zone (m) and sRPE (au) per match, by athlete number with global data including regression line, equation, and r values, as colour coded by r value and opacity indicating frequency of reported value.

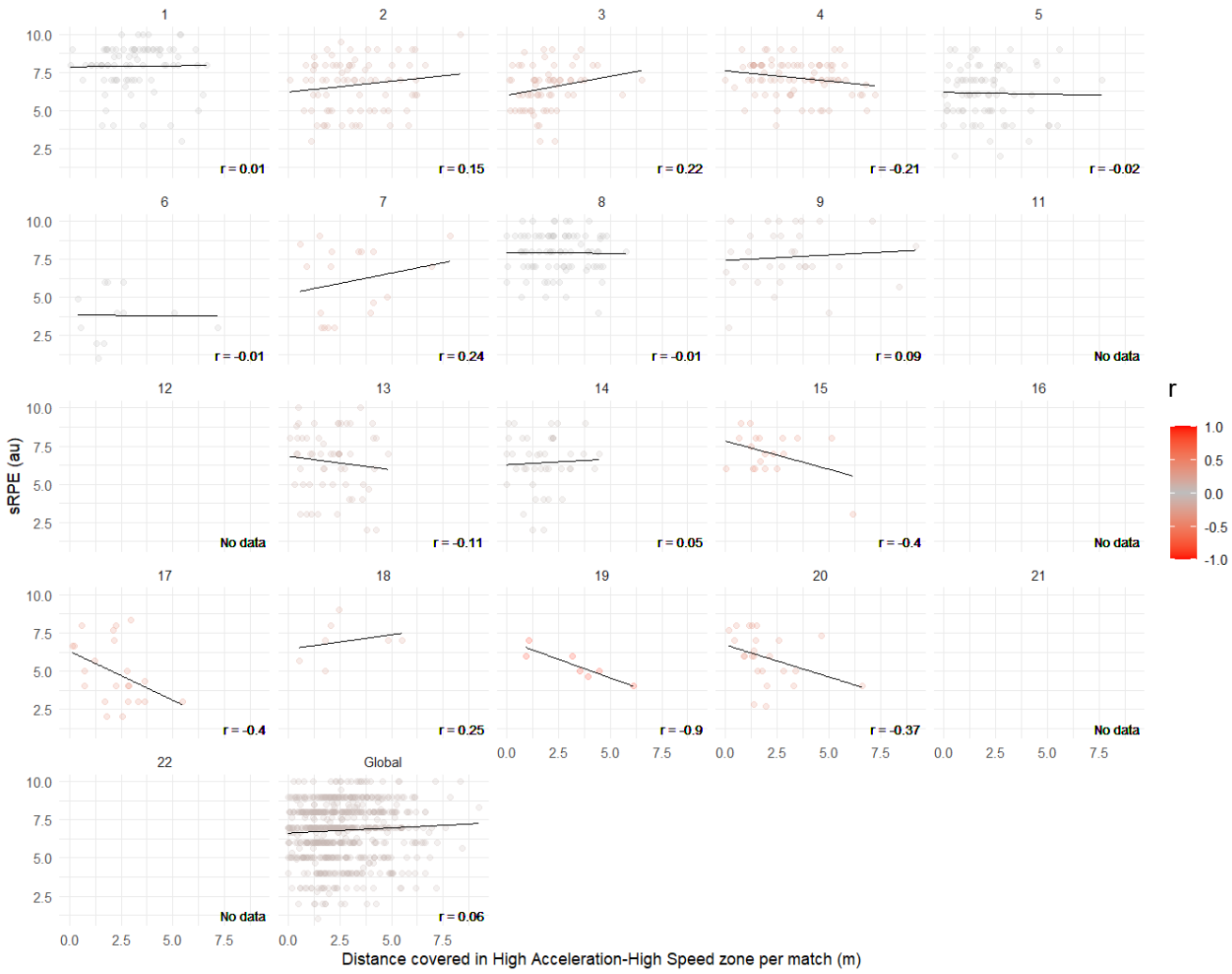

Figure A.1.18: Distance covered in high acceleration-high speed zone (m) and sRPE (au) per match, by athlete number with global data including regression line, equation, and r values, as colour coded by r value and opacity indicating frequency of reported value.

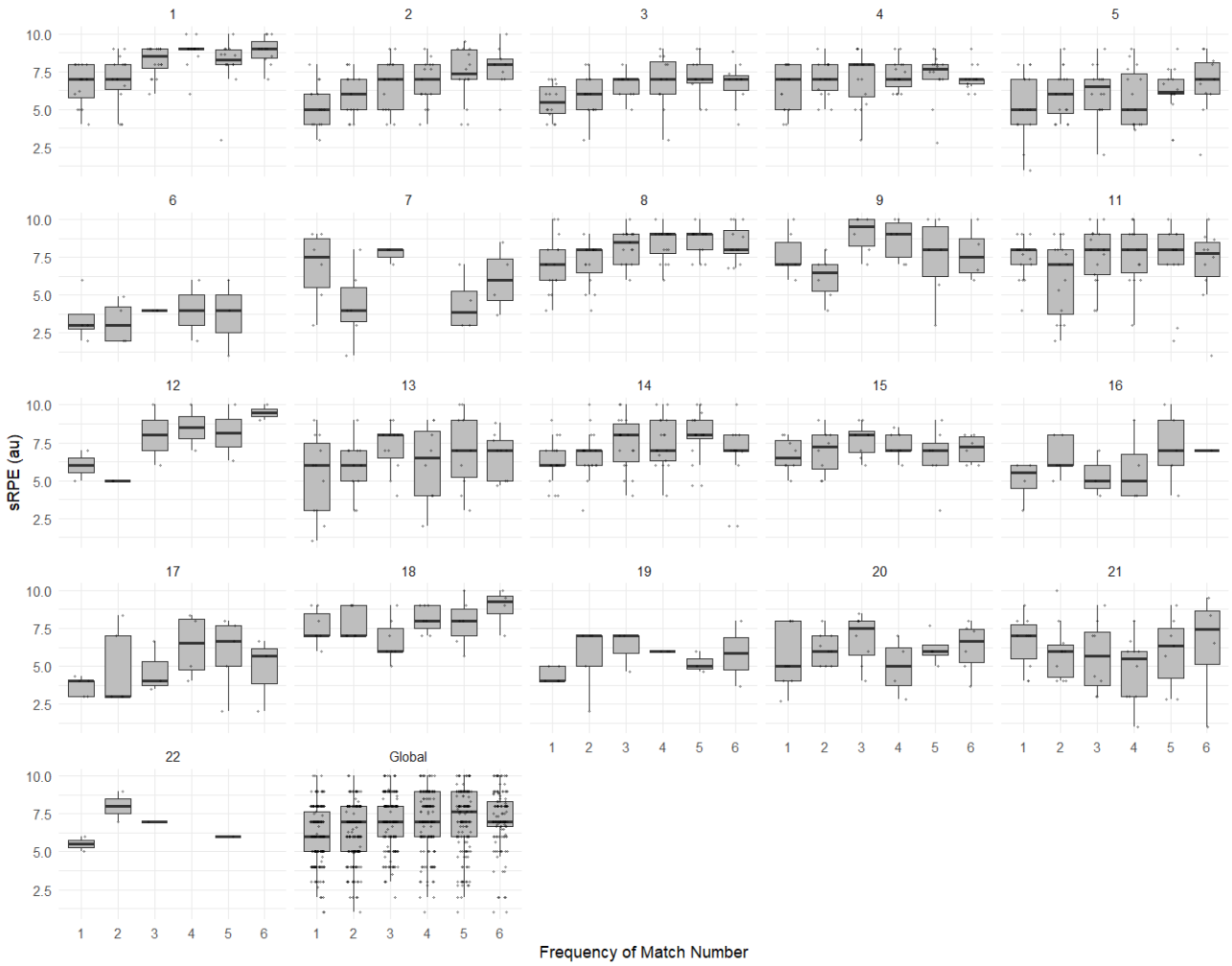

Figure A.1.19: Frequency of matches by match number and sRPE (au) per match, by athlete number with global data.

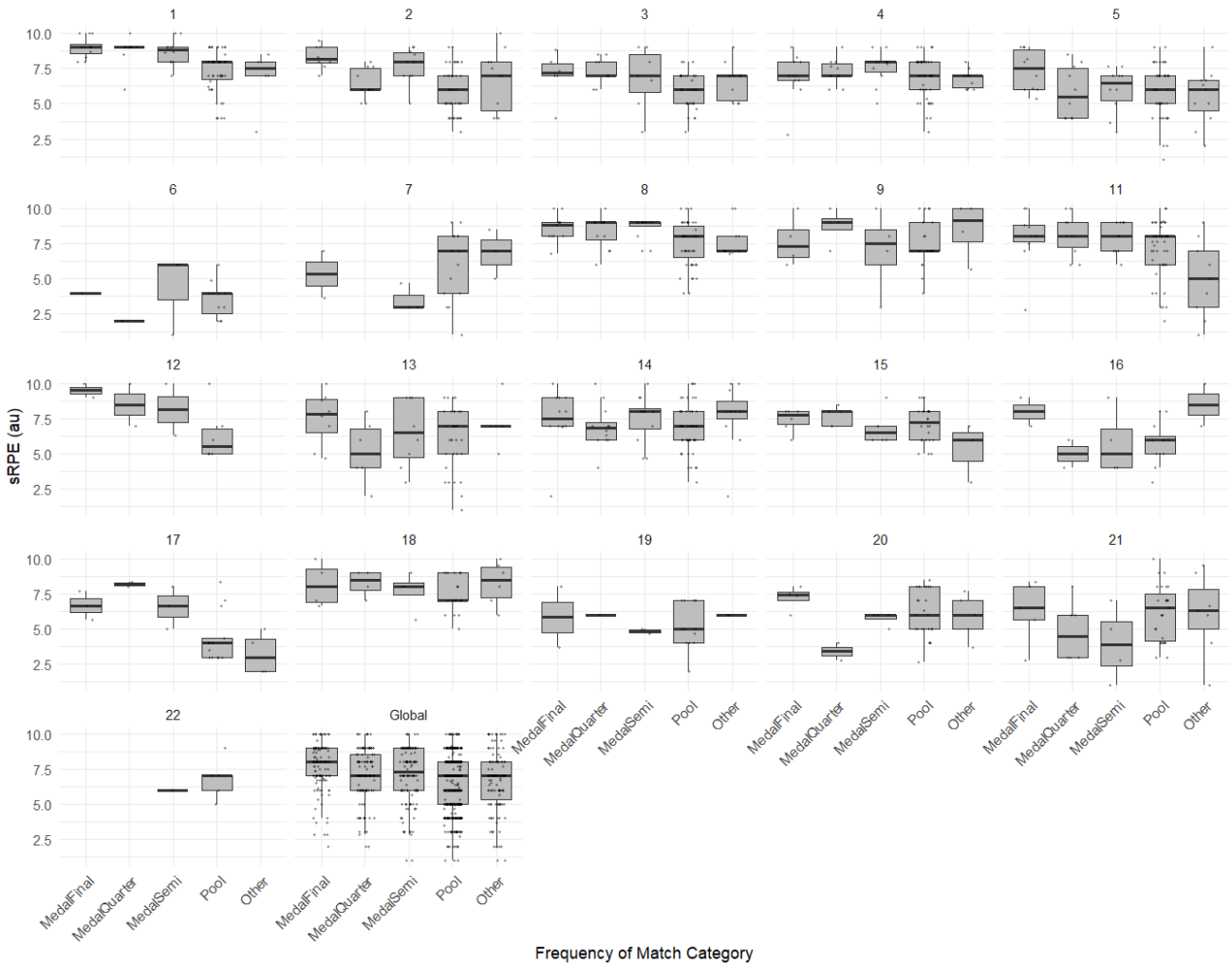

Figure A.1.20: Frequency of matches by match categories and sRPE (au) per match, by athlete number with global data.

