Expanding the Statistical Toolkit of Sports Scientists

Table of contents

Preface

Timothy J. Newans

B. Exercise Science, M. Medical Research

Griffith Health - Griffith University

March 2023

GRIFFITH UNIVERSITY

Submitted in fulfilment of the requirements of the degree of Doctor of Philosophy.

Supervisors: Clare Minahan, PhD & Phillip Bellinger, PhD, Brent Richards MB ChB

One of the most appealing aspects of watching sport is encountering world-class athletes compete against each other. While these performances create enthralling experiences, their exclusivity unintentionally creates some pitfalls when robustly analysing their relative performances and changes to their performance. With the increasing professionalism of sport and increasing availability of wearable technologies, Sports Scientists can gather a prolific amount of data about athletes' physical, physiological, psychological, and tactical state.

This thesis aims to expand the statistical toolkit of Sports Scientists by providing novel research using statistical techniques that are underutilised by the sports science community to provide real-world examples of how correct statistical methodologies and principles can help overcome commonly seen pitfalls in sports science data sets. While each of these pitfalls are not unique to sports science, the combination of multiple pitfalls can hinder performing robust statistical analyses. Sports Scientists are required to utilise statistical methods that can account for an individual's baseline level before an intervention, especially when there are imbalanced amounts of data for each athlete. Additionally, with the continual development of testing batteries and methodologies when understanding an athlete's profile, it is easy to simply take the players that are exceptional in each attribute; however, more consideration is required to identify 'all-rounder' players, that perform well in numerous testing batteries without being exceptional in any attribute. Furthermore, given these athletes are at, or close to, the peak of their ability, the levels of improvement required for athletes are substantially lower than that which would be deemed successful at a general population. Similarly, given there is only a limited number of athletes compete at the highest level of competition in their given sport, consideration needs to be had for statistical methods that can accommodate for limited sample sizes.

This thesis calls for an increased adoption of mixed models for data sets in which there are repeated measures of athletes and it is necessary to account for the athlete's inherent baseline in an imbalanced data set. This thesis presents Pareto frontiers as a statistical principle of identifying the 'extreme' athletes that excel in multiple attributes of interest, even if they may not be exceptional in any single attribute. Finally, this thesis highlights Bayesian inference as a statistical framework, in which prior subject matter expert knowledge of what the data typically seen can be incorporated into the statistical model to identify beneficial interventions more easily when dealing with small effect sizes and small sample sizes.

The first three studies provide methodological overviews of these three statistical aspects (i.e., mixed models, Pareto frontiers, and Bayesian inference). The final two studies are more applied studies to show how more rigorous statistical methods can become more commonplace throughout sports science research. In particular, the final study serves as a 'capstone' study, in which the uncertainty around the Pareto frontier can be estimated by a mixed model run in a Bayesian framework. Each of the five studies are presented in manuscript form in the format required by the individual journal guidelines (e.g., abstract length and structure); however, to assist with thesis flow, some slight formatting changes have been made. A single reference style was used to assist with readability, with all references placed together at the end of the thesis, rather than keeping the references in the style of each journal.

Throughout the thesis, all the lines of code required to execute the statistics required in each study is published alongside the study to provide readers with access, with the intent to encourage Sports Scientists to adopt similar practices in their own research. By publishing this thesis as an eBook, it is intended that the barrier for entry for Sports Scientists to such statistical methods and frameworks can be lowered and more accessible by reading research using such statistical methods and frameworks published within their own discipline.

Acknowledgements

To my primary supervisors, **Clare Minahan** and **Phil Bellinger**, thank you for all your support, advice, and guidance over the past six years' worth of research project, Masters, and now PhD candidature. Thanks for welcoming me into the Griffith Sports Science team and putting up with my rants about statistics in sports. I didn't think I'd be able to stand getting through my planned 5 years at uni, let alone this 9-year stint; however, Griffith has always felt like a place where I belonged, and the Griffith Sports Science team has been a large part of that. To **Brent Richards**, thanks for our 30,000-ft view conversations of the health and sports science landscape and always looking further down the track of where we could be and working out how we can get there. To **Chris Drovandi**, your statistical insights and catchups and have meant you have felt like a supervisor to me and I'm so thankful for your patience with me in explaining statistical concepts.

To **Brad and Bucko from the NRL**, it's been a crazy journey, starting with identifying talent for the inaugural NRLW premiership and finishing with a World Cup win at Old Trafford! Thanks for always making me feel part of the team and encouraging me in my academic pursuits. It's been a huge honour to have been even a small part of the evolution of female rugby league and I'm so excited to see what it becomes in the future.

To the **Queensland Academy of Sport**, thanks for your partnership in this project, allowing opportunities to present my research, and bounce ideas back and forth with like-minded sports scientists.

To my parents, **Jamie and Jenny**, and to **Josh and Brodie**, thank you for all your support and encouragement as I moved away from home. Even when I was 7 years old and said I wanted to be a sports statistician when we thought it wasn't a real thing, thanks for encouraging me to explore opportunities and make it into a reality. To my in-laws, **Ian and Jayne**, thanks for the countless meals and for finally coming to terms with what I do is actually a real job.

Finally, to my wife **Breeana**, I could not have finished this without you. Thank you for just being you. Pushing me when I need to be pushed but also telling me to step back when I needed a break. It has been a whirlwind 4 years through this PhD: from getting married, to COVID dramas, and finally with the birth of our beautiful son, **Theodore**. Bringing Theo into the world has brought me so much joy and you have held our family together through these final few months throughout my submission period and just showed me unconditional love, even when I showed my rare signs of stress.

Abstract

Given the proliferation of data regarding an athlete's physical, physiological, psychological, and tactical state, Sports Scientists in the applied setting are increasingly being required to provide statistics, both descriptive and inferential, to coaches and other support staff to provide decision-making insight. However, these data sets can be imbalanced (i.e., more/less data on some athletes), contain many variables, include small sample sizes, and display only small individual/group differences. While these properties are not unique to sports science, these properties are often a barrier to performing robust statistical analysis. Therefore, the overall aim of this thesis is to provide Sports Scientists with access to applications of statistical methods that will expand their statistical toolkit to accommodate data sets regularly seen in a sports science context.

As sports science research consistently contains repeated measures and imbalanced data sets, Study 1 calls for further adoption of mixed models when analysing longitudinal sports-science data sets. Mixed models were used to understand whether the level of competition affected the intensity of women's rugby league match play recorded during club-, state-, and international-level competitions. As athletes featured in all three levels of competition and there were multiple matches within each competition (i.e., repeated measures), mixed models were shown to be the appropriate statistical method for these data. If a repeated-measures ANOVA were used for the statistical analysis in this study, at least 48.7% of the data would have been omitted to meet ANOVA assumptions. However, using a mixed model, it was determined that mean speed recorded during Test matches was 73.4 $m \cdot min^{-1}$, while the mean speed for NRLW

and Origin matches were 77.6 and 81.6 m · min^{-1}, respectively. Study 1 demonstrates how to use mixed models with typical data sets acquired in the professional sports setting and calls for mixed models to be more readily used within sports science, especially in observational, longitudinal data sets such as movement pattern analyses.

As athletes often require a mix of physical, physiological, psychological, and skill-based attributes, multiple variables need to be considered in tandem when identifying talent. Study 2 introduces Pareto frontiers as a technique that can identify the observations that possess an optimal balance of the desired attributes, especially when these attributes are negatively correlated. The study explores the trade-off relationship between batting average and strike rate as well as bowling strike rate, economy, and average in Twenty 20 cricket. Batting and bowling data from both the men's (MBBL) and women's (WBBL) Australian Big Bash Leagues were compiled to determine the best batting and bowling performances, both within a single innings and across each player's Big Bash career. Each Pareto frontier identified players that were not the highest ranked athlete in any metric when analyzed univariately and yet possess an optimal 'trade-off' between attributes. Study 2 concludes that Pareto frontiers can be used when assessing talent across multiple metrics, especially when these metrics may be conflicting or uncorrelated and that Pareto frontiers can identify athletes that may not have the highest ranking on a given metric but have an optimal balance across multiple metrics that are associated with success in a given sport.

As sporting success can be determined by the smallest margins, detecting small, worthwhile effects is notoriously difficult, especially given the small sample sizes seen in sports science.

Study 3 demonstrates how utilising a Bayesian framework when conducting research can incorporate prior information to assist in decision making when sample sizes and effect sizes are small. The study revisits a paper published in the European Journal of Sport Science investigating the effects of -alanine on 4-km cycling TT performance through a randomised placebo-controlled trial with 14 trained cyclists. By analysing the data in both a frequentist framework and in a Bayesian framework with priors varying in terms of informativeness, the study demonstrates why incorporating prior information can improve the quality of sports science research and reinforces that -alanine supplementation may be beneficial for cycling time trials of ~6 min in duration.

Study 4 builds on the call in Study 1 for mixed models to be further adopted, by using mixed models to quantify the position-specific demographics, technical match statistics, and movement patterns of the National Rugby League Women's (NRLW) Premiership. As women's rugby league grows, the need for understanding the movement patterns of the sport is essential for coaches and Sports Scientists. Global positioning system, demographic, and match statistics collected from all NRLW clubs across the full 2018 and 2019 seasons were compared between the ten positions using generalised linear mixed models. By understanding the load of NRLW matches, coaches, high-performance staff, and players can better prepare as the NRLW Premiership expands. These movement patterns and match statistics of NRLW matches can lay the foundation for future research as women's rugby league expands.

Finally, Study 5 brings together the mixed models discussed in Study 1, Pareto frontiers discussed in Study 2, and Bayesian inference discussed in Study 3. Team sport athletes require both speed and endurance to perform in their given sport; however, it is difficult to excel at an elite level in both attributes. Univariate analysis of these attributes is commonplace in sports science whereby speed is typically evaluated independently of endurance. While this methodology readily identifies athletes excelling in either speed or endurance, it fails to highlight athletes who possess 'the best compromise' in speed and endurance. Study 5 presents an innovative approach to evaluating running intensity during short (10 s; anaerobic power/speed) and long (20 min; aerobic power/endurance) periods across three seasons of elite female football matches by using the Pareto frontier to visualize athlete characteristics simultaneously. Given the differing number of observations for each athlete, the study uses samples drawn from the Bayesian posterior distribution of both rolling averages estimated by a multivariate mixed model to calculate the probability an athlete sits on the true population Pareto frontier. The 10-s and 20-min rolling averages were calculated (325.1 and 104.7 $m \cdot min^{-1}$, respectively) for the 18 elite female footballers to provide coaches with a holistic view of their athletes' speed and endurance capabilities and enable a more informed decision when identifying the athletes with the optimal balance of these polarizing attributes.

This thesis encourages Sports Scientists to develop in their statistical literacy to ensure validity and robustness in their statistics being presented for decision-making. As repeated-measures, imbalanced data sets, multiple variables of interest, small samples, and small effect sizes are commonplace in sports science, this thesis is written to urge Sports Scientists to develop a deeper understanding and improved competency in statistical methods. The thesis culminates with Study 5 illustrating how the three different statistical concepts explored in this thesis (i.e., mixed models, Pareto frontiers, and Bayesian inference) can be used in tandem to produce novel research. The thesis provides all the data and R code required to run all the analyses within this thesis to ensure Sports Scientists can replicate the analyses with their own data, as well as to provide an example to the sports science community of open and transparent research practices. By compiling this thesis as an eBook, it is intended that Sports Scientists can utilise these studies as a resource when conducting their own research.

Statement of Originality

This work has not been submitted previously for a degree or diploma in any university. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

29/03/2023

Timothy Newans

Acknowledgement of Papers included in this Thesis

Section 9.1 of the Griffith University Code for the Responsible Conduct of Research ("Criteria for Authorship"), in accordance with Section 5 of the Australian Code for the Responsible Conduct of Research, states: To be named as an author, a researcher must have made a substantial scholarly contribution to the creative or scholarly work that constitutes the research output, and be able to take public responsibility for at least that part of the work they contributed. Attribution of authorship depends to some extent on the discipline and publisher policies, but in all cases, authorship must be based on substantial contributions in a combination of one or more of:

- conception and design of the research project

- analysis and interpretation of research data

- drafting or making significant parts of the creative or scholarly work or critically revising it so as to contribute significantly to the final output.

Section 9.3 of the Griffith University Code ("Responsibilities of Researchers"), in accordance with Section 5 of the Australian Code, states: Researchers are expected to:

- Offer authorship to all people, including research trainees, who meet the criteria for authorship listed above, but only those people.

- accept or decline offers of authorship promptly in writing.

- Include in the list of authors only those who have accepted authorship

- Appoint one author to be the executive author to record authorship and manage correspondence about the work with the publisher and other interested parties.

- Acknowledge all those who have contributed to the research, facilities or materials but who do not qualify as authors, such as research assistants, technical staff, and advisors on cultural or community knowledge. Obtain written consent to name individuals.

Included in this thesis are papers in Chapters 2-6 and Appendix A which are co-authored with other researchers. My contribution to each co-authored paper is outlined at the front of the relevant chapter. The bibliographic details (if published or accepted for publication)/status (if prepared or submitted for publication) for these papers including all authors, are:

Chapter 2: **Newans, T.**, Bellinger, P., Drovandi, C., Buxton, S., & Minahan, C. The utility of mixed models in sports science: A call for further adoption in longitudinal datasets. International Journal of Sports Physiology and Performance. 17:8 p. 1289-1295.

Chapter 3: **Newans, T.**, Bellinger, P., & Minahan, C. The balancing act: Identifying multivariate sports performance using Pareto frontiers. Frontiers in Sports and Active Living. 4:918946.

Chapter 4: **Newans, T.**, Bellinger, P., Drovandi, C., & Minahan, C. The role of informative priors: A new look at the role of -alanine on 4-km time-trial performance in cyclists. European Journal of Sports Science. Submitted for publication.

Chapter 5: **Newans, T.**, Bellinger, P., Buxton, S., Quinn, K., & Minahan, C. Movement patterns and match statistics in the National Rugby League Women's (NRLW) Premiership.

Frontiers in Sports and Active Living. 3:618913.

Chapter 6: **Newans, T.**, Bellinger, P., Drovandi, C., Griffin, J. & Minahan, C. Bayesian approximation of the trade-off relationship between running intensity measured during short and long periods using Pareto frontiers. Prepared for publication.

Appendix A: **Newans, T.**, Bellinger, P., & Minahan, C. Identifying multivariate cricket performance using Pareto frontiers. MathSport Conference 2022.

Appropriate acknowledgements of those who contributed to the research but did not qualify as authors are included in each paper.

29/03/2023

Timothy Newans

29/03/2023

Supervisor: Clare Minahan

Additional publications not included in this thesis that have been made during this candidature

Minahan, C., **Newans, T.**, Quinn, K., Parsonage, J., Buxton, S., & Bellinger, P. Strong, Fast, fit, lean, and safe: a positional comparison of physical and physiological qualities within the 2020 Australian women's rugby league team. The Journal of Strength & Conditioning Research. 35 p. S11-S19. https://doi.org/10.1519/JSC.0000000000004106

Griffin, J., **Newans, T.**, Horan, S., Keogh, J., Andreatta, M., & Minahan, C. Acceleration and high-speed running profiles of women's international and domestic football matches. Frontiers in Sports and Active Living. 3:604605. https://doi.org/10.3389/fspor.2021.604605

Quinn, K., **Newans, T.**, Buxton, S., Thomson, T., Tyler, R., & Minahan, C. Movement patterns of players in the Australian Women's Rugby League team during international competition. Journal of Science and Medicine in Sport. 23:3 p. 315-319. https://doi.org/10.1016/j.jsams.2019.10.009

Bellinger, P., **Newans, T.**, Whalen, M., & Minahan, C. Quantifying the activity profile of female beach volleyball tournament match-play. Journal of Sports Science & Medicine. 20:1 p. 142-148. https://doi.org/10.52082%2Fjssm.2021.142

Bellinger, P., Ferguson, C., **Newans, T.**, & Minahan, C. No influence of prematch subjective wellness ratings on external load during elite Australian Football match play. International Journal of Sports Physiology and Performance. 15:6 p. 801-807. https://doi.org/10.1123/ijspp.2019-0395

1 Review of the Literature

1.1 Introduction

Sports Scientists are increasingly being required to collect a lot of data and subsequently perform complex statistical analysis to support their decision making (Garfield & Ben-Zvi, 2008). Indeed, wearable technologies, such as receivers using the Global Navigation Satellite System (GNSS), local positioning systems, heart-rate monitors, and inertial movement unit sensors have provided the opportunity to collect large amounts of data (Crang et al., 2021; Seshadri et al., 2021). Recognising the need for Sports Scientists to possess the statistical analysis skills necessary to robustly analyse sports science data, Exercise & Sports Science Australia (ESSA), have now included 'Data Handling and Management' as one of their six standards for Sports Science accreditation (Exercise and Sports Science Australia, 2019). This standard for accreditation comprises three requirements, including: i. Assesses data critically to identify meaningful effects, ii. Uses data to evaluate and develop programs for service users, and iii. Translates the outcomes of data analysis into meaningful information for service users and other relevant stakeholders (Exercise and Sports Science Australia, 2019). While these requirements for Sports Science accreditation may emphasize the importance of statistical literacy, it is unclear whether simply meeting these requirements will result in a substantial improvement in the statistical literacy of Sports Scientists. Additional training, practical experience, and ongoing professional development may be necessary to truly enhance statistical literacy in this field.

1.2 Defining statistical literacy

While the etymological root of the word literacy refers as 'an ability to read and write letters', the Queensland Department of Education and Training (2023) defines literacy as:

> "… the ability to read, view, write, design, speak and listen in a way that allows us to communicate effectively and to make sense of the world."

This definition moves past the idea of simply reading and writing alphabetic letters, allowing the word literacy to be used to describe a level of competency in other concepts (Queensland Department of Education and Training, 2023). For example, Whitehead (2010) argued that the concept of physical literacy comprised more than skill *per se*, stating that:

> "Physical literacy describes the motivation, confidence, physical competence, understanding and knowledge that individuals develop in order to maintain physical activity at an appropriate level throughout their life."

(Whitehead, 2010)

Whitehead's articulation of physical literacy is now a key component of physical education globally (Scott et al., 2021) and captures four key attributes to physical literacy: i. **Knowledge and understanding** of why physical activity is necessary, ii. **Competence** in performing a physical activity, iii. **Confidence** in knowing they are proficient in the physical activity and the associated benefits, and iv. **Motivation** to develop and improve in physical activity. In a similar vein, the idea of 'statistical literacy' was developed Gal (2002) who proposed that:

> "
>
> *Statisticalliteracyis*
>
> people's ability to interpret and critically evaluate statistical information … and their ability to discuss or communicate their reactions to such statistical information such as their understanding of the meaning of the information, their opinions about the implications of this information, or their concerns regarding the acceptability of the given conclusions."

Within this definition there are two key components: i. The ability to interpret and critically evaluate statistical information and ii. The ability to discuss or communicate their reactions to statistical information. As Gal writes, these two abilities move past the minimal standard required for what used to ratify as literacy, to a deeper understanding built on inter-related knowledge bases.

As Sports Scientists are seeking to utilize data sets to make meaningful inferences, there is a minimum requisite level of statistical literacy required to correctly analyse sports science data sets. However, Sports Scientists may be lacking in statistical literacy if they have not received adequate formal training in appropriate statistical methods. Having adopted the four key attributes of Whitehead's physical literacy framework and considering Gal's statistical literacy definition, we propose the following framework for statistical literacy for Sports Scientists:

1. **Knowledge and understanding** of statistical principles and robust modeling techniques;
2. **Competence** in performing the required statistical analyses;
3. **Confidence** in performing statistical methods on unfamiliar new data sets and the subsequent interpretation of the analysis;
4. **Motivation** to ensure robustness and validity in the statistical analysis.

As undergraduate sports science programs typically contain an introductory statistics course, 'Statistics for Sports and Exercise Science' (Newell et al., 2014) provided Sports Scientists with complete data sets and software guides to assist readers through appropriate execution of more complex statistical analyses in a variety of situations. These resources are vital to provide Sports Scientists with tutorials on how to execute robust statistical analysis. Similarly, it also promotes the idea of 'reproducible research', in which the code used to run the analysis can be published as a supplementary material that can critiqued for statistical rigour. However, due to some idiosyncrasies seen in commonly found sports science data sets, Sports Scientists have been utilising alternative methods, such as 'customised spreadsheets' (Batterham & Hopkins, 2006), to provide their statistical inferences for their decision making. By relying on customised spreadsheets, a Sports Scientist needs to trust the underlying mathematics by the authors of the spreadsheets which, of recent times, have occasionally been shown to be flawed (Barker & Schofield, 2008; Sainani, 2018; Welsh & Knight, 2015). The subsequent conversations and rebuttals have challenged the status quo within sports science and have encouraged Sports Scientists to focus on understanding the data set and correctly identifying the suitable robust statistical methods.

This thesis presents the analyses of data collected in an applied setting using statistical principles, techniques, and methods that are not commonplace in sports-science research literature. Its intention is to contribute to the statistical literacy of Sports Scientists by presenting, and encouraging the further adoption of, statistical principles, techniques, and methods that could be utilised by Sports Scientists.

1.3 Common characteristics of sports science data sets

Data obtained by Sports Scientists often cannot conform to the rigid assumptions required to perform traditional parametric statistics. There are five key characteristics of data commonly collected by Sports Scientists that require careful consideration when conducting statistical analyses:

Repeated observations are measures taken on multiple occasions of the same participant either within a discrete session such as a single match (Newans et al., 2019) or across multiple sessions (Quinn et al., 2020). For example, high-speed locomotive data using the GNSS is routinely collected on athletes during match play, in training, and across multiple seasons (Griffin et al., 2021). In this instance, the assumption of 'independence of observations', a requirement for many parametric inferential statistics, is not met (Hopkins et al., 2009). Consequently, statistical methods that can accommodate for repeated observations are required.

Missing data points can be incurred when measurements are taken on several occasions in a participant group (i.e., repeated observations), but not all participants have the same number of observations (Nakai & Ke, 2011). For example, some athletes may play all matches across one season ensuring their GNSS data set is complete. However, other athletes may have missing GNSS data for one or more games due to injury or squad selection. In which case, the data is classified as Missing at Random (MAR), where these missing data could bias the results (Borg et al., 2021). Alternatively, there could be equipment malfunction leading to data that is Missing Completely at Random (MCAR), which is less likely to bias the results (Borg et al., 2021). Consequently, this creates an imbalanced data set, in which some athletes have more observations than other athletes, ruling out statistical methods, such as the repeated-measures ANOVA which require a 'complete case analysis' data set (Nakai & Ke, 2011). Thus, Sports Scientists require methods that can accommodate data sets with irregular missing data points.

Small sample sizes refers to a small number of participants included in a study. In sport, the number of athletes able to participate at the highest-level of competition are relatively few and sports scientists conducting studies at the elite level typically only having access to one or two teams/squads (Bernards et al., 2017; Glassbrook et al., 2019), thus small samples sizes are commonplace in sports science research. Indeed, Abt et al. (2020) randomly sampled 120 studies from the Journal of Sports Sciences and found that the median sample size was 19 participants. Importantly, small samples may lack sufficient statistical power to extrapolate the statistical analysis results to the overall population (Abt et al., 2020; Speed & Andersen, 2000); therefore, methods to accommodate for such small sample sizes are required.

Small effect sizes refers to the small true between-individual differences in competitive performances (Mengersen et al., 2016). As the highest-level of competition within a given sport is necessarily exclusive, the athletes within these competitions are at, or near, the peak of their sporting abilities. Consequently, when conducting research on these athletes, the improvements being sought after through interventions or ergogenic aids are relatively small compared to improvements that could be seen in a general population. For example, in the 2020 Tokyo Olympics 100-metre men's final, only 0.18 s separated the field. While a change of 0.18 s would be deemed relatively immaterial for a person from the general population in improving their 100-metre sprint time, it would be the difference between a gold medal or no medal at all for an Olympic athlete. As a result, statistical methods that can detect small differences within a sample are also required.

1.4 Recurrent flaws in the statistical approach to sports science data sets

1.4.1 Inappropriate analysis of imbalanced data

One of the most-commonly violated assumptions in general linear models when analysing elite athletes' data is 'independence of observations'. This can be violated when describing the movement patterns of a sport where there are repeated measurements of the same participants (i.e., the athletes) across multiple matches (i.e., time points). The 'independence of observations' assumption of general linear models states that within each condition, an individual can only be included once. This is often seen where one player has been recorded in more matches than another athlete and, as such, the data set is weighted more heavily towards the athletes with more observations. These MAR data points creates an 'imbalanced' data set, further compounded by possible MCAR data points from errors during data collection (e.g., sensors malfunctioning, user error) (Borg et al., 2021). Consequently, for general linear models, unless the Sports Scientist is stringent in the inclusion of data, assumptions can easily be violated.

Repeated-measures analysis of variance (RM-ANOVA) is a commonly used general linear model applied in these studies (Dalton-Barron et al., 2020) due to the repeated measurements of the same participants (i.e., the athletes); however, unless strict guidelines are adhered to, it can produce misconstrued results. RM-ANOVA requires a 'complete' data set to perform the analysis, that is, every participant must have a value in every observation, and within a condition (e.g., position) every participant must have contributed an equal number of observations

(Kenny & Judd, 1986). This is troublesome, as elite sports' data sets are rarely 'complete' due to the frequent missing data points arising from injuries, player selections, and the previously mentioned lack of time, resources, and abilities. As a result, this limitation severely hampers the data set's abilities to make inferences due to the frequent missing data points.

To account for missing data, data sets need to be manipulated to correctly perform general linear models. Whenever there is missing data within an RM-ANOVA, either the time-point or the athlete needs to be excluded or an imputed data point needs to be inserted to retain a 'complete case analysis' data set (Nakai & Ke, 2011). While removing an athlete if they do not have an observation in every time point or removing a time point if many athletes are missing is statistically appropriate if the missing data is proved to be random (i.e., no bias in missing data), this method can cause unnecessary deletion of large amounts of data. Further, if there are many participants and many time points, it can sometimes eliminate so much data that no analysis can be performed on the remaining data set. Therefore, it is improper to use just complete cases and different methods needs to be explored.

The second method of handling missing data that is commonly used in sports science is data imputation (Borg et al., 2021). Sports Scientists employ different methods of 'imputing' estimated data in place of missing data, with some utilising the time point mean, some using the athlete's mean, while others will employ the "last observation carried forward" method, where the last observed value is imputed for the missing data. While imputing data can be useful when there was an error in data collection (e.g., GNSS receiver loses signal during a match), imputing data when an athlete never played in a match is inappropriate, Similarly, the 'last observation carried forward' method is also dangerous as it carries the strong assumption that there was no change in conditions between the previously observed time-point and the missing data time-point. For example, if the previously-observed time-point was in the middle of the day, while the missing data time-point was during the night, it would be improper to assume that the same conditions were present in both time points. Therefore, it is necessary for Sports Scientists to utilise a method that can appropriately account for missing data.

1.4.2 Univariate analyses of multivariate concepts

When conducting inferential statistics, the goal is to translate information learned about a sample to make inferences about a given population. As a theoretical framework, the population parameters are never fully known, thus inferences are made based on the sample statistics. Typically, the mean and the standard deviation of the sample are the two key statistics gathered from the sample for inferences regarding the population mean to be made via the use of a sampling distribution. This suits most research applications which focus on the population mean as the parameter of interest to answer a given research question. However, in sports science and talent identification, practitioners are rarely focused on finding the mean of a population; rather, they are focused on identifying extreme values that may possess superior performance than the average athlete within a cohort (K. Johnston et al., 2018). Consequently, statistical methods that can elucidate information regarding the extreme values of a given data set are also of use within a Sports Scientist's statistical toolbox.

A stumbling block for Sports Scientists in identifying talent is the proliferation of testing protocols, leading to an abundance of metrics concerning multiple attributes of an athlete's talent (e.g., physical, physiological, psychological, tactical etc.) (Dodd & Newans, 2018). While factor-reduction techniques can be used to develop a conjugate for these various attributes (Vaeyens et al., 2008), there is still a need to balance each of these attributes if each is equally desirable in an athlete. For example, assume there are three attributes of interest and that each attribute has a normal distribution and is completely independent of the other attributes. If an athlete that is at least 1 standard deviation above the mean (i.e., approximately top 16%) is deemed 'good', then the probability a randomly selected athlete is 'good' in all three attributes is approximately 0.4%. However, if we define that an athlete that is at least 2 standard deviations above the mean (i.e., approximately top 2.5%) is deemed 'exceptional', the likelihood that a randomly selected athlete would be exceptional in all three attributes drops to 0.002%. Unless a sporting organisation had the testing capability to identify the 2 in 100,000 that possess an exceptional standard in all three attributes, then they will necessarily have to compromise on some metrics in the process of recruitment. As a result, athletes possess the best compromise of the attributes and are only 'good' in all three attributes may get missed in talent identification processes if the scouts and coaches screen using a univariate analysis for each attribute.

1.4.3 Under-powered studies

Given the exclusiveness of the highest-level of competition in sports and that athletes are of similar standards within these competitions, Sports Scientists are often faced with small sample sizes and/or small effect sizes when conducting research to improve these athletes (Atkinson et al., 2012). As classical statistics (commonly referred to 'frequentist statistics') has revolved around traditional null hypothesis statistical testing (NHST), its ability to provide meaningful inferences in decision making for Sports Scientists is inherently difficult (Batterham & Hopkins, 2006). Statistical methods learned in undergraduate statistics courses use moderate-to-large size samples which are amply powered, Sports Scientists' sample sizes are typically much smaller than that, due to the aforementioned exclusivity of elite sports and their association with typically only one team (Bernards et al., 2017; Glassbrook et al., 2019). For example, in team sports such as basketball, teams are only permitted limited to squads of 15 athletes, drastically limiting a Sports Scientist's sample size. While the sample size and effect size issues are not limited to sports science (Bacchetti et al., 2011; Ploutz-Snyder et al., 2014), these issues provide a barrier to the ability to perform robust statistical tests. Similarly, as the slightest differences in score, time, or distance could affect the result of a sport, any marginal gains are seen as a necessity (Batterham & Hopkins, 2006). However, a small effect size in the true population may also not be detected (Borenstein, 2009; Mengersen et al., 2016; Speed & Andersen, 2000). Consequently, under-powered studies are pervasive within sports science research (Speed & Andersen, 2000) and methods to accommodate for such small sample sizes and effect sizes are required.

1.4.4 Probabilistic Statements of Inference

The use of probabilistic statements of inference within the sports science community has been of increasing interest over the past 15 years (Batterham & Hopkins, 2006; Borg et al., 2018; Mengersen et al., 2016; Sainani, 2018; Welsh & Knight, 2015). Of particular note, Batterham & Hopkins (2006) began compelling readers that NHST is unnecessarily restrictive because NHST has the assumption that the parameter of interest is 'fixed' (e.g., the population mean) and the probability that the data has arisen from this fixed parameter is calculated. If it is unlikely that the data has originated from this parameter, the null hypothesis is deemed to be rejected and that the parameter is different to what was hypothesised. Given the small samples and small effect sizes seen in sports science, there has been a push for more clarity around probabilistic distributions than what NHST offers. Consequently, Batterham and Hopkins provided an alternative to NHST named 'magnitude-based inferences' (Batterham & Hopkins, 2006). This newly established method aiming to provide probabilistic inference statements gained traction quickly, amassing over 1700 cites and was published in guidelines for authors for some journals (Hopkins et al., 2009). While this method received early criticism (Barker & Schofield, 2008), this criticism gained little traction. More attempts to discredit the concept of magnitude-based inferences were vocalized by Welsh & Knight (2015) who provided a statistical review of the method and illustrated the method's unacceptably high levels of Type I error rates (Welsh & Knight, 2015). However, it was not until Sainani (2018) weighed into the statistical debate which triggered the sports science community to critically assess whether magnitude-based inferences should be accepted within literature (Sainani, 2018). As a consequence, the

Medicine in Sports and Exercise (MSSE) editorial board has even explicitly stated that studies should not use magnitude-based inferences for any statistical inferences (Medicine & Science in Sports & Exercise, 2023). However, this has left a need for robust probabilistic statements of inference to help inform sports science practitioners and researchers.

1.4.5 Transparency in data

An additional concept to which statistical literacy is required is amidst the reproducibility crisis seen in sports science research (Caldwell et al., 2020). In a standard sports science research paper, researchers are expected to give a thorough explanation of the methods used to gather their data, or otherwise cite an already-published explicit methods study. This process allows another researcher to gather data in conditions as homogenous as possible to the original research. While this is seen as traditional practice, the same rigour is not currently applied to the implementation of statistical methods in which merely mentioning the statistical method used can pass through editorial review. This lack of academic rigour results in a replication and a reproducibility issue (John et al., 2012). With manipulation of hypotheses after results are known, p-hacking, data dredging, cherry picking, and the file drawer problem all resulting in a distorted view of the reality (Franco et al., 2014), the need for greater transparency in the way researchers explain their statistical approach within studies is paramount.