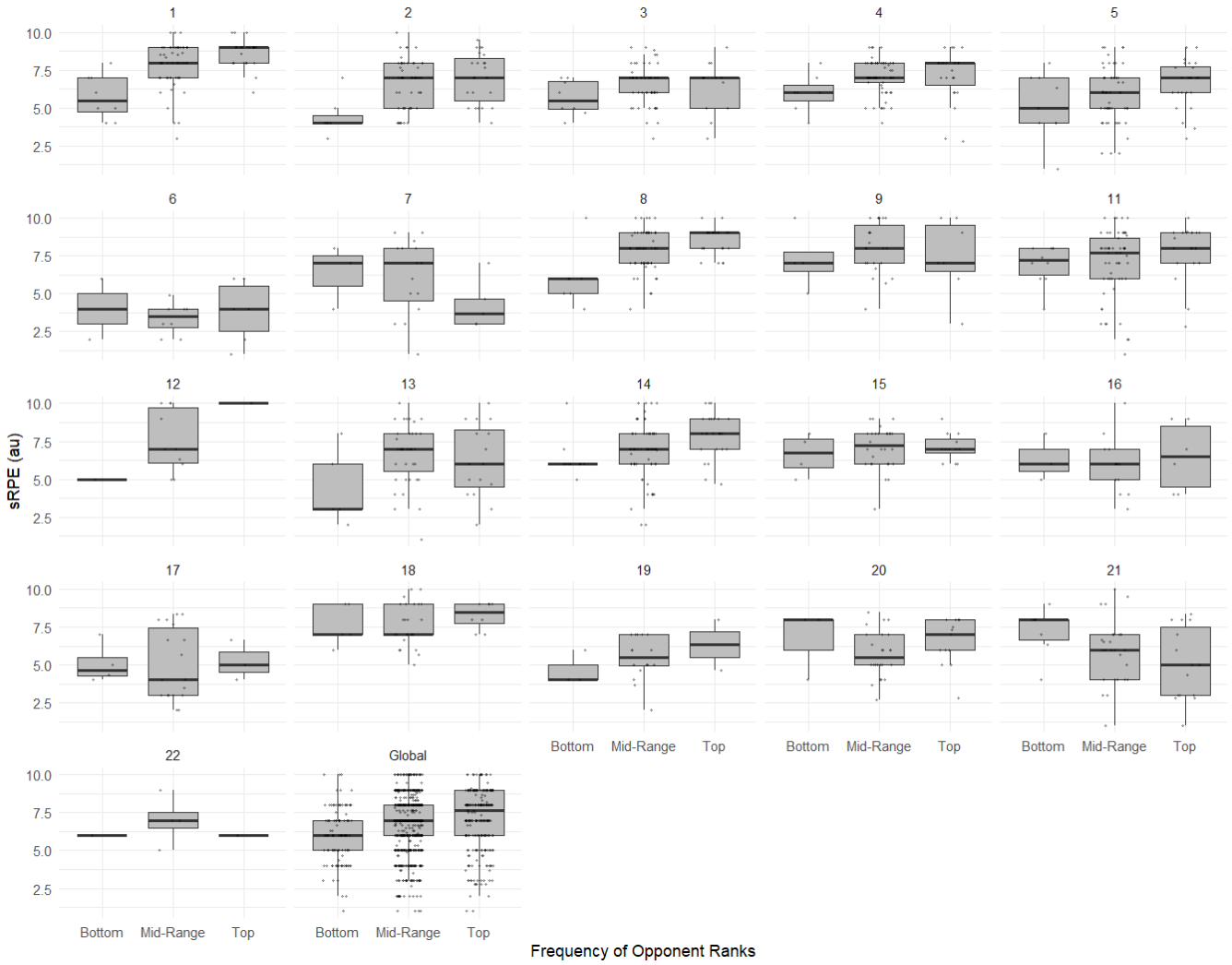

Figure A.1.21: Frequency of matches by opponent rank and sRPE (au) per match, by athlete number with global data.

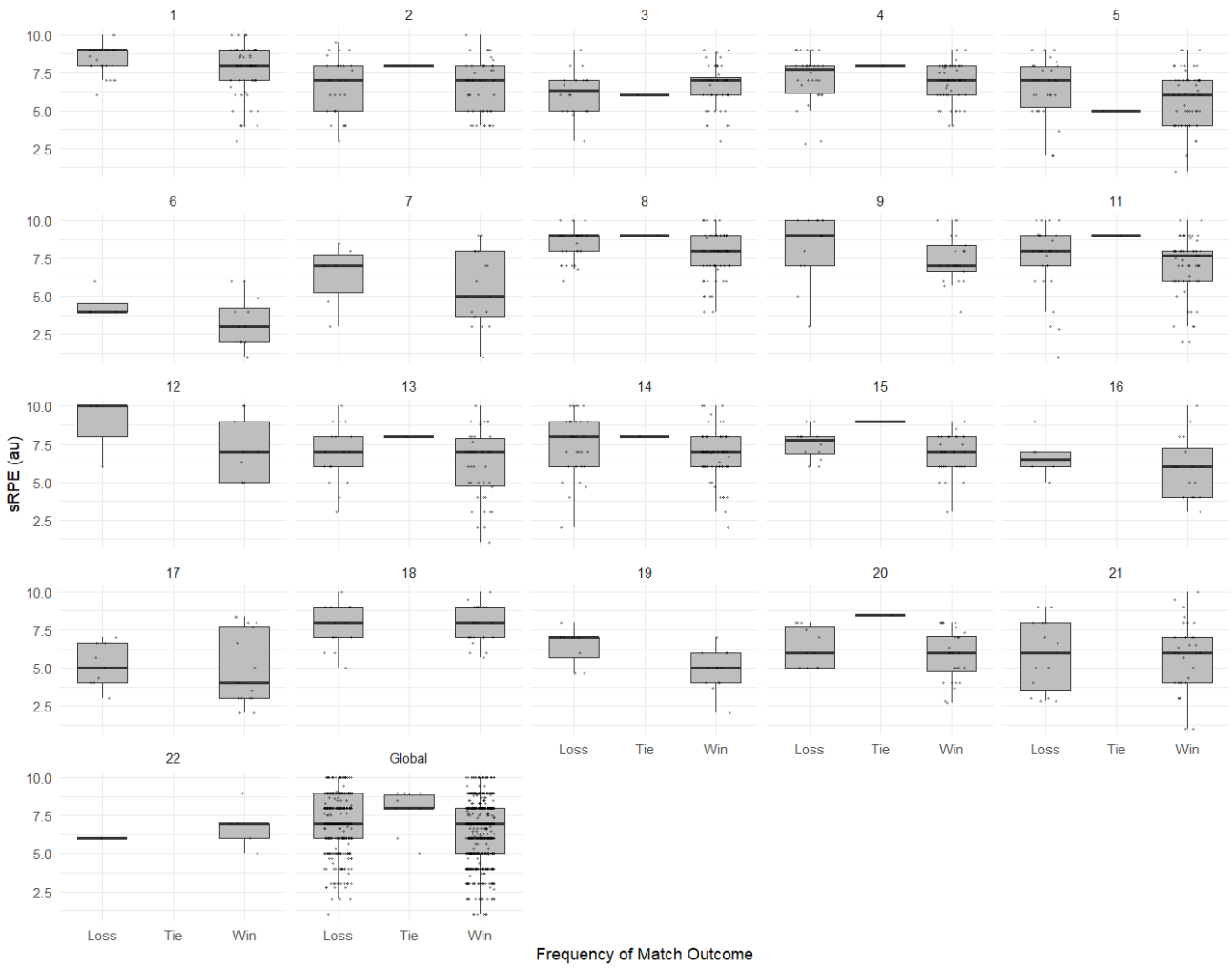

Figure A.1.22: Frequency of matches by match outcome and sRPE (au) per match, by athlete number with global data.

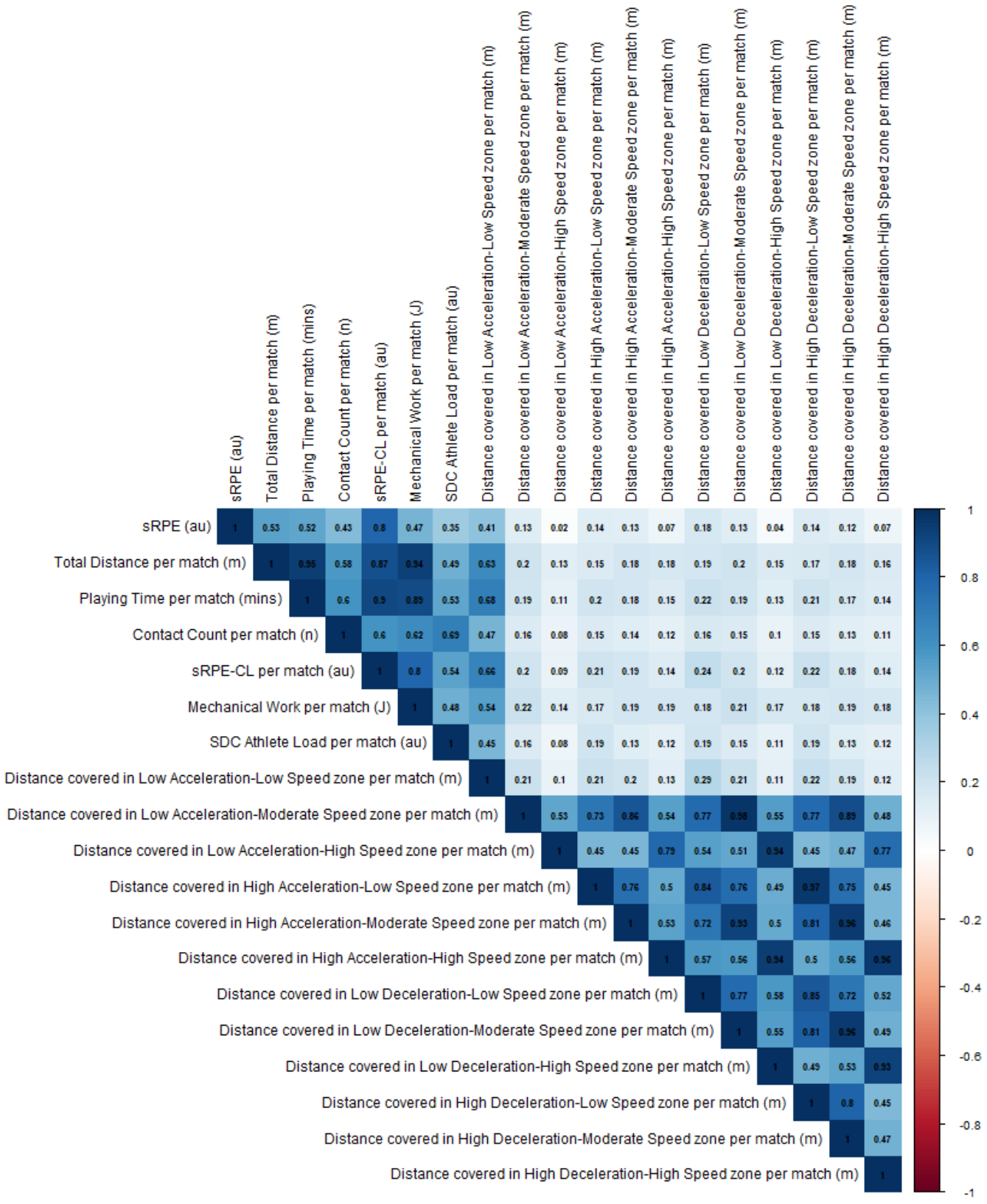

Figure A.1.23: Correlation graph of key exploratory variables, colour coded by level of correlation.

Appendix 2

This appendix contains information relating to Chapter 2.

Table A.2.1: Variance inflation factors for linear regression.

Variable	VIF Value	Interpretation (James et al., 2017)	Action (James et al., 2017)
Player	1.03	Low multicollinearity	Retain
Game	1.03	Low multicollinearity	Retain
Opponent	1.01	Low multicollinearity	Retain
Contact Count	1.69	Low multicollinearity	Retain
Total Distance	8.10	Moderate multicollinearity	Retain – domain relevant
Duration	8.34	Moderate multicollinearity	Retain – domain relevant

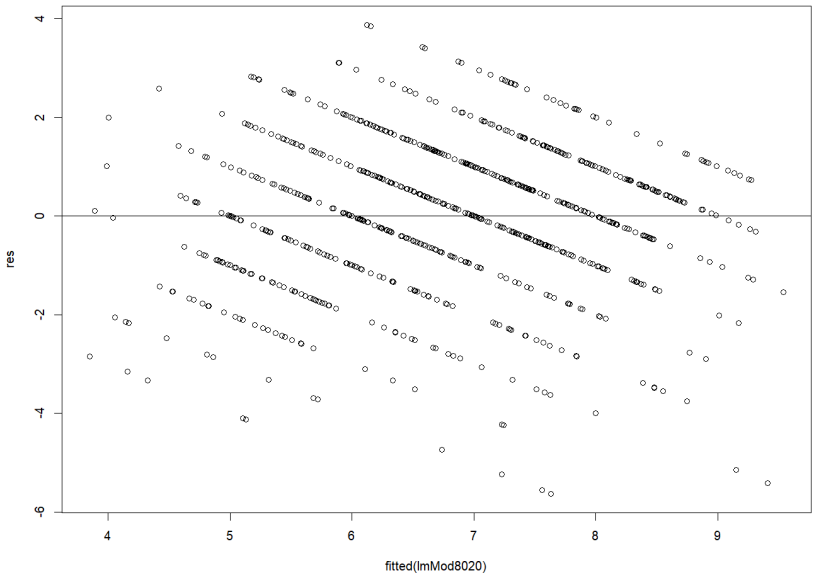

Figure A.2.1: Residual plot of linear regression data.