Borg et al. (2020) sampled 299 'sports science' studies to understand the availability of data and the associated computer codes used in these studies and found that none of the articles provided the code used to perform the statistical tests and only 4.3% shared the data used in the study (Borg et al., 2020). This shift in ideology within sports science can help alleviate the commonly faced issues previously mentioned. While researchers may only have access to the data of one or two teams (Bernards et al., 2017; Glassbrook et al., 2019), by adopting a more transparent mindset to research, researchers can collaborate with other teams working on similar projects, thus increasing their sample size, to strengthen their research impact by alleviating struggles observed with small sample sizes (Batterham & Hopkins, 2006).

1.5 Being comfortable with variability and uncertainty

Within sports, the winner is typically determined using a quantitative scale, whether that be a tally of points, a judge's score, a duration, or a distance. While declaring the winner can be determined through simple mathematics (e.g., one team scored higher than the other team), it does not take into consideration the events leading to the outcome. That is, if each of the events within a match were to be replicated, would it be predicted that the outcome would be the same? Consequently, there is a need for statistics to provide the context around the outcome of the sport. This difference between mathematics and statistics has been discussed by Cobb & Moore (1997) who state that:

> "Statistics is a methodological discipline. It exists not for itself, but rather to offer to other fields of study a coherent set of ideas and tools for dealing with data. The need for such a discipline arises from the omnipresence of variability."

Statistical literacy is required to appropriately handle data that exhibits inherent variability and, thus, understanding the uncertainty regarding inferences arising from the data. Statistical literacy is required to identify the correct statistical approach (knowledge), skillfully code and interpret the output (competence), and tackle an unfamiliar data set (confidence). The recognition and acceptance of variability in a data set is at the forefront of the GAISE report (Franklin et al., 2007). This document outlines that the objective of statistics education is for students to develop the ability to deal with the ubiquity of variability and the decision-making ability of interpreting the variability in the data (Franklin et al., 2007).

As a result, for Sports Scientists to be statistically literate, they need to understand there is an inherent variability surrounding that quantitative scale, often which can be the difference between winning and losing (Batterham & Hopkins, 2006). Similarly, when quantifying the response to a given external load, there will be an inherent variability due to the plethora of external factors contributing to the load within a session (Dalton-Barron et al., 2020). Consequently, there is a level of uncertainty with decision making as a function of the variability within the data being collected, most plainly seen in the standard error of the mean formula:

$$SE_x = \frac{s_x}{\sqrt{n}}$$

The level of uncertainty (in this case, standard error denoted by SE_x is determined by two variables: a level of variability (the standard deviation denoted by s_x), and a sample size (denoted by n). As a result, there are two contributing factors to a high level of uncertainty: a) a high level of variability and b) a small sample size. Unfortunately, sports science data sets often feature both factors which suggests inferences are intrinsically difficult to deduce. As a result, statistical methods that can robustly account for the variability within sports science data sets and can provide estimates that reflect the uncertainty arising from the data are required for Sports Scientists to perform high-quality research.

1.6 Mixed Models

When quantifying the external load of an athlete within a match, researchers will often gather data from the GNSS for multiple athletes across multiple matches. When the data set is compiled, it is highly unlikely that the same athletes played for the team every week (due to injury, non-selection etc.). As a result, three possible options exist to ensure that the assumptions of a RM-ANOVA are not violated:

1. Eliminate any athlete who did not play every match;
2. Eliminate any match that contained players who didn't play many matches;
3. Impute missing data for any athlete that did not play in a given match.

Both option 1 and 2 unnecessarily eliminate crucial data, while option 3 is improper to impute data into matches that a player never played in. Despite these limitations, RM-ANOVA remains a common statistical method to analyse GNSS-derived data (Dalton-Barron et al., 2020; Mara et al., 2017; Russell et al., 2016; Vigh-Larsen et al., 2018). These previously mentioned limitations severely hamper a general linear model such as an RM-ANOVA to make inferences due to the frequent missing data points. In contrast, there are few examples of sports scientists who have incorporated the use of mixed models into their research to accommodate the challenges of sports science data sets (Delaney, Thornton, et al., 2016; Kempton et al., 2017; Newans et al., 2019; Quinn et al., 2020).

Mixed models provide an improved alternative to general linear models within sports science data sets. Mixed models can estimate the population mean by accounting for the variability within each athlete as well as the variability in each time-point. Given the differing numbers of observations for each athlete, the mixed model 'shrinks' the estimates for each athlete based on their number of observations. This concept, known as 'partial pooling', means that athletes with few observations are shrunk closer to the mean intercept and slope than athletes with more observations (McElreath, 2018). Similarly, it can also account for correlations both between and within athletes (Kwon et al., 2014). For example, within rugby league there are nine distinct positions (plus interchange players); however, some of these positions have more similar movement patterns than other positions. If a five-eighth were to play at halfback for some matches, there would be minimal change in movement patterns; however, if they were to play fullback or lock, a marked change in their movement patterns would be apparent (Glassbrook et al., 2019). Consequently, the properties of mixed models are particularly useful in such data sets in which athletes differ substantially in baseline physiological measures, as well as fluctuate in their physiological measures based on differing training conditions. By accounting for the athlete's baseline, the condition (e.g., position in rugby league) can be more easily interrogated. As a result, mixed models can more accurately and validly provide information on athletes, time points, and conditions much more robustly than general linear models.

1.7 Pareto Frontiers

Statistical methods that can evaluate the trade-off relationship between equally desirable attributes are required when identifying talent. In statistics, this is called 'multi-objective optimisation'; that is, we are aiming to optimise the value of each variable at the least expense of another variable. Consequently, the solution is a series of values in which it is not possible to improve one variable without detracting from another variable. That is, the series of values that possess the best compromise between the variables of interest. This series of values is deemed the 'Pareto frontier'. This series is easily visualised as seen by the red line in Figure 1.1.

```
library(tidyverse)
library(rPref)
set.seed(99.94)
df <- runif(35, 40, 80) %>%
  data.frame(x = .) %>%
  mutate(y = (1 / x) * rnorm(35, 100, 50)) %>%
  psel(high(x) * high(y), top_level = 99) ## Generate Pareto frontier for data set of 35 r

ggplot(df, aes(x = x, y = y, color = as.factor(`.level`), alpha = as.factor(`.level`))) +
```

```r
geom_point(size = 3) +
geom_line(data = df %>% filter(`.level` == 1),
          color = "red",
          linewidth = 1) +
scale_color_manual(values = c("red", rep("grey20", 10))) +
scale_alpha_manual(values = c(1.0, rep(0.4, 10))) +
coord_cartesian(xlim = c(45, 80),
                ylim = c(1, 4.5)) +
labs(x = "Variable 1",
     y = "Variable 2") +
theme_minimal() +
theme(
  legend.position = "none",
  axis.text = element_blank(),
  axis.title = element_text(size = 10, color = "black", face = "bold")
) ## Plot the Pareto frontier of the points
```

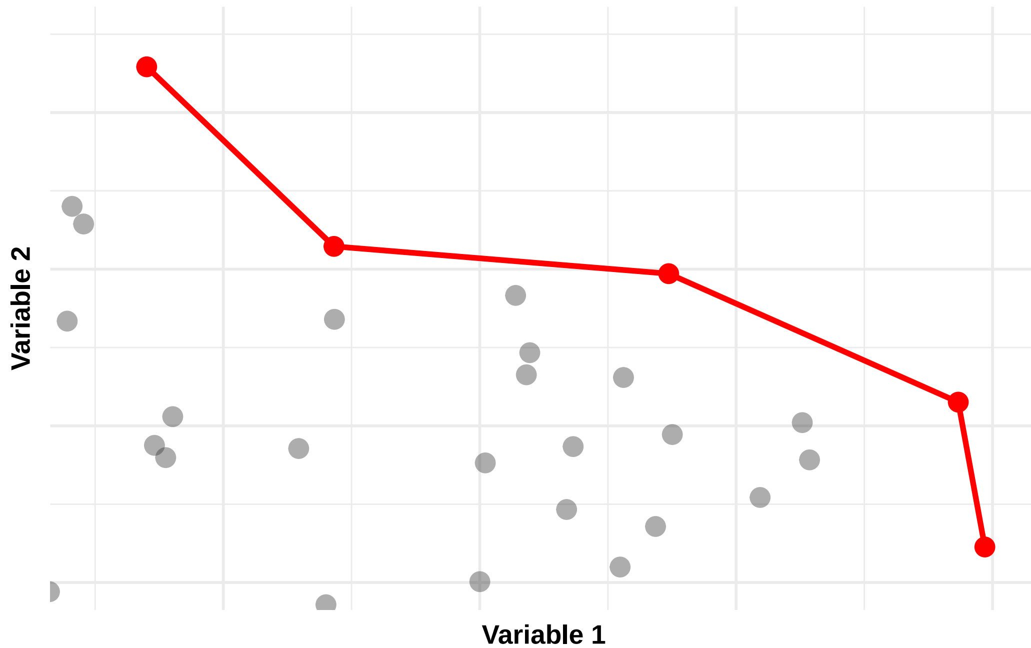

Figure 1.1: Illustration of the Pareto frontier, highlighted in red.

The Pareto frontier allows decision-makers to assess the series of values and determine what trade-offs would be required if an individual were to improve in a certain variable. While originally explored in economics by Vilfredo Pareto in the early 1900's to understand the wealth distribution in societies, the applications have been wide-reaching, with extensive use-cases within economics (Horn et al., 1994; Ponsich et al., 2012; Tapia & Coello, 2007) and engineering disciplines (Deb & Datta, 2012; Gunantara, 2018; Marler & Arora, 2004; Mastroddi & Gemma, 2013).

For example, Ottosson et al. (2009) explored the trade-off problem when providing radiation therapy for cancer treatment. There are two key factors: i. maximising the radiation dose on the targeted area of cancerous cells and ii. minimising the radiation damage to nearby healthy cells. Given the inter-dependency of these variables, no one solution is achievable; rather, a series of possible solutions were all returned as optimal solutions (that is, no other solution is superior in both variables of interest). Consequently, this set of observations can be provided to decision-makers to determine which solution is most suitable for the given use-case.

With the abundance of metrics concerning multiple attributes of an athlete's talent, understanding the Pareto frontier of the relationship between these metrics allow Sports Scientists to identify more easily those with the best compromise between the attributes. For example, when batting in Twenty20 cricket, the objective is to score as many runs as possible; but, given the limited number of balls possible (120 balls in a standard Twenty20 innings), there is an impetus for the batter to score their runs as fast as possible. Consequently, this creates a trade-off relationship between the number of runs scored per dismissal (referred to as the batting average) and the numbers of runs scored per 100 balls faced (referred to as the batting strike rate). With univariate analyses, only the batters with either the highest batting average or highest strike rate would be identified; however, this disregards the batters that possess the optimal trade-off between these metrics (i.e., the 'hybrid' athlete). By constructing the Pareto frontier on the relationship between these metrics, not only will the best batter for batting average and best batter for strike rate be identified, but also the set of batters in which no other batter scores more runs per dismissal at a faster strike rate will also be identified.

1.8 Bayesian Modeling

Throughout all the rebuttals of magnitude-based inferences (Barker & Schofield, 2008; Borg et al., 2018; Mengersen et al., 2016; Sainani, 2018; Welsh & Knight, 2015), the recurring rhetoric is for Sports Scientists to adopt a fully Bayesian approach to provide probabilistic statements when inferring conclusions. In response to the growing use of magnitude-based inferences, Mengersen and colleagues (2016), provided a worked example and template to perform Bayesian statistics in exercise and sports science data sets. The paper clearly articulated the need for a Bayesian analysis for providing probabilistic statements, similar to those provided by magnitude-based inferences (Mengersen et al., 2016). With some tutorials being developed to provide readers with the skills in executing Bayesian modelling (Kruschke, 2014; Quintana & Williams, 2018), there appears to be a growing trend in the use of Bayesian statistical methods in sports science (Santos-Fernandez et al., 2019). However, it seems that there is still a lack of accessibility for Sports Scientists to grasp the knowledge of why Bayesian inference is necessary (Mengersen et al., 2016). Therefore, the pathway from realizing the lack of statistical rigour in these customized spreadsheets to perform fully Bayesian analysis needs to be continually improved.

In a frequentist framework, the parameter is fixed and the probability that the data fits the null parameter is calculated. As a result, only interpretations with respect to the null parameter can be made. However, in a Bayesian framework, the data is fixed and, therefore, the probabilities and subsequent distributions for each parameter can be calculated, providing much richer information to interpret. Consequently, a key benefit that Bayesian inference provides is the ability to incorporate prior information to curate the model and provide estimates with more certainty than frequentist models can provide. For example, consider the example used earlier in the 2020 Tokyo Olympics Men's 100-metre running. When establishing the true population mean race time for elite male 100-m sprinters, a frequentist model has no assumed knowledge; therefore, assumes any race time from negative infinity to positive infinity is a possible result. While a different distribution could limit the sample space to only positive times, that is still untenable, as the margins of victory (i.e., 0.18 s separating 1st and last place) are so minute comparative to the sample space. However, as Sports Scientists have expert knowledge on the possible sample space (e.g., 8 to 12 s to be overly conservative), by inserting this prior knowledge into a Bayesian model will limit the possible sample space so that the Bayesian model will only need to iterate over a narrower range to establish where the true population mean lies.

1.9 Thesis Aims

The aim of this thesis is to provide Sports Scientists with access to applications of statistical methods that will expand their statistical toolkit to accommodate data sets regularly seen in a sports science context.

The specific aims of the experimental chapters were to:

1. Demonstrate the utility of mixed models for Sports Scientists when analysing longitudinal data sets;

2. Introduce Pareto frontiers to the sports science community and illustrate how they can identify players with the optimal balance of attributes that can be obfuscated when performing univariate analysis across each metric;
3. Explore how various prior distributions in a Bayesian framework can affect the inferences relating to the research hypothesis;
4. Examine the position-specific demographics, technical match statistics, and movement patterns of the National Rugby League Women's (NRLW) Premiership;
5. Utilise Pareto frontiers to visualise the trade-off relationship between short- and long-duration running intensities.

1.10 Thesis Structure

There are five experimental chapters to this thesis. The first three of which are more methodology-focused, with each chapter highlighting a different statistical principle. Namely, Chapter 3 explores mixed models, Chapter 4 explores Pareto frontiers, and Chapter 5 explores Bayesian inference. Following this are another two chapters that are more application-focused, building off the three methodology chapters. Chapter 6 presents a relatively simple mixed model application study, while Chapter 7 serves as the 'capstone' study, presenting a Pareto frontier of conditional effects arising from a mixed model in Bayesian framework.

2 The Utility of Mixed Models in Sports Science

2.1 Abstract

Title: The utility of mixed models in sports science: A call for further adoption in longitudinal data sets

Purpose: Sports science research consistently contains repeated measures and imbalanced data sets. This study calls for further adoption of mixed models when analysing longitudinal sports-science data sets. Mixed models were used to understand whether the level of competition affected the intensity of women's rugby league match play.

Methods: A total of 472 observations were used to compare the mean speed of female rugby league athletes recorded during club-, state-, and international-level competition. As athletes featured in all three levels of competition and there were multiple matches within each competition (i.e., repeated measures), we demonstrated that mixed models are the appropriate statistical approach for these data.

Results: We determined that if a repeated-measures ANOVA were used for the statistical analysis in the present study, at least 48.7% of the data would have been omitted to meet ANOVA assumptions. Using a mixed model, we determined that mean speed recorded during Test matches was 73.4 $m \cdot min^{-1}$, while the mean speed for NRLW and Origin matches were 77.6 and 81.6 $m \cdot min^{-1}$, respectively. Random effects of team, athlete, and match all accounted for variations in mean speed, that otherwise could have concealed the main effects of position and level of competition had less-flexible ANOVAs been used.

Conclusion: These data clearly demonstrate the appropriateness of applying mixed models to typical data sets acquired in the professional sports setting. Mixed models should be more readily used within sports science, especially in observational, longitudinal data sets such as movement pattern analyses.

The following chapter is a copy of the published manuscript:

Newans, T., Bellinger, P., Drovandi, C., Buxton, S., & Minahan, C. The utility of mixed models in sports science: A call for further adoption in longitudinal datasets. International Journal of Sports Physiology and Performance. 17:8 p. 1289-1295.

As co-author of the paper "The utility of mixed models in sports science: A call for further adoption in longitudinal datasets", I confirm that Timothy Newans has made the following contributions:

- Study concept and design
- Data collection
- Data analysis and interpretation
- Manuscript preparation

Name: Clare Minahan

Date: 29/03/2023

2.2 Introduction

Highly-controlled, longitudinal, or cross-over designed research experiments are difficult to conduct in the high-performance sport environment. Researchers are regularly met with hesitation from both coaching staff and the athletes due to the potential disruption 'best-practice' research methodology poses to competition and their carefully configured training programs. Nonetheless, to advance the understanding of performance and adaptation in elite athletes, Sports Scientists, and researchers often interrogate routinely collected health (Bartlett et al., 2020; Ferris et al., 2018; Halson, 2014), wellbeing (Bellinger et al., 2020; Gallo et al., 2017; Govus et al., 2018), physical and physiological performance metrics, and locomotive movement data (i.e., movement patterns) (Griffin et al., 2021; Newans et al., 2021; Quinn et al., 2020; Thornton et al., 2020) over multiple days, weeks, and years. Indeed, these observational studies have played a critical role in the understanding and development of sports performance and related disciplines over the last 20 years. However, these data sets are often characterized by multiple dependent observations (not only across matches and competitions, but within each athlete) and imbalanced data (e.g., athletes missing due to injuries, team selections, and/or rescheduling) that pose a significant challenge for this type of research. Nonetheless, the nuisances associated with dependent observations and imbalanced data are among the least acknowledged when conducting sports-science research.

Analysis of variance (ANOVA) with repeated measures has dominated as the most frequently utilized statistical method with which to analyse repeated measures data sets. However, ANOVA requires every participant to have a value in every observation and within a condition (e.g., position) every participant must have contributed an equal number of observations to avoid violating the relatively stringent assumptions (Kenny & Judd, 1986). Despite these assumptions, ANOVAs are still readily used within longitudinal sports science research (Hannon et al., 2021; Haugnes et al., 2019; McCaskie et al., 2019; Rago et al., 2020; Tierney et al., 2021).

A systematic review of contextual factors on match running in rugby league (Dalton-Barron et al., 2020) reported that only seven of fifteen studies using the Global Navigational Satellite System (GNSS) to quantify movement patterns in athletes during match-play accounted for repeated measures within athletes (i.e., dependent observations). Of these seven, two studies eliminated data to run their ANOVA, while another study did not account for the athletes' natural variation in performance, leaving only four studies utilising their full data set. Clearly, the characteristics of routinely collected health, wellbeing, and performance data sets in the high-performance sport environment can involve the manipulation of data, and depending on its severity, can undermine the confidence of analysis.

Obvious to the statistician and data analyst is that mixed models are likely a more appropriate statistical methodology for routinely collected health, wellbeing, and performance data such as the data in the present study. Indeed, within statistics literature, McElreath argues that mixed models deserve to be the default form of regression and that experiments that do not use a mixed model should justify not using this approach (McElreath, 2018). In his argument, he lists four reasons to use mixed models: i. To adjust estimates for repeat sampling, ii. To adjust estimates for imbalance in sampling, iii. To model variation among individuals or other groups, and iv. To avoid averaging, as some researchers pre-average data before running analyses (McElreath, 2018). Therefore, it is reasonable to suggest that mixed models are the most appropriate statistical methodology to analyse longitudinal data sets often acquired by Sports Scientists. This aligns with previous guidance by Hopkins and colleagues (2009) in encouraging sports science researchers to utilise mixed models rather than a repeated-measures

ANOVA (Hopkins et al., 2009). While mixed models have become more popular within sport-science research (Bellinger et al., 2020; Newans et al., 2021; Quinn et al., 2020), the use of mixed models are often perceived as technical and requiring statistical expertise. Thus, the aim of this study was to make Sports Scientists aware of the utility of mixed models when analysing longitudinal data sets. To illustrate this, we described and compared the movement patterns of female rugby league athletes across three levels of competition. We hypothesised that match intensity, as determined by the mean speed, would be higher in Origin matches due to the higher quality players compared to NRLW matches and shorter duration compared to Test matches.

2.3 Methods

2.3.1 Subjects

Over the 2018-2019 seasons, we were provided access to athlete positioning and timing data (i.e., movement patterns) transmitted by the GNSS for Australian female rugby league athletes including: club- (i.e., National Rugby League Women's; NRLW), state- (i.e., State of Origin; Origin), and international-level (i.e., Trans-Tasman Test; Test) competition. These data provided a unique opportunity to identify differences in movement patterns across three levels of rugby league competition previously observed in female athletes of other football codes (Andersson et al., 2010; Griffin et al., 2020). Importantly, the rugby-league data examined in the present study included repeated match data for multiple athletes which varied in number and level of competition. That is, some athletes playing NRLW played more matches than others, and some also played Origin and Test matches. This created a data set of repeated and dependent measures as well as an imbalance in sampling among athletes and competition level. Therefore, careful consideration of the statistical approach was paramount to derive meaningful conclusions.

Figure 2.1 shows the sample data set that contained 109 athletes in 12 NRLW matches during the 2018 and 2019 seasons, 49 athletes in the 2018 and 2019 Origin matches (2 matches), and 26 Australian 'Jillaroos' athletes in 2018 and 2019 Trans-Tasman Test matches (2 matches). Collectively, the present study included 115 unique athletes (age = 26.8 ± 5.4 yr; height = 1.68 ± 0.07 m; body mass = 76.7 ± 11.9 kg) and 472 match entries.

2.3.2 Design

The present study was a retrospective, observational, cohort analysis using GNSS-derived data routinely collected by each team's Sports Scientists. The National Rugby League coordinated the distribution of receivers, amalgamation of the data, and provision of the data sets to the authors.

2.3.3 Methodology

The movement patterns of athletes competing in NRLW matches were collected using 10 Hz Optimeye S5 receivers (Catapult Sports, Victoria, Australia), while the Origin and Test movement pattern data were collected using 10 Hz GPSports EVO receivers (Catapult Sports).

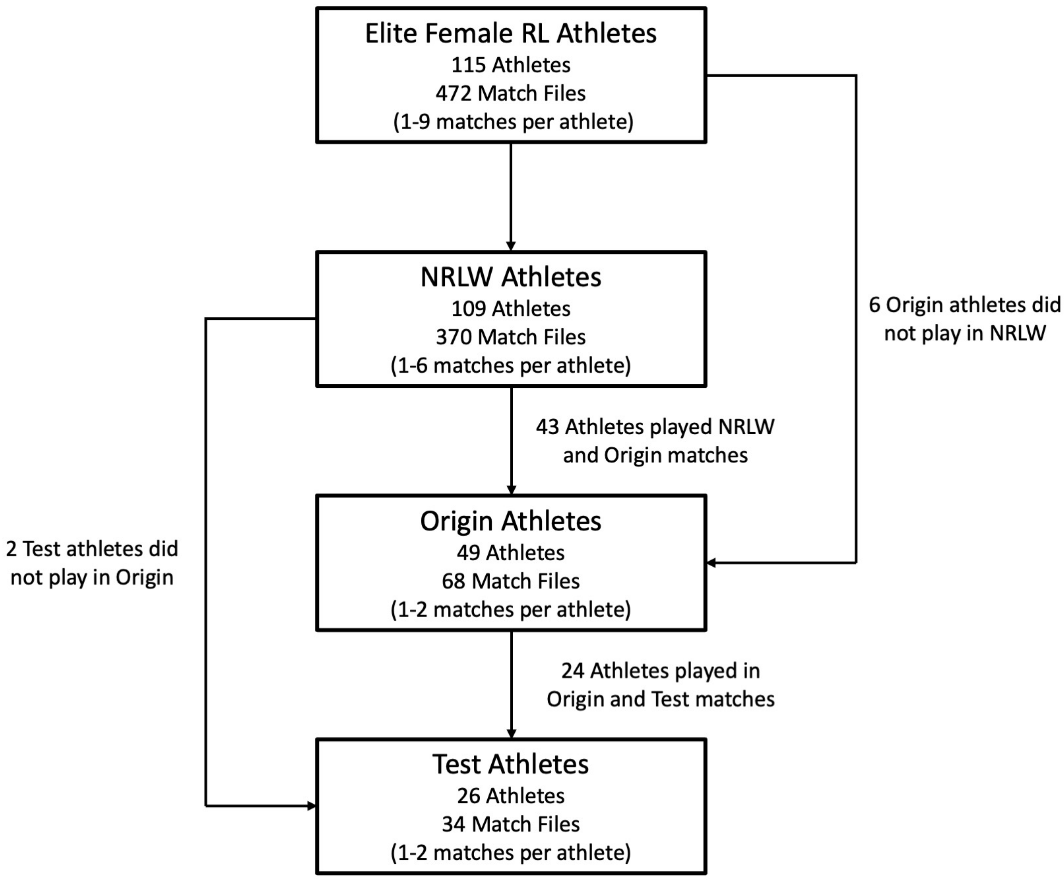

Figure 2.1: Flowchart of women's rugby league athletes playing at various competition levels indicating the levels of dependency within the data set. N.B. NRLW = National Rugby League Women's; Origin = State of Origin; Test = Trans-Tasman Test.

To measure match intensity, the mean speed (i.e., total distance divided by the time spent on the field) of each athlete was recorded. While it was not ideal that units by different manufacturers were used, only the total distance was required for this analysis which has been found to differ by 1.8% between these two devices (Thornton, Nelson, et al., 2019). As the mean speed of NRLW athletes has been found to decrease as function of the minutes played, a minimum threshold of 20 min of game time was applied to ensure the athlete spent an adequate time on the field. Additionally, this filter removed athletes playing for a short duration and therefore recording a high mean speed (i.e., metres/minute) if there were no stoppages in the short time they were on the field.

```
library(lme4)

library(effects)

library(dplyr)

library(sjPlot)

library(ggplot2)

library(performance)

library(see)

library(patchwork)

library(kableExtra)

options(scipen = 999)

NRLW <- read.csv("www/data/Study_1_mm_nrlw.csv") ## Read in data

NRLW$FieldTime <- chron::chron(times = as.character(NRLW$FieldTime)) ## Turn field time in
```

```
NRLW <- NRLW %>%
  filter(FieldTime > chron::chron(times = "00:20:00")) ## Filter only those playing more t
NRLW$MMIN <- NRLW$TotalDistance / NRLW$FieldTime / 60 / 24 ## Calculate running intensity
```

Consequently, of the 472 match files, 380 were deemed eligible for analysis. The Griffith University Human Ethics Committee approved this study (GU Ref No: 2019/359).

2.3.4 Statistical Analysis

To compare the differences in analysis, both a repeated-measures ANOVA and a mixed model were fitted to the present study data set. In both analyses, the effect of a player's position and level of competition on mean speed was assessed. For the ANOVA to be performed, we restructured the data to achieve a 'complete-case analysis' (Nakai & Ke, 2011). The mixed models were built using the *lme4* (Bates et al., 2015) package in R version 4.0.2 (R Core Team, 2019), while the *emmeans* (Lenth, 2020) and *sjPlot* (Lüdecke, 2020) packages were used for pairwise comparisons and model diagnostics respectively. The *see* (Lüdecke, Patil, et al., 2021) and *performance* (Lüdecke, Ben-Shachar, et al., 2021) packages were used to determine which model was the best performing model. The data set was arranged in 'long form' with each observation for each athlete on a new row.

```
full_model <- lmer(MMIN ~ Competition*Position + (1|Player)  + (1|Team) + (1|Match), data
```

After loading in the *lme4* package, the first model was built and the following components were outlined:

1. Dependent Variable; the metric we were interested in explaining. In this case, mean speed which as expressed as a continuous variable.

2. Fixed Effects; the variables of interest that could explain variation in the dependent variables. Here, we are looking at ‘level of competition’ and ‘playing position’. An interaction effect was also assessed, which would determine whether the difference in mean speed because of level of competition was uniform across all positions or whether the difference in mean speed because of level of competition is different for each position.

3. Random Effects; the variables that contain variation purely from the random sampling which could hide the influence of the fixed effects and would vary if the study were to be replicated (Yarkoni, 2019). Here, they are:

   i. ‘athlete’ as an individual athlete could have played up to 12 matches;

   ii. ‘team’ as the team that an athlete plays for (e.g., Broncos, Roosters, QLD Maroons, etc.) could have a confounding effect and is nested within the levels of competition;

   iii. ‘match’ as there is variation match-to-match in the intensity of play, perhaps due to weather, for example.

   These can be included as random intercepts and/or as random slopes, in this example, all three variables were included as random intercepts. By including these random effects, we can account for the variability associated with these effects to reveal the underlying effect of ‘level of competition’ and ‘position’ on the dependent variable.

4. Distribution Shape; in this case, the data adequately met the assumptions for a normal (Gaussian) mixed model. This example used a dependent variable (mean speed) that met the assumptions of a normal (Gaussian) distribution which meant that a simpler linear mixed model could be used in this case. However, this may not always be the case and, therefore the dependent variable may need to be expressed as a factor or percent, or log-transformed to reduce non-normality of the data (Hopkins et al., 2009). Alternatively, if there was frequency data (e.g., tackle count) or probability data, a Poisson distribution or binomial distribution respectively can be specified as the distribution family in the mixed model (using the *glmer* function rather than the simpler *lmer* function).

Once the full model (i.e., all fixed and random effects included) was established, the full model was compared to models without the fixed and random effects to assess the different model fits.

```
full_model1 <- lmer(MMIN ~ Competition*Position + (1|Team) + (1|Match), data = NRLW) ## Bu
full_model2 <- lmer(MMIN ~ Competition*Position + (1|Player) + (1|Match), data = NRLW)  ##
full_model3 <- lmer(MMIN ~ Competition*Position + (1|Player) + (1|Team), data = NRLW)  ##
reduced_model1 <- lmer(MMIN ~ Competition + Position + (1|Player)  + (1|Team) + (1|Match),
reduced_model2 <- lmer(MMIN ~ Competition + (1|Player)  + (1|Team) + (1|Match), data = NRL
reduced_model3 <- lmer(MMIN ~ Position + (1|Player)  + (1|Team) + (1|Match), data = NRLW)
model_comparison <- compare_performance(full_model,full_model1,full_model2,full_model3, re
```

When comparing models using the Akaike Information Criterion (AIC), it was determined that the full model was the preferred model as it displayed the lowest AIC. AIC was chosen as it includes a model complexity term (twice the number of model parameters) to penalise models containing variables that did not contribute to the model. The R^2 and Root-Mean-Square Error (RMSE) were also reported for the reader but were not used for model selection. As the models were nested, a likelihood ratio test could also have been used.

To meet the assumptions of a mixed model, a histogram of the residuals was generated to assess the normality of the residuals, Q-Q plots for each random effect were generated to assess the normality of the random effects, and the model's residuals were plotted against its fitted value to assess homoscedasticity.

```
diag <- plot_model(full_model, type = "diag") ## Assess model diagnostics
```

2.3.4.1 QQ Plot of residuals

```
plot_model(full_model, type = "diag")[[1]] +
  theme_minimal()+
  theme(plot.title = element_blank(),
        plot.subtitle = element_blank(),
        axis.title = element_text(size = 10, color = "black", face = "bold")) ## Generate
```

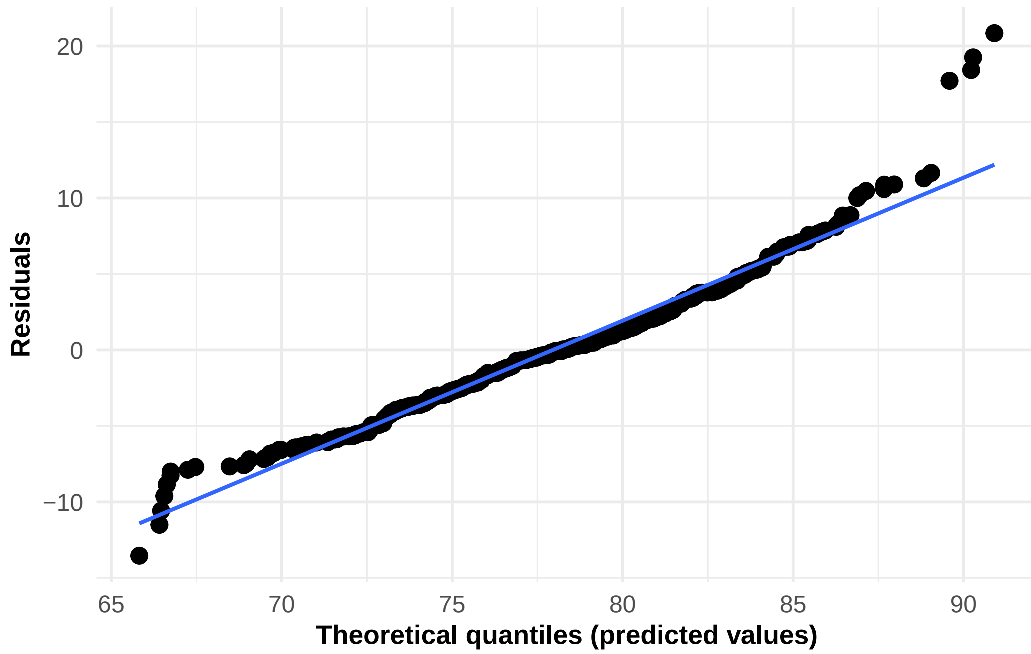

Figure 2.2: Q-Q Plot of residuals for full mixed model.