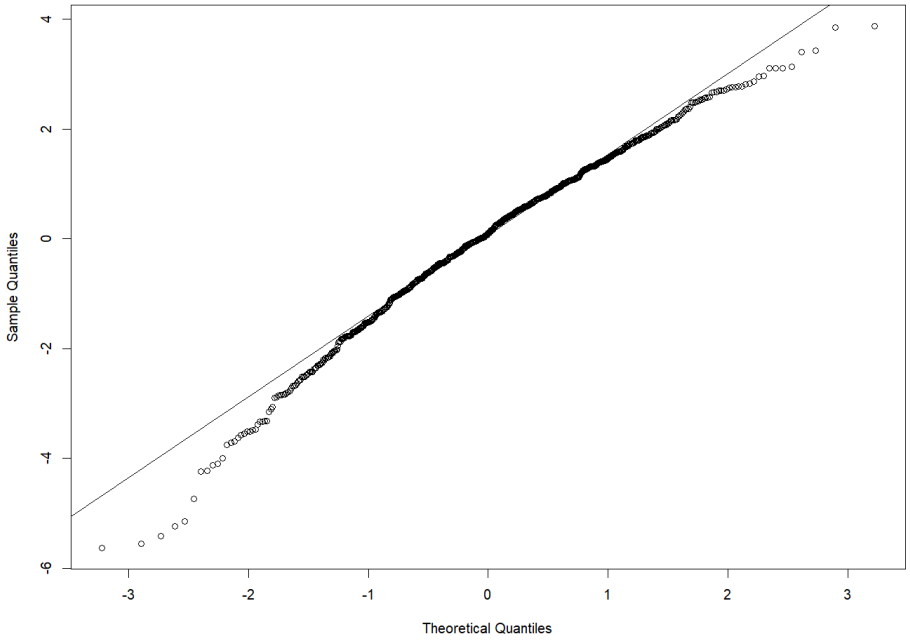

Figure A.2.2: Q-Q plot of linear regression data.

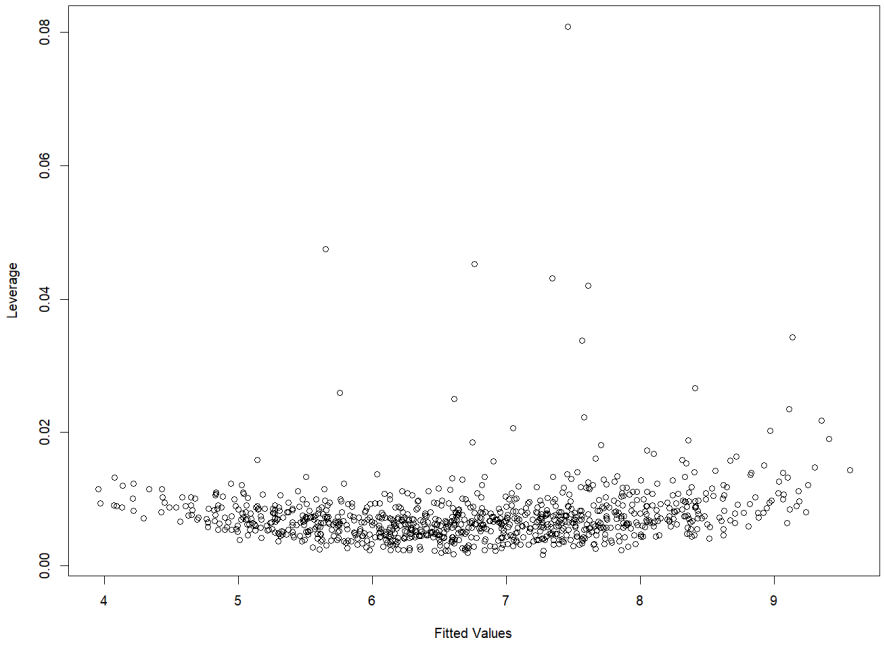

Figure A.2.3: Leverage plot of linear regression data.

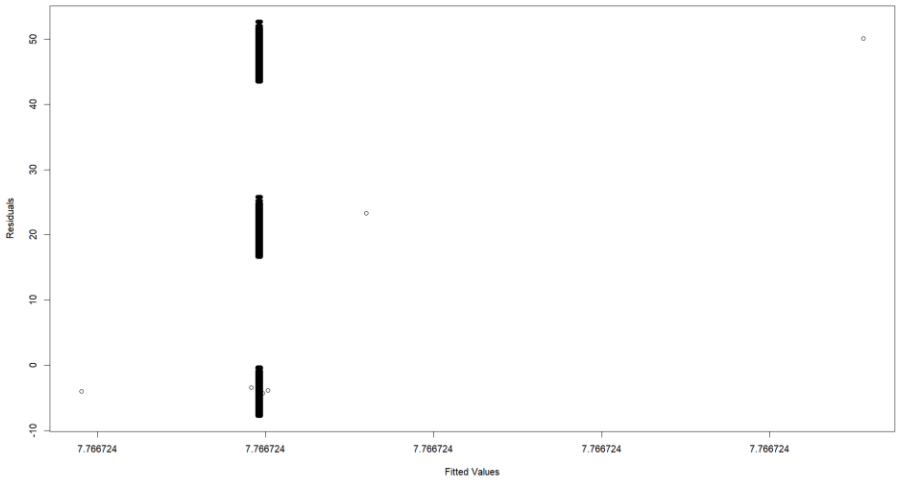

Figure A.2.4: Residual plot of unsupervised ANOVA data.

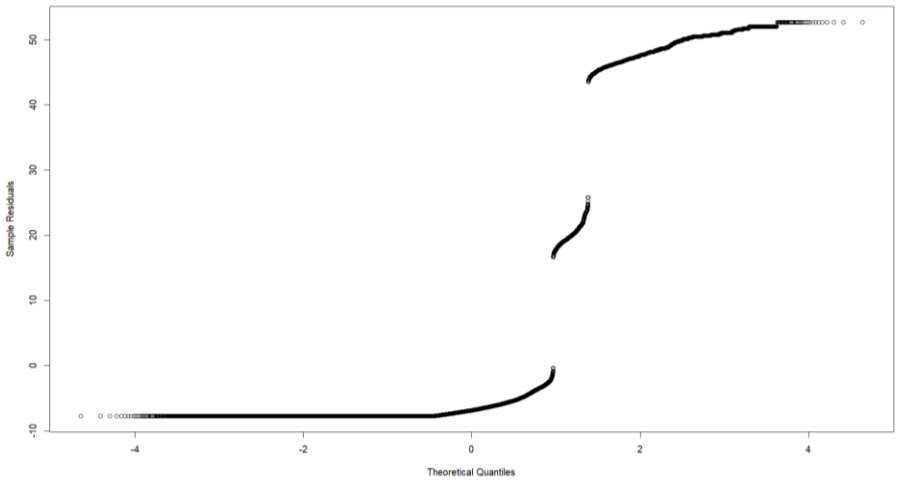

Figure A.2.5: Q-Q plot of unsupervised ANOVA data.

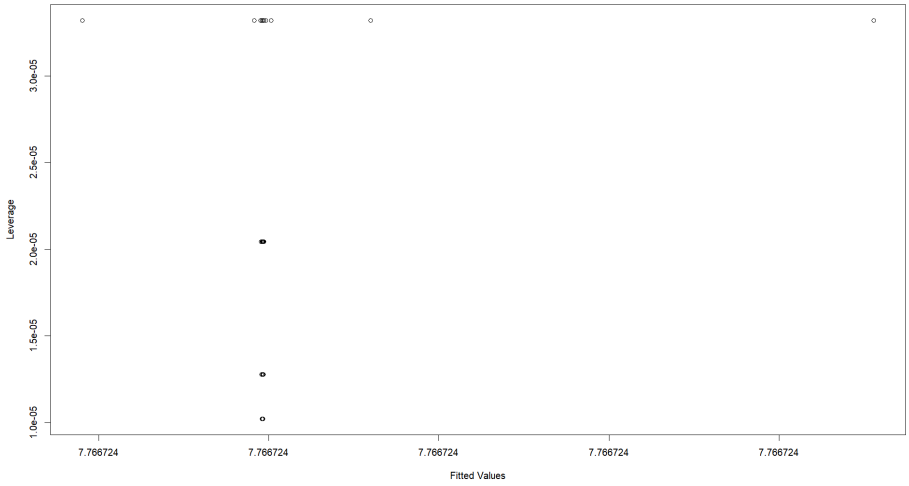

Figure A.2.6: Leverage plot of unsupervised ANOVA data.

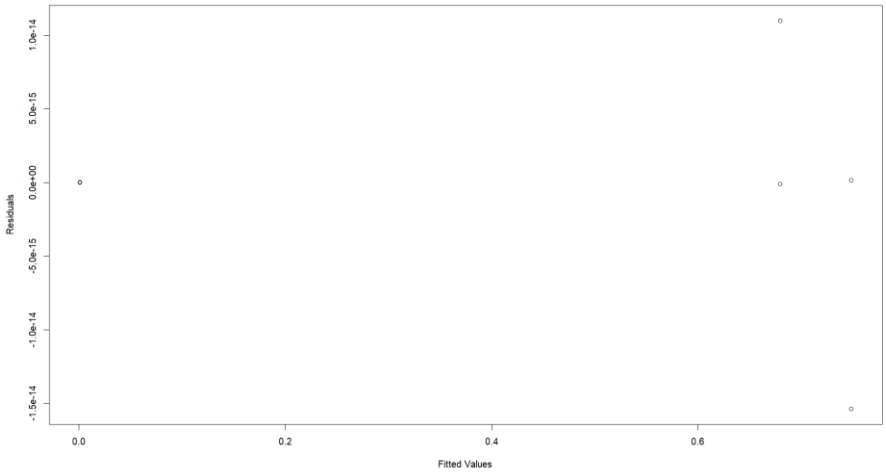

Figure A.2.7: Residual plot of supervised ANOVA data.

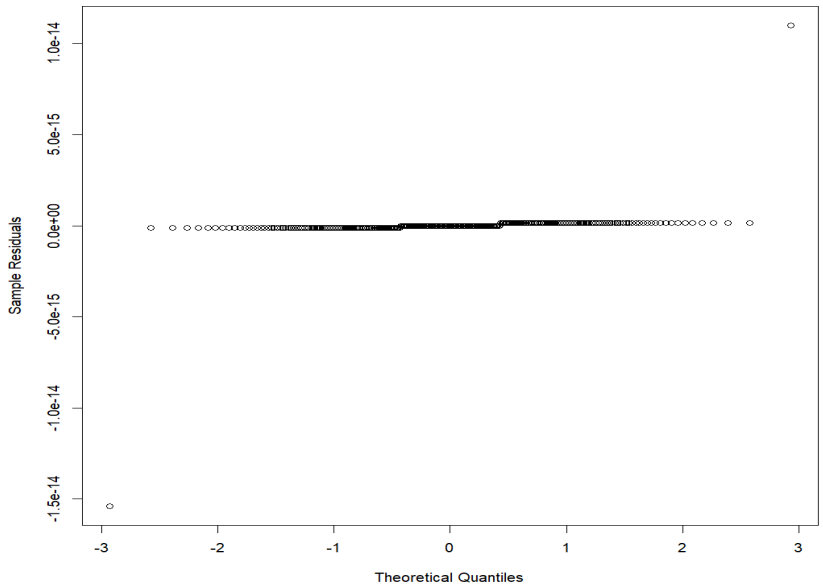

Figure A.2.8: Q-Q plot of supervised ANOVA data.

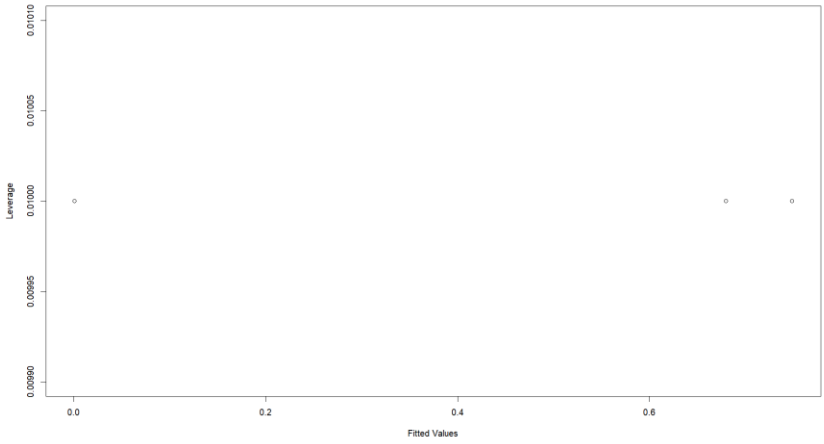

Figure A.2.9: Leverage plot of supervised ANOVA

Appendix 3

This appendix contains information relating to Chapter 3.

Table A.3.1: Variance inflation factors for linear regression.

Variable	**VIF Value**	**Interpretation** (James et al., 2017)	**Action** (James et al., 2017)
Duration	3.73	Low multicollinearity	Retain
Contact Count	9.36	Moderate multicollinearity	Retain – domain relevant
Duration* Contact Count	15.77	High multicollinearity	Adjust – transform variable (group contacts)

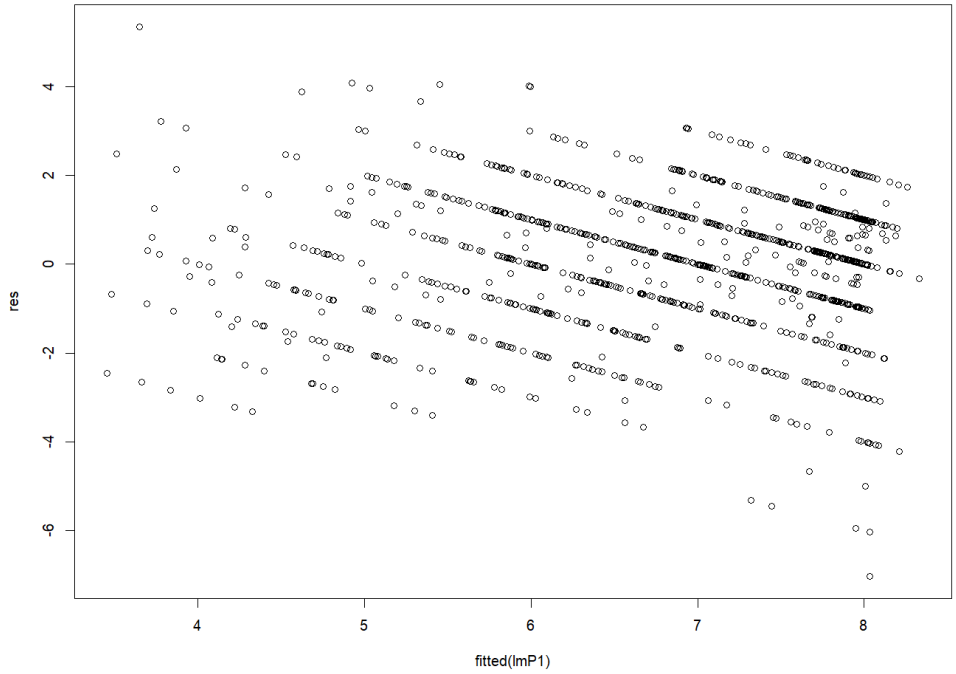

Figure A.3.1: Residual plot of linear regression data.

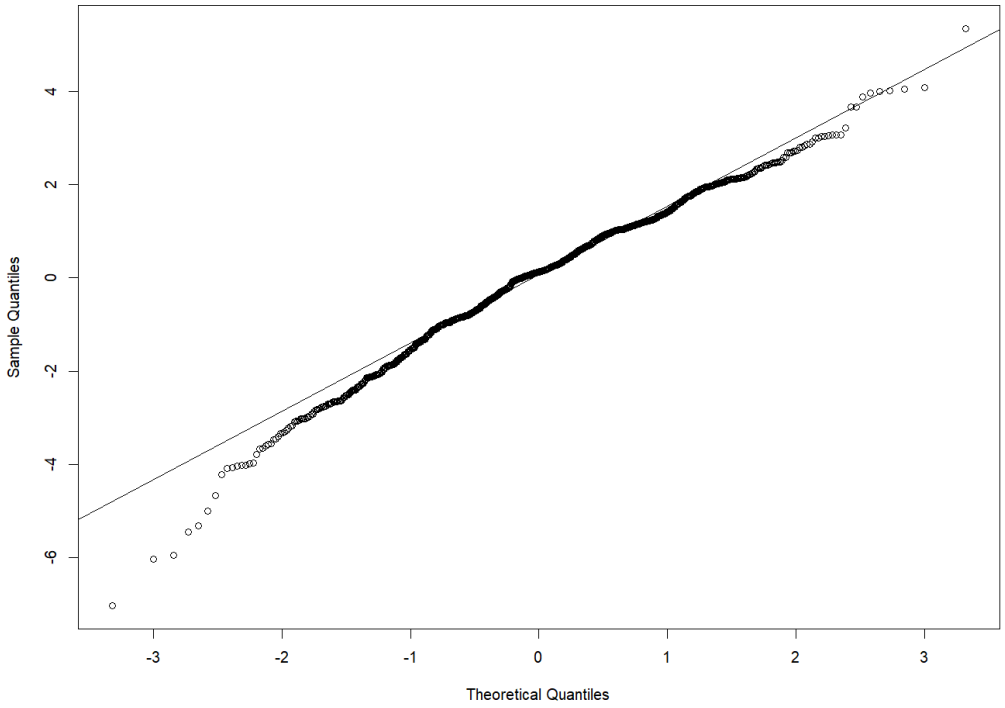

Figure A.3.2: Q-Q plot of linear regression data.

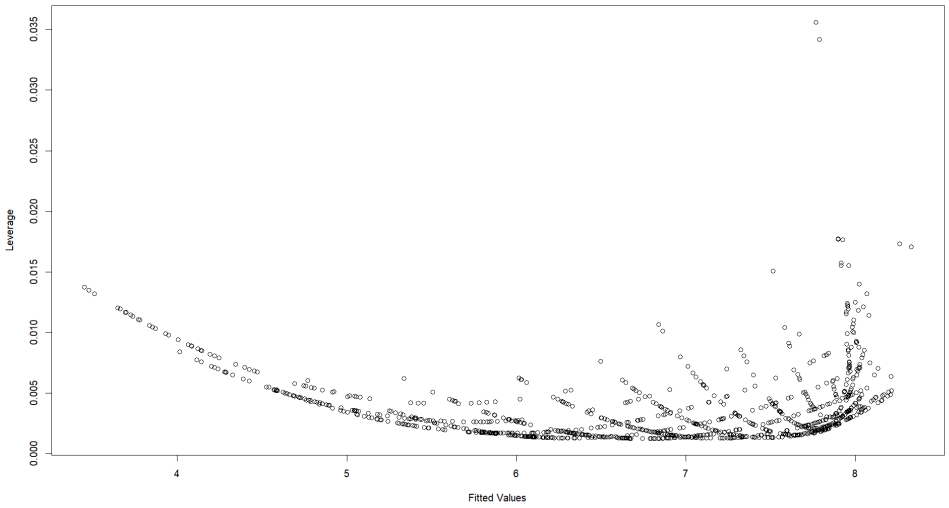

Figure A.3.3: Leverage plot of linear regression data.

Appendix 4

This appendix contains information relating to Chapter 5.

Table A.4.1: Variance inflation factors for linear regression.

Variable	**VIF Value**	**Interpretation** (James et al., 2017)	**Action** (James et al., 2017)
Player	1.03	Low multicollinearity	Retain
Match Category	1.03	Low multicollinearity	Retain
Opponent Rank	1.02	Low multicollinearity	Retain
Contact Count	1.67	Low multicollinearity	Retain
Total Distance	8.25	Moderate multicollinearity	Retain – domain relevant
Duration	8.50	Moderate multicollinearity	Retain – domain relevant

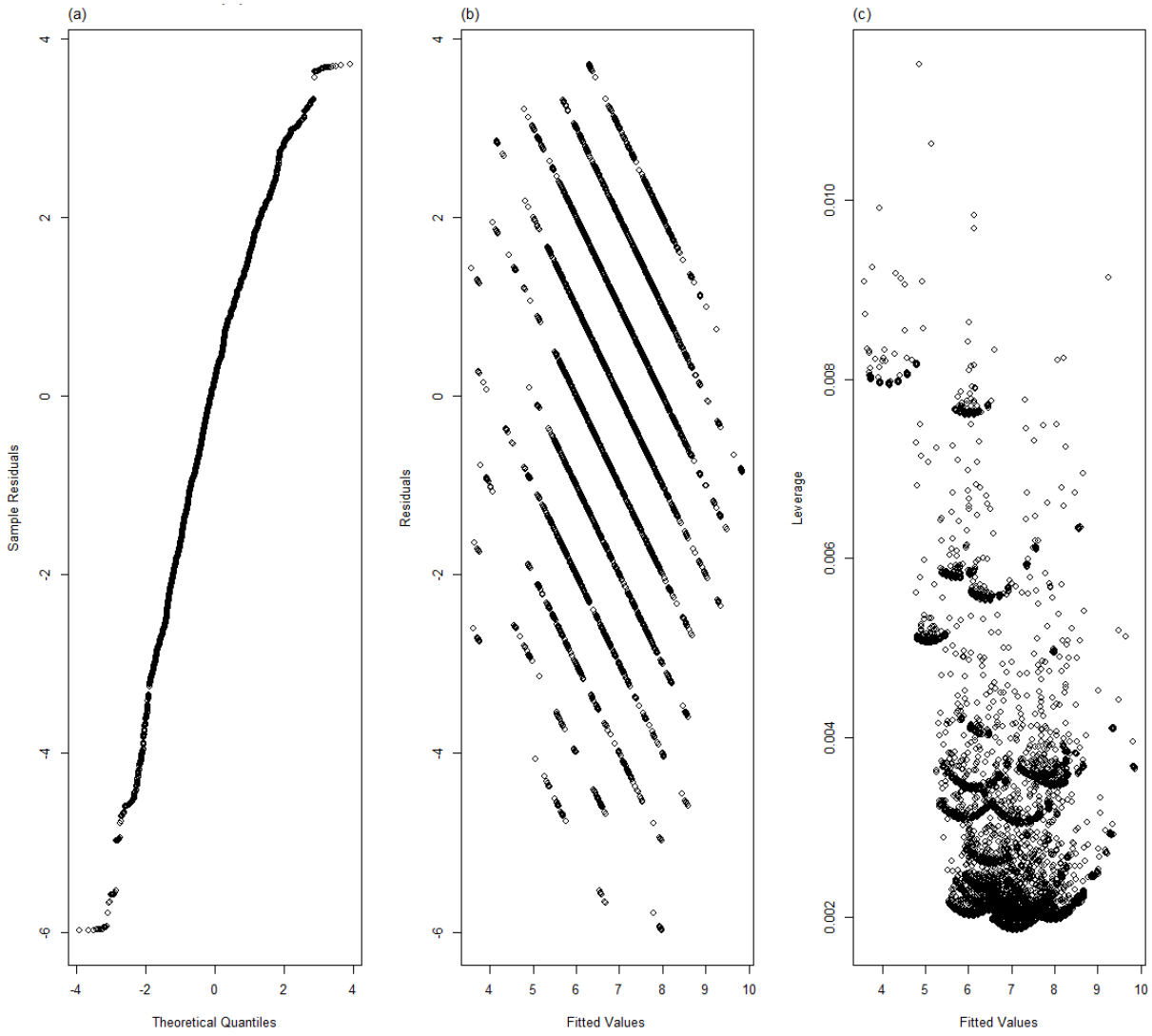

Figure A.4.1: (a) Q-Q, (b) residual, and (c) leverage plots for linear mixed model of SDC data.

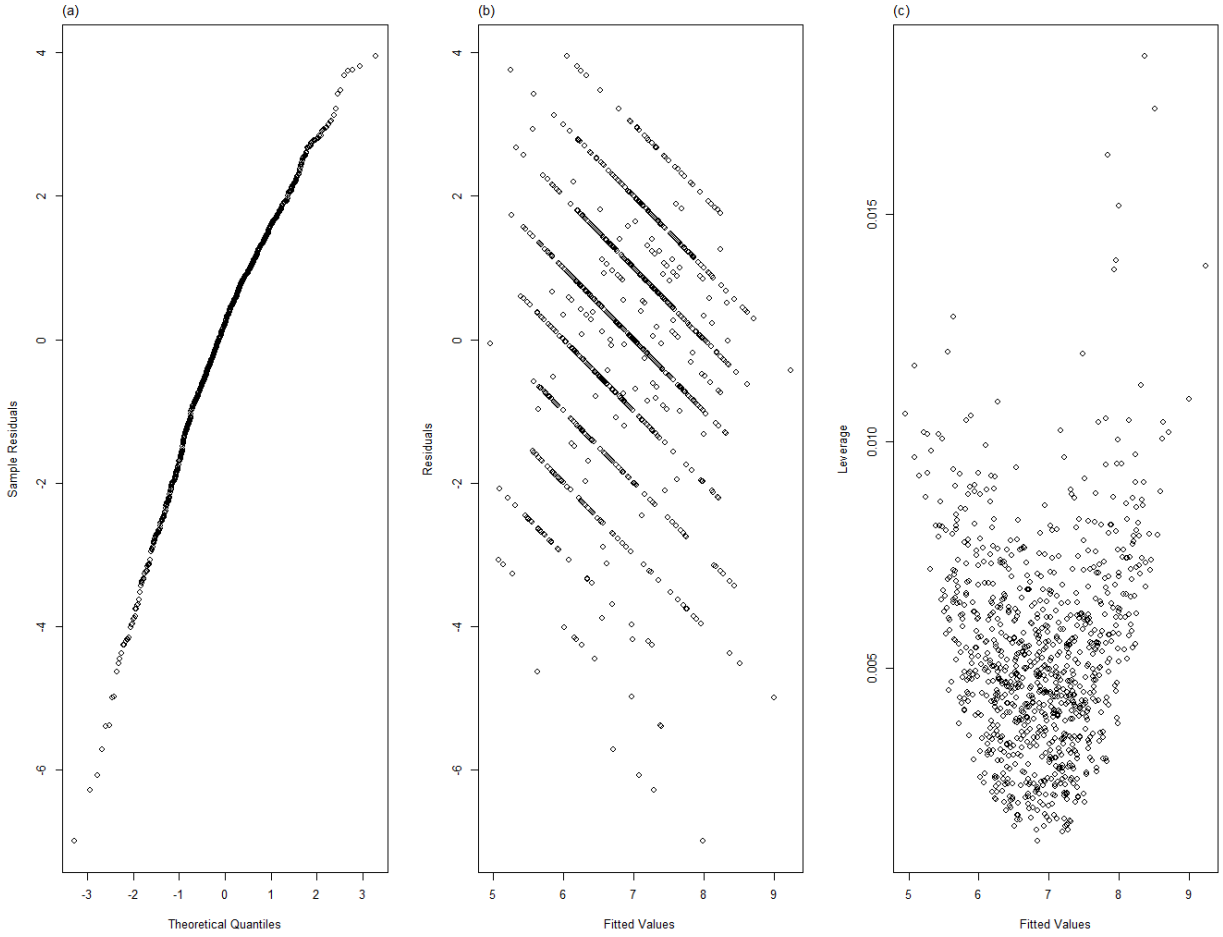

Figure A.4.2: (a) Q-Q, (b) residual, and (c) leverage plots for linear mixed model of mechanical work data.