2.3.4.2 Normality Plot of residuals

```
plot_model(full_model, type = "diag")[[3]] +

  theme_minimal() +

  theme(plot.title = element_blank(),

        plot.subtitle = element_blank(),

        axis.title = element_text(size = 10, color = "black", face = "bold")) ## Generate
```

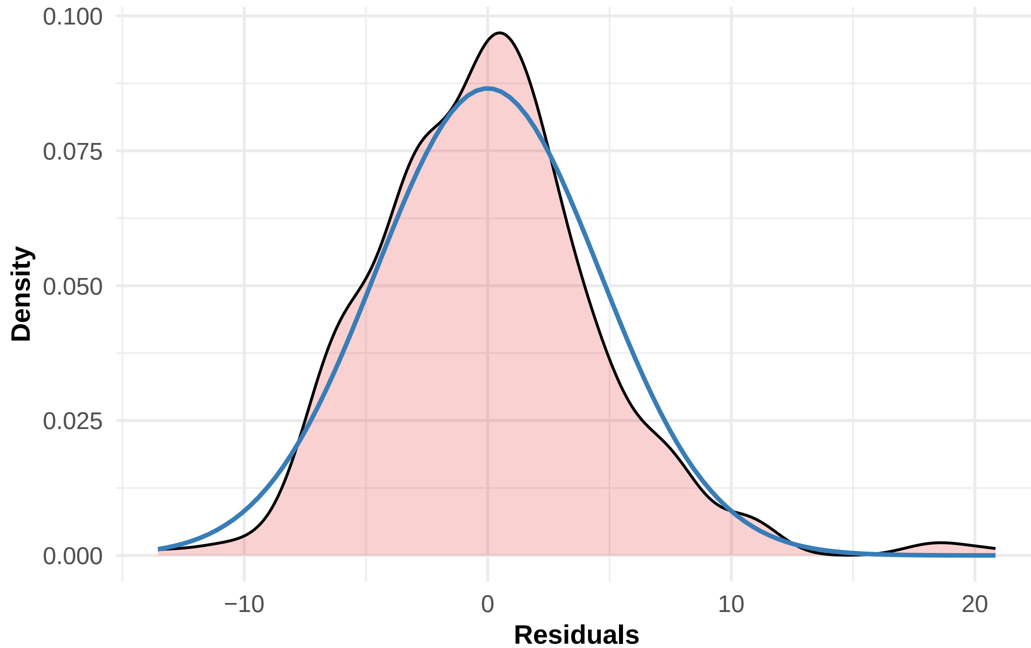

Figure 2.3: Histogram of residuals for full mixed model.

2.3.4.3 Homoscedasticity Plot of residuals

```
plot_model(full_model, type = "diag")[[4]] +
  theme_minimal() +
  theme(plot.title = element_blank(),
        plot.subtitle = element_blank(),
        axis.title = element_text(size = 10, color = "black", face = "bold")) ## Generate
```

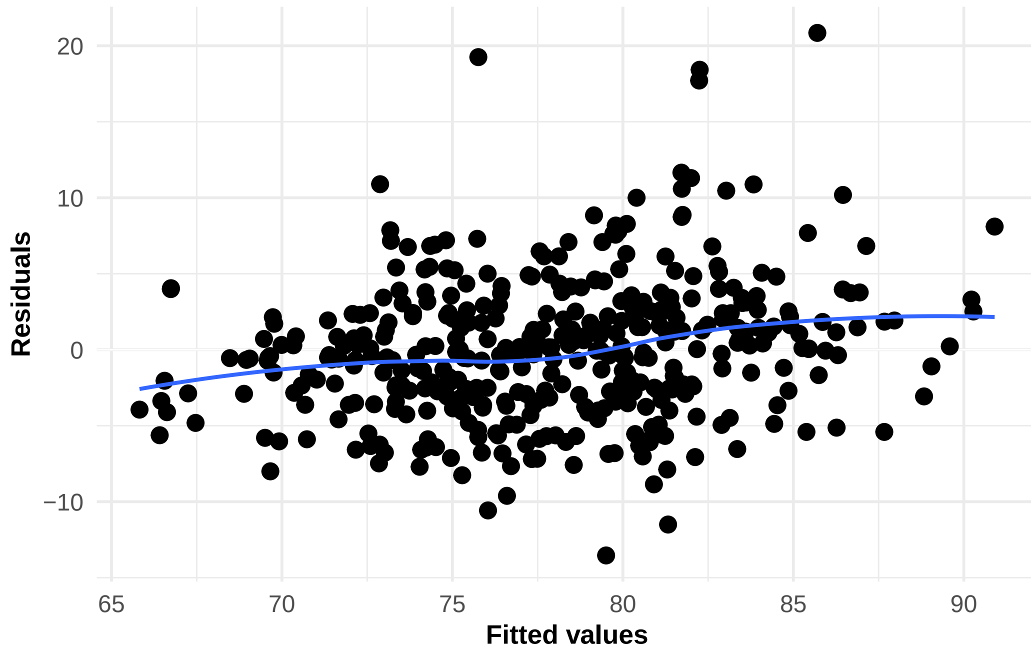

Figure 2.4: Homoscedasticity of residuals for full mixed model.

None of these assumptions were violated so, therefore, it was deemed appropriate to proceed with the analysis.

2.4 Results

In order to meet the assumptions of a repeated-measures ANOVA, we restructured the data to achieve a 'complete-case analysis' (Nakai & Ke, 2011), which eliminated any athlete that did not feature in all three levels of competition. This initial filter eliminated 48.7% of the available data. Of the remaining data, each athlete varied from one to seven observations within each level of competition which continued to threaten the validity of the analysis as multiple observations of one athlete could bias the results. Secondly, it would not be suitable to answer the hypothesis as eliminating any athlete that did not play in the Test matches would result in the average NRLW mean speed being compiled of only athletes that had played in both Origin and Test matches, rather than the collective NRLW cohort. Alternatively, if we used the mean for each athlete in each level of competition, this reduced the number of observations to 157 (i.e., eliminated 58.7% of the available data). However, when averaging each athlete's values, it would not account for the fact that there is less uncertainty in the mean speed of an athlete that had seven observations compared to an athlete that only has one observation. Therefore, it was determined that it was not appropriate to run an ANOVA across this data set. Due to the use of ANOVA being deemed inappropriate to adequately analyse these data, only the results of the mixed model are presented here.

2.4.1 Model Comparisons

Seven mixed models were assessed: the full model (i.e., containing all fixed and random effects), three models with each model removing one random effect, and three models removing the interaction effect, the competition level, and position respectively. The full model displayed the lowest AIC, the equal-highest R^2, and the lowest RMSE and, therefore, was chosen as the preferred model. The results are displayed in Table 2.1. We can write the selected model in short form as:

$$MeanSpeed_{pmti} = \beta_0 + \beta_1 \times Competition_{ptmi} + \beta_2 \times Position_{ptmi} + \beta_3 \times Competition_{ptmi} \times Position_{ptmi} + {\gamma_p}^{player} -$$

where $\gamma_p{}^{player} \sim N(0, \sigma_p{}^2)$, $\gamma_m{}^{match} \sim N(0, \sigma_m{}^2)$, and $\gamma_t{}^{team} \sim N(0, \sigma_t{}^2)$ are the random intercepts for player, match, and team respectively, and $\varepsilon_{ptmi} \sim N(0, \sigma^2)$ is the residual term. All these terms are normally distributed with zero mean and variance given by the second term in the parentheses. The random effects are crossed (i.e., not nested) since, for example, a player can play for multiple teams. The model is parameterised so that the intercept term β_0 corresponds to Test level of competition and the Adjustable position. Since there are three competitions and four positions in the data, the number of elements in β_1, β_2 and β_3 is two, three and six, respectively, and are to be interpreted relative to the baseline categories in β_0. Correspondingly, Competition$_{ptmi}$ ε {"Origin", "NRL"} and Position$_{ptmi}$ ε {"Backs", "Forwards","Interchange"}. Here the subscripts p, m, t and i refer to the pth player, mth match, tth team, and the ith observation for each combination of player, match, and team.

```
model_comparison %>%
  select(Name, AIC, R2 = R2_conditional, RMSE) %>%
  mutate(AIC = round(AIC, 1),
         R2 = round(R2, 2),
         RMSE = round(RMSE, 2)) ## Generate table comparing models with and without random
```

Table 2.1: Model comparison of mixed models with and without random and fixed effects.

Model	AIC	R^2	RMSE
Full Model	2379.4	0.65	3.96
Model without Player random effect	2425.9	0.44	5.47

Model	AIC	R^2	RMSE
Model without Match random effect	2481.2	0.38	5.38
Model without Team random effect	2384.5	0.62	4.00
Model without Interaction fixed effect	2398.1	0.65	4.01
Model without Competition fixed effect	2406.2	0.64	4.01
Model without Position fixed effect	2407.0	0.64	4.03

N.B. AIC = Akaike's Information Criterion, RMSE = Root Mean Squared Error

2.4.2 Random Effects

When considering the model comparison, the full model with player, team, and match random effects was preferred over the reduced models that removed each of the random effects, as evidenced by the full model having the lowest AIC value. Similarly, when assessing the variance explained by each of the random effects, it was determined that all three random effects should remain in the model. The estimated random effects variances are displayed in Table 2.2, together with the estimated residual variance. The conditional means for each level of random effect are displayed in Figure 2.5, Figure 2.6, and Figure 2.7.

```
VarCorr(full_model) %>%
  as.data.frame() %>%
  mutate(`Variance ± SE` = paste0(round(vcov, 1), " ± ", round(sdcor, 1))) %>%
```

```
  select(`Source of Variance` = grp,
         `Variance ± SE`)  ## Generate table of random effect variances and residual
```

Table 2.2: Estimated variances for random effects and the model residual variance.

Source of Variance	Variance ± SE
Player	13.66 ± 3.70
Match ID	14.61 ± 3.82
Team	4.31 ± 2.08
Residual	20.61 ± 4.54

The player conditional means differed between -6.9 to 5.9 m · min^{-1} from the marginal mean, the team conditional means differed between -2.7 to 2.1 m · min^{-1}, and the match conditional means differed between -8.7 to 5.8 m · min^{-1} from the marginal mean.

2.4.2.1 Player

```
ggplot(diag[[2]]$Player$data, aes(x = ID, y = y)) +
  geom_errorbar(aes(ymin = y - ci, ymax = y + ci), color = "#333333") +
  geom_point(color = "black") +
  scale_x_discrete(labels = NULL) +
  coord_cartesian(ylim = c(-11, 11)) +
  theme_minimal() +
```

```
  labs(x = "Player",
       y = "Random Effect Intercept") +
  theme(panel.grid.major.x = element_blank(),
        axis.title = element_text(size = 10, color = "black", face = "bold")) ## Generate
```

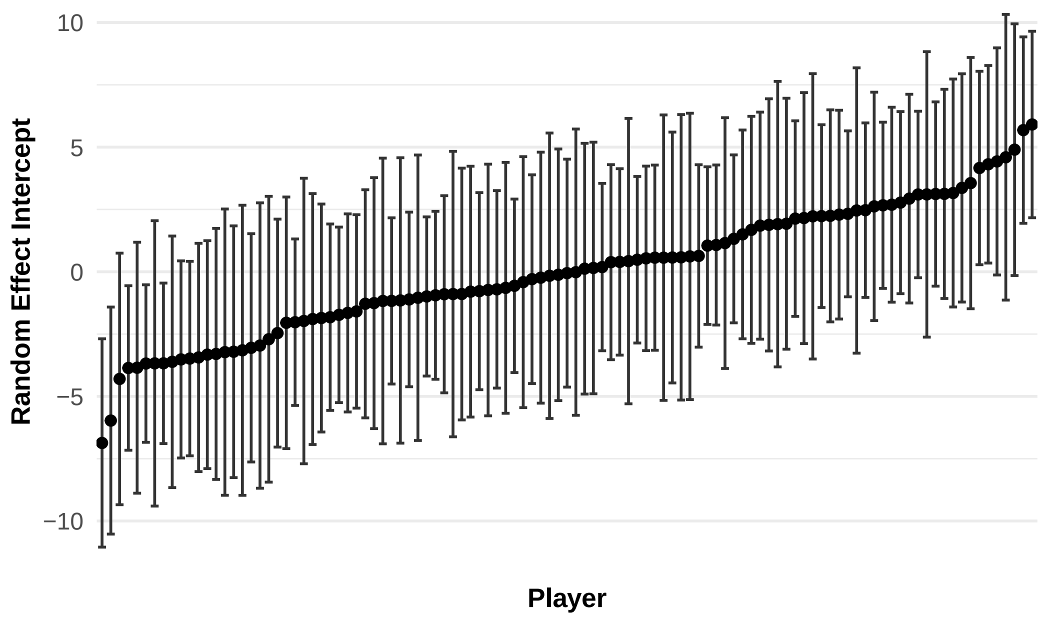

Figure 2.5: Random effects for each level of player.

2.4.2.2 Team

```
ggplot(diag[[2]]$Team$data, aes(x = ID, y = y)) +
  geom_errorbar(aes(ymin = y - ci, ymax = y + ci), color = "#333333") +
  geom_point() +
  theme_minimal() +
  scale_x_discrete(labels = NULL) +
```

```
  coord_cartesian(ylim = c(-11, 11)) +
  labs(x = "Team",
       y = "Random Effect Intercept") +
  theme(panel.grid.major.x = element_blank(),
        axis.title = element_text(size = 10, color = "black", face = "bold")) ## Generate
```

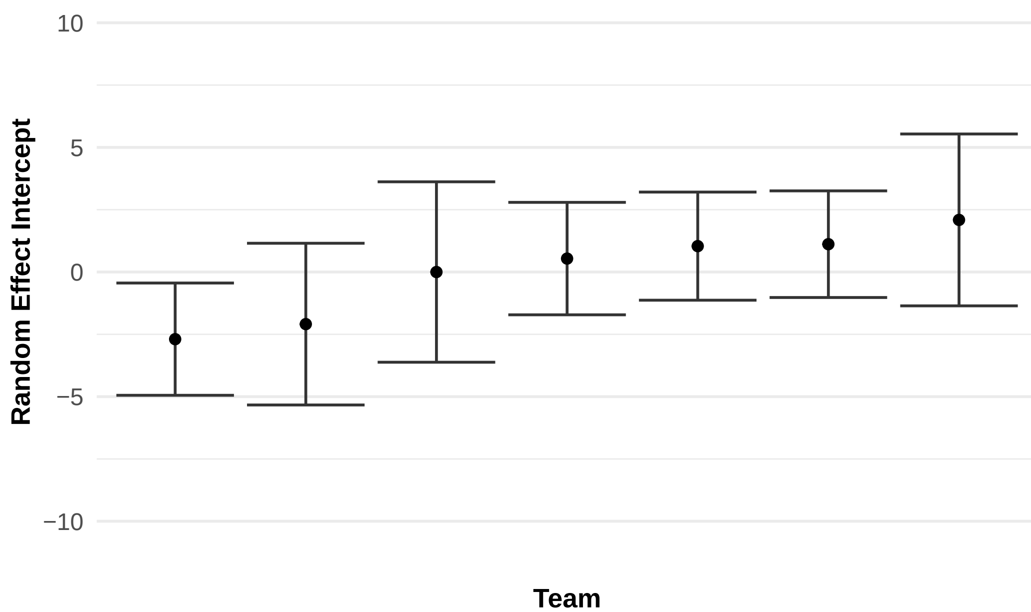

Figure 2.6: Random effects for each level of team.

2.4.2.3 Match

```
ggplot(diag[[2]]$Match$data, aes(x = ID, y = y)) +
  geom_errorbar(aes(ymin = y - ci, ymax = y + ci), color = "#333333") +
  geom_point() +
  theme_minimal() +
```

```
    scale_x_discrete(labels = NULL) +
    coord_cartesian(ylim = c(-11, 11)) +
    labs(x = "Match",
         y = "Random Effect Intercept") +
    theme(panel.grid.major.x = element_blank(),
          axis.title = element_text(size = 10, color = "black", face = "bold")) ## Generate
```

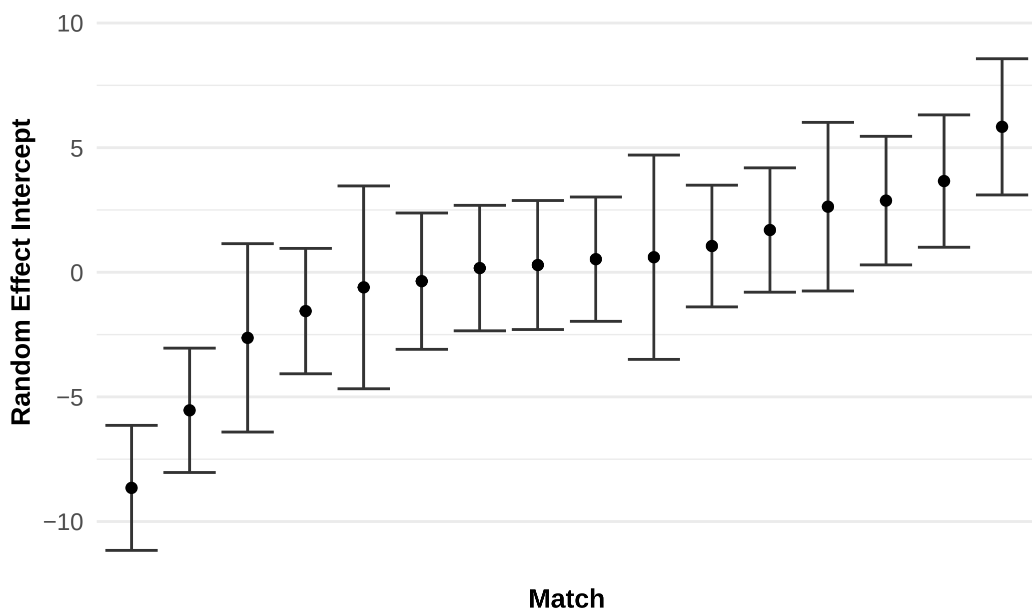

Figure 2.7: Random effects for each level of match.

2.4.3 Fixed Effects

Table 2.3 displays the additive interaction effects to calculate each permutation within the data set. The intercept (i.e., 74.5 m · min^{-1}) represents the mean speed for a Test adjustable player on average. From here, the coefficients can be added depending on what level of competition and what position was being played. For example, for the mean speed of an Origin forward athlete, you would start with the intercept of 75.0 $m \cdot min^{-1}$. From here, you will add 8.7 $m \cdot min^{-1}$ for an Origin-level athlete, then add another 1.9 $m \cdot min^{-1}$ for a forward, and then subtract 2.2 $m \cdot min^{-1}$ for the interaction effect of an Origin-level forward, resulting in an estimated mean of 83.4 $m \cdot min^{-1}$.

```
mm_fixed_effect_summary <- summary(full_model)$coef %>%
  data.frame() %>%
  mutate(`Fixed Effect` = rownames(.)) ## Generate fixed effect coefficients

mm_confint <- confint(full_model) %>%
  data.frame() %>%
  mutate(`Fixed Effect` = rownames(.)) ## Generate fixed effect confidence intervals

mm_fixed_effect_summary %>%
  left_join(mm_confint) %>%
  mutate(`95% CI` = paste0(round(X2.5.., 1), " to ", round(X97.5.., 1)),
         Estimate = round(Estimate, 1)) %>%
  select(`Fixed Effect`, Estimate, `95% CI`)  ## Compile table of fixed effect coefficient
```

Table 2.3: Estimated means for fixed effects and interactions.

Fixed Effect	Estimate (95% CI)
Intercept	74.5 (67.0 – 82.0)
Competition - Test	-
Competition - Origin	8.7 (-1.1 – 18.5)
Competition - NRLW	5.1 (-2.9 – 13.1)
Position - Adjustables	-
Position - Backs	-2.6 (-8.2 – 3.0)
Position - Forwards	1.9 (-3.5 – 7.1)
Position - Interchange	-3.9 (-9.2 – 1.4)
Interaction: NRLW-Backs	-1.6 (-7.0 – 3.9)
Interaction: Origin-Backs	1.1 (-5.3 – 7.7)
Interaction: NRLW-Forwards	-3.9 (-9.2 – 1.3)
Interaction: Origin-Forwards	-2.2 (-8.5 – 4.1)
Interaction: NRLW-Interchange	2.2 (-3.2 – 7.6)
Interaction: Origin-Interchange	-0.9 (-7.7 – 6.0)

Figure 2.8 presents the estimated marginal means for mean speed recorded during NRLW, Origin, and Test matches. Origin matches recorded, on average the highest mean speed, followed by NRLW matches, and then Test matches. When considering position, in both NRLW and Origin matches, adjustables recorded the highest mean speed; however, forwards

recorded the highest mean speed in Test matches. Meanwhile, backs recorded the lowest mean speed in NRLW matches, with interchange recording the lowest mean speed in both Origin and Test matches. As the changes in mean speed across competitions were not uniform across all competition, this confirmed the presence of an interaction effect.

```
Position_Estimates <- effect("Competition*Position", full_model) %>%
  data.frame() %>%
  select(Competition, Position, fit, lower, upper) %>%
  mutate_at(vars(fit, lower, upper), round, 1) ## Create parameter estimates based on posi

ggplot(Position_Estimates, aes(x = 1, y = fit, shape = Position)) +
  geom_jitter(position = position_dodge(width = 0.5), size = 3) +
  geom_errorbar(aes(ymin = lower,
                    ymax = upper),
                width = 0.5,
                alpha = .75,
                position = position_dodge(width = 0.5)
  ) +
  theme_bw() +
  scale_y_continuous(breaks = seq(60, 100, 10),
                     limits = c(60, 100)) +
  facet_wrap( ~ Competition, ncol = 3) +
```

```r
labs(y = bquote(bold('Mean Speed (m·min ' ^ -1 * ')'))) +

theme(

  panel.grid.major.x = element_blank(),

  panel.grid.minor.x = element_blank(),

  panel.grid.major.y = element_line(colour = "grey50",

                                    linewidth = 0.2),

  panel.grid.minor.y = element_blank(),

  axis.ticks = element_blank(),

  axis.title.x = element_blank(),

  axis.text.x = element_blank(),

  axis.title.y = element_text(size = 10, color = "black"),

  axis.text.y = element_text(colour = "black", size = 12),

  strip.text.x = element_text(colour = "black", size = 12, face = "bold"),

  legend.position = "bottom",

  legend.title = element_text(colour = "black", size = 12, face = "bold"),

  legend.text = element_text(colour = "black", size = 10) ## Generate marginal mean plot

)
```

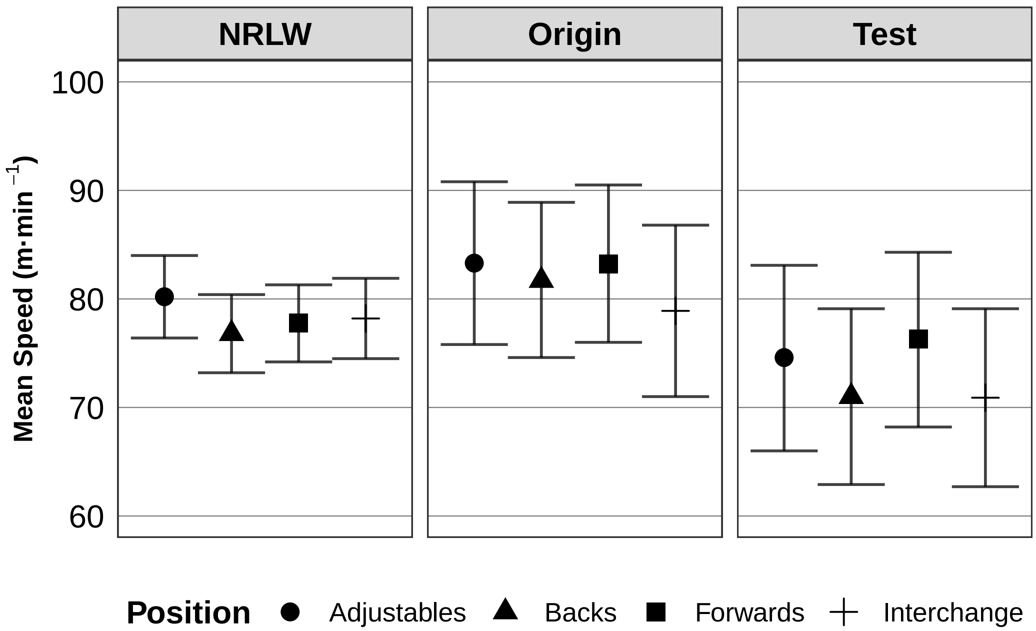

Figure 2.8: Mean speed of women's rugby league athletes by competition level and position. N.B. NRLW = National Rugby League Women's; Origin = State of Origin; Test = Trans-Tasman Test.

2.5 Discussion

The present study displayed why mixed models should be the more heavily-adopted when analysing sports science field-based data sets with repeated measures, like the rugby league women's data set used in this study. Mixed models were able to account for multiple observations of the same individuals in the data set, players changing positions between matches, and account for inter-player, inter-match, and inter-team dependencies to extract the true effects of position and level of competition on mean speed and the appropriate quantification of the uncertainty of these estimates. One strength of the mixed model in analysing data was its ability to account for the differing number of observations of athletes. In the data set, athletes ranged between one and nine matches played with 24 athletes playing only one match, while five athletes played nine matches each. This is not ideal when applying repeated-measures ANOVA as either some athletes or some time points (i.e., games) would need to be excluded from the analysis. While removing an athlete if they do not have an observation in every time point or removing a time point if many athletes are missing is statistically appropriate if the missing data points are proved to be missing completely at random (i.e., no bias in missing data) (Borg et al., 2021), this method can cause unnecessary deletion of large amounts of data. If there are many participants and many time points, it can sometimes eliminate so much data that no analysis can be performed on the remaining data set; for example, 48.7% of the data collected for the present study would have been omitted when filtering out players that did not participate in Origin or Test matches. However, mixed models can attribute a level of uncertainty to each athlete dependent on their sample size and, therefore, the model can more accurately quantify the true mean speed of a player with nine observations compared to a single observation of another athlete. In doing so, mixed models create flexibility in the data sets that can be utilised that more closely align to data sets seen in sports due to injuries, squad selection, and access to athletes on a given day. The mixed model provided an analysis that could retain all available data points and did not require the elimination of data points to retain a 'complete case analysis' data set (Nakai & Ke, 2011), nor did it require data imputation to estimate missing data (Borg et al., 2021).

Another strength of the mixed model was its ability to use the dependency within the data set to increase the power of the statistics, rather than detract like in general linear model applications. For example, in the present study, since there were only two matches at an international level and only two matches at Origin level, these sample sizes would be underpowered when using a linear model; however, when merged with the twelve NRLW matches, the model can draw on the variance attributed to players and positions to estimate the effects of level of competition more robustly. Similarly, two players played in three of the different positional groups which would typically require the athlete to have matches in which they were not in their 'primary' position to run parametric general linear models. However, this information is actually very useful as it enables the model to observe the same athlete, in the same team, in the same level of competition but in a different position which provides more information than if the two positional groups were completely independent of each other. That is, by enabling players to be their own control group, they can more accurately interrogate the between-position differences in mean speed. As a result, mixed models can more accurately and more robustly determine the variation attributable to athletes, time points, and conditions (e.g., position) much more than any general linear model. Additionally, by using a mixed model, it was evident that the random effect for player is larger than the fixed effect of position, which would not be able to be established when using a repeated-measures ANOVA. Therefore, this reinforces that individualised training for an athlete, rather than the position, is more important from a training prescription perspective.

When considering the findings of this study, the reduced overall mean speed recorded for the Test (80 min), compared with the Origin (60 min) and NRLW (60 min) matches in the present study could be explained by the longer Test-match duration. We have previously reported that the mean speed when travelling >12 km $\cdot$ h^{-1} of athletes recorded during international matches declines by 40% within the first half of the match (Quinn et al., 2020). We also previously demonstrated that there were no significant differences in the relative distances covered in any of the speed zones, as well as the overall mean speed when comparing the first and second half of NRLW matches (Newans et al., 2021). These results contrast with those seen in other codes, with significantly increased total distance and high-speed running in international football compared to domestic football (Griffin et al., 2020). On closer examination of mean speed in the present study, it was evident that the position contributed to the model with forwards recording the highest intensity and the interchange athletes delivering the lowest intensity. This could be due to a disparity in playing ability between the starting forwards and the interchange replacements. We previously established that the mean speed of interchange athletes significantly reduced as their playing duration increased (Newans et al., 2021), which could explain why at the international level, when interchange athletes are required to play increased minutes due to the longer format, they cannot sustain the same intensity as the starting forwards.

2.6 Conclusion

The requirement to account for repeated measures and imbalanced data is pertinent in longitudinal sports science data sets. As previous studies have demonstrated a lack of statistical literacy to correctly understand dependency within data sets and the consequent violations of parametric statistical assumptions, the present study provides a more thorough account of the process, the associated R script, and the resultant interpretations to inform Sports Scientists on mixed models. It is anticipated that the present study will empower Sports Scientists to assess the various dependencies more critically within their data sets.

2.7 Practical Applications

- Mixed models should be a more-heavily adopted statistical method for analysing sports science data sets with repeated measures as they are more flexible than repeated-measures ANOVA
- Mixed models can accommodate differing frequency of observations of athletes and players swapping positions in between matches
- If Test matches continue to be 80 min in duration, the physical and physiological capacity of athletes should be improved to maintain running intensity at the international level
- NRLW matches should be increased to 70 min in 2022 to gradually bridge the gap between domestic- and international-level competition

3 Pareto Frontiers for Multivariate Sports Performance

3.1 Abstract

Athletes often require a mix of physical, physiological, psychological, and skill-based attributes that can be conflicting when competing at the highest level within their sport. When considering multiple variables in tandem, Pareto frontiers is a technique that can identify the observations that possess an optimal balance of the desired attributes, especially when these attributes are negatively correlated. This study presents Pareto frontiers as a tool to identify athletes who possess an optimal ranking when considering multiple metrics simultaneously. This study explores the trade-off relationship between batting average and strike rate as well as bowling strike rate, economy, and average in Twenty 20 cricket. Eight hundred ninety-one matches of Twenty 20 cricket from the men's (MBBL) and women's (WBBL) Australian Big Bash Leagues were compiled to determine the best batting and bowling performances, both within a single innings and across each player's Big Bash career. Pareto frontiers identified 12

and seven optimal batting innings performances in the MBBL and WBBL respectively, with nine and six optimal batting careers respectively. Pareto frontiers also identified three optimal bowling innings in both the MBBL and WBBL and five and six optimal bowling careers in MBBL and WBBL, respectively. Each frontier identified players that were not the highest ranked athlete in any metric when analyzed univariately. Pareto frontiers can be used when assessing talent across multiple metrics, especially when these metrics may be conflicting or uncorrelated. Using Pareto frontiers can identify athletes that may not have the highest ranking on a given metric but have an optimal balance across multiple metrics that are associated with success in a given sport.

The following chapter is a copy of the published manuscript:

Newans, T., Bellinger, P., & Minahan, C. The balancing act: Identifying multivariate sports performance using Pareto frontiers. Frontiers in Sports and Active Living. 4:918946.

As co-author of the paper "The balancing act: Identifying multivariate sports performance using Pareto frontiers", I confirm that Timothy Newans has made the following contributions:

- Study concept and design
- Data collection
- Data analysis and interpretation
- Manuscript preparation

Name: Clare Minahan

Date: 29/03/2023

3.2 Introduction

Many sports require athletes to possess well-developed physiological or mechanical characteristics (e.g., endurance, maximal sprint speed) and/or skill-related attributes (i.e., shot speed and shot accuracy) that are often opposing and/or are not associated. These disparate attributes appear to be more obvious between athletes in team sports when compared to specific disciplines in individual sports that require highly specific attributes for success (e.g., track sprint cycling (Lievens et al., 2021) or marathon running (A. Jones et al., 2021)). Furthermore, it is not unusual for team sports to include both specialists, such as athletes who possess superior maximal sprint speed and athletes who possess superior endurance capacity, and individual athletes who have a wide range of physiological/mechanical and skill-related attributes. For instance, in stick or racquet sports such as golf and tennis, each athlete is required to balance the accuracy and speed of their strokes (Maquirriain et al., 2016; Wells et al., 2009); while in most team sports, an athlete should develop a balance of speed and endurance (Stølen et al., 2005). Moreover, in sports like cricket, an athlete is required to balance their batting/bowling average with their batting/bowling strike rate or economy (Barr & Kantor, 2004; Patel et al., 2017). In each of these examples, a case could be made for which is the preferred attribute; however, the preferred attribute may differ given the other athletes within a particular team or given a particular situation within a single match. A cohesive team may require a squad of players that differ in their balance of attributes and, therefore, it is apparent that performance analysis requires a multi-faceted approach when selecting prospective players.