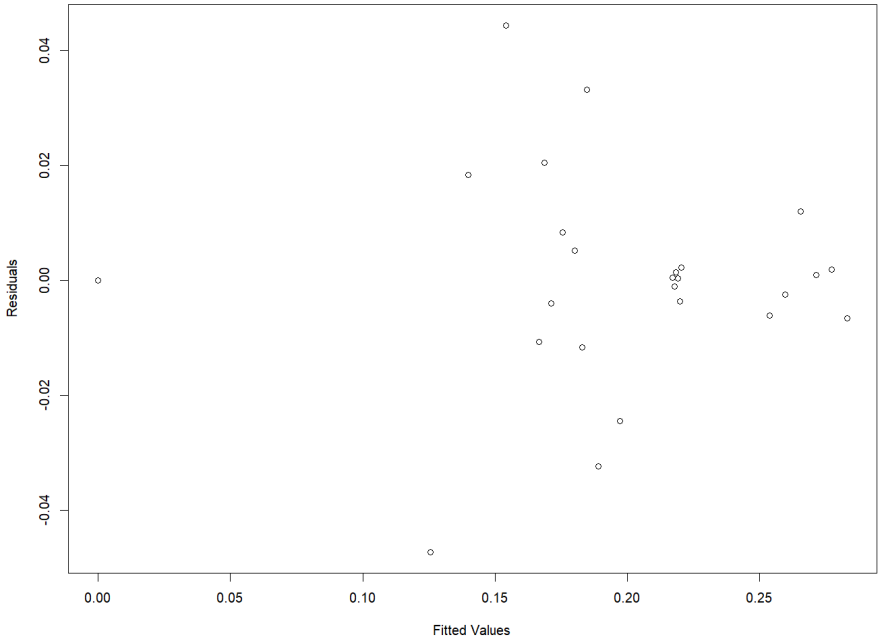

Figure A.4.3: Residual plot of ANOVA data.

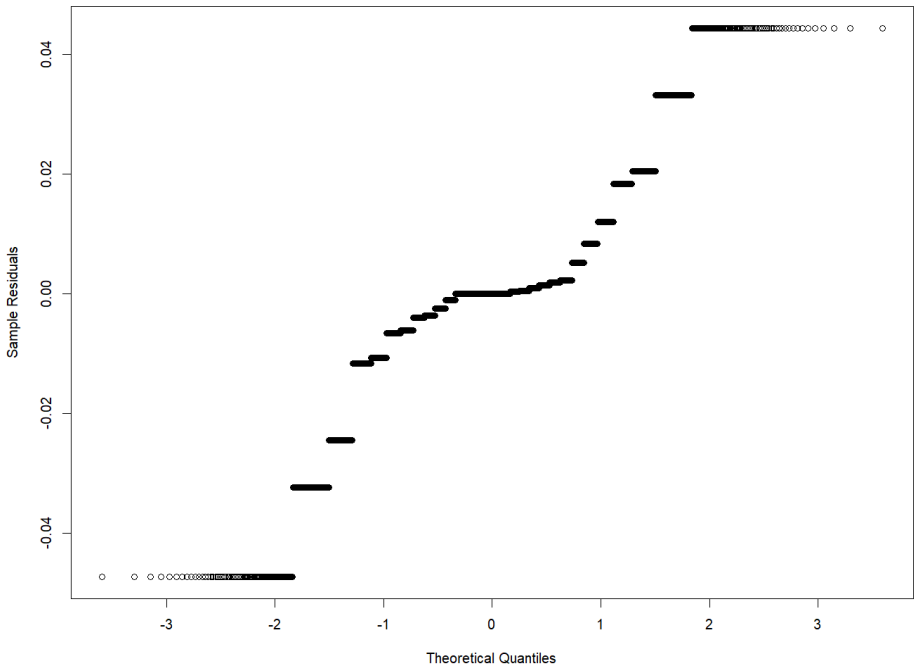

Figure A.4.4: Q-Q plot of ANOVA data.

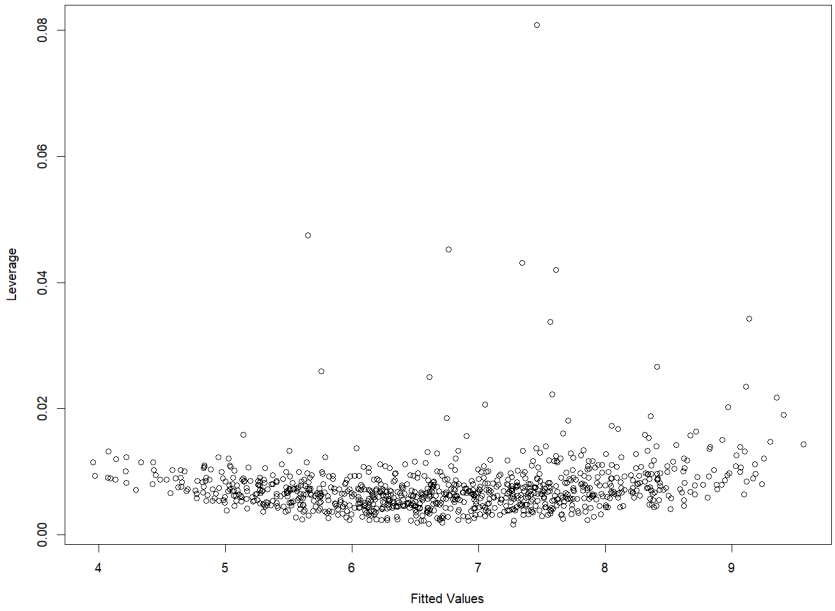

Figure A.4.5: Leverage plot of ANOVA data.

Appendix 5

This appendix contains information relating to Chapter 6.

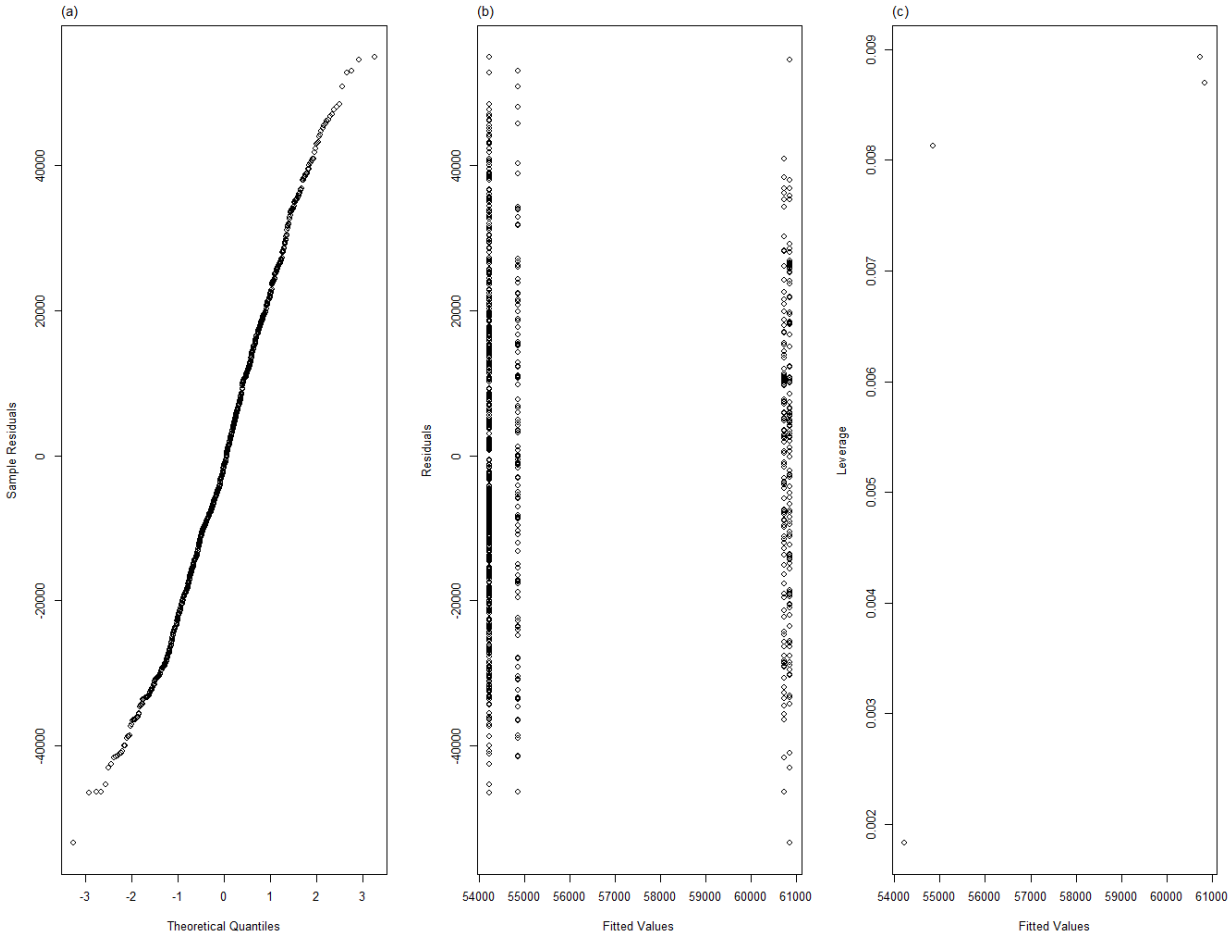

Figure A.5.1: (a) Q-Q, (b) residual, and (c) leverage plots for ANOVA data for match category using mechanical work.

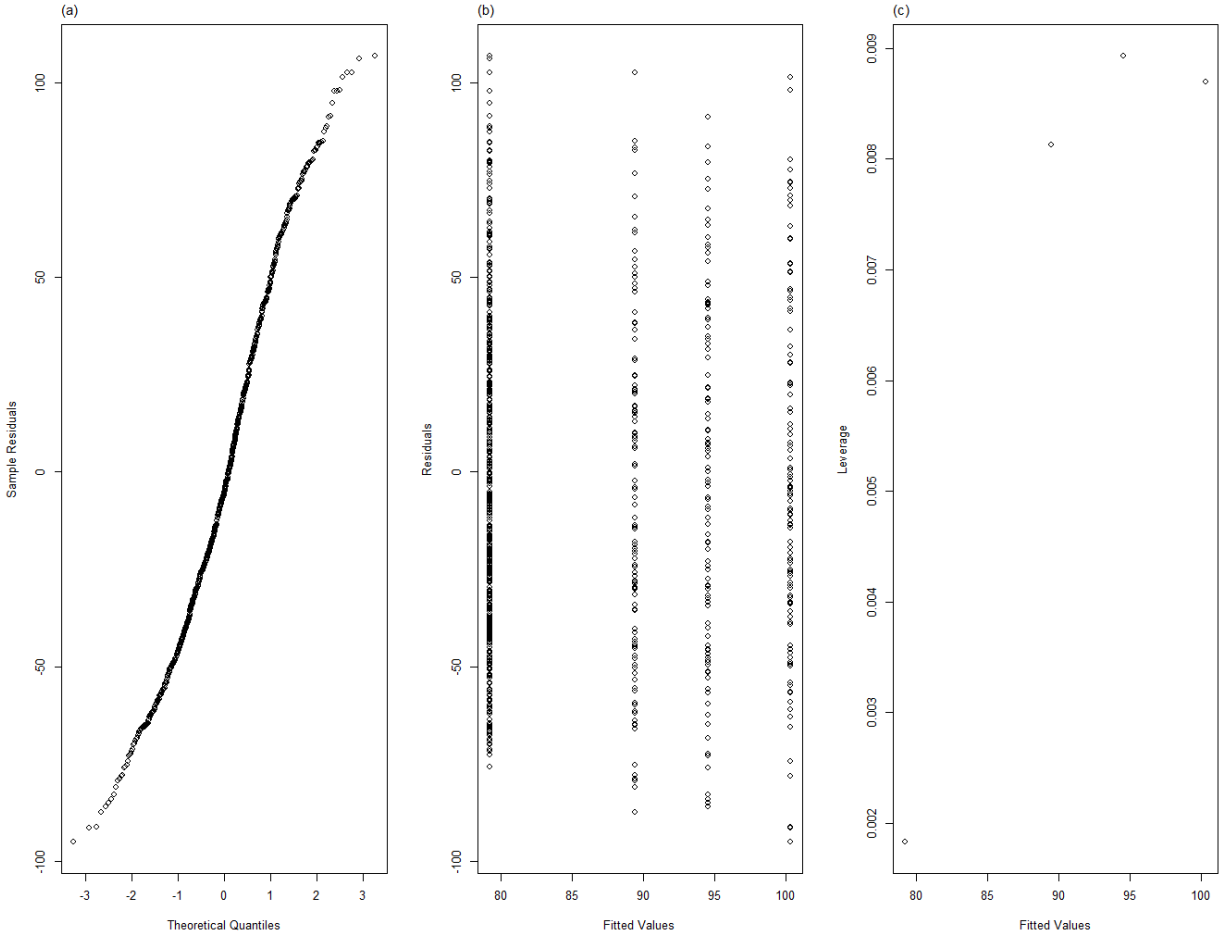

Figure A.5.2: (a) Q-Q, (b) residual, and (c) leverage plots for ANOVA data for match category using sRPE.