While talent identification processes have been extensively reported (Dodd & Newans, 2018; Falk et al., 2004; K. Johnston et al., 2018; Pion et al., 2015; Pyne et al., 2005; Till et al., 2016), standards and benchmarks are typically reported univariately; that is, each attribute is assessed in isolation. For example, players could be standardized within each attribute (i.e., z-scores) to determine where the athlete sits with respect to the rest of the athletic population

(Turner et al., 2019). However, this method has a flaw, in that some attributes are negatively correlated, as well as physiological/mechanical characteristics such as maximal sprint speed and endurance capacity (Sánchez-García et al., 2018). Therefore, if an athlete excels in one attribute, it is likely that this would come at the expense of a conflicting attribute. By assessing talent identification univariately, athletes are identified when they have specialist skills (e.g., strongest, fastest, fittest, leanest etc.) (Minahan et al., 2021). Although a given sport often requires a balance of these attributes, it is reasonable to suggest that talent identification processes should assess talent multivariately, that is, multiple variables in tandem, rather than univariately.

The process of optimizing the balance of multiple attributes is termed "multi-objective optimization". This technique is becoming increasingly of interest with recent developments in machine learning algorithms; however, their origins are quite simple mathematically in that they aim to create the perfect balance of the attributes of interest. If a data point was defined as: $\vec{x_1} \,\varepsilon\, X$, it is, therefore, better than another data point defined by: $\vec{x_2} \,\varepsilon\, X$ if $f_i(\vec{x_1}) \leq f_i(\vec{x_2})$ for all metrics $i \;\varepsilon\; \{1,\ 2,\ ...,\ k\}$ and $f_i(\vec{x_1}) < f_i(\vec{x_2})$for at least one metric $j \;\varepsilon\; \{1,\ 2,\ ...,\ k\}$. Once these conditions have been met, the remaining points are deemed Pareto-optimal and form what is called the Pareto frontier. There are two key strengths to using Pareto frontiers. Firstly, when balancing multiple attributes, Pareto-optimal observations can be identified with just few lines of computer code. Secondly, when visualizing a limited number of attributes (i.e., three or less), the Pareto frontier is intuitive and can be clearly identified, assisting in the translation and interpretation of the results to coaches and other support staff. This has been used in other fields such as designing aircrafts with maximum aerodynamic efficiency, maximum range, and minimum weight (Mastroddi & Gemma, 2013). However, there is very limited use of Pareto frontiers within sport (Pérez-Toledano et al., 2019). By using Pareto frontiers, the optimal balance of all these attributes can be identified rather than guessing through siloed univariate analyses.

To illustrate the concept of Pareto frontiers, the present study used batting and bowling data in Twenty 20 (T20) cricket. Like other forms of cricket (i.e., Test and one-day matches), T20 cricket requires players to score as many runs as possible within the allotted 20 overs without being dismissed (i.e., being bowled, caught out, run out etc.). In addition, players of T20 cricket also need to score runs in as few deliveries as possible. Therefore, it is difficult to determine whether 80 runs "off" (i.e., from) 60 balls or 40 runs off 20 balls is of more benefit to a team as their differing risk profiles contribute differently to the formation of the team (Bukiet & Ovens, 2006). For example, early on in an innings the risk-return of attempting to hit six runs off a ball is significantly different than in the final over of an innings. Similarly, a bowler needs to balance taking wickets while also conceding as few runs as possible. For instance, when bowling four overs, it is again difficult to determine whether taking three wickets for 50 runs is of more worth than taking no wickets but only conceding eight runs as the three wickets may not have been worth conceding 50 runs. Therefore, when assessing the quality of players, it is necessary to utilize tools that can analyse these data sets without favoring one metric over another. The concept of Pareto frontiers is one such tool and, therefore, the present study aimed to introduce Pareto frontiers to the sports science community and illustrate how they can identify players with the optimal balance of attributes that can be obfuscated when performing univariate analysis across each metric.

3.3 Methods

The present study comprised all 489 matches of the first 11 editions of the Men's Big Bash League (MBBL) and all 402 matches of the first seven editions of the Women's Big Bash League (WBBL), Australia's domestic T20 cricket competition. The MBBL data set contained 423 batters and 313 bowlers, while the WBBL data set contained 214 batters and 159 bowlers. All scorecards were freely available online. Collectively, there were 13,764 individual batting innings with observations ranging from 1 to 113 innings per batter, while there were 10,796 individual bowling innings with observations ranging from 1 to 106 innings per bowler.

```r
library(tidyverse)
library(rPref)
library(patchwork)
library(scatterplot3d)
options(scipen = 999)
mbblbat <- read_csv('www/data/Study_2_pareto_mbblbat.csv') # Men's batting scorecards
mbblbowl <- read_csv('www/data/Study_2_pareto_mbblbowl.csv') # Men's bowling scorecards
wbblbat <- read_csv('www/data/Study_2_pareto_wbblbat.csv') # Women's batting scorecards
wbblbowl <- read_csv('www/data/Study_2_pareto_wbblbowl.csv') # Women's bowling scorecards
```

To summarize the data, two summary statistics were generated for batting and three summary statistics were generated for bowling. The summary statistics were as follows:

- Batting Average: runs scored divided by frequency of dismissal
- Batting Strike Rate: runs scored divided by balls faced multiplied by 100
- Bowling Average: runs conceded divided by wickets taken
- Bowling Strike Rate: balls bowled divided by wickets taken
- Bowling Economy: runs conceded divided by overs (i.e., 6 balls) bowled

```
dismissals_men <- mbblbat %>%
  group_by(id) %>%
  filter(Dismissed == T) %>%
  summarise(Dismissals = n()) ## Calculate number of dismissals

notouts_men <- mbblbat %>%
  group_by(id) %>%
  filter(Dismissed == F) %>%
  summarise(NotOuts = n()) ## Calculate number of not outs

sumBat_men <- mbblbat %>%
  group_by(id, Batter, LastName) %>%
```

```
  summarise(
    TotalRuns = sum(R),
    TotalBalls = sum(B),
    Innings = n()
  ) %>%
  ungroup() %>%
  left_join(dismissals_men) %>%
  left_join(notouts_men) %>%
  mutate(
    Dismissals = case_when(is.na(Dismissals) ~ as.integer(0),
                           T ~ Dismissals),
    NotOuts = case_when(is.na(NotOuts) ~ as.integer(0),
                        T ~ NotOuts),
    Average = TotalRuns / Dismissals,
    StrikeRate = TotalRuns / TotalBalls * 100) ## Calculate career batting average and str

filtBat_men <- sumBat_men %>%
  filter(Innings >= 15) ## Filter only those with 15 or more batting innings

sumBowl_men <- mbblbowl %>%
  group_by(id, Bowling, LastName) %>%
```

```
  summarise(
    Innings = n(),
    Balls = sum(Balls),
    Wickets = sum(W),
    Runs = sum(R)
  ) %>%
  mutate(
    Average = Runs / Wickets,
    Economy = Runs / Balls * 6,
    StrikeRate = Balls / Wickets
  ) ## Calculate bowling average, economy, and strike rate

filtBowl_men <- sumBowl_men %>%
  filter(Balls >= 200) %>%
  ungroup() ## Filter only those with 200 or more bowling deliveries

dismissals_women <- wbblbat %>%
  group_by(id) %>%
  filter(Dismissed == T) %>%
  summarise(Dismissals = n()) ## Calculate number of dismissals
```

```
notouts_women <- wbblbat %>%
  group_by(id) %>%
  filter(Dismissed ==  F) %>%
  summarise(NotOuts = n()) ## Calculate number of not outs

sumBat_women <- wbblbat %>%
  group_by(id, Batter, LastName) %>%
  summarise(
    TotalRuns = sum(R),
    TotalBalls = sum(B),
    Innings = n()
  ) %>%
  ungroup() %>%
  left_join(dismissals_women) %>%
  left_join(notouts_women) %>%
  mutate(
    Dismissals = case_when(is.na(Dismissals) ~ as.integer(0),
                           T ~ Dismissals),
    NotOuts = case_when(is.na(NotOuts) ~ as.integer(0),
                           T ~ NotOuts),
    Average = TotalRuns / Dismissals,
```

```
    StrikeRate = TotalRuns / TotalBalls * 100) ## Calculate career batting average and str

filtBat_women <- sumBat_women %>%
  filter(Innings >= 15) ## Filter only those with 15 or more batting innings

sumBowl_women <- wbblbowl %>%
  group_by(id, Bowling, LastName) %>%
  summarise(
    Innings = n(),
    Balls = sum(Balls),
    Wickets = sum(W),
    Runs = sum(R)
  ) %>%
  mutate(
    Average = Runs / Wickets,
    Economy = Runs / Balls * 6,
    StrikeRate = Balls / Wickets
  ) ## Calculate bowling average, economy, and strike rate

filtBowl_women <- sumBowl_women %>%
  filter(Balls >= 200) %>%
```

```
ungroup()  ## Filter only those with 200 or more bowling deliveries
```

To understand optimal batting performance, the number of runs scored as well as the rate at which these runs were scored are both of importance in Twenty20 cricket (Barr & Kantor, 2004). As it was expected that there would be a trade-off relationship between these variables, there would be multiple batters that possess an optimal balance of these attributes. Therefore, it was deemed appropriate that a Pareto frontier could be established to identify these batters. Similarly, to identify optimal bowling performance, the number of wickets, runs conceded, and rate of which the wickets and runs are recorded are all of importance (Patel et al., 2017). As the number of wickets taken can come at the expense of runs conceded, it was also expected that no bowler would be optimal in every attribute and therefore, a Pareto frontier would also be required to identify the bowlers that possess the optimal balance of these bowling attributes.

Consequently, four Pareto frontiers for each competition were established within the data set:

- *Pareto-optimal Batting Innings*

  This analysis outlined the highest runs scored within an innings at the highest strike rate.

  ```
  BatInnPareto_men <- psel(mbblbat %>% filter(R > 0), high(R) * high(SR), top_level = 9
  BatInnPareto_women <- psel(wbblbat %>% filter(R > 0), high(R) * high(SR), top_level =
  ```

- *Pareto-optimal Batting Career*

  This analysis outlined the highest batting average across a career at the highest strike rate. To provide a more accurate career report, batters required to have played a minimum of 15 innings which left 158 eligible male batters and 116 eligible female batters.

```
BatCarPareto_men <- psel(filtBat_men,high(Average) * high(StrikeRate), top_level = 99
  filter(Average > 20 | StrikeRate > 100) ## Identify Pareto frontier for men's BBL b
BatCarPareto_women <- psel(filtBat_women,high(Average) * high(StrikeRate), top_level
  filter(Average > 20 | StrikeRate > 100) ## Identify Pareto frontier for women's BBL
```

- *Pareto-optimal Bowling Innings*

  This analysis outlined the most wickets taken in an innings at the lowest economy.

```
BowlInnPareto_men <- psel(mbblbowl, high(W) * low(Econ), top_level = 999) ## Identify
BowlInnPareto_women <- psel(wbblbowl, high(W) * low(Econ), top_level = 999) ## Identi
```

- *Pareto-optimal Bowling Career*

  This analysis outlined the lowest bowling average across a career at the lowest economy and lowest strike rate. To provide a more accurate career report, bowlers required to have bowled a minimum of 200 balls which left 137 eligible male bowlers and 98 eligible female bowlers.

```
BowlCarPareto_men <- psel(filtBowl_men, low(Average) * low(StrikeRate) * low(Economy)
  filter(Average < 50) ## Identify Pareto frontier for men's BBL bowling career
BowlCarPareto_women <- psel(filtBowl_women, low(Average) * low(StrikeRate) * low(Econ
  filter(Average < 50) ## Identify Pareto frontier for women's BBL bowling career
```

All data was analyzed using R (v 4.1.0) statistical software (R Core Team, 2019). Firstly, the *dplyr* (Wickham et al., 2021) and *tidyr* (Wickham, 2021) packages were used for data manipulation to format the data into the correct structure to identify the Pareto frontiers. The *rPref* package (Roocks, 2016) was used to determine the Pareto frontiers using the *psel* function with the "top_level" argument set to 999 to ensure every athlete was assigned to a frontier (see Line 15 of the attached script for an example). Once the frontiers were established, the *ggplot2* (Wickham, 2016) and *scatterplot3D* (Ligges & Mächler, 2003) packages were used to visualize the data and subsequent Pareto frontiers.

3.4 Results

Men's Pareto-optimal batting

As seen by the red line in Figure 3.1, 11 Pareto-optimal innings were identified as Pareto-optimal innings. That is, no other batter has scored more runs at a faster strike rate than these 12 innings. These innings ranged from 6 runs off 1 ball (i.e., strike rate = 600) to 154 off 64 balls (i.e., strike rate = 240.63). Additionally, the solution of 6 runs off 1 ball has been attained six times. In Figure 3.2, nine Pareto-optimal batting careers were identified (in red) with Andre

Russell achieving the highest strike rate (164.07) and Brad Hodge achieving the highest average (42.78), while Joe Clarke, Alex Hales, Glenn Maxwell, Chris Lynn, Ben McDermott, Kevin Pietersen and Mitchell Marsh were all deemed Pareto-optimal due to varying combinations of both metrics.

3.4.0.1 Men's Pareto Batting Innings

```
ggplot(mapping = aes(x = R, y = SR)) +
    geom_point(data = BatInnPareto_men %>%  filter(R > 50 | SR > 100), alpha=0.05, size =
    geom_point(data = BatInnPareto_men %>% filter(.level == 1), alpha = 0.1, color = "red"
    geom_text(data = BatInnPareto_men %>% filter(.level == 1 & R != 6 & !(LastName %in% c(
    geom_text(data = BatInnPareto_men %>% filter(.level == 1 & LastName == "Rashid"), aes(
    geom_text(data = BatInnPareto_men %>% filter(.level == 1 & LastName == "Coulter-Nile")
    geom_text(data = BatInnPareto_men %>% filter(.level == 1 & LastName == "Cutting"), aes
    geom_text(data = BatInnPareto_men %>% filter(.level == 1 & LastName == "McAndrew"), ae
    geom_text(data = BatInnPareto_men %>% filter(.level == 1 & LastName == "Maxwell"), aes
    geom_line(data = BatInnPareto_men %>% filter(.level == 1), alpha = 0.5, colour = "red"
    annotate(geom = "text",x = 15,y = 600, label = "6 players", color = "red")+
    scale_y_continuous(breaks = seq(100,600,100))+
    coord_cartesian(xlim = c(0,160))+
    theme_minimal() +
    labs(x = "Runs Scored in an Innings",
```

```
    y = "Innings Batting Strike Rate") +
  theme(axis.title = element_text(size = 12,face = "bold"),
      panel.grid.minor.y = element_blank(),
      axis.text = element_text(size = 12, color = "black"))
```

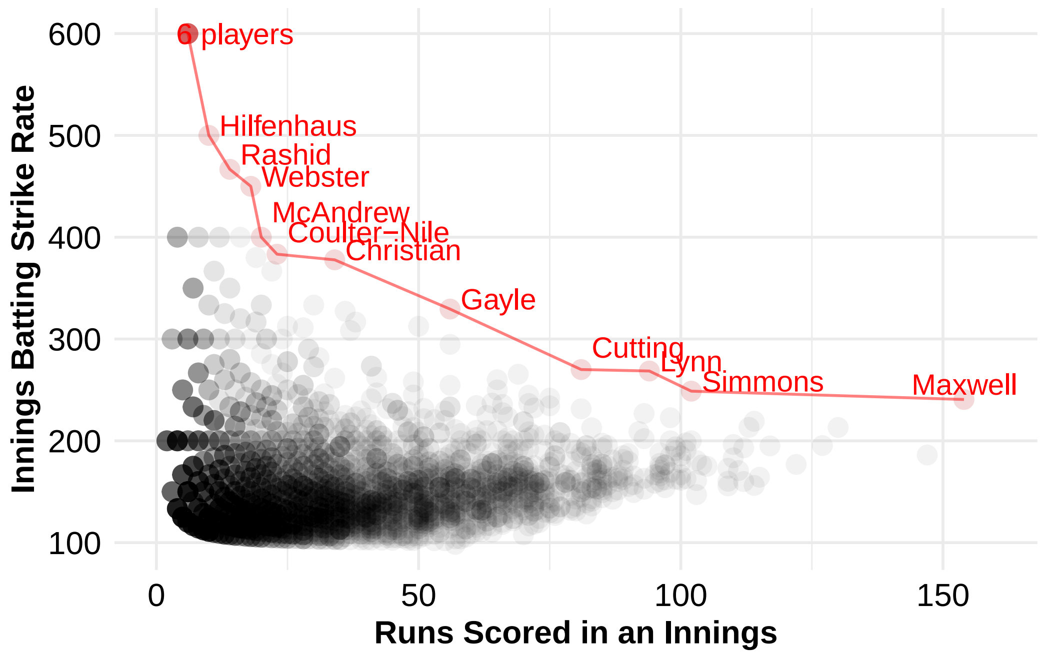

Figure 3.1: Men's Pareto-optimal batting within an innings with the Pareto frontier highlighted in red. N.B. For illustrative purposes, points were filtered out if both their runs scored was below 50 and their strike rate was below 100.

3.4.0.2 Men's Pareto Batting Career

```
ggplot(BatCarPareto_men, aes(x = Average, y = StrikeRate)) +
    geom_point(alpha=0.3, size = 3) +
    geom_point(data = BatCarPareto_men %>% filter(.level == 1), alpha = 0.3, color = "red"
    geom_line(data = BatCarPareto_men %>% filter(.level == 1), color = "red")+
    geom_text(data = BatCarPareto_men %>% filter(.level == 1 & LastName == "Clarke"), aes(
    geom_text(data = BatCarPareto_men %>% filter(.level == 1 & LastName == "Hales"), aes(l
    geom_text(data = BatCarPareto_men %>% filter(.level == 1 & LastName == "Marsh"), aes(l
    geom_text(data = BatCarPareto_men %>% filter(.level == 1 & LastName == "McDermott"), a
    geom_text(data = BatCarPareto_men %>% filter(.level == 1 & LastName == "Pietersen"), a
    geom_text(data = BatCarPareto_men %>% filter(.level == 1 & !LastName %in% c("Clarke","
    theme_minimal()  +
    coord_cartesian(xlim = c(0,46))+
    labs(x = "Career Batting Average",
         y = "Career Batting Strike Rate") +
    theme(axis.title = element_text(size = 12,face = "bold"),
        panel.grid.minor.y = element_blank(),
        axis.text = element_text(size = 12, color = "black"))
```

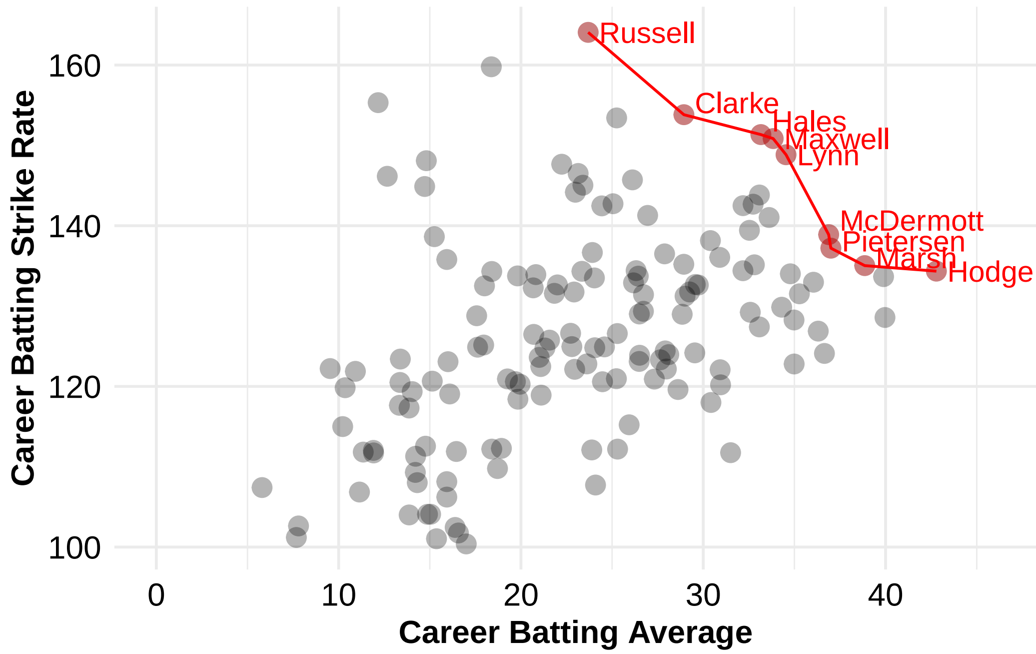

Figure 3.2: Men's Pareto-optimal batting across a career with the Pareto frontier highlighted in red. N.B. For illustrative purposes, points were filtered out if both their average was below 20 and their strike rate was below 100.

Men's Pareto-optimal bowling

Three Pareto-optimal bowling innings in Figure 3.3 were identified: 1/0 (i.e., 1 wicket for 0 runs conceded) by Jhye Richardson, 3/3 by Mitchell Johnson, and 6/7 achieved by Lasith Malinga. While there were three occurrences of 0/0, none of these are deemed Pareto-optimal as 1/0 by Richardson supersedes this combination. Five Pareto-optimal bowling careers were identified in Figure 3.4 as having the best balance of bowling average, strike rate, and economy, with Adil Rashid achieving the lowest average (14.13), Lasith Malinga achieving the lowest economy (5.40), Mitchell Starc achieving the lowest strike rate (11.25), while Rashid Khan and Hayden Kerr were deemed Pareto-optimal due to a combination of the three metrics.

3.4.0.3 Men's Pareto Bowling Innings

```
ggplot() +
  geom_jitter(data = BowlInnPareto_men %>% filter(.level != 1),aes(x = W, y = Econ), alpha
  geom_point(data = BowlInnPareto_men %>% filter(.level == 1), aes(x = W, y = Econ), alpha
  geom_text(data = BowlInnPareto_men %>% filter(.level == 1), aes(x = W, y = Econ, label =
  geom_line(data = BowlInnPareto_men %>% filter(.level == 1), aes(x = W, y = Econ), alpha
  theme_minimal() +
  labs(x = "Wickets in an innings",
       y = "Innings Bowling Economy") +
    coord_cartesian(xlim = c(0,6.3))+
    theme(axis.title = element_text(size = 12,face = "bold"),
        panel.grid.minor.y = element_blank(),
        axis.text = element_text(size = 12, color = "black"))
```

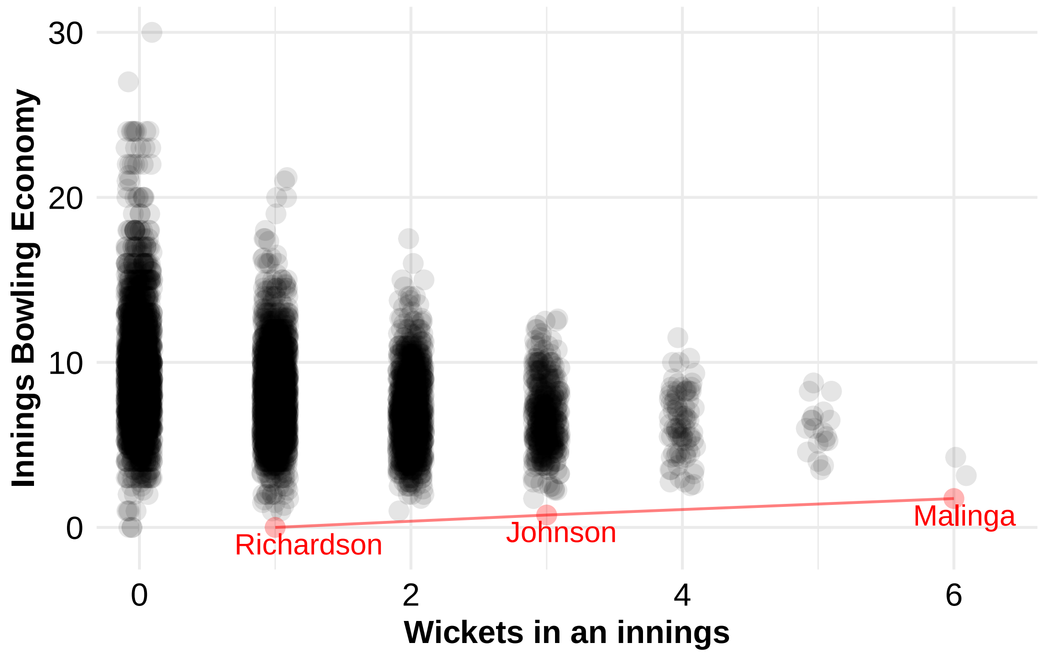

Figure 3.3: Men's Pareto-optimal bowling within an innings with the points on the Pareto frontier highlighted in red.

3.4.0.4 Men's Pareto Bowling Career

```
BowlCarPareto_men$color <- case_when(BowlCarPareto_men$.level == 1 ~ 2,
                                     BowlCarPareto_men$.level > 1 ~ 1)
BowlCarPareto_men$Label[BowlCarPareto_men$.level == 1] <- BowlCarPareto_men$LastName[BowlC
MenBowlCarParetoPlot <-scatterplot3d(BowlCarPareto_men[c("Economy","Average","StrikeRate")
                                  xlab="Career Bowling Economy",
                                  ylab="Career Bowling Strike Rate",
                                  zlab="Career Bowling Average")
zz.coords <- MenBowlCarParetoPlot$xyz.convert(BowlCarPareto_men$Economy, BowlCarPareto_men
```

```r
text(zz.coords$x,
     zz.coords$y,
     labels = BowlCarPareto_men$Label,
     cex = .8,
     pos = 2,
     col = "red")
```

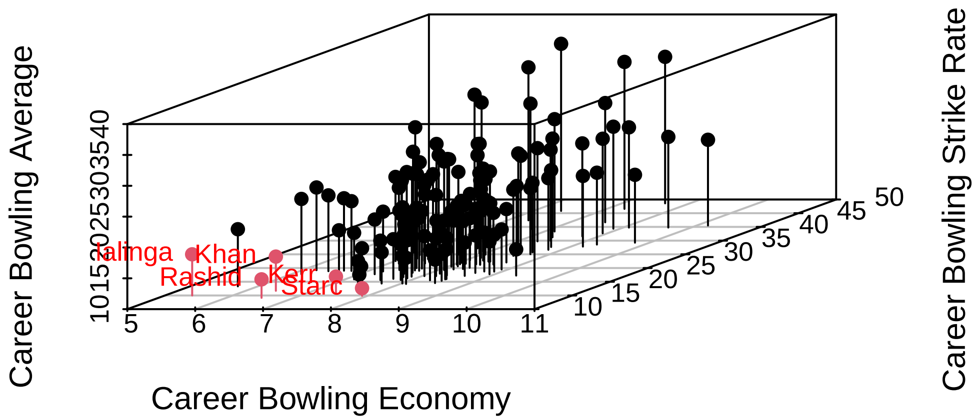

Figure 3.4: Men's Pareto-optimal bowling across a career with the points on the Pareto frontier highlighted in red.

Women's Pareto-optimal batting

Seven Pareto-optimal innings were identified in Figure 3.5 with extremities ranging from 6 runs off 1 ball (i.e., strike rate = 600) to 114 off 52 balls (i.e., strike rate = 219.23). In Figure 3.6, six Pareto-optimal batting careers were also identified with Laura Kimmince achieving the highest strike rate (144.08), Ellyse Perry achieving the highest average (50.15), with other Pareto-optimal solutions due to varying combinations of both metrics.

3.4.0.5 Women's Pareto Batting Innings

```
ggplot(BatInnPareto_women %>%  filter(R > 50 | SR > 100), aes(x = R, y = SR)) +
    geom_point(alpha=0.05, size = 3) +
    geom_point(data = BatInnPareto_women %>% filter(.level == 1), alpha = 0.1, color = "re
    geom_text(data = BatInnPareto_women %>% filter(.level == 1 & !(LastName %in% c("Nitsch
    geom_text(data = BatInnPareto_women %>% filter(.level == 1 & LastName == "Nitschke"),
    geom_text(data = BatInnPareto_women %>% filter(.level == 1 & LastName == "Molineux"),
    geom_text(data = BatInnPareto_women %>% filter(.level == 1 & LastName == "Devine"), ae
    geom_text(data = BatInnPareto_women %>% filter(.level == 1 & LastName == "Harris"), ae
    geom_line(data = BatInnPareto_women %>% filter(.level == 1), alpha = 0.5, colour = "re
    scale_y_continuous(breaks = seq(100,600,100))+
    coord_cartesian(xlim = c(0,125))+
    theme_minimal()  +
    labs(x = "Runs Scored in an Innings",
         y = "Innings Batting Strike Rate") +
    theme(axis.title = element_text(size = 12,face = "bold"),
        panel.grid.minor.y = element_blank(),
        axis.text = element_text(size = 12, color = "black"))
```

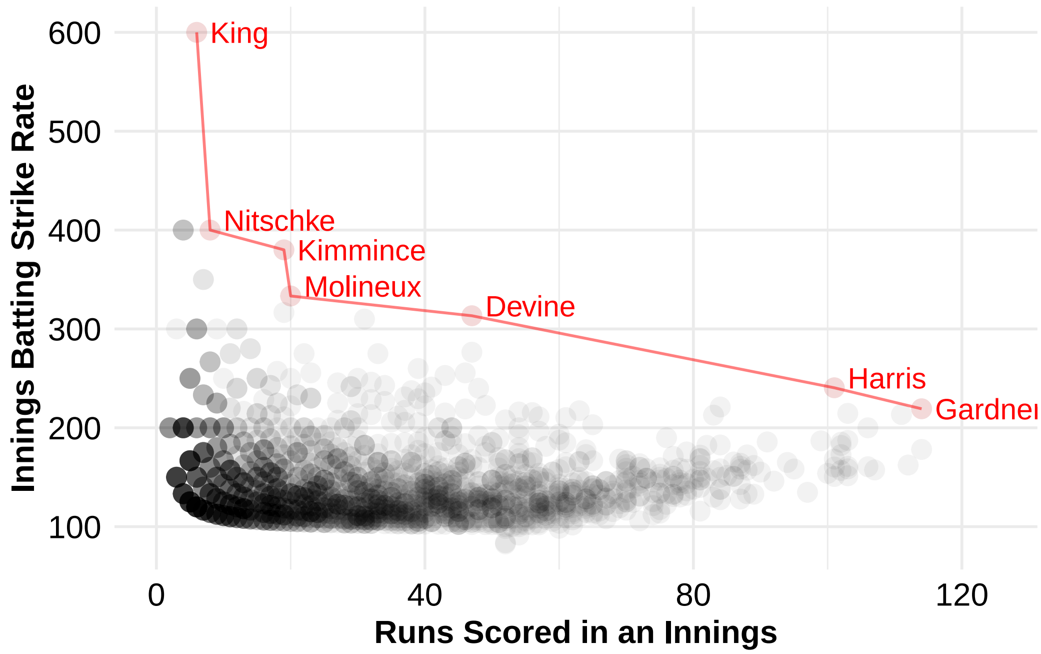

Figure 3.5: Women's Pareto-optimal batting within an innings with the Pareto frontier highlighted in red. N.B. For illustrative purposes, points were filtered out if both their runs scored was below 50 and their strike rate was below 100.

3.4.0.6 Women's Pareto Batting Career

```
ggplot(BatCarPareto_women, aes(x = Average, y = StrikeRate)) +
    geom_point(alpha=0.3, size = 3) +
    geom_point(data = BatCarPareto_women %>% filter(.level == 1), alpha = 0.3, color = "re
    geom_line(data = BatCarPareto_women %>% filter(.level == 1), color = "red")+
    geom_text(data = BatCarPareto_women %>% filter(.level == 1 & !LastName %in% c("Kimminc
    geom_text(data = BatCarPareto_women %>% filter(.level == 1 & LastName == "Kimmince"),
    geom_text(data = BatCarPareto_women %>% filter(.level == 1 & LastName == "Lanning"), a
    geom_text(data = BatCarPareto_women %>% filter(.level == 1 & LastName == "Mooney"), ae
```

```
coord_cartesian(xlim = c(0,53))+
theme_minimal()  +
labs(x = "Career Batting Average",
     y = "Career Batting Strike Rate") +
theme(axis.title = element_text(size = 12,face = "bold"),
    panel.grid.minor.y = element_blank(),
    axis.text = element_text(size = 12, color = "black"))
```

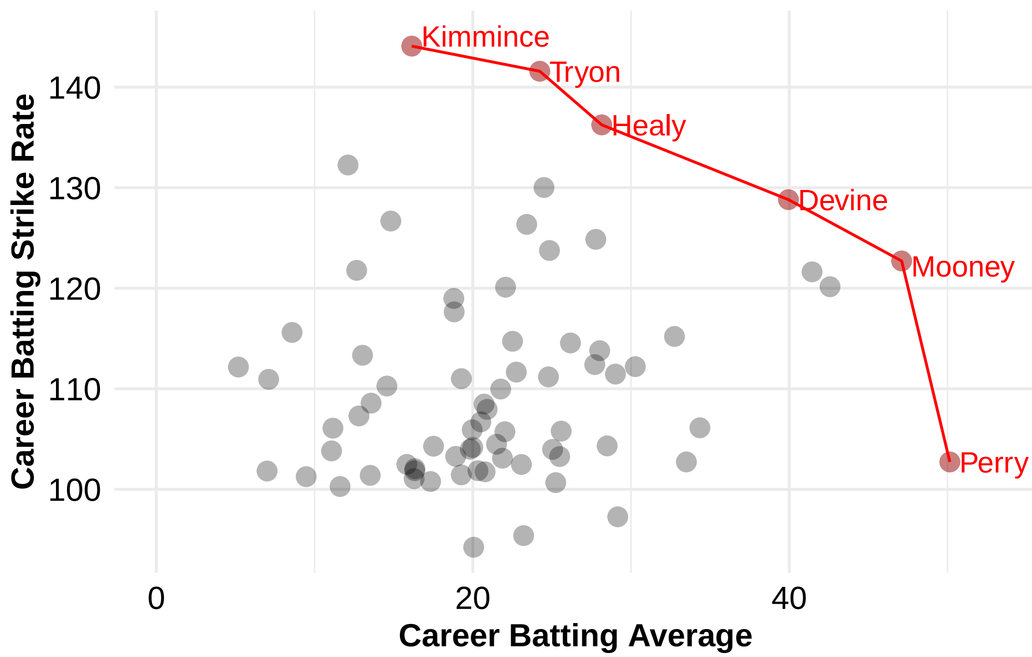

Figure 3.6: Women's Pareto-optimal batting across a career with the Pareto frontier highlighted in red. N.B. For illustrative purposes, points were filtered out if both their average was below 20 and their strike rate was below 100.

Women's Pareto-optimal bowling

Three Pareto-optimal bowling innings in Figure 3.7 were identified: 2/0 by Samantha Bates, 4/2 by Jemma Barsby, and 5/8 achieved by Amanda-Jade Wellington. Six Pareto-optimal bowling careers were identified in Figure 3.8 as having the best balance of bowling average, strike rate, and economy, with Julie Hunter achieving both the lowest average (16.38) and lowest economy (5.16), Harmanpreet Kaur achieving the lowest strike rate (16.00), while Sarah Aley, Darcie Brown, Ruth Johnston, and Hannah Darlington are deemed Pareto-optimal due to a combination of the three metrics.

3.4.0.7 Women's Pareto Bowling Innings

```
ggplot(BowlInnPareto_women, aes(x = W, y = Econ)) +
    geom_jitter(data = BowlInnPareto_women %>% filter(.level != 1),aes(x = W, y = Econ), a
    geom_point(data = BowlInnPareto_women %>% filter(.level == 1), aes(x = W, y = Econ), a
    geom_text(data = BowlInnPareto_women %>% filter(.level == 1), aes(x = W, y = Econ, lab
    geom_line(data = BowlInnPareto_women %>% filter(.level == 1), aes(x = W, y = Econ), al
    theme_minimal() +
    coord_cartesian(xlim = c(0,5.5))+
    scale_x_continuous(breaks = c(0:5))+
    labs(x = "Wickets in an innings",
         y = "Innings Bowling Economy") +
    theme(axis.title = element_text(size = 12,face = "bold"),
        panel.grid.minor.y = element_blank(),
        axis.text = element_text(size = 12, color = "black"))
```

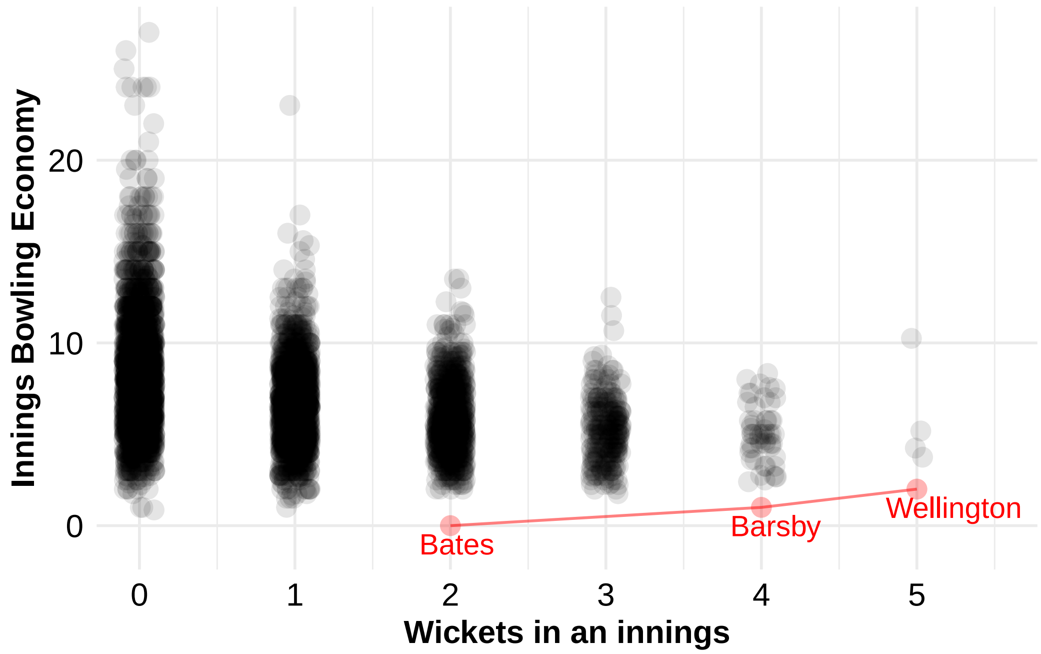

Figure 3.7: Women's Pareto-optimal bowling within an innings with the points on the Pareto frontier highlighted in red.

3.4.0.8 Women's Pareto Bowling Career

```
BowlCarPareto_women$color <- case_when(BowlCarPareto_women$.level == 1 ~ 2,
                                        BowlCarPareto_women$.level > 1 ~ 1)
BowlCarPareto_women$Label[BowlCarPareto_women$.level == 1] <- BowlCarPareto_women$LastName
WomenBowlCarParetoPlot <-scatterplot3d(BowlCarPareto_women[c("Economy","Average","StrikeRa
                                      xlab="Career Bowling Economy",
                                      ylab="Career Bowling Strike Rate",
                                      zlab="Career Bowling Average")
zz.coords <- WomenBowlCarParetoPlot$xyz.convert(BowlCarPareto_women$Economy, BowlCarPareto
```

```
text(zz.coords$x,
     zz.coords$y,
     labels = BowlCarPareto_women$Label,
     cex = .8,
     pos = 2,
     col = "red")
```

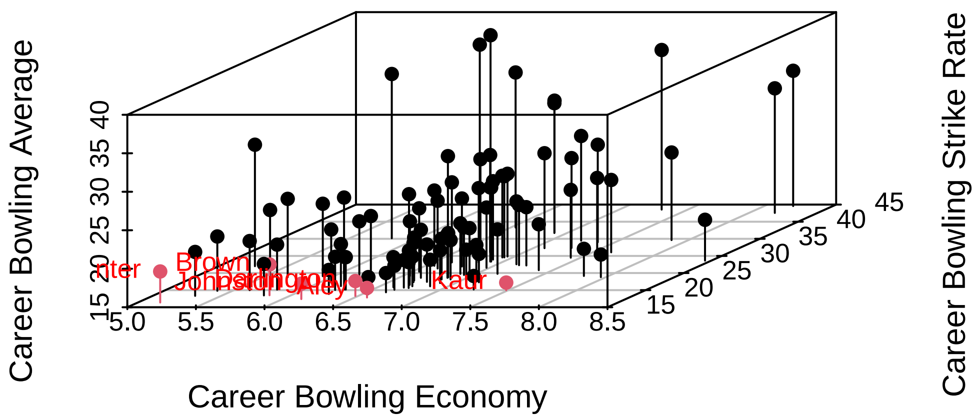

Figure 3.8: Women's Pareto-optimal bowling across a career with the points on the Pareto frontier highlighted in red.

3.5 Discussion

The present study aimed to introduce Pareto frontiers to sports scientists and its application in identifying and visualizing the most extraordinary players when considering multiple variables. While it is intuitively recognized that multiple attributes are required for success in sport, by identifying the Pareto frontier between these attributes is a simple, yet effective method to identify all the players that possess an optimal balance of these attributes relative to other individuals in the cohort. By analyzing talent multivariately, rather than simply analyzing multiple variables univariately, players can be deemed optimal despite not being objectively highest in a single variable. The present study highlights that when there are conflicting attributes that are of equal interest, the attributes should be viewed in tandem using Pareto frontiers, or else there is a risk that the expectations of an individual to attain the highest level in both attributes univariately may be unfeasible. This was evident in all eight Pareto frontiers, as at least one athlete was identified in each example that was not identified as highest ranked athlete in any metric when analyzed univariately, and yet was deemed Pareto-optimal due to their balance in the metrics of interest. For instance, when observing the career batting average Pareto frontier in the WBBL Figure 3.6, there is an expansive continuum of batters that are all deemed Pareto-optimal as they each have slightly different average/strike rate profiles.

The main advantage of Pareto frontiers highlighted in the present study is identifying athletes who are optimal across multiple metrics even when they are not the highest ranked in any metric. This was most evident in the MBBL where, Chris Lynn when viewed univariately, has the 14th-highest career batting average (34.54), which is 8.24 runs per innings lower than the highest Figure 3.2. Similarly, he has the eighth-highest strike rate, striking at 148.84 which is 15.22 runs per 100 balls lower than the highest. However, when considering both metrics simultaneously and visualizing these metrics, it is clear that he is one of the best batters across the 11 seasons of the MBBL.

Pareto frontiers can also be used to provide benchmarks when conflicting attributes are both desirable and by analyzing these attributes in tandem, more realistic expectations can be set for each athlete depending on whether they sit in the cartesian plane. The resulting benchmarks can be more individualized than what can be expected when viewing metrics univariately. Consequently, by viewing these metrics multivariately, different levels of quality within differing squad roles can be better expressed. Pareto frontiers can be used not only for extreme values, but can also be used to visualize the second, third, and n-th frontiers to further interrogate where an individual lies within the cartesian space. For example, in Figure 3.5, it was evident that there is a substantial gap between the first and second WBBL career batting Pareto frontiers.

In this study we chose to observe batting and bowling as purely independent roles within cricket; however, there are also avenues for Pareto frontiers to be established for all-rounders within cricket (i.e., players that are picked for both their batting and bowling ability). However, it should be noted that if an all-rounder Pareto frontier were to be established with both batting average and strike rate as well as bowling average, economy, and strike rate, the resulting five-dimensional outputs, while valid and executable, become increasingly difficult to interpret and visualize. To do such an analysis, a factor-reduction technique such as principal components analysis should be considered and the Pareto frontier could be built from the extracted components (e.g., batting and bowling).

This technique could also be used to identify "maximal" efforts when multiple observations of an individual are recorded which is common in sports science practice. For example, Duthie et al. (2021) sought to define maneuverability by identifying the maximum tortuosity within each

0.5 $m \cdot s^{-1}$ increment of speed and plotting the line of best fit through these points (Duthie et al., 2021). In this instance, a Pareto frontier could have been fitted to the data to eliminate the need for an arbitrary selection of 0.5 $m \cdot s^{-1}$ speed bins. Similarly, Morin et al. (2021) sought to develop an *in-situ* acceleration-speed profile by identifying the maximum acceleration within each 0.2 $m \cdot s^{-1}$ of speed and plotting the line of best fit through these points (Morin et al., 2021). By using a Pareto frontier, the need for an arbitrating 3 $m \cdot s^{-1}$ threshold and 0.2 $m \cdot s^{-1}$ speed bins could have been avoided. Finally, Rudsits et al. (2018) sought to identify the torque-cadence and power-cadence profile of cyclists by identifying maximal torque or power values in each 5 rpm bin (Rudsits et al., 2018). If a Pareto frontier were used, the model would not require any additional filtering to remove "non-maximal" efforts as the frontier will have already deemed those points non-optimal.

The present study also illustrated how Pareto frontiers can be used to visualize talent in more than 2 dimensions. For example, while Darcie Brown has the seventh-lowest bowling average, 11th-lowest economy, and the 18th-lowest strike rate (see Figure 3.8), she can be deemed a Pareto-optimal bowler as there are no other bowlers who supersede her across all three metrics. Similarly, in the MBBL (Figure 3.4), while Rashid Khan has the sixth-lowest average, sixth-lowest economy and the 18th-lowest strike rate, he can be deemed a Pareto-optimal bowler as there are no other bowlers who supersede him across all three metrics. While there will be some correlations between the three bowling metrics (i.e., average, economy, and strike rate) as the metrics are related (e.g., wickets taken is the denominator of average and numerator of strike rate), visualizing the third dimension is still necessary as the reader would still need to multiply the x and y values to understand where they would sit in the third dimension. This could be further expanded into higher dimensions; however, these dimensions become increasingly difficult to visualize.

It should also be considered that there is some level of uncertainty surrounding each observation in the career Pareto frontiers due to the differing number of observations. For example, Joe Clarke is deemed Pareto-optimal as he is currently striking at 153.82 at an average of 28.94 after 16 innings; however, it is right to assume that it is more uncertain that he lies on the frontier than Chris Lynn who has 100 observations. Therefore, future research could consider providing confidence or credible intervals around the probability that an individual lies on the Pareto frontier. Consequently, it is then feasible that a probability that an individual sits on the first, second, or third frontier could be calculated.

While the present study used Twenty20 cricket to illustrate the power and usefulness of Pareto frontiers, the concept can be widely applied within sports science data sets, especially when the variables of interest are uncorrelated or negatively correlated. Pareto frontiers can still be established between two positively correlated metrics; however, it is likely that there will be less "hidden" athletes on this frontier as naturally the athletes who are high in one metric will be high in the other metric. Future research should apply Pareto frontiers across different avenues within sports performance analysis, such as repeated-sprint ability and dynamic strength index which have multi-faceted determinants. In addition, there are many other possibilities within sports science whereby Pareto frontiers can reveal athletes who possess the optimal balance of the metrics of interest.

4 The Role of Informative Priors in Bayesian Inference

4.1 Abstract

Sporting success can be determined by the smallest margins; however, with small sample sizes, detecting these small worthwhile effects is notoriously difficult. Consequently, the call to utilise a Bayesian framework when conducting research to address these issues has increased recently. This study illustrates how Bayesian models can incorporate prior information and how models with more prior information can assist in decision making when margins are small. This study revisits a paper published in this journal investigating the effects of -alanine on 4-km cycling TT performance through a randomised placebo-controlled trial with 14 trained cyclists. The data set was analysed again in both a frequentist framework and in a Bayesian framework with priors varying in terms of informativeness. In a frequentist framework, there was no significant effect of -alanine supplementation on 4-km TT performance compared to a placebo. However, a Bayesian model with a very informative prior revealed an estimated 97.8% probability that

-alanine improved performance compared to a placebo, and an estimated 99.0% probability that -alanine supplementation improved 4-km TT performance compared to baseline. On average, the 4-km TT time decreased by an estimated 6.64 s (-12.20 to -1.19 s), resulting in a small (~2%) but beneficial effect that was more difficult to identify in the previous study using frequentist statistics. While the present study demonstrates why incorporating prior information can increase the power in small-sample research, it also reinforces that -alanine supplementation may be beneficial for cycling time trials of ~6 min in duration. Using prior predictive distributions can assist researchers in understanding the informativeness of a given prior distribution and should be reported when publishing.

The following chapter is a copy of the submitted manuscript:

Newans, T., Bellinger, P., Drovandi, C., & Minahan, C. The role of informative priors: A new look at the role of -alanine on 4-km time-trial performance in cyclists. European Journal of Sports Science. Under review.

As co-author of the paper "The role of informative priors: A new look at the role of -alanine on 4-km time-trial performance in cyclists", I confirm that Timothy Newans has made the following contributions:

- Study concept and design
- Data analysis and interpretation
- Manuscript preparation

Name: Clare Minahan

Date: 29/03/2023

4.2 Introduction

Sports Scientists are often faced with small sample sizes and/or small effect sizes (Atkinson et al., 2012) and, while these issues are not limited to sports-science research (Bacchetti et al., 2011; Baldwin & Fellingham, 2013; McNeish, 2016; Nomura et al., 2011), they limit the ability to infer changes in the population parameter from sample statistics. Consequently, there has been on-going discussion surrounding the use of probabilistic statements within the sports-science community (Batterham & Hopkins, 2006; Borg et al., 2018; Mengersen et al., 2016; Sainani, 2018; Welsh & Knight, 2015). As such, alternatives to traditional null hypothesis statistical testing have been proposed to provide more insightful inferences in decision making for Sports Scientists (Batterham & Hopkins, 2006). Magnitude-based inferences (MBI), a newly-established method to provide probabilistic inference statements, gained traction quickly, becoming highly cited, and was published in guidelines for authors for some journals (Hopkins et al., 2009). Although MBI received early criticism (Barker & Schofield, 2008), it was still widely adopted. More recently, additional attempts to discredit the concept of MBI were vocalized by Welsh & Knight (2015) who provided a statistical review of the method illustrating unacceptably high levels of Type I errors (Welsh & Knight, 2015). Next, Sainani (2018) weighed into the statistical debate which triggered the sport-science community to critically assess whether MBI should be accepted as a statistical method (Sainani, 2018). Throughout all the rebuttals of MBI (Barker & Schofield, 2008; Borg et al., 2018; Mengersen et al., 2016; Welsh & Knight, 2015), the recurring rhetoric was for Sports Scientists to adopt a 'fully Bayesian' approach to provide probabilistic statements when inferring conclusions (Mengersen et al., 2016; Welsh & Knight, 2015).

In 2016, Mengersen and colleagues (Mengersen et al., 2016), provided a worked example and template to perform Bayesian inference in exercise- and sports-science data sets. The paper clearly articulated the need for Bayesian inference for providing probabilistic statements (Mengersen et al., 2016). A growing trend in the use of Bayesian inference in sports science (Santos-Fernandez et al., 2019) has benefited from numerous tutorials being developed to provide readers with the skills in executing Bayesian analyses (Kruschke, 2014; Quintana & Williams, 2018). Nonetheless, there is still a gap in knowledge regarding the appropriateness and benefit of Bayesian inference for Sports Scientists (Mengersen et al., 2016).

Bayesian inference incorporates prior information ('priors') to curate the model and can provide estimates with more certainty when compared with frequentist statistical analyses, particularly when informative priors are used (Mengersen et al., 2016). For example, when considering the mean 100-m race time of a sample of elite track sprint runners, a frequentist approach does not impose any constraints on what would be considered realistic sprint times despite our knowledge from extensive historical data, whereas such information can be readily incorporated into a Bayesian approach via a prior distribution. As a result, Bayesian statistics can be more sensitive to small, but real, differences between samples. However, as prior distributions can be used on each parameter within a model (i.e., the intercept, coefficients, variance terms etc.), it is difficult to decipher the interactions between these priors and may unintentionally generate an 'overly-informative' prior. Consequently, by sampling parameter values from the prior distribution, and for these parameter values, simulating data from the statistical model, a 'prior predictive distribution' can be generated, which can then be compared to the *a priori* likely values expected in the population to determine if the prior distributions being used are reasonable.

In the present study, we revisited data from an experiment published in 2016 in this journal (Bellinger & Minahan, 2016). In the study by Bellinger & Minahan (2016), fourteen highly-trained cyclists were supplemented with either 6.4 $g \cdot day^{-1}$ of -alanine or a placebo for 4 wk.

They completed a 1-, 4- and 10-km time trial (TT) as well as a supramaximal cycling bout before and after the supplementation period. Of those performance tasks, the 4-km TT was of most interest given that the duration of the 4-km TT (~6 min) fits within the timeframe most likely to be amenable by -alanine supplementation (Hobson et al., 2012). However, employing frequentist statistics, -alanine supplementation was deemed to not significantly alter performance when assuming a 5% significance level ($p = 0.060$), while a MBI approach demonstrated that there was a 94% likelihood of a beneficial effect of -alanine supplementation on 4-km TT performance time. Given the shortcomings of the MBI approach (Mengersen et al., 2016) and the discretised significance/non-significance in frequentist statistics (given a $p = 0.060$ is close to the 5% significance level), we suggest that a Bayesian approach with an informative prior distribution may provide a more precise understanding of the effects of -alanine supplementation in light of the small sample size.

Consequently, the authors have given permission to re-visit this study and re-analyse this experiment using a Bayesian framework with various prior distributions to illustrate how the different priors result in different interpretations. It is hypothesised that, when using an informative prior, the model could show an improvement (>95% probability) in 4-km TT performance (i.e., reduced time to complete 4 km) with -alanine.

4.3 Methods

Fourteen trained male cyclists (age = 24.8 ± 6.7 yr, mass = 71.1 ± 7.1 kg, $VO_{2max} = 65.4 \pm 10.2$ mL $\cdot$ kg $\cdot$ min^{-1}) who were currently cycling 250-600 km $\cdot$ wk^{-1} and competing in either local A-grade criterion or national-road series racing completed the study. For further information regarding the testing protocol, please refer to the previous study (Bellinger & Minahan, 2016).

```
library(tidyverse)
library(brms)
library(tidybayes)
library(bayestestR)
library(patchwork)
set.seed(99.94)
ba_df <- read_csv("www/data/Study_3_bayesian_beta_alanine.csv") ## Read in data
ba_df <- ba_df %>%
  mutate(Diff4k = `Post-4km`-`Pre-4km`) ## Calculate difference in 4km TT
```

To provide a reproducible example, we made the data from this study publicly available as well as the statistical analyses. While more statistics software programs are providing access to point-and-click Bayesian analyses (e.g., SPSS and Jamovi), the present study chose to conduct all analyses in R for Statistical Computing (R Core Team, 2019) to demonstrate the minimal increase in complexity to progress from frequentist linear models to Bayesian linear models.

As the study was a randomised control trial, the -alanine effect was assessed compared to 0 (i.e., no effect) as well as the placebo group (i.e., placebo effect). Consequently, the -alanine supplementation group was selected as the reference group (thus the intercept in the model summary can show the effect compared to 0) and the placebo was selected as the alternate treatment (thus the coefficient in the model summary can show the -alanine effect compared to placebo).

To compare the results of the frequentist and Bayesian frameworks, the frequentist analysis was re-run to confirm the findings of Bellinger & Minahan (2016). This was performed using the *lm* function in the base R package *stats*. To provide the greatest accessibility and readability for the Bayesian analyses, the *brms* package was chosen due to its similarity in structure to the typical *lm* formula structure. The *brms* package (Bürkner, 2017) provides an interface to Stan (a very fast programming language written in C++) within R and outputs the results in a similar format to the *stats* package.

```
## Frequentist Model
freq <- lm(Diff4k ~ Treat, data = ba_df) ## Frequentist linear model


summary(freq) ## Summary of Frequentist model


confint(freq) ## Confidence Interval of Frequentist model
```

The model has two regression parameters: the intercept b_0 which here corresponds to the mean difference in 4-km times under the treatment, and b_1 is the effect of the placebo relative to the treatment, so that $b_0 + b_1$ is the mean difference in 4-km times under the placebo. Here we consider imposing different levels of informativeness on the prior predictive distribution of the mean response (difference in 4-km times). The prior predictive distribution of the mean response under the treatment is equivalent to the prior distribution of b_0 for this model.

Firstly, the flat priors for b_0 and b_1 were considered; however, as the brms package does not allow for flat priors, a normal prior was assigned to b_0 and a normal prior assigned to b_1, both with a mean of 0 and a SD of 10000 to act as an uninformative prior. For this prior configuration, 48.6% of the prior predictive values of the mean response were theoretically impossible under the treatment, as a 4-km TT lasting on average 360 s cannot have, on average, an improvement of more than 360 s and, as a result, no model was built using this prior.

```
## Uninformative Prior Model
pnorm(-360,0,10000) # Probability of -360 s or larger decrease in 4km TT given a mean = 0
```

Consequently, the first Bayesian model (Model 1) used was built with a normal prior distribution (mean = 0, SD = 72) for the mean response of both b_0 and b_1. Here, 1 SD (i.e., 72 s) represented a 20% improvement or decrement in performance under the treatment. The -alanine treatment was set as the default baseline in the model so that the intercept represented the -alanine treatment effect (i.e., is the -alanine treatment effect different from zero), while the b coefficient represented the placebo level (i.e., is the -alanine treatment effect different from a placebo). The prior for the normal response standard deviation, sigma, was kept controlled across all models to allow for consistency, in which the default-selected brms prior was used, in this case a half t-distribution with a mean of 0 and SD of 6.3 with 3 degrees of freedom.

```
## Model 1 - Vague Prior Model
mean_v <- 0 ## Set vague prior mean to 0
sd_v <- 72 ## Set vague prior mean to 72

pnorm(-100,mean_v,sd_v) + 1 - pnorm(100,mean_v,sd_v) ## Calculate probability the treatmen

vague <- brm(Diff4k ~ Treat, iter = 10000, chains = 8, data = ba_df, seed = 99.94,
             prior = c(set_prior(paste0("normal(",mean_v,",",sd_v,")"), class = "Intercept
                       set_prior(paste0("normal(",mean_v,",",sd_v,")"), class = "b"))) ##
```

The second Bayesian model (Model 2) used was built with a normal prior distribution (mean = 0, SD = 18) for both b_0 and b_1, with 1 SD (i.e., 18 s) representing a 5% improvement or decrement in performance under the treatment.

```
## Model 2 - Informative Prior Model
mean_i <- 0 ## Set informative prior mean to 0
sd_i <- 18 ## Set informative prior SD to 18
```

```r
pnorm(-50,mean_i,sd_i) + 1 - pnorm(50,mean_i,sd_i)  ## Calculate probability the treatment
1 - pnorm(-10,mean_i,sd_i) - (1 - pnorm(10,mean_i,sd_i))  ## Calculate probability the tre

informative <- brm(Diff4k ~ Treat, data = ba_df,  iter = 10000, chains = 8, seed = 99.94,
                   prior = c(set_prior(paste0("normal(",mean_i,",",sd_i,")"), class = "Int
                             set_prior(paste0("normal(",mean_i,",",sd_i,")"), class = "b")
```

The final Bayesian model (Model 3) was built with a normal prior distribution (mean = -10.26, SD = 18) for b_0, representing a 2.85% improvement in performance, the estimated effect of -alanine supplementation as outlined in the previously-completed meta-analysis (Hobson et al., 2012). A normal prior distribution (mean = 10.26, SD = 18) was chosen for b_1, with 1 SD (i.e., 18 s) to again represent a 2.85% improvement in performance because of -alanine supplementation.

```r
## Model 3 - Very Informative Prior Model
mean_vi <- -10.28 ## Set very informative prior mean to -10.28
sd_vi <- 18 ## Set very informative prior SD to 18

1 - pnorm(-37.76,mean_vi,sd_vi) - (1 - pnorm(1.33,mean_vi,sd_vi)) ## Calculate probability

veryinformative <- brm(Diff4k ~ Treat, data = ba_df, iter = 10000, chains = 8, seed = 99.9
                       prior = c(set_prior(paste0("normal(",mean_vi,",",sd_vi,")"), class
```

```
                                        set_prior(paste0("normal(",-mean_vi,",",sd_vi,")"), class
```

4.4 Results

To illustrate the differences in the priors, the density curves of the prior predictive distribution of the mean response under the treatment were constructed, which for this model, corresponds to the prior distribution for b_0. The varying prior predictive distributions for the mean effect of -alanine are displayed in Figure 4.1. As expected, the first two distributions are centred around the mean of 0 with differing variances, with the third distribution displaying the same variance as the third distribution, just left-shifted to a mean of -10.26, as specified in the model. The density curves in Figure 4.1 show the sample space that each Bayesian model was "constrained to", illustrating the more-informative priors (e.g., Model 2 and Model 3).

```
vpriorplot <- ggplot(data.frame(x = c(-300, 300)), aes(x)) +
    stat_function(fun = dnorm, n = 600, args = list(mean = mean_v, sd = sd_v)) +
    geom_text(aes(x = -250, y = 0.003), label = "Vague", fontface = "bold")+
    scale_x_continuous(breaks = c(seq(-300,300, by = 100)))+
    theme_minimal()+
    labs(y = "Density")+
    theme(axis.title.x = element_blank(),
          panel.grid.minor = element_blank(),
          axis.title.y = element_text(size = 10, color = "black", face = "bold"),
```

```
            axis.text = element_text(size = 10, color = "black")) ## Visualise the prior pre
infpriorplot <- ggplot(data.frame(x = c(-300, 300)), aes(x)) +
    stat_function(fun = dnorm, n = 600, args = list(mean = mean_i, sd = sd_i)) +
    geom_text(aes(x = -62.5, y = 0.015), label = "Informative", fontface = "bold")+
    theme_minimal()+
    labs(y = "Density")+
    scale_x_continuous(breaks = c(seq(-75,75, by = 25)))+
    scale_y_continuous(breaks = c(seq(0,0.02, by = 0.01)))+
    coord_cartesian(xlim = c(-75,75))+
    theme(axis.title.x = element_blank(),
          panel.grid.minor = element_blank(),
          axis.title.y = element_text(size = 10, color = "black", face = "bold"),
          axis.text = element_text(size = 10, color = "black")) ## Visualise the prior pre
vinfpriorplot <- ggplot(data.frame(x = c(-75, 75)), aes(x)) +
    stat_function(fun = dnorm, n = 600, args = list(mean = mean_vi, sd = sd_vi)) +
    geom_text(aes(x = -60, y = 0.015), label = "Very Informative", fontface = "bold")+
    theme_minimal()+
    labs(x = "Difference in 4km TT from Placebo",
         y = "Density")+
    scale_x_continuous(breaks = c(seq(-75,75, by = 25)))+
    scale_y_continuous(breaks = c(seq(0,0.03, by = 0.01)))+
```

```
    coord_cartesian(xlim = c(-75,75),

                    ylim = c(0,0.022))+

    theme(panel.grid.minor = element_blank(),

          axis.title = element_text(size = 10, color = "black", face = "bold"),

          axis.text = element_text(size = 10, color = "black")) ## Visualise the prior pre


vpriorplot / infpriorplot / vinfpriorplot
```

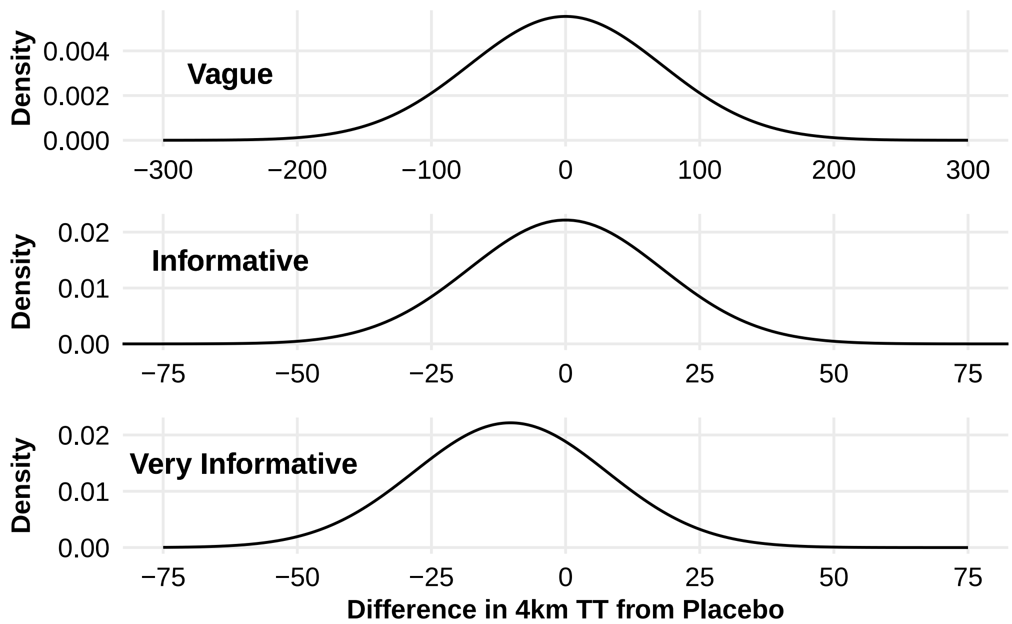

Figure 4.1: Prior predictive distributions of the mean pre-post change in 4-km time trial (TT) difference under the -alanine supplementation based on three prior distributions that differ in terms of level of informativeness.

The results from the frequentist model and the three Bayesian models are displayed in Table 4.1.

In the frequentist model, at the 5% significance level, the intercept (i.e., -alanine effect) was significantly different from zero ($t(12)$ = -2.45, p = 0.031), although the coefficient (i.e., -alanine compared to a placebo) was not significant ($t(12)$ = 2.08, p = 0.060), confirming previous findings by the authors (Bellinger & Minahan, 2016). Model 3, the Bayesian model with the very informative prior, displayed a 99.0% probability that -alanine decreases the mean 4-km TT time, and that this decrease was, on average, by 6.64 s.