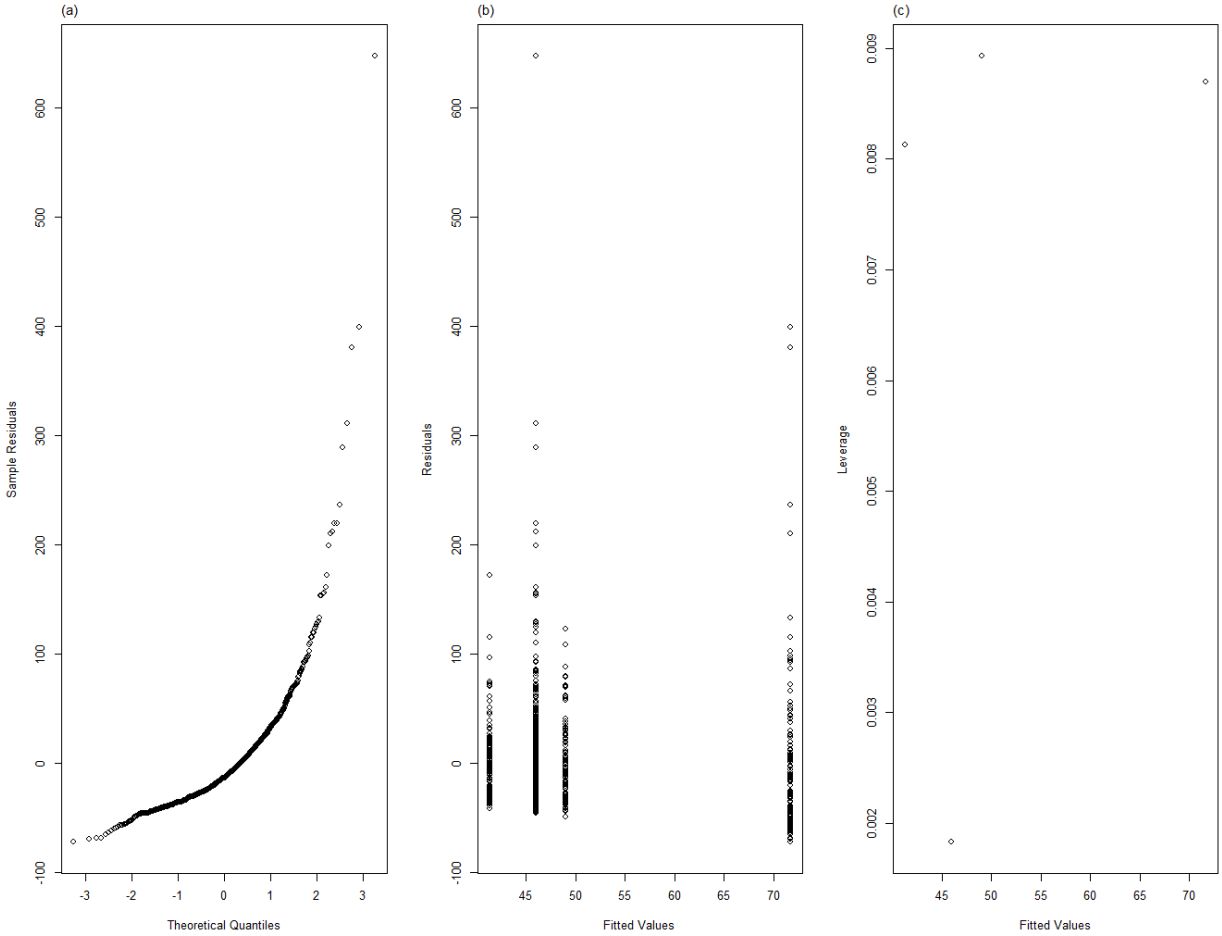

Figure A.5.3: (a) Q-Q, (b) residual, and (c) leverage plots for ANOVA data for match category using the SDC model.

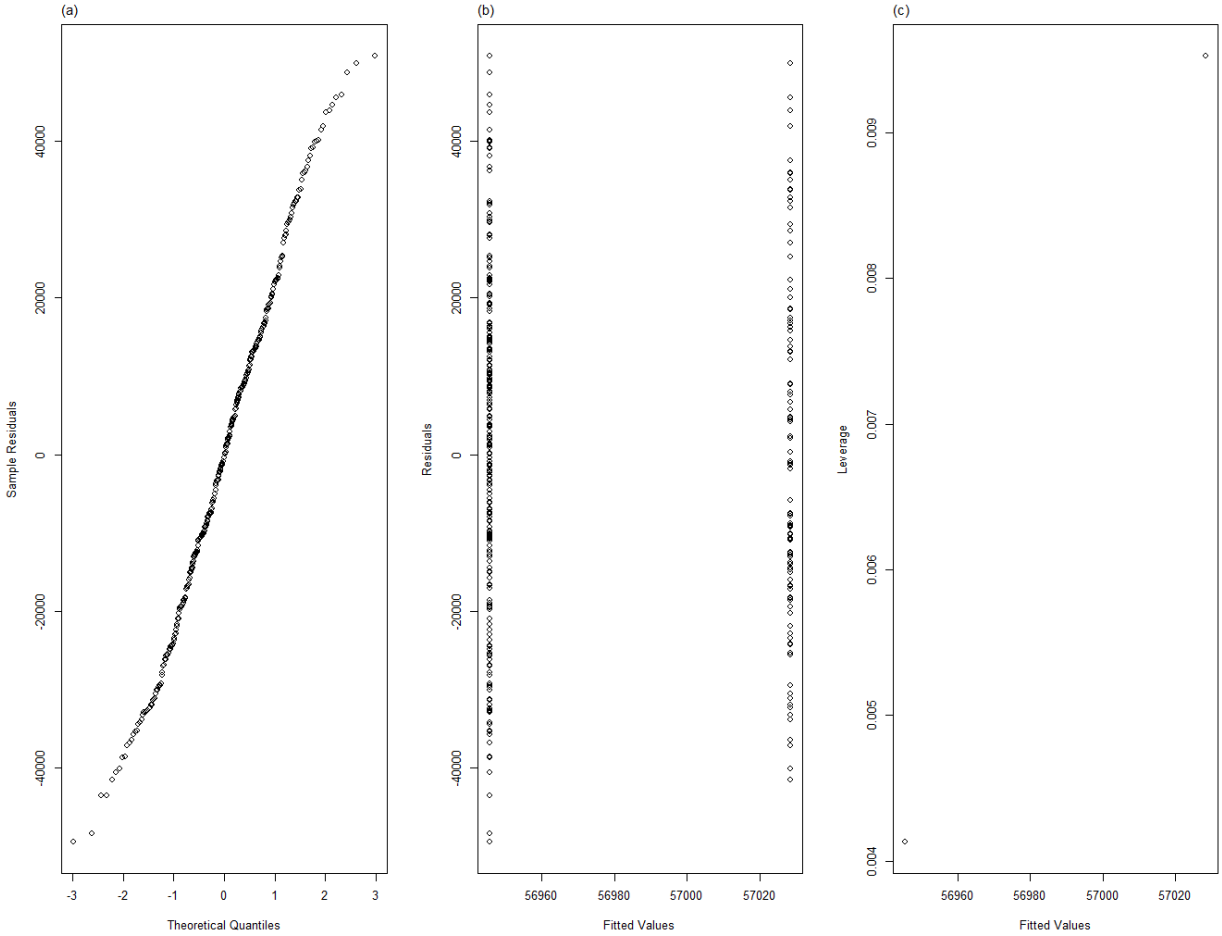

Figure A.5.4: (a) Q-Q, (b) residual, and (c) leverage plots for ANOVA data for opponent rank using mechanical work.

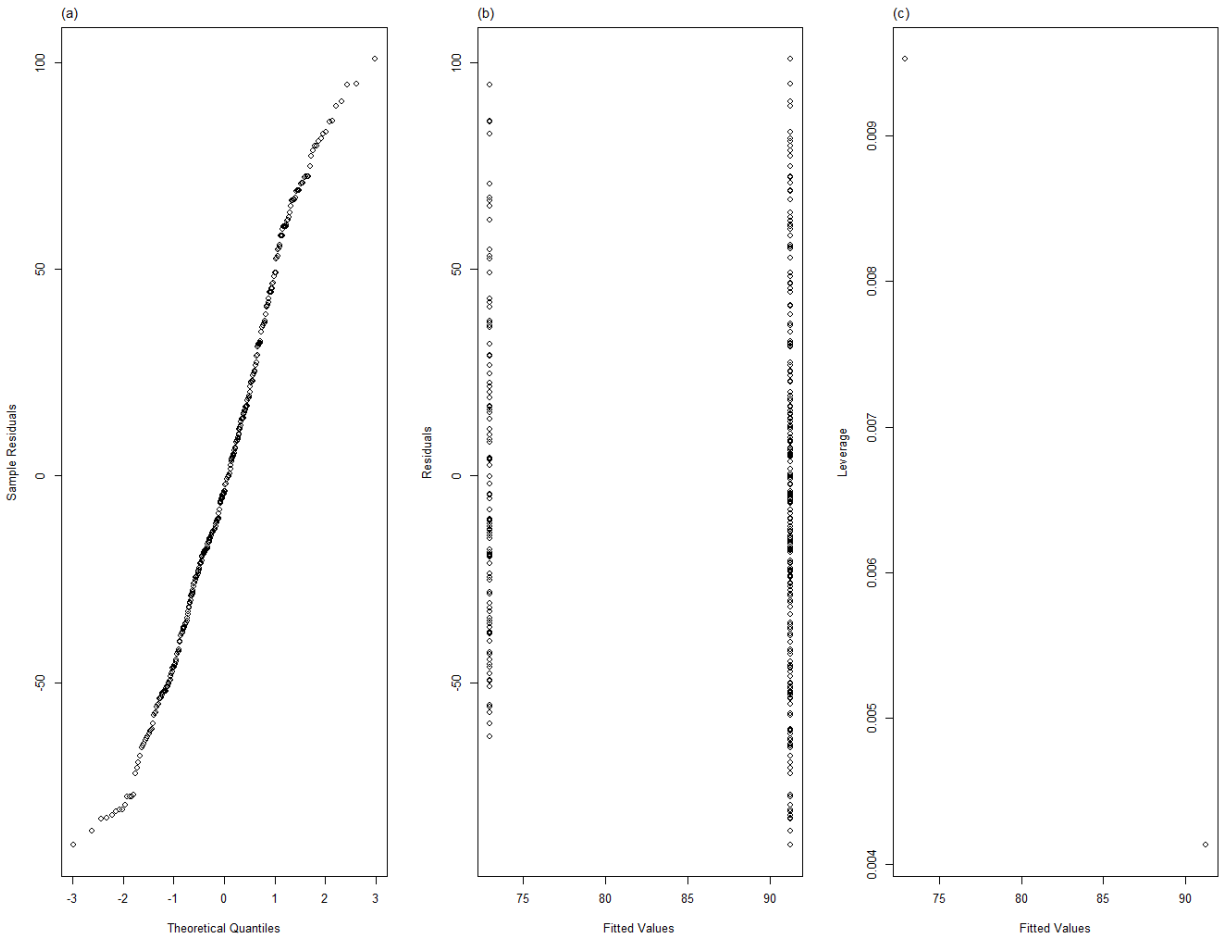

Figure A.5.5: (a) Q-Q, (b) residual, and (c) leverage plots for ANOVA data for opponent rank using sRPE.