Table 4.1: Comparison of frequentist and various Bayesian models with different priors.

Model	Intercept (i.e., -alanine effect)	Coefficient (i.e., -alanine vs Placebo)
Frequentist	-6.47 (-12.23 to -0.72)	7.76 (-0.38 to 15.91)
Bayesian - Vague	-6.46 (-12.09 to -0.84)	7.71 (-0.22 to 15.69)
Bayesian - Informative	-6.25 (-11.73 to -0.72)	7.39 (-0.44 to 14.99)
Bayesian - Very Informative	-6.64 (-12.20 to -1.19)	7.88 (0.30 to 15.64)

N.B. Frequentist values are presented as mean (95% confidence interval), while Bayesian values are presented as mean (95% credible interval).

When comparing -alanine with a placebo, there is a 97.8% probability that -alanine improves mean 4-km TT performance compared to a placebo. The posterior distributions of the pre-post difference in mean 4-km TT performance from each of the three Bayesian models are displayed in Figure 4.2.

```
## Draws from posterior distribution
vagueplot <- vague %>%
  spread_draws(b_Intercept,b_TreatPL) %>%
  mutate(PL = b_TreatPL-b_Intercept) ## Calculate placebo effects for Model 1
informativeplot <- informative %>%
  spread_draws(b_Intercept,b_TreatPL) %>%
  mutate(PL = b_TreatPL-b_Intercept)  ## Calculate placebo effects for Model 2
veryinformativeplot <- veryinformative %>%
  spread_draws(b_Intercept,b_TreatPL) %>%
  mutate(PL = b_TreatPL-b_Intercept) ## Calculate placebo effects for Model 3


## Posterior Predictive Distribution Figure


ggplot()+
  geom_density(data = vagueplot, aes(x = b_Intercept), color = "red")+
  geom_density(data = informativeplot, aes(x = b_Intercept), color = "darkgreen")+
  geom_density(data = veryinformativeplot, aes(x = b_Intercept), color = "darkblue")+
  geom_rect(aes(xmin = -18,xmax = -12,
              ymin = 0.11, ymax = 0.14), fill = "white",color = "black")+
  geom_text(aes(x = -15,y = 0.136), label = "Types of Priors", fontface = "bold",color = "
  geom_text(aes(x = -15,y = 0.129), label = "Vague", color = "red")+
```

```
geom_text(aes(x = -15,y = 0.122), label = "Informative", color = "darkgreen")+

geom_text(aes(x = -15,y = 0.115), label = "Very Informative", color = "darkblue")+

geom_vline(xintercept = 0)+

labs(x = "Difference in 4km TT from Placebo",

     y = "Density")+

theme_minimal()+

scale_x_continuous(limits = c(-18,5))+

theme(legend.position = "bottom",

      axis.title = element_text(size = 12,face = "bold"),

      axis.text = element_text(size = 12),

      panel.grid.minor.y = element_blank())
```

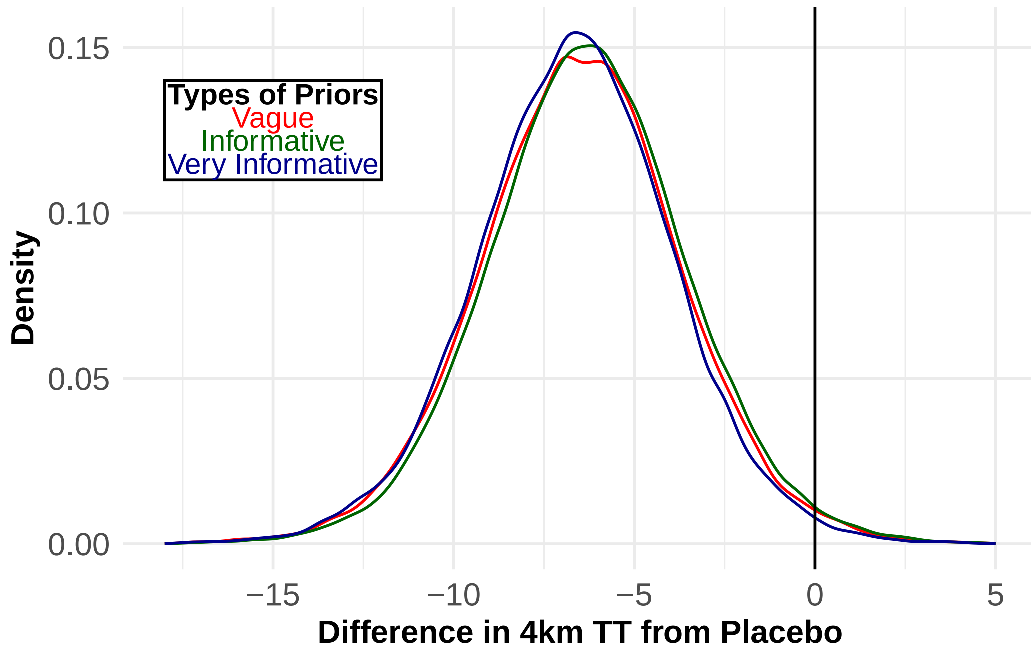

Figure 4.2: Posterior distributions of the mean pre-post change in 4-km time trial (TT) difference between -alanine supplementation and placebo from three Bayesian models with differing priors.

4.5 Discussion

This study updates the literature on the effect of -alanine on a 4-km cycling TT performance and outlines how informative priors can provide more meaningful statistical analyses in a Bayesian framework compared to a frequentist framework. The present study showed that, with the most-informative prior, there was an estimated 99.0% probability that -alanine improves the 4-km TT performance and an estimated 97.8% probability that -alanine improves performance more than a placebo. On average, 4-km TT performance decreased by an estimated 6.64 s, confirming similar findings by the authors in the previously published paper

(Bellinger & Minahan, 2016). Consequently, it is reinforced that -alanine supplementation is likely to provide an ergogenic effect to maximal effort time trials with a duration of ~6 min. While the exact mechanisms underpinning how increased carnosine contributes to improved exercise performance requires further clarification, the ergogenic effect is likely due to enhanced muscle contractile properties (Dutka et al., 2012; Everaert et al., 2013), and/or an increase in intracellular buffering capacity (Baguet et al., 2010).

A main feature of the present study was the use of prior predictive distributions to ensure the priors encapsulated the possible sample space for the mean of the response variable (in this case, change in 4-km TT performance). While this study sought to keep the methodology simple by only generating prior predictive distributions for the mean of the response variable, the prior predictive distribution can often be impacted by the interaction between the different priors (e.g., intercept, levels in conditions, variances, covariances etc.). Therefore, sampling from the prior distribution is essential to ensure the prior predictive distribution both captures all the practically possible values of the mean, as well as minimises practically impossible values. As humans struggle to conceptualise probabilities (Gillies, 2000), it is helpful to visualise this prior predictive distribution to ensure the Bayesian model does not contain too-informative priors that unintentionally bias the model. In the first instance of using Bayesian analyses, it is natural to start with uninformative priors in a bid to minimise the bias of running an analysis and align as close to a frequentist framework as possible. However, by providing such wide priors as seen in the uninformative prior, it unintentionally introduces a bias towards unrealistic values. Consequently, providing priors based on previously published studies in similar populations or subject matter expert knowledge is required (Zondervan-Zwijnenburg et al., 2017).

After visualising the prior predictive distributions of the mean for the vague and informative models (i.e., Models 1 & 2), it is evident that more-constrained prior distributions are required due to the elite nature of the athletes involved. That is, in the vague prior predictive distribution, 16.5% of the distribution still predicted that -alanine would, on average, affect the 4-km TT by more than 100 s. When narrowing the prior for the informative model, only 0.5% of the distribution predicted that -alanine would, on average, affect the 4-km TT by more than 50 s. However, being a normal distribution, it also disadvantaged the model. As it was centred around 0, 42.1% of the prior distribution predicted that -alanine would not affect the 4-km TT by more than 10 s (i.e., that is the mean effect is between -10 and 10 s). This is now deemed “too informative” as a prior meta-analysis had shown that the median effect of -alanine was 2.85% which would be approximately 10 s. This was evident in the resultant models, with the coefficient reducing from the vague (7.75 s) to the informative model (7.28 s). Consequently, by translating the informative prior to a mean of -10.26 s in line with the previous meta-analysis, 67.7% of the prior predictive distribution contained values within a 0.3% decrement in performance (1.33 s) and 10.49% improvement (37.76 s) in performance, the two limits of the confidence interval from the meta-analysis.

When comparing the results to the frequentist model previously used, the informative priors were able to not only show that -alanine supplementation improved performance, but the interpretation of the results is also more translatable than in a frequentist framework. The frequentist interpretation of $p = 0.06$ would be that, if the experiment were to be replicated infinitely many times, there would be a 6% chance that the true population mean difference between -alanine and a placebo was 0, and 95% of the time, the interval (-0.38 to 15.91 s) would include the true population parameter. However, in a Bayesian framework, the probability that -alanine is better than a placebo can be estimated (i.e., 97.8%) and that there is a 95% probability that the true population mean is between 0.30 to 15.64 s can also be calculated.

While this study is only one example of how prior distributions can impact on the statistical model in a Bayesian context, it shows researchers and Sports Scientists the process of selecting informative priors by generating prior predictive distributions and critiquing the suitability of those priors given the knowledge of the data being collected. Narrowing the standard deviation of the prior distribution can adjust the prior predictive distribution to encapsulate the valid range of values seen in the population. Similarly, by shifting the mean according to prior information (in this case, a meta-analysis), smaller effects can be identified by utilising that prior information. One potential risk of informative priors is that, if not specified appropriately, it may conflict with the information provided in the data (Evans & Moshonov, 2006). This potential prior-data conflict can be examined by comparing the prior and posterior distributions and verifying that the posterior distribution does not lie in the tail of the prior. Therefore, it is recommended to visualise both the prior and posterior distributions. There is also scope for Bayesian models to address similar issues where coaches may only prescribe training for one athlete. Given athletes are being observed on a daily basis (Thornton, Delaney, et al., 2019), utilising previous daily measurements may allow the for the assessment of individual changes in response to a particular intervention.

The present study has demonstrated that Sports Scientists could adopt Bayesian models in research to apply appropriately sensitive statistical methodologies when measuring small effect sizes and small samples. The use of prior predictive distributions allows the researcher to visualise the constraints each prior has on the model to ensure it reflects information that may be available from domain experts or previous studies. It should also be noted, that just because a more-informative prior is used, a significant effect may not be present in the true population and therefore assessment that the informative priors are sound is required. Consequently, future studies being published using a Bayesian framework should outline the priors used for model components as well as the prior predictive distribution generated from the prior distributions. In doing so, readers can understand the sample space prior to the model and provides greater context for the posterior distribution to ensure transparency in the results being produced.

5 NRLW Movement Patterns and Match Statistics

5.1 Abstract

As women's rugby league grows, the need for understanding the movement patterns of the sport is essential for coaches and Sports Scientists. The aims of the present study were to quantify the position-specific demographics, technical match statistics, and movement patterns of the National Rugby League Women's (NRLW) Premiership and to identify whether there was a change in the intensity of play as a function of game time played. A retrospective observational study was conducted utilising global positioning system, demographic, and match statistics collected from 117 players from all NRLW clubs across the full 2018 and 2019 seasons and were compared between the ten positions using generalised linear mixed models. The GPS data were separated into absolute (i.e., total distance, high-speed running distance, and acceleration load) and relative movement patterns (i.e., mean speed, mean high speed ($> 12\ \text{km} \cdot \text{h}^{-1}$), and mean acceleration). For absolute external outputs, fullbacks covered the greatest distance (5504 m),

greatest high-speed distance (1081 m), and most ball-carry metres (97 m), while five-eighths recorded the greatest acceleration load (1697 m · s^{-2}). For relative external outputs, there were no significant differences in mean speed and mean high speed between positions, while mean acceleration only significantly differed between wingers and interchanges. Only interchange players significantly decreased in mean speed as their number of minutes played increased. By understanding the load of NRLW matches, coaches, high-performance staff, and players can better prepare as the NRLW Premiership expands. These movement patterns and match statistics of NRLW matches can lay the foundation for future research as women's rugby league expands. Similarly, coaches, high-performance staff, and players can also refine conditioning practices with a greater understanding of the external output of NRLW players.

The following chapter is a copy of the published manuscript:

Newans, T., Bellinger, P., Buxton, S., Quinn, K., & Minahan, C. Movement patterns and match statistics in the National Rugby League Women's (NRLW) Premiership. Frontiers in Sports and Active Living. 3:618913.

As co-author of the paper "Movement patterns and match statistics in the National Rugby League Women's (NRLW) Premiership", I confirm that Timothy Newans has made the following contributions:

- Study concept and design
- Data collection
- Data analysis and interpretation
- Manuscript preparation

Name: Clare Minahan

Date: 29/03/2023

5.2 Introduction

The National Rugby League Women's (NRLW) Premiership is the highest level of domestic rugby league competition for women in Australia. The NRLW was introduced in 2018 involving a three-round competition between four teams, culminating in a grand final between the two highest-ranked teams. The NRLW follows the international rugby league rules set by the Rugby League International Federation with the exception that NRLW matches are 60 min comprised of two 30-min halves, half time is 15 min, allow ten interchanges in each match, and observe a 40/30 kick advantage.

Describing the movement patterns of a given sport lays the foundations for future research to examine the intricacies of match play. While female soccer movement patterns have been established as seen in a recent review (Griffin et al., 2020), Clarke and colleagues have articulated the movement patterns in both Australian-rules football (Clarke et al., 2019; Clarke et al., 2018) and rugby sevens (Clarke et al., 2017, 2015). Similarly, a review of movement patterns in field-based sports spanning maximal speed, high-speed thresholds and movement patterns has been conducted in women (Hodun et al., 2016); however, the movement patterns of players competing in elite-level women's rugby league has been rarely explored in the scientific literature (Quinn et al., 2020). With an 18% year-on-year growth in women's rugby league participation (National Rugby League, 2020), the NRLW Premiership is set to expand in the number of teams over the coming years. To ensure that the level of competition does not regress as a function of including more teams, there is a need for emerging players to be sufficiently trained and conditioned to compete at the elite level, such as those described in the present study.

The movements of players in the Australian women's rugby league team during international match-play have previously been recorded using global positioning system (GPS) technology (Quinn et al., 2020). The findings of this initial study reported that players covered 6500 m in total distance, with ~500 m covered above 15 km · h^{-1}. This study also displayed the relative change in the intensity of match play across each half, whereby mean high speed (>12 km · h^{-1}) was reduced by approximately 40% from the start to the end of the first half (Quinn et al., 2020). While backs were shown to exhibit greater absolute distance covered, there were no significant differences in the relative distances between forwards and backs. In female soccer, it was shown that domestic movement patterns were significantly lower than in international-level matches (Andersson et al., 2010); however, it remains to be seen whether this is present in rugby league. As there is a disparity in match duration from the NRLW Premiership to international matches and as there is varying quality of opposition in women's international rugby league (Quinn et al., 2020) which can affect movement patterns (Hulin et al., 2015), the proportional change in movement patterns from domestic to international rugby league may not imitate the change seen in soccer (Andersson et al., 2010).

There is no doubt that the study by Quinn et al. (2020) provides important information regarding the movement patterns of elite, international rugby league players. However, a common critique in the literature when describing player movement patterns is that studies typically only gather data from one or two teams (Glassbrook et al., 2019). By gathering data from only a few teams, team tactics and strategy could obscure the true movement patterns inherent across the entire competition (Glassbrook et al., 2019). As the only previous investigation in women's rugby league was in international rugby league and focused only on a single team (Quinn et al., 2020), this study will examine the GPS data from the four domestic NRLW teams during the 2018 and 2019 seasons to describe the movement patterns of elite, domestic women's rugby league players. Similarly, in studies with only one or two teams, there is usually only enough power to segregate players into three positional groups; however, there have now been studies in men's rugby league that have segregated players into more specific positional groups (Delaney, Duthie, et al., 2016; Delaney et al., 2015). Therefore, in the present study, we expand the comparisons to further quantify the effect of playing position on the match statistics (tackles, runs, etc.) and GPS metrics (distance, velocity, acceleration).

5.3 Methods

The present study included 117 players from the four NRLW clubs (age = 26.8 ± 5.4 yr; height = 1.68 ± 0.07 m; body mass = 76.7 ± 11.9 kg). The Griffith University Human Ethics Committee approved this study (GU Ref No: 2019/359). There were 475 match entries as one player did not enter the field during one match. The mean number of matches played by each player was 4.1 ± 2.2 matches.

All match statistics and demographics (i.e., age, height, and body mass) were publicly sourced from NRL.com. The definitions of match statistics were determined by STATS, the National Rugby League's statistics provider. 'All runs' were defined as any time the ball carrier went into contact with a defender, 'all run metres' was the cumulative distance of all runs, 'tackles' were when the defender successfully executed a tackle, 'missed tackles' were when the defensive player could not bring the attacking player to the ground or successfully complete the tackle, and 'tackle breaks' were when the attacking player was able to continue running after a missed tackle. These five statistics were chosen by the authors to reflect the contact nature of rugby league and, therefore, to highlight the relative position differences in these events.

The movement patterns of all players during NRLW matches were collected using 10 Hz Optimeye S5 GPS units (Catapult Sports, Victoria, Australia). The GPS data were routinely collected at each match by the sport scientists at each club, while the National Rugby League oversaw the collection, amalgamation and provision of the data sets to the authors in '.cpf' form. Of the 475 match entries that had match statistics recorded, 370 were available for GPS analysis. This discrepancy was due to several reasons, with poor satellite coverage within the stadium being the main reason for why files were not provided to the authors for analysis. Regardless of whether an athlete's GPS data was available, the match statistics were still included in the analysis. Match files were segmented into halves within the proprietary software and interchanges were recorded in the software as per the match footage and GPS tracing. Velocity zones (V1-6) were aligned with those previously reported in women's rugby league (Quinn et al., 2020), with V1 set at 0 - 6 km h^{-1}, V2 set at 6.01 - 9 km h^{-1}, V3 set at 9.01 - 12 km h^{-1}, V4 set at 12.01 - 15 km h^{-1}, V5 set at 15.01 - 18 km h^{-1}, and V6 set at greater than 18 km h^{-1}. Acceleration load was calculated as the summation of absolute acceleration and deceleration values across the duration of the match. Three absolute output metrics (total distance, distance greater than 12 km h-1 (i.e., high-speed running; HSR), and acceleration load (Delaney, Duthie, et al., 2016)) were divided by the total time spent on the field to calculate the mean speed (MS), mean speed when travelling >12 km $\cdot$ h^{-1} (MS_{12}), and mean acceleration (Delaney, Duthie, et al., 2016) respectively to represent the relative outputs.

The term 'absolute' was used to describe the sum of the metric throughout a match, while the term 'relative' was used when dividing the absolute metric by time. 12 km $\cdot$ h^{-1} was used as the threshold for high-speed running to draw comparisons to the previous women's rugby league literature (Quinn et al., 2020).

To assess the positional differences for each tactical match statistic, GPS data metric, and the age of each position, GLMM's were employed. All GLMM's were built using the *lme4* package (Bates et al., 2015) in R version 3.5.2 (R Core Team, 2019), the *afex* package (Singmann et al., 2020) was used to determine significance at = 0.05 for all analyses, the *emmeans* package (Lenth, 2020) was used for pairwise comparisons, while the *sjPlot* package (Lüdecke, 2020) was used for model diagnostics. The data set was arranged in 'long form' with each observation for each player on a new row. After loading in the *lme4* and *afex* packages, the first model was built.

```
library(tidyverse)

library(afex)

library(effects)

library(emmeans)

library(chron)

library(sjPlot)

NRLW_match <- read_csv("www/data/Study_4_NRLW_match.csv")

options(scipen = 999)
```

```r
NRLW_match$FieldTime <- chron(times = as.character(NRLW_match$FieldTime))

NRLW_match$MeanSpeed <- NRLW_match$TD / NRLW_match$FieldTime / 24 / 60

NRLW_match$MeanHSR <- NRLW_match$HSR / NRLW_match$FieldTime / 24 / 60

NRLW_match$MeanAcc <- NRLW_match$Acc_Load / NRLW_match$FieldTime / 24 / 60
```

For each model, the dependent variable was the metric we were interested in explaining ('age', 'tackles made', 'total distance', etc.), a fixed effect of position was inserted, with random effects of 'player', 'match', and 'team' inserted as well to remove the variability attributed to these variables. Finally, the shape of the distribution was identified for each model with most models following a Gaussian distribution except those with count data which followed a Poisson distribution.

```r
demo_vars <- c("Age","Height","Mass","Game_Minutes")

gps_vars <- c("TD","HSR","Acc_Load","MeanSpeed","MeanHSR","MeanAcc")

tactical_vars <- c("All_Runs","All_Run_Metres","Tackles","Missed_Tackles","Tacklebreaks")

full_model <- glmer("Variable" ~ Position + (1| Team) + (1 | Match_ID) + (1 | Player_ID),
```

Once the full model (i.e., all fixed and random effects included) was established for each dependent variable, each random effect was consecutively removed to identify which random effects were required in the model. If the Bayesian Information Criterion (BIC) was lower after removing a random effect and an analysis of variance between the full and the reduced model was significant, it was deemed a significant contributor to the model, otherwise it would be subsequently dropped from that model. Once only random effects that significantly contributed to the model were remaining, the reduced model was deemed the most parsimonious.

```
reduced_model <- glmer("Variable" ~ Position +  (1 | Match_ID) + (1 | Player_ID), data = N
anova(full_model, reduced_model) # Determine if model is significantly affected by random
```

In all models, the match ID was deemed a significant random effect, while the team that each player competed for was not deemed a significant random effect in any of the models. Once the reduced model was identified, the null model was established (i.e., the reduced model without the predictor included. If the calculated X^2 statistic from an analysis of variance between the reduced model and the null model was significant ($p < 0.05$), then the predictor variable was deemed to significantly contribute to the model.

```
null_model <- glmer("Variable" ~ (1 | Match_ID) + (1 | Player_ID), data = NRLW_match, fami
anova(reduced_model, null_model) # Determine if model is significantly affected by positio
Position_Estimates <- as.data.frame(effect("Position", reduced_model)) # Create parameter
Pairwise_Positions <- as.data.frame(pairs(emmeans(full_model, "Position"))) # Create pairw
```

To check the assumptions, the reduced model's residuals were plotted against its fitted value to determine if there were any patterns emerging (Harrison et al., 2018).

```
plot_model(reduced_model, type = "diag") # Assess model diagnostics
```

GLMM's were also used to determine whether the duration of play affected the MS (i.e., players who play shorter duration perform at a higher intensity). Due to the nature of interchange in rugby league, some positions were more likely to play the full 60 min than other positions.

Consequently, 93% of backs (i.e., fullbacks, wingers, and centres), 80% of halves (i.e., five-eighths and halfbacks), 59% of second-rowers, 46% of hookers, 21% of locks, and 4% of props played the full duration of the match. If the regression coefficient calculated was significantly different ($p < 0.05$) from zero, it displayed a relationship between the amount of time played and the relative output of the position. For the positions where there was a significant effect of duration on MS, the slopes were reported alongside the MS.

```
full_model <- glmer("Variable" ~ Minutes + (1 | Match_ID) + (1 | Player), data = NRLW_matc
summary(full_model) # Inspect fixed and random effects
null_model <- glmer("Variable" ~ (1 | Match_ID) + (1 | Player), data = NRLW_match, family
anova(full_model, null_model) # Determine if model is significantly affected
plot_model(full_model, type = "diag") # Assess model diagnostics
```

Finally, GLMM's were used to determine the change in output in different velocity bands across the two 30-min halves. Like the positional analysis, a full model was developed, with non-significant random effects removed in the reduced model. An analysis of variance was performed between the reduced model and the null model (i.e., the reduced model with 'half' excluded) to calculate a X^2 statistic to determine whether there was a significant difference ($p < 0.05$) in external output measures between the two halves.

```
NRLW_match_Halves <- read_csv("www/data/Study_4_nrlw_half.csv") # Load in half-by-half dat


NRLW_match_Halves <- NRLW_match_Halves %>%
```

```
  mutate(
    AveSpeed = TotalDistance / (FieldTime * 1440),
    V1Speed = V1Distance / (FieldTime * 1440),
    V2Speed = V2Distance / (FieldTime * 1440),
    V3Speed = V3Distance / (FieldTime * 1440),
    V4Speed = V4Distance / (FieldTime * 1440),
    V5Speed = V5Distance / (FieldTime * 1440),
    V6Speed = V6Distance / (FieldTime * 1440),
    MeanHSR = HSR / (FieldTime * 1440),
    MeanAccel = Acc_Load / (FieldTime * 86400)
  ) # Create new speed metrics

full_model <- glmer("Variable" ~ Half + (1 | Team) + (1 | Match_ID) + (1 | Player_ID), dat
summary(full_model) # Inspect fixed and random effects
reduced_model <- glmer("Variable" ~ Half + (1 | Match_ID) + (1 | Player_ID), data = NRLW_m
anova(full_model, reduced_model) # Determine if model is significantly affected
reduced_model <- glmer("Variable" ~ (1 | Match_ID) + (1 | Player_ID), data = NRLW_match_Ha
anova(reduced_model, null_model) # Determine if model is significantly affected
plot_model(reduced_model, type = "diag") # Assess model diagnostics
Half_Estimates <- as.data.frame(effect("Half", full_model)) # Create parameter estimates b
```

5.4 Results

The demographics and minutes played across each playing position are displayed in Table 5.1. Second-row players were the tallest, while locks and hookers were the shortest. Props were considerably heavier than all other positions, with fullbacks and hookers the lightest of the positions. There was only one significant difference in age between the positions (X^2 (9) = 17.59, p = 0.040), with hookers significantly older than fullbacks.

Table 5.1: Demographics and technical data of NRLW players by position

Position	Unique Players (n)	Match Entries (n)	Height (cm)	Body mass (kg)	Age (yr)	Game Time (min)	All Run Metres (m)	All Runs (n)	Tackles (n)	Missed Tackles (n)	Tackle Breaks (n)
Fullback	8	28	168.4 ± 4.2	66.8 ± 5.2	26.1 (24.9 - 27.3)	60 (60 - 60)	97.2 (82.0 - 112.4)	10.3 (8.5 - 12.5)	4.8 (1.9 - 7.7)	1.3 (0.9 - 2.0)	3.5 (2.4 - 4.7)
Winger	19	56	168.0 ± 5.5	69.4 ± 5.0	26.6 (25.4 - 27.7)	60 (60 - 60)	66.4 [a] (55.9 - 76.9)	7.5 (6.5 - 8.8)	4.1 (2.1 - 6.1)	1.3 (0.9 - 1.7)	3.1 (2.2 - 3.9)

Position	Unique Players (n)	Match Entries (n)	Height (cm)	Body mass (kg)	Age (yr)	Game Time (min)	All Run Metres (m)	All Runs (n)	Tackles (n)	Missed Tackles (n)	Tackle Breaks (n)
Centre	17	56	168.3 ± 6.9	71.4 ± 4.4	26.2 (25.1 - 27.3)	60 (60 - 60)	82.9 (72.1 - 93.7)	9.4 (8.2 - 10.8)	9.4 [b] (7.3 - 11.5)	2.0 (1.6 - 2.6)	3.0 (2.2 - 3.8)
Five-Eighth	10	28	168.1 ± 6.4	70.9 ± 7.8	27.2 (26.2 - 28.3)	60 (60 - 60)	38.6 [a,c] (24.4 - 52.8)	5.2 [a,c] (4.1 – 6.6)	14.3 [a,b] (11.6 - 16.9)	2.5 (1.8 - 3.5)	0.9 [a] (-0.2 – 2.0)
Halfback	8	28	168.7 ± 7.1	70.8 ± 11.4	26.6 (25.5 - 27.7)	60 (59 - 60)	44.8 [a,c] (29.9 - 59.8)	6.3 [a] (5.1 – 7.8)	12.5 [a,b] (9.6 - 15.3)	3.0 [b] (2.2 - 4.0)	1.5 (0.4 - 2.5)
Hooker	6	28	164.5 ± 6.5	67.5 ± 7.0	27.6 (26.5 - 28.7)	55 (41 - 60)	42.9 [a,c] (27.3 - 58.4)	5.3 [a,c] (4.1 - 6.9)	26.4 [a,b,c,d,e] (23.5 - 29.4)	3.4 [a,b] (2.5 - 4.8)	1.1 (-0.2 - 2.4)

Position	Unique Players (n)	Match Entries (n)	Height (cm)	Body mass (kg)	Age (yr)	Game Time (min)	All Run Metres (m)	All Runs (n)	Tackles (n)	Missed Tackles (n)	Tackle Breaks (n)
Prop	23	56	171.1 ± 6.2	90.6 ± 8.5	27.0 (26.0 – 28.0)	35 (30 - 39)	69.8 [a,d] (59.9 - 79.7)	7.4 (6.4 - 8.6)	16.1 [a,b,c,f] (14.3 - 17.9)	1.6 [f] (1.2 - 2.1)	1.1 [a,b,c] (0.4 – 1.9)
Second Row	15	56	171.6 ± 6.2	81.5 ± 6.8	26.8 (25.8 - 27.9)	60 (53 - 60)	65.1 [a] (54.6 - 75.6)	8.0 (6.9 – 9.3)	17.8 [a,b,c,f] (15.8 - 19.8)	2.6 [b] (2.0 - 3.2)	2.6 (1.8 – 3.3)
Lock	9	28	164.4 ± 8.2	75.4 ± 4.8	27.0 (26.0 - 28.1)	44 (37 - 55)	53.2 [a,c] (39.0 - 67.5)	7.2 (5.9 - 8.9)	22.9 [a,b,c,d,e,g,h] (20.3 - 25.5)	3.2 [a,b,g] (2.4 - 4.4)	2.3 (1.2 – 3.4)
Interchange	48	111	167.5 ± 7.1	80.2 ± 13.4	26.9 (25.9 – 28.0)	24 (18 - 33)	47.7 [a,c,g] (40.7 - 54.6)	5.5 [a,b,c,g,h] (4.9 - 6.2)	11.9 [a,b,f,g,h,i] (10.6 - 13.2)	1.6 [e,f,i] (1.3 - 2.0)	1.7 (1.2 – 2.3)

N.B. All values are displayed as mean (95% CI) except height and body mass are displayed as mean ± SD and game time which is median (IQR), superscript indicates significantly different from [a]fullback, [b]winger, [c]centre, [d]five-eighth, [e]halfback, [f]hooker, [g]prop, [h]second-row, [i]lock ($p < 0.05$).

Table 5.1 also displays the tactical match statistics for each playing position. Most of the backs, halves, and second-row players completed the full 60 min of match play. Fullback players recorded the most runs (X^2 (9) = 50.64,< 0.001), run metres (X^2 (9) = 49.96, $p < 0.001$), and tackle breaks (X^2 (9) = 32.81, $p < 0.001$), while hookers recorded the most tackles (X^2 (9) = 153.02, $p < 0.001$).

Table 5.2: External workload in the NRLW by player position

Position	Mean Speed by Minutes	Mean Speed ($m \cdot min^{-1}$)	Mean High Speed ($m \cdot min^{-1}$)	Total Distance (m)	High-Speed Running (>12 $km\ h^{-1}$) (m)	Acceleration Load ($m \cdot s^{-2}$)	Mean Acceleration ($m \cdot s^{-3}$)
Fullback	-	80.1 (75.4 - 84.8)	15.3 (12.1 - 18.6)	5504.3 (5008.4 - 6000.3)	1080.5 (915.4 - 1245.6)	1529.9 (1373.8 - 1686.1)	0.37 (0.34 - 0.40)
Winger	-	75.5 (71.8 - 79.3)	13.1 (10.4 - 15.8)	5133.5 (4763.1 - 5503.8)	916.1 (780.3 - 1052.0)	1548.9 (1433.9 - 1663.8)	0.38 (0.35 - 0.40)

Position	Mean Speed by Minutes	Mean Speed (m · min-1)	Mean High Speed (m · min-1)	Total Distance (m)	High-Speed Running (>12 km h-1) (m)	Acceleration Load (m · s-2)	Mean Acceleration (m · s-3)
Centre	-	75.9 (72.3 - 79.6)	13.5 (10.9 - 16.2)	5116.4 (4759.9 - 5472.9)	882.3 (749.0 - 1015.6)	1582.2 (1471.5 - 1692.8)	0.39 (0.37 - 0.42)
Five-Eighth	-	79.8 (75.3 - 84.3)	13.3 (10.3 - 16.3)	5243.8 (4754.4 - 5733.3)	883.3 (725.1 - 1041.5)	1697.1 (1544.5 - 1849.8)	0.42 (0.39 - 0.45)
Halfback	-	80.5 (76.1 - 84.9)	14.3 (11.4 - 17.2)	5212.0 (4741.2 - 5682.7)	900.0 (745.0 - 1055.1)	1593.0 (1445.8 - 1740.1)	0.41 (0.39 - 0.44)
Hooker	-0.18 (-0.41 - 0.12)	82.7 (77.5 - 87.9)	15.2 (11.7 - 18.7)	4844.5 (4289 - 5400)	817.4 (636.3 - 998.5)	1472.1 (1296.2 - 1648.1)	0.42 (0.39 - 0.45)
Prop	-0.11 (-0.37 - 0.16)	78.6 (75.1 - 82.1)	13.4 (11.0 - 15.8)	2908.1 [a,b,c,d,e,f] (2557.4 - 3258.9)	431.9 [a,b,c,d,e,f] (304.9 - 558.8)	886.7 [a,b,c,d,e,f] (779.4 - 994.0)	0.41 (0.39 - 0.43)

Position	Mean Speed by Minutes	Mean Speed ($m \cdot min^{-1}$)	Mean High Speed ($m \cdot min^{-1}$)	Total Distance (m)	High-Speed Running (>12 km h^{-1}) (m)	Acceleration Load ($m \cdot s^{-2}$)	Mean Acceleration ($m \cdot s^{-3}$)
Second-Row	-0.12 (-0.31 - 0.06)	78.0 (74.3 - 81.6)	12.7 (10.2 - 15.1)	4627.2 ^{g} (4270.6 - 4983.8)	692.6 ^{a,b,g} (562.6 - 822.6)	1361.7 ^{d,g} (1251.8 - 1471.6)	0.39 (0.37 - 0.41)
Lock	-0.22 (-0.53 - 0.16)	78.4 (74.0 - 82.9)	13.9 (11.1 - 16.8)	4044.0 ^{a,b,c,d,e,g} (3562.1 - 4525.9)	697.3 ^{a,g} (542.7 - 851.8)	1227.9 ^{b,c,d,e,g} (1078.1 - 1377.7)	0.40 (0.38 - 0.43)
Interchange	-0.41 * (-0.57 - -0.26)	80.1 (77.0 - 83.1)	14.5 (12.4 - 16.7)	2514.2 ^{a,b,c,d,e,f,h,i} (2246.2 - 2782.3)	438.9 ^{a,b,c,d,e,f,h,i} (326.5 - 551.3)	775.3 ^{a,b,c,d,e,f,h,i} (695.2 - 855.4)	0.41 ^{b} (0.40 - 0.43)

N.B. All figures are presented as mean (95% CI), - indicates insufficient data to determine regression coefficient, * indicates sig. different from zero, superscript indicates significantly different from ^{a}fullback, ^{b}winger, ^{c}centre, ^{d}five-eighth, ^{e}halfback, ^{f}hooker, ^{g}prop, ^{h}second-row, ^{i}lock ($p < 0.05$).

As most fullback, wing, centre, five-eighth, and halfback players completed the full 60 minutes, it was not necessary to determine the coefficient for the relationship between MS and minutes played. This relationship was non-significant for hooker, prop, second-row, and lock players; however, it was significant for interchange players (X^2 (1) = 24.41, $p < 0.001$) and therefore needed to be accounted for when calculating MS. For relative outputs, there were no significant differences in MS (X^2 (9) = 14.61, $p = 0.102$) and MS_{12} (X^2 (9) = 8.04, $p = 0.530$), between any of the positions, while mean acceleration (X^2 (9) = 21.26, $p = 0.012$) was significantly different between wingers and interchange. For absolute output, total distance (X^2 (9) = 195.41, $p < 0.001$), HSR (X^2 (9) = 115.16, $p < 0.001$), and acceleration load (X^2 (9) = 183.14, $p < 0.001$) were significantly different between positions (Table 5.2). When comparing the first and second halves, there were no significant differences in the relative distances covered in any of the speed zones, as well as the overall MS. The parameter estimates can be seen in Table 5.3.

Table 5.3: Half-by-half analysis of external outputs obtained from during domestic women's rugby league matches

Metric	1st Half	2nd Half
Relative Total Distance ($m \cdot min^{-1}$)	79.6 (76.8 - 82.5)	79.3 (76.5 - 82.2)
Relative V1 Distance ($m \cdot min^{-1}$)	36.7 (33.0 - 40.4)	36.9 (33.2 - 40.7)
Relative V2 Distance ($m \cdot min^{-1}$)	15.1 (14.4 - 15.8)	15.0 (14.2 - 15.7)
Relative V3 Distance ($m \cdot min^{-1}$)	14.5 (12.8 - 16.1)	14.1 (12.5 - 15.8)
Relative V4 Distance ($m \cdot min^{-1}$)	7.7 (4.3 - 14.1)	7.6 (4.2 - 13.8)
Relative V5 Distance ($m \cdot min^{-1}$)	2.9 (1.6 - 5.3)	2.9 (1.6 - 5.2)

Metric	1st Half	2nd Half
Relative V6 Distance ($m \cdot min^{-1}$)	2.0 (1.1 - 3.7)	1.8 (1.0 - 3.3)

N.B. Values presented as mean (95% CI). V1 = 0 – 6 $km \cdot h^{-1}$, V2 = 6.01 – 9 $km \cdot h^{-1}$, V3 = 9.01 – 12 $km \cdot h^{-1}$, V4 = 12.01 – 15 $km \cdot h^{-1}$, V5 = 15.01 – 18 $km \cdot h^{-1}$, and V6 = >18 $km \cdot h^{-1}$.
* indicates significantly different from the first half ($p < 0.05$).

5.5 Discussion

The present study describes the absolute and relative (to time) movement patterns, player demographics, and match statistics of the 2018-2019 NRLW Premiership in Australia. With respect to differing absolute movement patterns, backs covered between 5100 - 5500 m with between 850 - 1100 m above 12 $km \cdot h^{-1}$, five-eighths and halfbacks covered approximately 5200 m with 900 m above 12 $km \cdot h^{-1}$, while forwards covered between 2900 - 4900 m with between 430 - 820 m above 12 $km \cdot h^{-1}$. Similar patterns were reflected in acceleration load where there were no significant differences between any back or halves position, while props were significantly lower than all other starting positions. However, when comparing movement patterns expressed relative to time, there were no significant differences in the relative distance metrics (MS and MS_{12}) between any of the positions. The only significant difference in relative movement patterns was mean acceleration between wingers and interchange players. While the absolute movement patterns display a disparity in the typical match profile across positions of elite women's rugby league, the lack of difference in movement patterns relative to time means when designing training and conditioning protocols the intensity of play is matched across all positions and varying levels of total work is required between positions.

When assessing whether the MS changed as a function of the time on field, there were also no significant relationships for any of the starting positions, with a decrease of 0.4 $m \cdot min^{-1}$ for each minute played in the interchange players. While this decrease may seem insignificant practically, when comparing an interchange player competing for 10 min compared to 35 min, it equates to a decrease of 10 $m \cdot min^{-1}$, which is approximately 13% decrease in MS. These results reflect a similar pattern seen in the men's game where transient fatigue was attributed to the decline in intensity as an interchange player's bout was prolonged (Waldron et al., 2013). These results also display a need to further understand the role and requirement of interchange players in women's rugby league, similar to those explored in the men's game (Delaney, Thornton, et al., 2016).

When comparing relative movement patterns between domestic (i.e., the present study) and international women's rugby league (Quinn et al., 2020), NRLW backs covered between 75.5 - 80.1 $m \cdot min^{-1}$ compared to approximately 75 $m \cdot min^{-1}$ in international matches, NRLW halves covered between 79.8 - 80.5 $m \cdot min^{-1}$ compared to approximately 78 $m \cdot min^{-1}$ in international matches, and NRLW forwards covered between 78.0 - 82.7 $m \cdot min^{-1}$ compared to approximately 71 $m \cdot min^{-1}$ in international matches. This contradicts figures seen in other codes of football played by women, with female soccer players recording a higher intensity when playing at the international level compared to at the domestic level (Andersson et al., 2010). These comparisons show that while the backs and halves roughly display similar MS, it is evident that forwards in the shorter-duration NRLW may maintain their output better than in the longer-duration international level. Conversely, it could be that the international players did not require as high output as they were of a higher quality than their opponents (Hulin et al., 2015).

While the values for MS and MS_{12} for the full match were comparable to Quinn et al. (2020) (who also found no significant differences in the 80-min players between any of the positions), when comparing half-to-half (Table 5.3), Quinn et al. (2020) found a significant decline in certain velocity bands from the first to second half, where the present study showed no such difference. This could be due to the differing inclusion criteria where Quinn et al. (2020) only included those playing the full 80-min in the half-to-half comparison. The MS in the present study was still lower than those seen in other sports played by women (Hodun et al., 2016), which could be attributed to the energetic demand of the tackling nature of rugby league which is not present in sports such as soccer and hockey. Although, the MS was higher than that recorded in women's rugby union (Suarez-Arrones et al., 2014) which would feasibly be the most similar sport code, which reflects the difference in movement patterns between men's rugby league (Cummins et al., 2018) and rugby union (M. Jones et al., 2015). However, due to the low threshold for mean high speed (i.e., 12 $km \cdot h^{-1}$), it is difficult to compare high-intensity efforts to other sports (Hodun et al., 2016). While the mean high-speed threshold was chosen to make direct comparisons with international women's rugby league, further investigations could use higher intensity thresholds to further interrogate the differences in movement patterns with other sports.

Another noteworthy finding was the unique characteristics of the lock position. While locks within modern men's rugby league are most commonly paired with second-rowers (i.e., the back row) or with props (i.e., the 'middle forwards') (Glassbrook et al., 2019), the present study showed that locks in NRLW were different in their role. NRLW locks were, on average, 8 and 7 cm shorter than second-rowers and props respectively, weighed 6.1 and 15.2 kg less respectively, ran for 11.9 and 16.8 m less with the ball respectively, and made 5.1 and 6.8 more tackles respectively. Consequently, these findings reinforce the limitations in simply transferring male rugby league conditioning research into the female game (Emmonds et al., 2019).

The data from the present study is crucial as it is the first week-by-week examination of women's rugby league. Quinn et al. (2020) provided an overview of the movement patterns of the Australian Women's Rugby League team during international competition (Quinn et al., 2020); however, these movement patterns could be adversely affected by cumulative fatigue, as previously seen in male rugby league (R. Johnston et al., 2013). As the NRLW Premiership expands, it will conceivably shift from a semi-professional to a professional competition enabling players to commit full-time to their career in rugby league. This shift, with players unconstrained by holding secular jobs, may also see the movement patterns of NRLW match play increase above the values seen in this study.

Another notable feature of the present study was the dissemination of knowledge, skills, and awareness around the use of the mixed models for sports science practitioners. As noted in a recent call to Sports Scientists (Sainani et al., 2021), there is a growing need for Sports Scientists to be familiar with statistical methods that can properly account for variation within their data sets. While the requirement to account for repeated measures and missing data is pertinent in longitudinal GPS movement pattern studies, it is still lacking within a substantial proportion of the sports science community. For instance, in a recent systematic review of rugby league (Dalton-Barron et al., 2020), only seven of the 15 studies correctly accounted for the repeated observations within GPS data sets. Throughout this study, the authors have highlighted the importance of using appropriate statistical modelling for repeated measures of individuals and teams in longitudinal data sets. We explained the methodology and application of a GLMM as well as provide the relevant computer scripts required to perform a GLMM. We anticipate that this approach will increase the transparency and reproducibility of the findings, as well as encourage statistical literacy for Sports Scientists.

5.6 Conclusion

This study was the first study to describe the movement patterns and match statistics in domestic elite-level women's rugby league. Two seasons of the NRLW Premiership (13 matches) were analysed, with all clubs contributing to the data set. While backs and halves covered significantly greater absolute distances than forwards, the relative distances covered were not different across any of the positions. The lock position was also a substantially different position to men's rugby league. This study also provided more clarity on the application of mixed models for sports science data sets, with a specific focus on team-sport data with repeated observations.

6 Bayesian Approximation of the Pareto Frontier

6.1 Abstract

Elite team sports are challenged by the need to possess multiple, often conflicting, athlete characteristics (e.g., sprint ability and endurance) to maximise match running performance and tactics. Univariate analysis of athlete characteristics is commonplace in sports science whereby speed is typically evaluated independently of endurance. While this methodology readily identifies athletes excelling in either speed or endurance, it fails to highlight athletes who possess ‘the best compromise’ in speed and endurance. This study presents an innovative approach to evaluating running intensity during short (10 s; anaerobic power/speed) and long (20 min; aerobic power/endurance) periods during football matches by using the Pareto frontier to visualize athlete characteristics simultaneously. Importantly, this study was the first to quantify the uncertainty around the Pareto frontier, using samples drawn from the Bayesian posterior distribution of both rolling averages to calculate the probability an athlete sits on

the true population Pareto frontier. The 10-s and 20-min rolling averages were calculated (325.1 and 104.7 $m \cdot min^{-1}$, respectively) for 18 elite female footballers across three seasons using a Bayesian mixed model. No linear relationship between the two periods was identified when considering the entire sample (n = 18, r = 0.03). However, a moderate negative linear relationship was observed when only considering athletes that were on the first three (i.e., outer most) Pareto frontiers (n = 12, r = -0.49). The Pareto frontier provides coaches with a holistic view of their athletes' speed and endurance capabilities and enables a more informed decision when identifying the athletes with the optimal balance of these polarizing attributes.

The following chapter is a copy of a manuscript in preparation:

Newans, T., Bellinger, P., Drovandi, C., Griffin, J. & Minahan, C. Bayesian approximation of the trade-off relationship between running intensity measured during short and long periods using Pareto frontiers. In preparation.

As co-author of the paper "Bayesian approximation of the trade-off relationship between running intensity measured during short and long periods using Pareto frontiers", I confirm that Timothy Newans has made the following contributions:

- Study concept and design
- Data collection
- Data analysis and interpretation
- Manuscript preparation

Name: Clare Minahan

Date: 29/03/2023

6.2 Introduction

Many team-based ball sports (e.g., soccer, rugby, Australian football) comprise intermittent high-intensity locomotion (i.e., high-speed running and sprinting) interspersed among continuous low-intensity locomotion (Coutts et al., 2010; Deutsch et al., 2007; Newans et al., 2021;

Stølen et al., 2005). Defined 'speed zones' (km/h) have been previously used to examine the total number of efforts or the total distance covered by an athlete at various locomotive intensities across a match (Cummins et al., 2013; Dalton-Barron et al., 2020; Quinn et al., 2020). However, by simply aggregating the number of efforts or total distance performed within each speed zone, Sports Scientists were unable to identify the most intense period of a match, making it difficult to prescribe match-specific training interventions (Mohr et al., 2003). In recent years, new analysis techniques, such as the rolling-average method (Delaney et al., 2015; Varley et al., 2012), have been used to quantify the locomotive intensity of athletes more precisely during match play expressing intensity as distance covered per minute (i.e., m/min (Cunningham et al., 2018)). Indeed, rolling-average analyses can quantify athletes' locomotive intensity over both short (e.g., < 1 min) and long (e.g., > 10 min) intervals which provide sophisticated data about each athlete's unique running ability which in turn can assist sports scientists and conditioning coaches in prescribing individualized match-specific training interventions (Delaney et al., 2015; Fereday et al., 2020). While new metrics have been devised to calculate the metabolic costs of both these tasks (Brown et al., 2016; Buchheit et al., 2015), little consideration has been given to which task is the larger contributor toward these new metrics for athletes with differing movement patterns. Despite the documented usefulness of independently quantifying duration-specific running intensities in athletes during match play (Delaney et al., 2015; Fereday et al., 2020), there may be a trade-off or compromise between the running performance during short and long periods that remains unexplored.

It is true that athletes who possess a high maximal anaerobic power (i.e., speed) are unlikely to also possess a high aerobic power (i.e., endurance) (McKay et al., 2021) when compared with other elite athletes. Therefore, it is reasonable to assume that elite team-sport athletes who can maintain a relatively high intensity during short locomotive efforts (e.g., <1 min) will be unable to maintain a relatively high intensity over long locomotive efforts (e.g., >10 min), and vice versa. While univariate statistics can easily identify athletes at both extremes (i.e., those with high speed or high endurance profiles), it is more difficult to ascertain athletes who are uniquely proficient at both, possessing the best compromise between speed and endurance. No previous statistical approach has been used in sports science to meaningfully identify the 'extreme all-rounder' of short- and long-duration running performance in elite team-sport athletes.

Presented with a similar quandary in T20 cricket, whereby the number of runs scored per dismissal (i.e., batting average) and the rate at which the athlete scores runs relative to the number of balls faced (i.e., batting strike rate) are negatively correlated (Newans, Bellinger, & Minahan, 2022). We utilised Pareto frontiers to: i. Athletes with the highest batting average, ii. Athletes with the highest strike rate, and iii. Athletes possessing the best compromise between batting average and strike rate (Newans, Bellinger, & Minahan, 2022). Indeed, Pareto frontiers are a relatively new statistical approach to analyzing sports science data that may assist in the identification of athletes who are not necessarily specialists in one metric but possess the best compromise on two or more opposing abilities.

However, when dealing with longitudinal data sets such as locomotive movement patterns, the uncertainty around an athlete's movement patterns when each athlete has multiple observations needs to be accounted for (Newans, Bellinger, Drovandi, et al., 2022). In this instance, mixed models are required to account for this longitudinal data set (Newans, Bellinger, Drovandi, et al., 2022) and the regression framework also allows the model to adjust for other variables in the data. Furthermore, when executing the mixed model in a Bayesian framework, the probability of an athlete being a member of the Pareto frontier for each athlete by resampling from the posterior distribution of the athletes' effects can be estimated. While Bayesian Pareto frontiers have previously been explored in portfolio construction in economics (Bauder et al., 2021), we do not believe that Bayesian Pareto frontiers have been constructed while accounting for the uncertainty due to the repeated-measures nature of the data.

Therefore, in the present study, we analyse running intensities in short- (i.e., 10 s) and long- (i.e., 20 min) efforts during match play to identify athletes at both extremes as well as athletes possessing the best compromise between running intensity over short and long durations. To accommodate multiple observations from each athlete, this study proposes to use a mixed model within a Bayesian framework to quantify the uncertainty around each athlete's locomotion. Specifically, Pareto frontiers generated from the posterior distribution of each running intensity were established to highlight the trade-off nature between the two intensities, having accounted for uncertainty through a Bayesian mixed model.

6.3 Methods

The locomotive activities of eighteen female football athletes competing in the Australian women's domestic football league were analysed. All athletes were from the same team and thirty-three matches across three seasons (2017-2019) were analysed. Only data where an athlete completed the full match (i.e., 90-min) was included in the study. A minimum of three matches was required for each athlete to ensure that the means were not too heavily impacted by individual match demands. Goalkeepers were excluded from the data set due to their vastly different locomotion patterns. Collectively, there were 154 unique match files in the data set. This study was approved by the Griffith University (GU) Human Ethics Committee and Football Federation Australia (GU Ref: 2019/359).

Locomotion patterns were measured using 10 Hz VX GPS units (Visuallex Sport International, Wellington, New Zealand) with the 10 Hz velocity data exported for each match file. The units in the present study have been reported to have acceptable accuracy and both between- and within-manufacturer reliability for quantifying locomotion patterns during team sport (Delaney et al., 2019; Varley et al., 2012).

To summarise the locomotion patterns of each athlete, a customised R-script was used to calculate the rolling averages. To define the high-intensity locomotion, a 10-second rolling average was iterated over each match file to identify the peak short-duration running intensity period, while a 20-minute rolling average was used to define the peak long-duration running intensity period. Given the inherent variability within athletes and given the imbalanced nature of the data set (range: 1-20 matches per athlete), mixed models were used to address this issue (Newans, Bellinger, Drovandi, et al., 2022).

```r
library(tidyverse)

library(brms)

library(tidybayes)

library(rPref)


bayespareto <- read_csv("www/data/Study_5_bayesian_pareto_mm.csv") # Read in data


bayespareto <- bayespareto %>%

  mutate(R10s = R10s*6,

         R20m = R20m/20) # Convert distances to intensity (metres per minute)
```

The Pareto frontier that was to be generated was to identify the athletes in which no other athlete is higher in both the 10-second (short duration) and 20-minute (long duration) rolling averages. For the Pareto frontier to be generated, the means in both the running intensities for each athlete were calculated. Consequently, a mixed model within a Bayesian framework was employed to estimate the posterior distribution for each athlete, having accounted for the uncertainty of each athlete's mean based on the number of observations. As the rolling 10-second and 20-minute periods are inherently related, a bivariate model was implemented and defined by the following equation:

$$\begin{bmatrix} S_{apm} \\ L_{apm} \end{bmatrix} = \begin{bmatrix} \beta_0^S \\ \beta_0^L \end{bmatrix} + \begin{bmatrix} \beta_a^S \\ \beta_a^L \end{bmatrix} + \begin{bmatrix} \beta_p^S \\ \beta_p^L \end{bmatrix} + \begin{bmatrix} \beta_m^S \\ \beta_m^L \end{bmatrix} + \begin{bmatrix} \epsilon_{apm}^S \\ \epsilon_{apm}^L \end{bmatrix}$$

where:

- S_{apm} is the response variable for the 10-second rolling average for the ath athlete, pth position, and mth match
- L_{apm} is the response variable for the 20-minute rolling average for the ath athlete, pth position, and mth match
- β_0^S is the intercept related to the 10-second rolling average response variable
- β_0^L is the intercept related to the 20-minute rolling average response variable
- β_a^S is the athlete fixed effect related to the 10-second rolling average response variable
- β_a^L is the athlete fixed effect related to the 20-minute rolling average response variable
- β_p^S is the position random effect related to the 10-second rolling average response variable
- β_p^L is the position random effect related to the 20-minute rolling average response variable
- β_m^S is the match random effect related to the 10-second rolling average response variable
- β_m^L is the match random effect related to the 20-minute rolling average response variable
- ϵ_{apm}^S is the residual related to the 10-second rolling average response variable
- ϵ_{apm}^L is the residual related to the 20-minute rolling average response variable

The random effects are assumed to be normally distributed, $\beta_p^S \sim N(0, \phi_p^S)$, $\beta_p^L \sim N(0, \phi_p^L)$, $\beta_m^S \sim N(0, \phi_m^S)$, $\beta_m^L \sim N(0, \phi_m^L)$ where ϕ_p^S, ϕ_p^L, ϕ_m^S, ϕ_m^L are variances to be estimated. The residuals are assumed to have a bivariate normal distribution:

$$\begin{bmatrix} \epsilon_{apm}^S \\ \epsilon_{apm}^L \end{bmatrix} \sim N(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma)$$

where Σ is a 2x2 covariance matrix which is estimated.

The mixed model allowed for the visualisation of the estimated conditional athlete effects for low and high intensity periods and utilised Pareto frontiers to identify the "extreme" athletes. As some athletes played in multiple positions across the three years, position was included as a random effect alongside the match ID to account for match-to-match variability in the team. A normal prior with a mean of 0 and SD of 30 was set for the regression coefficients related to both the rolling 10-second and the rolling 20-minute averages, which was derived from the standard deviations of the 'worst-case scenarios' identified in prior studies utilising rolling averages in football (Doncaster et al., 2020; Fereday et al., 2020). 'Worst-case scenario' analyses have recently come under scrutiny (Novak et al., 2021) to use as a prescriber for training programs as they do not consider the psycho-physiological responses to the given stimulus. However, there is still merit in utilising these analyses to better understand the locomotive requirements of game play comparative to whole-match or half totals or averages (Weaving et al., 2022).

```
univariate_model_R10s <- brm(R10s ~ (1|Player) + (1|Position) + (1|SeasonGame),
                             iter = 4000, chains = 8, data = bayespareto,
```

```r
                              seed = 99.94) # Build univariate model for rolling 10-second


univariate_model_R20m <- brm(R20m ~ (1|Player) + (1|Position) + (1|SeasonGame),
                              iter = 4000, chains = 8, data = bayespareto,
                              seed = 99.94) # Build univariate model for rolling 20-minute


multivariate_model <- brm(mvbind(R10s,R20m) ~ Player + (1|Position) + (1|SeasonGame),
                           iter = 8000, chains = 8, data = bayespareto,
                           seed = 99.94,
                           prior = c(set_prior("normal(0,30)", class = "b",resp = "R10s"),
                                     set_prior("normal(0,30)", class = "b",resp = "R20m")))
```

While the Pareto frontier for the sample could be established from the conditional means arising from the mixed model, the uncertainty around each athlete's mean because of match-to-match variability meant that other athletes could be on the 'true' population Pareto frontier as well. To solve this, the multivariate mixed model was implemented in a Bayesian framework, to allow samples to be drawn from the posterior predictive distribution of both the 10-second and 20-minute rolling averages, in which the Pareto frontier could be generated for each sample. To quantify the uncertainty of both the Pareto frontier, as well as the probabilities that an athlete was on the Pareto frontier, the following steps were used:

1. Draw a data set consisting of 10-second and 20-minute rolling averages by sampling

from the posterior predictive distribution arising from the marginal model of the fitted Bayesian bivariate mixed model.

```
posterior_mean_draws <- posterior_epred(multivariate_model,
                                        newdata = data.frame(Player = unique(bayespareto %
                                        re_formula = NA)
```

2. From the simulated data set generated in step 1, construct the Pareto fronts, specifically identifying the first, second, and third frontier.

```
combined_df <- posterior_mean_draws %>%
  reshape2::melt() %>%
  pivot_wider(names_from = Var3, values_from = value) %>%
  rename(Sim = Var1, Athlete = Var2)


combined_df_pareto <- psel(combined_df %>% group_by(Sim), high(R10s)*high(R20m), top_level
```

3. Repeat steps 1 and 2 4,000 times to generate 4,000 first, second and third Pareto frontiers.
4. Calculate the median and the central 50% interval of the posterior predictive distributions for the first, second and third Pareto frontiers for the 20-minute rolling average across the domain of possible rolling 10-second averages (i.e., 270, 271,…., 375 $m \cdot min^{-1}$).

```
frontiers <- combined_df_pareto  %>%
  ungroup() %>%
  group_by(Sim) %>%
  filter(.level <= 3) %>%
  ungroup()  %>%
  mutate(R10s = round(R10s / 3, 0) * 3) %>%
  filter(R10s >= 270 & R10s < 375) %>%
  group_by(R10s, .level) %>%
  point_interval(R20m, .width = c(.50)) %>%
  mutate(Frontier = case_when(.level == 1 ~ "First",
                              .level == 2 ~ "Second",
                              .level == 3 ~ "Third")) # Calculate the median and middl
```

5. Estimate the probability of each athlete being on the first, second, or third Pareto frontier by dividing the frequency that an athlete appeared on the corresponding Pareto frontier by the total number of posterior draws (i.e., 4,000 draws).

The mixed model allowed for the visualisation of the estimated athlete effects for low and high intensity periods and utilised Pareto frontiers to identify the "extreme" athletes. As some athletes played in multiple positions across the three years, position was included as a random effect alongside the match ID to account for match-to-match variability in the team. A normal prior with a mean of 0 and SD of 10 was set for the regression coefficients related to the rolling 10-second average and a normal prior with a mean of 0 and SD of 100 was set for the regression coefficients related to the rolling 20-minute average. Given the multivariate mixed model was implemented in a Bayesian framework, samples can be drawn from the posterior distribution in which the Pareto frontier can be generated for each sample. By summarising the results of all these posterior samples, the probability of each athlete being on the Pareto front can be estimated, thus quantifying the uncertainty.

All analysis was performed in R v4.1.1 (R Core Team, 2019). The *brms* package (an interface to Stan programming language) (Bürkner, 2017), was used to calculate the posterior distributions for each athlete, the *rPref* package (Roocks, 2016) was used to generate the Pareto frontiers, and the *tidyverse* suite of packages were used for all data manipulation, calculation of rolling-average blocks, and all visualisations.

6.4 Results

Using the conditional means generated from the multivariate mixed model, the average 10-second rolling mean speed was 325.1 $m \cdot min^{-1}$ (95% Credible Interval: 304.6 - 346.1 $m \cdot min^{-1}$) while the average 20-minute rolling mean speed was 104.7 (97.0 - 112.8) $m \cdot min^{-1}$. There was negligible correlation between the 10-second rolling mean speed and 20-minute rolling mean speed (r = 0.03). However, there was a moderate negative correlation between the rolling 10-second and 20-minute rolling averages for the athletes that were on the first three Pareto frontiers (r = -0.49), which are highlighted in Figure 6.1. Four athletes were identified on the first Pareto frontier, four athletes on the second frontier, and five athletes on the third frontier.

The three frontiers are visualised in Figure 6.1. Athletes G & L are on the Pareto frontier due to being the highest in the univariate analysis of 10-second rolling mean speed and 20-minute rolling mean speed respectively, athlete M is on the Pareto frontier as 2nd-highest for 20-minute rolling mean speed and 9th-highest for 10-second rolling mean speed, while athlete P is also on the Pareto frontier despite being 3rd-highest for 20-minute rolling mean speed and 6th-highest for 10-second rolling mean speed.

```
CondEffs <- conditional_effects(multivariate_model) # Extract conditional effects
R10sPlayer <- CondEffs$R10s.R10s_Player
R20mPlayer <- CondEffs$R20m.R20m_Player
df <- left_join(R10sPlayer,R20mPlayer, by = "Player", suffix = c("R10s","R20m")) # Bind co

conditional_pareto <- psel(df, high(estimate__R10s)*high(estimate__R20m), top_level=20) %>
  filter(.level <= 3) # Filter only players on the first three Pareto frontiers

ggplot(df, aes(x = estimate__R10s,y = estimate__R20m, label = Player))+
  geom_point(size = 2) +
  geom_line(data = conditional_pareto, aes(color = factor(.level))) +
  geom_point(data = conditional_pareto, aes(color = factor(.level)), size = 2) +
  scale_color_manual(values = c("darkred", "darkgreen", "darkblue")) +
  geom_errorbar(aes(ymin = estimate__R20m - se__R20m, ymax = estimate__R20m + se__R20m),
                alpha = 0.2) +
```

```r
geom_errorbarh(aes(xmin = estimate__R10s - se__R10s, xmax = estimate__R10s + se__R10s),
               alpha = 0.2) +
geom_text(hjust = 1.5, vjust = 1) +
theme_minimal() +
labs(x = "10-second running intensity (m/min)",
     y = "20-minute running intensity (m/min)") +
theme(
  legend.position = "none",
  axis.title = element_text(size = 10, color = "black", face = "bold"),
  axis.text = element_text(size = 10)
)
```

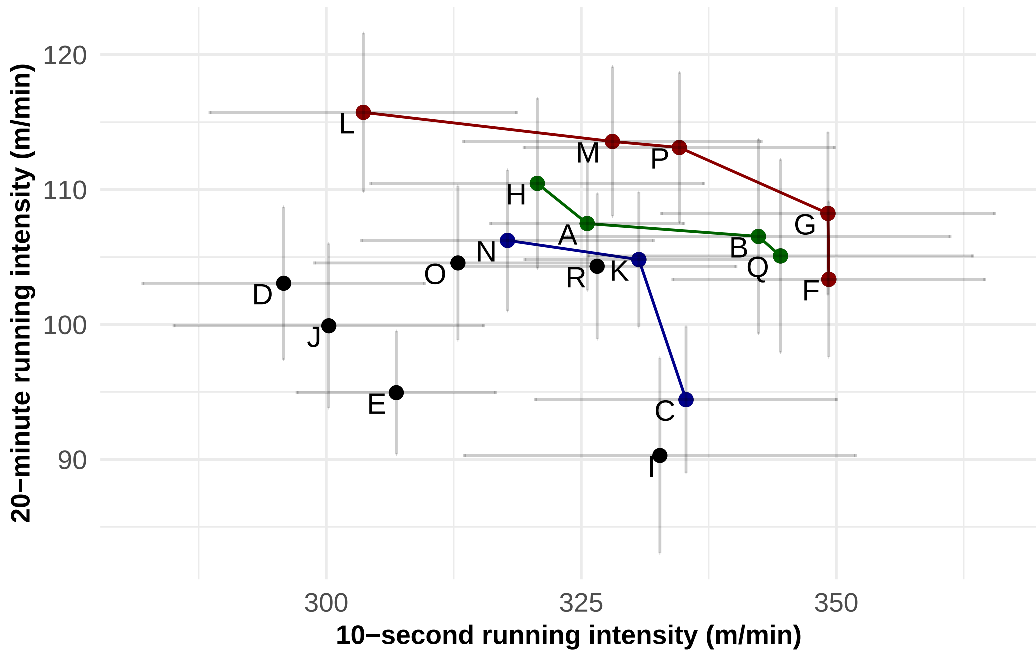

Figure 6.1: Conditional means of a 10-second rolling average and a 20-minute rolling average mean speed. N.B. Error bars represent the 95% credible interval.