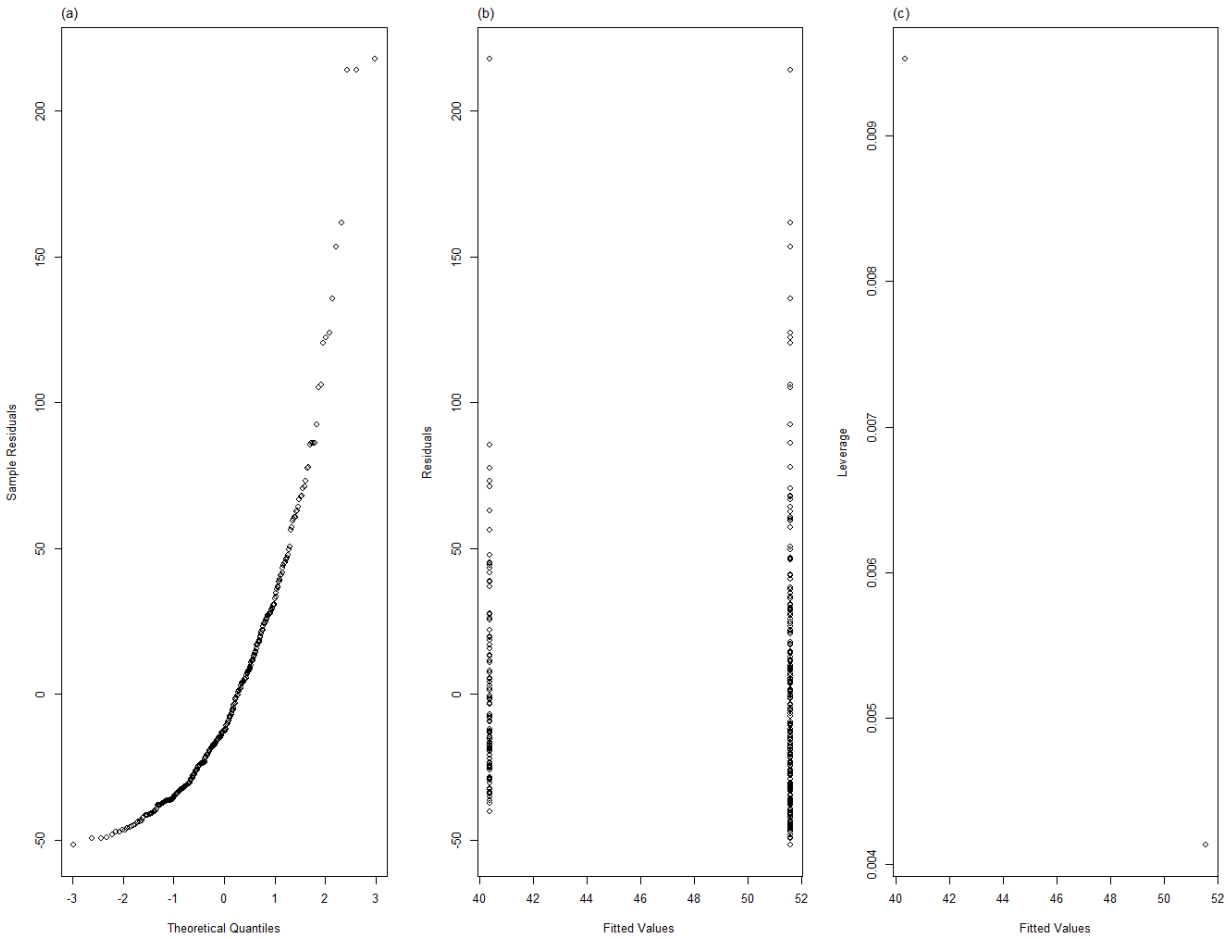

Figure A.5.6: (a) Q-Q, (b) residual, and (c) leverage plots for ANOVA data for opponent rank using the SDC model.

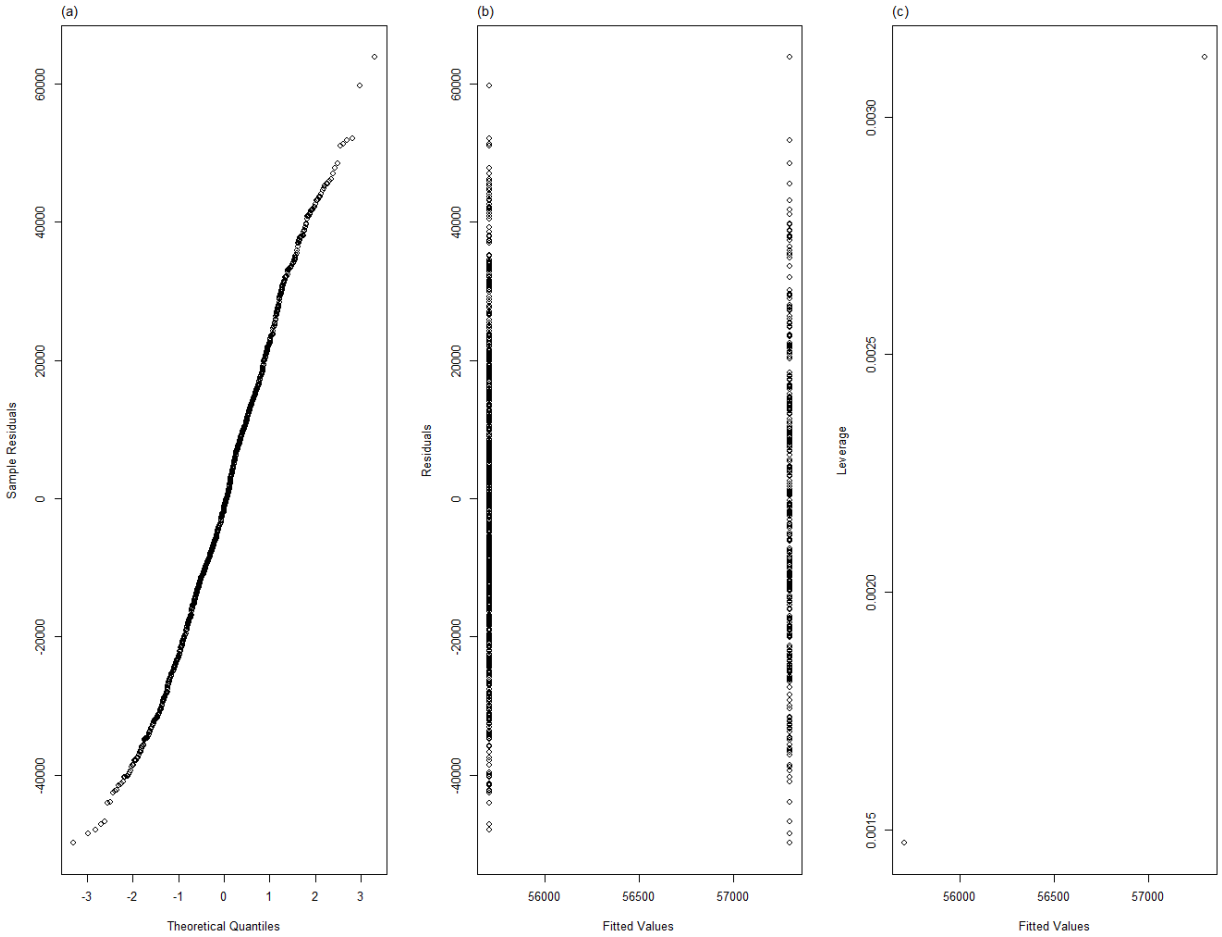

Figure A.5.7: (a) Q-Q, (b) residual, and (c) leverage plots for ANOVA data for match outcome using mechanical work.

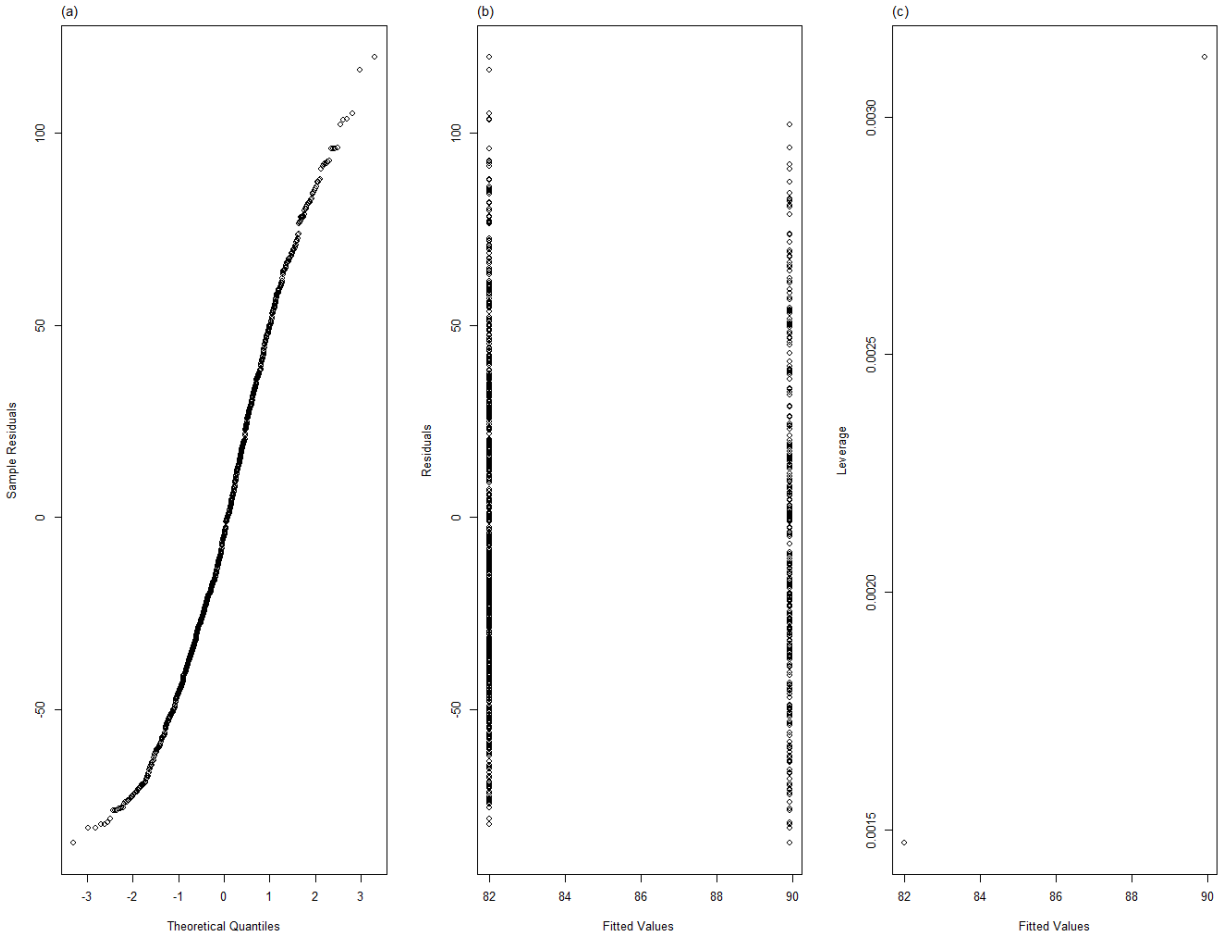

Figure A.5.8: (a) Q-Q, (b) residual, and (c) leverage plots for ANOVA data for match outcome using sRPE.

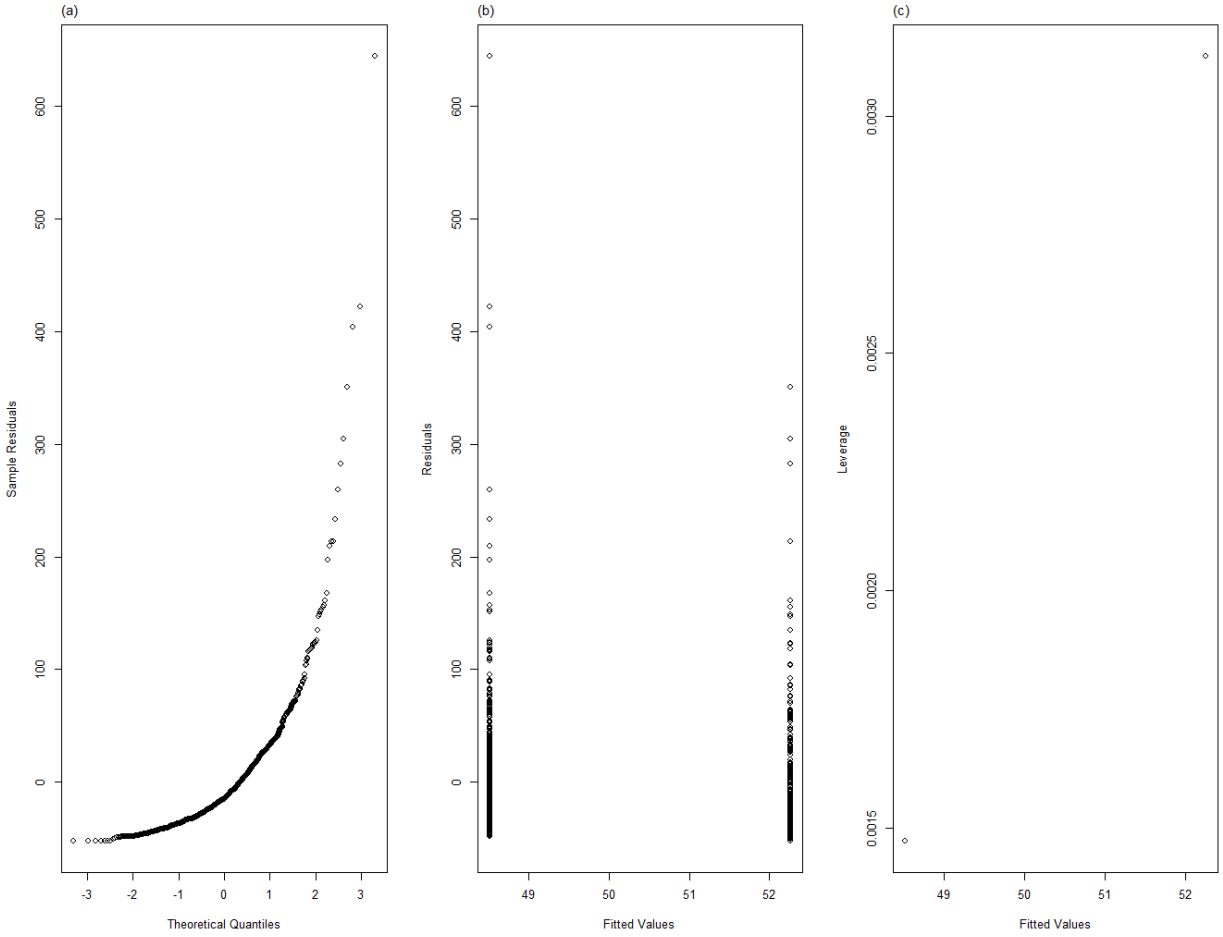

Figure A.5.9: (a) Q-Q, (b) residual, and (c) leverage plots for ANOVA data for match outcome using the SDC model.

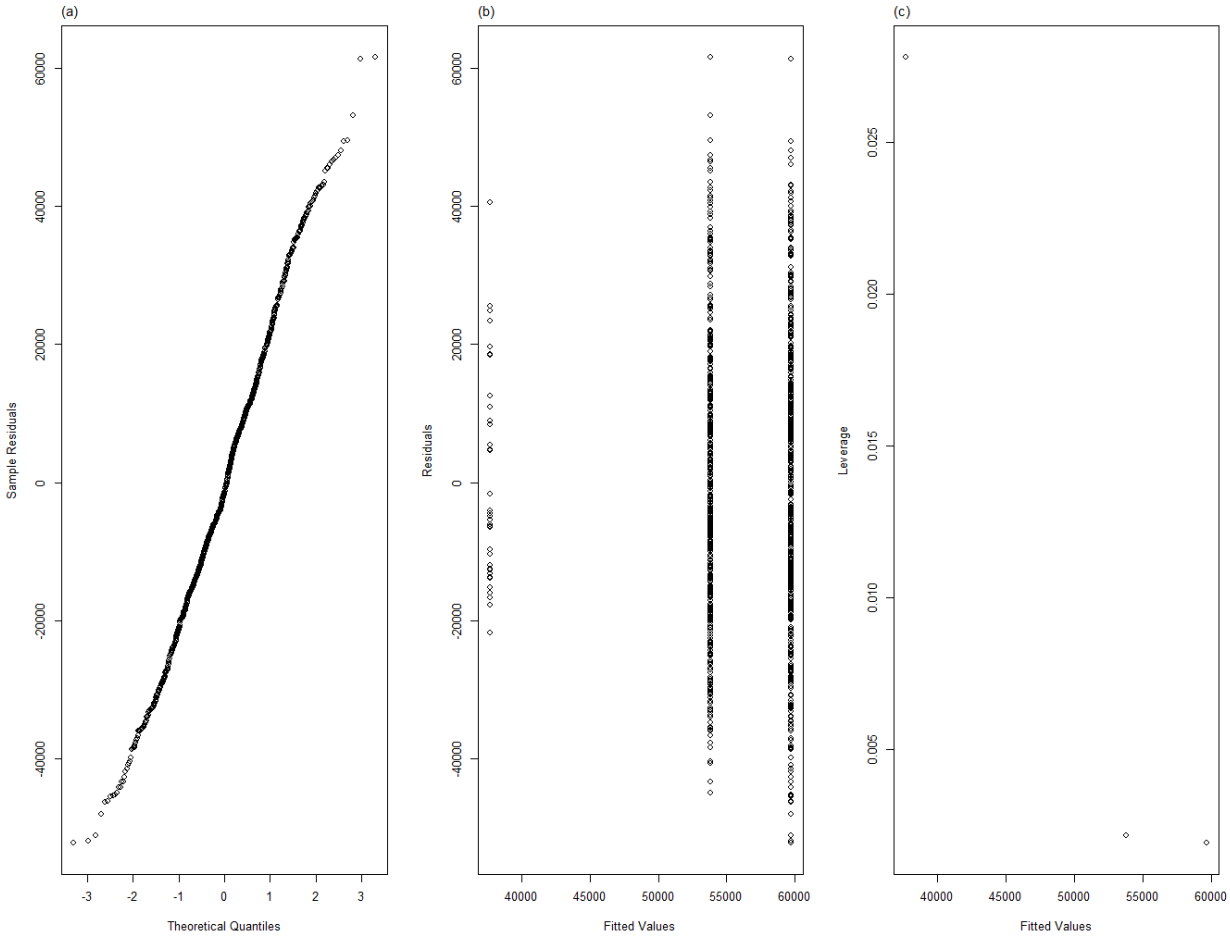

Figure A.5.10: (a) Q-Q, (b) residual, and (c) leverage plots for ANOVA data for player experience using mechanical work.

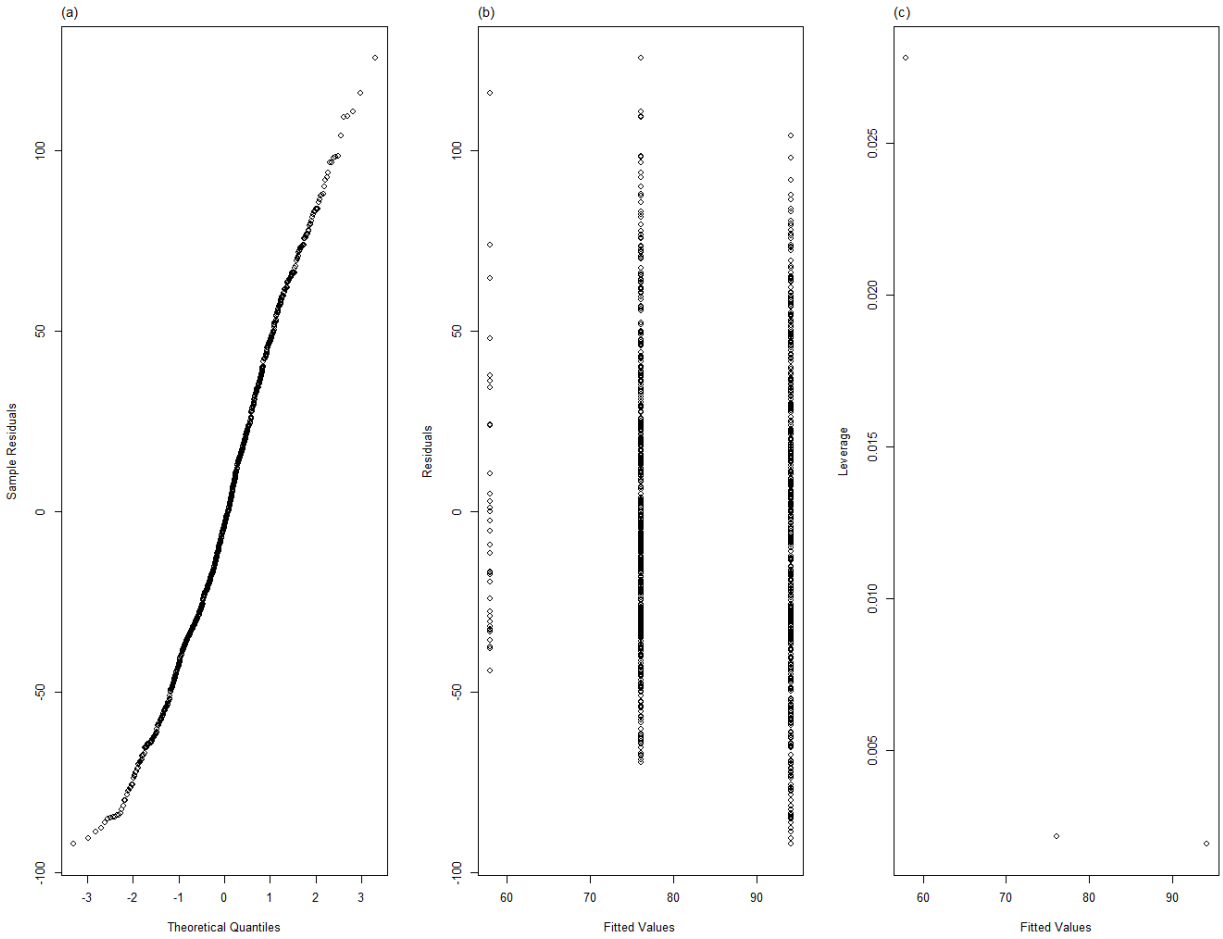

Figure A.5.11: (a) Q-Q, (b) residual, and (c) leverage plots for ANOVA data for player experience using sRPE.

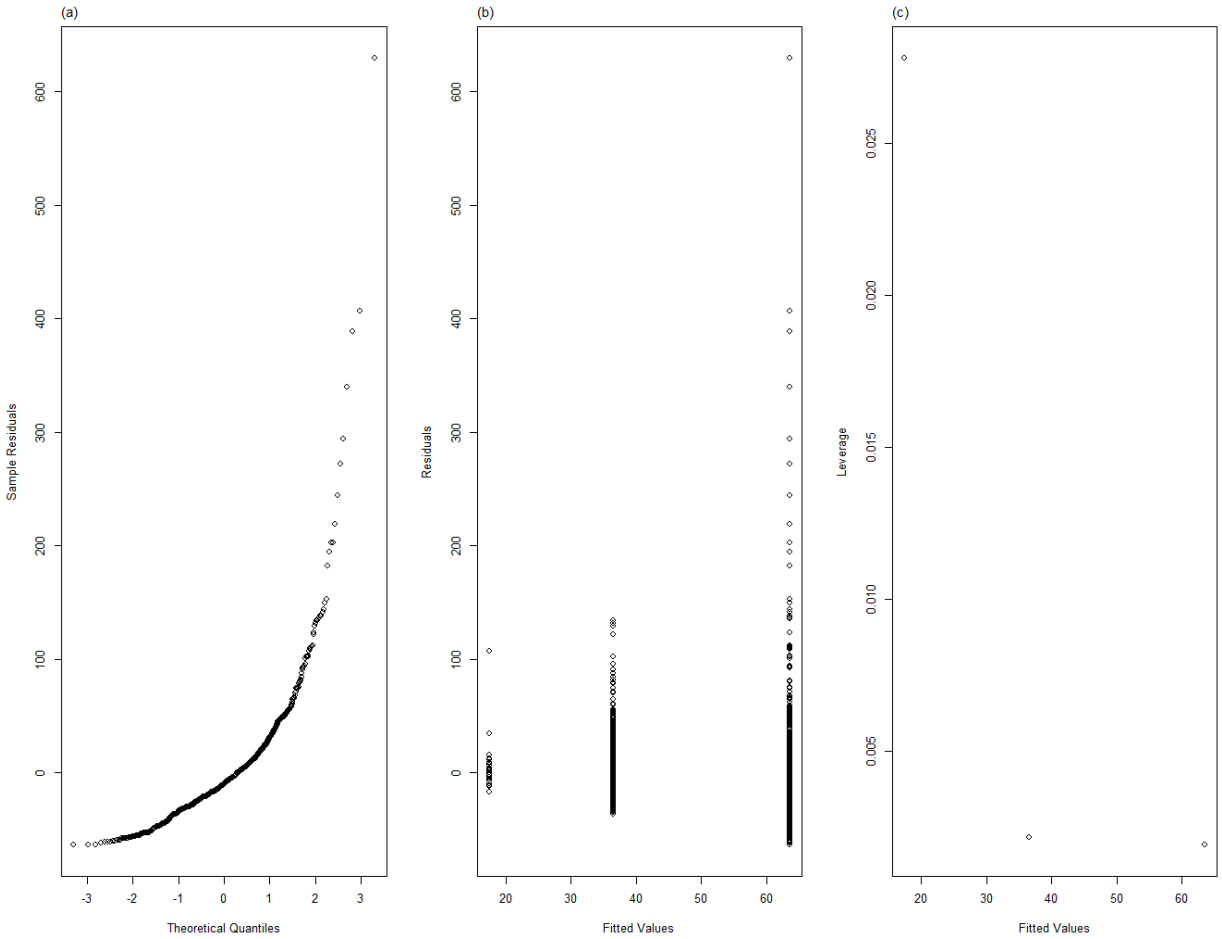

Figure A.5.12: (a) Q-Q, (b) residual, and (c) leverage plots for ANOVA data for player experience using the SDC model.