From the multivariate mixed model, 4,000 simulated data sets were drawn from the posterior predictive distribution of the 10-second rolling average and 20-minute rolling average mean speed. Figure 6.2 illustrates the uncertainty around the first, second, and third Pareto frontiers based on these 4000 posterior predictive data sets.

```
ggplot(data = frontiers, aes(x = R10s, y = R20m, ymin = .lower, ymax = .upper, fill = Fron
  geom_ribbon(alpha = 0.2, color = NA) +
  geom_line(linewidth = 1, alpha = 0.8) +
  scale_color_manual(values = c("darkred", "darkgreen", "darkblue")) +
  scale_fill_manual(values = c("darkred", "darkgreen", "darkblue")) +
```

```
scale_x_continuous(breaks = seq(270, 370, by = 20)) +
theme_minimal() +
labs(x = "10-second running intensity (m/min)",
     y = "20-minute running intensity (m/min)") +
theme(
  legend.position = "none",
  axis.title = element_text(size = 10, color = "black", face = "bold"),
  axis.text = element_text(size = 10)
)
```

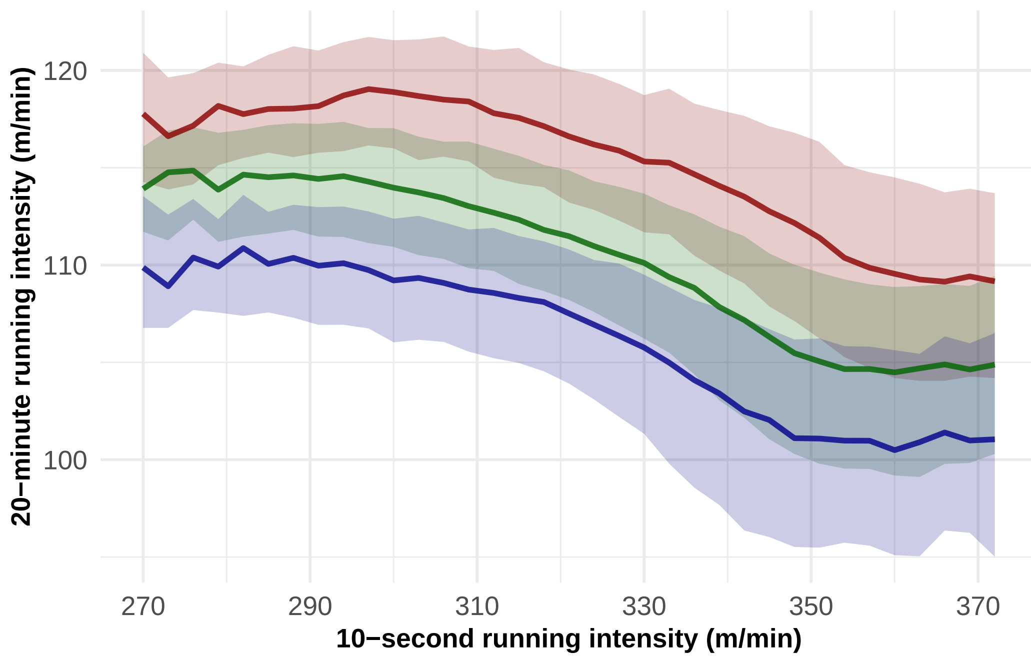

Figure 6.2: Estimation of the first three Pareto frontiers of the relationship between a 10-second rolling average and a 20-minute rolling average mean speed. The line represents the median Pareto frontier, while the ribbon represents the middle 50% of draws from the posterior predictive distribution of both the 10-second rolling average and 20-minute rolling average mean speed.

From the 4,000 simulated data sets of the 10-second rolling average and 20-minute rolling averages, the probabilities that an individual sits on each frontier were estimated. For this squad, the analysis is shown in Table 6.1.

```
combined_df_pareto %>%
  ungroup() %>%
  group_by(Athlete, .level) %>%
  summarise(n = n() / nrow(.) * 18) %>%
  pivot_wider(names_from = .level, values_from = n) %>%
  mutate(Plus4 = rowSums(across(`4`:colnames(.)[ncol(.)]), na.rm = TRUE),
         across(`1`:Plus4, round, 5),
         across(`1`:Plus4, ~ . * 100)) %>%
  select(Athlete:`3`, Plus4) ## Calculate the proportions of each athlete on each fron
```

Table 6.1: Example squad analysis of the probabilities for each athlete being located on the Pareto frontier between 10-second and 20-minute rolling averages.

Athlete	1st Frontier (%)	2nd Frontier (%)	3rd Frontier (%)	4th or worse Frontier (%)
A	5.4	24.0	34.1	36.5
B	30.7	28.5	19.8	20.9
C	4.6	15.3	26.0	54.2
D	0.2	2.1	8.9	88.9
E	0.0	0.1	1.4	98.5

Athlete	1st Frontier (%)	2nd Frontier (%)	3rd Frontier (%)	4th or worse Frontier (%)
F	30.4	32.5	22.4	14.7
G	46.2	31.5	13.9	8.5
H	20.7	27.8	23.2	28.3
I	6.2	13.1	19.0	61.7
J	0.2	2.0	6.2	91.6
K	5.4	21.3	32.7	40.6
L	38.0	31.9	17.5	12.6
M	44.1	29.45	15.5	10.9
N	5.7	15.2	23.1	56.0
O	2.5	11.3	18.8	67.5
P	46.5	31.1	14.6	7.8
Q	33.5	27.3	18.8	20.4
R	5.3	16.6	26.7	51.4

6.5 Discussion

The present study aimed to quantify the locomotion patterns of elite-level female footballers during matches and use Pareto frontiers to identify athletes with the highest running intensities in both short (10 seconds) and long (20 minutes) periods in a match. When analysing the data univariately, there was no athlete that, on average, recorded both the highest 10-second and 20-minute rolling average windows, meaning a multivariate approach was justified to explore the trade-off relationship. While there was no correlation between the 10-second and 20-minute rolling average windows across the whole cohort; when considering the multivariate analysis, there was a negative relationship between those were on the first three Pareto frontiers. The present study also quantified the uncertainty around the Pareto frontier through a Bayesian mixed model to provide coaches additional information for athletes with only limited data to start individualising prescription earlier after taking the uncertainty into consideration.

Given that no athlete recorded both the highest 10-second and 20-minute rolling average windows, a Pareto frontier consisting of four athletes (Athletes L, M, P, and G) was established. That is, no other athlete was higher in both the 10-second and 20-minute rolling averages than these four athletes. Although it had been hypothesised that there would be a negative relationship between the 10-second and 20-minute rolling averages based on previous research (Crielaard & Pirnay, 1981), no such relationship was identified. Previous research had identified a negative relationship between the initial sprint time and the resulting RSA decrement score (i.e., an athlete with a faster sprint speed experiences a more severe decrement) (Bishop et al., 2001). Similarly, there is also a negative relationship between the RSA decrement score and maximal oxygen uptake (i.e., VO2max) (i.e., an athlete with a higher VO2max experiences a less severe decrement) (Rampinini et al., 2009). Consequently, Rampinini et al. (2009) hypothesized that a positive relationship between initial sprint time and VO2MAX would exist (i.e., an athlete with a faster initial sprint time would have a lower VO2MAX); curiously however, no significant relationship was evident (Rampinini et al., 2009). These findings are akin to the present study where no relationship was identified between the 10-second and 20-minute rolling averages. Therefore, it is thought that a Simpson's paradox was present in the data, in which either the quality of an athlete or the opportunity to run is limited by strategic or tactical decisions (Abbott et al., 2018; Hennessy & Jeffreys, 2018) which led to no relationship being present. To test this, when looking at the relationship between only the first three Pareto frontiers, there was indeed a negative relationship between the two running intensities. That is, the six athletes that were not present on the first three Pareto frontiers (i.e., athletes located in the lower left quadrant of Figure 6.1) obfuscate the correlation between the two running intensities due to either tactical or position constraints (i.e., limited opportunities to perform maximal running efforts) or limited physical qualities (i.e., lower training status).

Unlike previous studies utilizing Pareto frontiers which have been purely descriptive (Gunantara, 2018; Mastroddi & Gemma, 2013; Newans, Bellinger, & Minahan, 2022), the present study extended Pareto frontiers into inferential statistics. While the predominant inferential statistics in sports science is frequentist (Santos-Fernandez et al., 2019), the use of a Bayesian mixed model was required to estimate the probabilities that an individual sits on each frontier. The frequentist framework returns a point estimate of the model parameter, typically the maximum likelihood estimate. In contrast, the Bayesian framework produces a posterior distribution for the model parameter, which can then be propagated into uncertainty analysis of the Pareto frontiers by generating plausible data sets from the posterior predictive distribution and constructing a Pareto front for each of these predictive data sets. For instance, the athlete labeled "P" on Figure 6.1, while sitting on the Pareto frontier when plotting the sample data, still only has a 47% probability that they are on the true Pareto frontier. Similarly, there is 31% probability that they are on the 2nd Pareto frontier, a 15% probability that they are on the 3rd Pareto frontier, with only a 8% probability that they are on the 4th or higher Pareto frontier.

This study used a mixed model within a Bayesian framework to quantify the uncertainty around the Pareto frontier. Consequently, the probabilities that each athlete was on the Pareto frontier could be estimated, having accounted for the uncertainty of each athlete. The approach of generating a credible interval for the Pareto frontier to determine the probability that an athlete sits on different frontiers was novel, with a similar approach only seen in portfolio construction in finance (Bauder et al., 2021); however, these did not require a mixed model to accounted for the imbalanced data like that seen in the present study. Given the variability in the data set and with some athletes only playing a limited number of matches, the mixed model generated large credible intervals around each of the conditional means. Consequently, the interpretation of the conditional mean Pareto frontier is limited as there is uncertainty whether the athlete truly lies on the population Pareto frontier. Using a Bayesian framework allows the uncertainty to be estimated for whether an athlete sits on the Pareto frontier, as well as calculating the uncertainty of where the Pareto frontier lies across the whole Cartesian plane.

While the present study illustrated how to analyse the balance between the two running intensities, the application of the statistical methods used in this study, namely, a Bayesian mixed model, to estimate the uncertainty around the Pareto frontier is wide-reaching and can be used in any sporting or other context where there are conflicting attributes that are equally desirable. By understanding which frontier an athlete is located in, coaches are provided additional information to devise training sessions that are targeted at improving both speed (i.e., 10 s) and endurance (i.e., 20 min) running performance that is relevant to match-specific running intensities. Similarly, when new squad members are introduced, identification of the various frontiers allows new squad members to be categorised into training groups with athletes displaying a similar locomotion profile within a given position.

7 Thesis Discussion

7.1 Interpretation of the results

Sports Scientists are inundated by data through wearables, testing equipment, motion-capture, and with automated and semi-automated annotated match statistics. Consequently, there is a requirement for more sports science research to focus on increasing the statistical capabilities of Sports Scientists, to ensure that the quality of statistical analysis can match the increasing quality of data being collected. The aim of this thesis is to provide Sports Scientists with access to applications of statistical methods that will expand their statistical toolkit to accommodate data sets regularly seen in a sports science context. This thesis explored mixed models for imbalanced data sets with repeated observations where there was inherent variability between athletes, introduced the sports science community to Pareto frontiers as a method to identify extreme values when considering multiple variables of interest, and explored inferential statistics within a Bayesian framework as an alternative to the traditional frequentist framework, especially when working with small samples and small effect sizes. It would have been simpler to simulate data sets to illustrate the statistical concepts explored in this thesis; however, it was decided that there would be some loss in authenticity and translatability into the field. Consequently, a secondary aim of this thesis, was to produce research that furthers not only the statistical education of the sports science community, but also contributes to the body of research within each sports discipline explored within this thesis.

This thesis firstly highlighted the need for mixed models when examining the differing movement patterns within the various levels of women's rugby league and describing the movement patterns and match statistics of the NRLW competition. Secondly, it highlighted the need for Pareto frontiers to visualise the trade-off relationship within batting and bowling in T20 cricket. Next it highlighted the need for Bayesian inference when examining the effect of -alanine supplementation of 4-km TT performance. Finally, it highlighted the need for mixed models, Pareto frontiers, and Bayesian inference when understanding the dynamics of short- and long-duration specific running intensities.

One feature of this thesis was the transparency in both the code and data used within this thesis. There has been calls for further adoption of 'open-science practices' over the past few years such as Registered reports (Caldwell et al., 2020), sharing of data and code (Borg et al., 2020). Indeed, some journals are now encouraging the use of data repositories (e.g., FigShare, Open Science Framework) to facilitate these open-science practices; however, this is still the minority within the sports science discipline (Borg et al., 2020). By publishing this thesis as an eBook, it is intended that the practices shown within this thesis serve as an example just as much as the actual code helps other researchers in performing their own analyses.

Another area of notable interest for this thesis, was the intentional use of data from women's sporting codes. While there is a growth in the professionalism of women's sport, the body of research within each sport in the women's game is still sparse. Consequently, this thesis presented an opportunity in which statistical concepts could be clearly articulated using quality data, while also addressing the inequality that is present in the frequency of publishing of research in women's sports. For example, prior to the start of this PhD candidature, there were no studies in women's rugby league examining the movement patterns at any level of competition. Therefore, programming and prescription for athletes was derived from men's rugby league research with no objective justification for the translation of research. Since the commencement of candidature, the movement patterns of women's international rugby league, movement patterns and match statistics of the Australian domestic NRLW competition, a comparison of the three levels of competition in Australia, and a study of the physiological characteristics of female rugby league players have all been published.

In Chapter 2 and Chapter 5, mixed models were able to provide insights into female rugby league that have previously been required to be assumed based on the research previously performed in male cohorts. By providing data relevant to the female cohorts, better decision making by both coaches, support staff, and the league administration can be achieved with the further knowledge elucidated by these studies. As the NRLW only contained four teams throughout the 2018-20 seasons and was played in 60-min matches, it was necessary to understand the movement patterns to guide the expansion of the competition. At the time of writing, the 2023 competition will feature 10 teams (National Rugby League, 2022) playing in

70-min matches (NRL.com, 2021), of which our research has provided a baseline understanding of the game to understand how the expansion has altered movement patterns of the athletes. The applications of these two studies can enable coaches and support staff to better program and prescribe training sessions suitable for these athletes, especially those recently recruited into the expansion teams. Furthermore, these baseline results can serve as a benchmark for lower divisions to provide information on the standards in physiology required to compete at the NRLW level.

Pareto Frontiers were also introduced to Sports Scientists in Chapter 3 of this thesis. While mixed models and Bayesian inference have been used, albeit somewhat uncommonly, in sports science; as far as we are aware, Pareto Frontiers have never been used within sports science applications. As sports are consistently focused on identifying athletes that could provide an added edge on their opposition, they require athletes that are substantially different from the 'average' athlete in a given sport. This concept has been shown with the introduction of statistics such as 'wins above replacement', 'points above replacement', 'runs above replacement', which describe the impact a player has relative to a 'replacement' player (i.e., if you substituted the player for the average player in the league). Consequently, sports science research should consider methodologies that do not revolve around only identifying the mean without extracting the player level effects to identify the extreme athletes.

This thesis endeavoured to encourage Sports Scientists in Chapter 4 to consider adopting inferential statistics in a Bayesian framework over the more commonly-used frequentist framework. There is a pitfall created by the arbitrary statistical significance threshold of $\alpha = 0.05$

typically seen in a frequentist framework. Due to the expected impact of a study, journals are more likely to publish results in which a significant difference is found, than a study in which no there were no significant findings. This generates a publication bias within sports science journals (Borg et al., 2023) in which the distribution of published results is skewed in favour towards those that present significant findings, leading to unintended consequences such as *p*-hacking. Therefore, sports science journals should promote the use of probabilistic statements generated from the posterior distribution of the variable of interest, resulting in less reason to reject papers solely based upon an arbitrary cut-off of $\alpha = 0.05$.

One curious observation seen in both Chapter 3 and Chapter 6, was that variables that were thought to be negatively correlated, were indeed positively correlated. While initially baffling, upon further investigation after identifying each athlete, it seems that a Simpson's paradox was present in each case. It was known that there is a negative relationship between the initial sprint time and the resulting repeated-sprint ability decrement score (i.e., an athlete with a faster sprint speed experiences worse decrement) (Bishop et al., 2001). Similarly, there is also a negative relationship between the repeated-sprint ability decrement score and maximal oxygen uptake (VO_{2MAX}) (i.e., an athlete with a higher VO_{2MAX} has a less severe decrement in repeated-sprint ability) (Rampinini et al., 2009). Consequently, there should be a positive relationship between initial sprint time and VO_{2MAX} (i.e., an athlete with a faster sprint speed displays a lower VO_{2MAX}); curiously however, no significant relationship is evident (Rampinini et al., 2009). Similarly, when comparing sprint performance and the Yo-Yo intermittent fitness test, there were either no correlation (Castagna et al., 2009; Lockie et al., 2017) or even positive relationships (Hermassi et al., 2015; Ingebrigtsen et al., 2014) found. While this has baffled researchers in the past (Hermassi et al., 2015), this is possibly due to a Simpson's paradox where athletes that are more highly-trained will be higher in both attributes than less highly-trained athletes. This is evident in one study (Ingebrigtsen et al., 2014), where athletes from a 3rd-division team recorded slower sprint times and decreased Yo-Yo intermittent fitness test results than a 1st-division team, yet the correlations calculated between the sprint times and the Yo-Yo intermittent fitness results did not account for the level of competition. To illustrate this point, some sample data was generated shown in Figure 7.1.

```
library(tidyverse)
library(patchwork)
set.seed(99.94)
a = c(runif(25, 60, 120), runif(25, 40, 100), runif(25, 20, 60)) ## Generate random number
b = (1 / a) * c(rnorm(25, 150, 30), rnorm(25, 80, 30), rnorm(25, 10, 20)) ## Generate rand
df <- data.frame(x = a,
                 y = b,
                 group = rep(c("A", "B", "C"), each = 25)) ## Form into a data frame

plot_1 <- ggplot(df, aes(x = x / 1.2, y = y * 40)) +
  geom_point(size = 3, alpha = 0.5) +
  geom_smooth(method = "lm",
              se = F,
```

```
              linetype = "dashed",
              color = "black") +
  theme_minimal() +
  labs(x = "Variable 1",
       y = "Variable 2",
       color = "Quality") +
  theme(axis.text = element_blank()) ## Generate plot with no grouping


plot_2 <- ggplot(df, aes(x = x / 1.2, y = y * 40, color = group)) +
  geom_point(size = 3, alpha = 0.6) +
  geom_smooth(method = "lm",
              se = F,
              linetype = "dashed") +
  theme_minimal()+
  labs(x = "Variable 1",
       y = "Variable 2",
       color = "Quality")+
  theme(legend.position = "right",
        axis.text = element_blank(),
        axis.title.y = element_blank()) ## Generate plot with grouping
```

```r
(plot_1 + plot_2) + plot_annotation(tag_levels = 'A') ## Combine plots
```

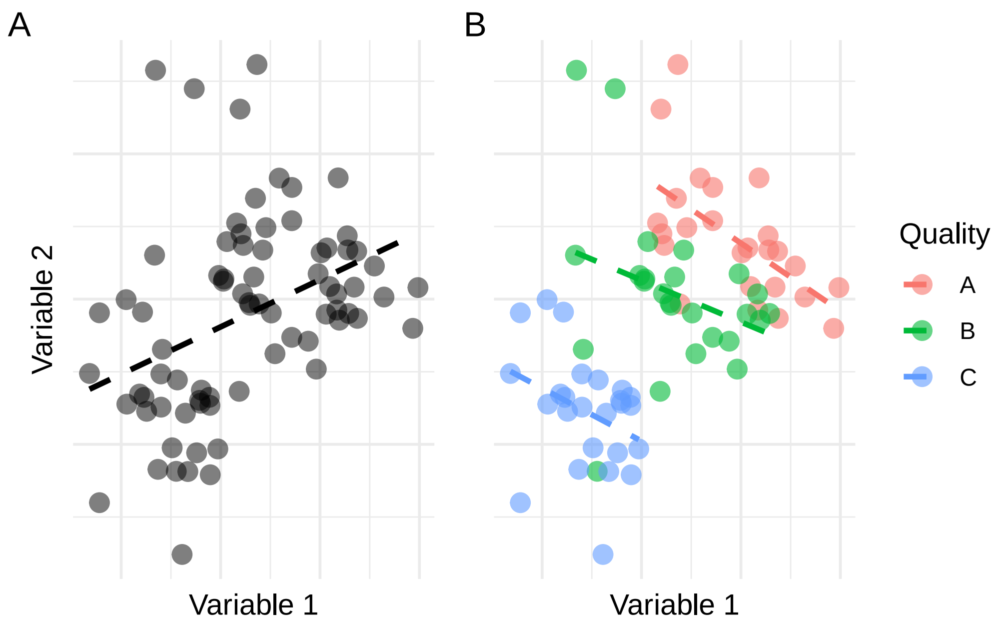

Figure 7.1: Simpson's paradox illustrated using sample data where a positive relationship is evident when the whole cohort is combined (A) yet a negative relationship is evident when the cohort is split by level of competition (B).

As seen in panel A of Figure 7.1, when the regression line is fitted across the whole cohort, a positive correlation is present. Consequently, if these variables are positively correlated, a Pareto Frontier may not be as useful given that it is likely that the individual that is highest in one variable is also likely high in the other variable. However, researchers and practitioners should be aware of other confounding variables that could influence the analysis. For example, in panel B of Figure 7.1, if the quality of player (or level of competition) is incorporated into the regression model, then a clear negative relationship is identified between the two variables and the Pareto Frontier could provide useful applications across each level of competition (e.g., finding players with movement patterns more representing that of a higher-level of competition). Consequently, when a relationship is unexpected (e.g., positive relationship when expected to be negative), it is recommended that additional demographic variables should also be tested in the regression model to ensure there is no Simpson's paradox and that the relationship is actually present.

7.2 Practical Applications

Even though mixed models are not new to sports science (Dalton-Barron et al., 2020), there is still valuable data being underutilised in some studies using alternative statistical methods (such as RM-ANOVA) that require complete data sets when dealing with longitudinal data. This requires the need to either discard data, summarise the data (i.e., invalidly take the mean for each participant without accounting for differing number of observations), or impute data which has its flaws too (Borg et al., 2021). By using a mixed model instead, Sports Scientists can retain the full data set and can provide flexibility for missing data due to injuries, squad selection, and access to athletes on a given day.

As Pareto frontiers have not been used within sports science before, we believe the use of Pareto frontiers is only in its infancy to identify athletes that possess the best compromise between attributes of interest. Consequently, the applications can be wide-ranging as it is not limited to one specific area of sports science. While this thesis predominantly framed Pareto frontiers in a talent identification context, the applications can be more widespread. As referred to in

Chapter 3, there have been multiple instances in the last few years (Duthie et al., 2021; Morin et al., 2021; Rudsits et al., 2018) in which researchers have attempted to identify "maximal" efforts when multiple observations of an individual are recorded across varying conditions of the independent variable. To eliminate any sub-maximal efforts, a Pareto frontier can be established to only extract the values that exhibit the best compromise between the variables of interest.

Of particular note, Chapter 6 explored Pareto frontiers using mixed models in a Bayesian framework, solidifying these three pillars of the thesis. As outlined as a limitation in Chapter 3, when multiple observations are present, quantifying the uncertainty around each individual and where they are with respect to the Pareto frontier is required. The mixed model can account for the dependency between observations, while the Bayesian framework enables the probability of an individual being located on the Pareto frontier can be calculated.

7.3 Further Research

While mixed models have become more common within sports science over the past five years, it seems that the uses of the inferential statistics generated by the mixed model are still relatively limited. However, the estimates arising from these mixed models (e.g., marginal means and conditional means) can also be used for further applications. For example, in Chapter 6, the estimated values from the mixed model were simulated thousands of times to estimate the true populate Pareto frontier. By sampling from the mixed model, the random effects can be accounted for and can be used to simulate data sets from the model (Borg et al., 2020). This data set can then be used for further research (e.g., building Pareto frontiers); therefore, a mixed model does not have to be the end of the analysis; rather, can serve a specific purpose in an analysis pipeline (in this case, providing partial pooling of estimates for an imbalanced data set) which can facilitate further analyses later in the pipeline.

When identifying the Pareto frontiers in different data sets, it become apparent that the shape of the frontier could differ substantially. Take, for instance, the following two Pareto frontiers:

```
library(rPref)
df <- data.frame(
  a = c(1, 2, 3, 4, 6, 8, 10, 6, 2, 2, 4, 2.5, 2),
  b = c(10, 8, 6, 4, 3, 2, 1, 2, 6, 4, 2, 2, 2.5)
) ## Generate concave data set
df <- psel(df, high(a) * high(b), top_level = 99) ## Retrieve Pareto frontier for concave

plot_1 <- ggplot() +
  geom_point(data = df %>% filter(.level != 1), aes(x = a, y = b)) +
  geom_point(data = df %>% filter(.level == 1),
             aes(x = a, y = b),
             color = "red") +
```

```
  geom_line(data = df %>% filter(.level == 1),
            aes(x = a, y = b),
            color = "red") +
  theme_minimal() +
  labs(x = "Variable 1",
       y = "Variable 2") +
  theme(axis.text = element_blank()) ## Plot concave Pareto frontier

set.seed(99.94)
df2 <- data.frame(a = c(1, 3, 5, 6, 7, 8, 9, 10, runif(20, 0, 7)),
                  b = c(10, 9, 8, 7, 6, 5, 3, 1, runif(20, 0, 7))) ## Generate convex data
df2 <- psel(df2, high(a) * high(b), top_level = 99) ## Retrieve Pareto frontier for convex

plot_2 <- ggplot() +
  geom_point(data = df2 %>% filter(.level != 1), aes(x = a, y = b)) +
  geom_point(data = df2 %>% filter(.level == 1),
             aes(x = a, y = b),
             color = "red") +
  geom_line(data = df2 %>% filter(.level == 1),
            aes(x = a, y = b),
            color = "red") +
```

```
  theme_minimal() +
  labs(x = "Variable 1",
       y = "Variable 2") +
  theme(axis.text = element_blank(),
        axis.title.y = element_blank()) ## Plot convex Pareto frontier

(plot_1 + plot_2) + plot_annotation(tag_levels = 'A')
```

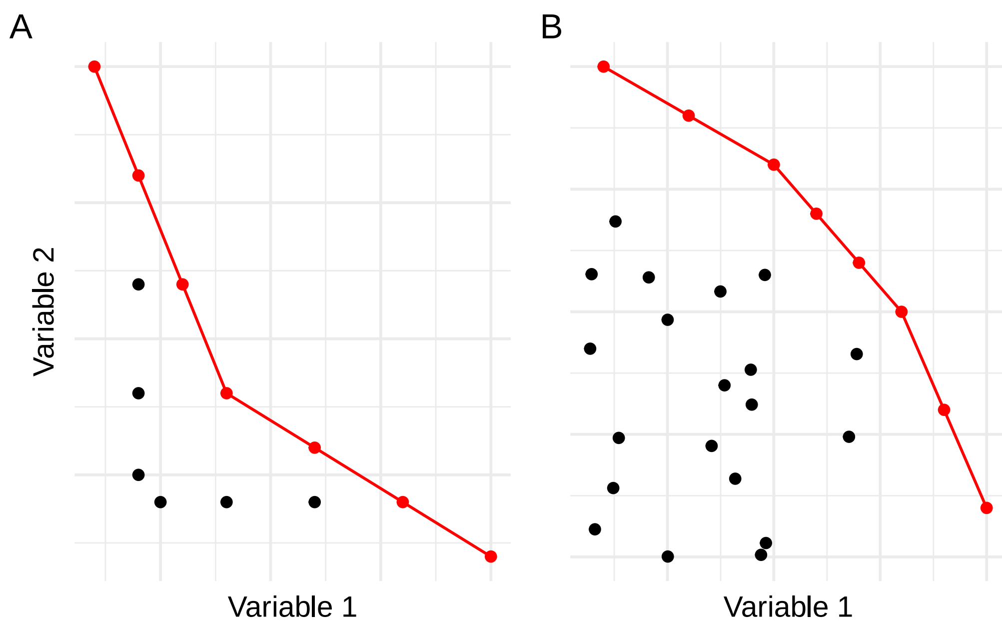

Figure 7.2: The difference between a concave (A) and a convex (B) Pareto frontier.

The Pareto frontier is concave (as seen in panel A of Figure 7.2) if the X axis variable increases, the Y axis variable declines sharply. Similarly, as the Y axis variable increases, the X axis variable declines sharply. However, when the Pareto frontier is convex (as seen in panel B of Figure 7.2), as the X axis variable increases, there is only a gradual decline in the Y axis variable. Additionally, as the Y axis variable increases, there is only a gradual decline in the X axis variable.

While initially overlooked, the applications of the shape of the Frontier could be wide-reaching. If the scatter plot features players within a team and only a limited number of athletes can be selected from the cohort, if there are multiple athletes on the Pareto frontier it may be difficult to decide which athletes should be selected. Consequently, in this case, it could be argued that if the data reflected that in panel A of Figure 7.2, the athletes at the extremities of the Pareto frontier are of greater value than those in the middle of the frontier as any gain in one variable comes at a great cost to the opposing variable. Juxtaposing this, if the data reflected that shown in panel B of Figure 7.2, it could be argued that athletes in the middle of the Pareto frontier are of the highest value, as their gains in each variable have come at a relatively low expense to the opposing variable. Further research could investigate the relative locations within the Pareto frontier and their contributions towards the team construction. Additionally, Pareto frontiers are only one of many approaches to multiobjective optimisation and other methods should be explored when applying into a sports science context.

Even though Bayesian inference is not new to sports science (Mengersen et al., 2016; Santos-Fernandez et al., 2019), there is still limited use, as well as limited resources on using Bayesian inference in sports science contexts. In this thesis, the use of informative priors (Chapter 4) was explored and it was shown how informative priors can assist in decision making for small samples and small effect sizes. Further research can build on this call for informative priors by developing catalogs of standardised, normative reference data to use as the prior distribution for a given variable. These databases could minimise the perceived effect of 'biasing the prior', by ensuring that all studies are utilising the same prior when determining their conclusions.

One area that this thesis intentionally did not explore was the application of machine learning techniques within a sports science context. It is noted that machine learning techniques are changing the way data is collected, extracted, and processed (Richter et al., 2021). However, the level of expertise required to understand these techniques may be beyond the scope of a 'typical' sports scientist. Within organisations, it may be that hybrid roles start to emerge in which there is crossover between a Sports Scientist and a data scientist; however, I do not believe that every applied Sports Scientist needs to understand and be competent to run their own machine learning models. Consequently, Sports Scientists should endeavour to learn general principles surrounding machine learning models to engage with data scientists that are building these models as the frequency in which Sports Scientists will utilise software that harness machine learning will only grow into the future. Therefore, statistical methodology readily available to applied sport scientists was the main focus of this thesis.

7.4 Thesis Conclusion

This thesis encourages Sports Scientists to develop in their statistical literacy to ensure validity and robustness in their statistics being presented for decision-making. As repeated-measures, imbalanced data sets, multiple variables of interest, small samples, and small effect sizes are commonplace in sports science, this thesis was written to urge Sports Scientists to develop a deeper understanding and improved competency in statistical methods. Additionally, Study 5 serves as a 'capstone' study, illustrating how the different statistical concepts explored in this thesis (i.e., mixed models, Pareto frontiers, and Bayesian inference) can be used in tandem to produce novel research. The thesis provides all the data and R code required to run all the analyses within this thesis to ensure Sports Scientists can replicate the analyses with their own data, as well as to provide an example to the sports science community of open and transparent research practices. By compiling this thesis as an eBook, it is intended that Sports Scientists can utilise these studies as a resource when conducting their own research.

References

Abbott, W., Brickley, G., & Smeeton, N. (2018). Positional differences in GPS outputs and perceived exertion during soccer training games and competition. *The Journal of Strength & Conditioning Research*, *32*(11), 3222–3231. https://doi.org/10.1519/JSC.0000000000002387

Abt, G., Boreham, C., Davison, G., Jackson, R., Nevill, A., Wallace, E., & Williams, M. (2020). Power, precision, and sample size estimation in sport and exercise science research. *Journal of Sports Sciences*, *38*(17), 1933–1935. https://doi.org/10.1080/02640414.2020.1776002

Andersson, H., Randers, M., Heiner-Møller, A., Krustrup, P., & Mohr, M. (2010). Elite female soccer players perform more high-intensity running when playing in international games compared with domestic league games. *The Journal of Strength and Conditioning Research*, *24*(4), 912–919. https://doi.org/10.1519/jsc.0b013e3181d09f21

Atkinson, G., Batterham, A., & Hopkins, W. (2012). Sports performance research under the spotlight. *International Journal of Sports Medicine*, *33*(12), 949. https://doi.org/10.1055/s-0032-1327755

Bacchetti, P., Deeks, S., & McCune, J. (2011). Breaking free of sample size dogma to perform innovative translational research. *Science Translational Medicine*, *3*(87), 87ps24. https://doi.org/10.1126/scitranslmed.3001628

Baguet, A., Koppo, K., Pottier, A., & Derave, W. (2010). B-alanine supplementation reduces acidosis but not oxygen uptake response during high-intensity cycling exercise. *European Journal of Applied Physiology*, *108*(3), 495–503. https://doi.org/10.1007/s00421-009-1225-0

Baldwin, S., & Fellingham, G. (2013). Bayesian methods for the analysis of small sample multilevel data with a complex variance structure. *Psychological Methods*, *18*(2), 151. https://doi.org/10.1037/a0030642

Barker, R., & Schofield, M. (2008). Inference about magnitudes of effects. *International Journal of Sports Physiology and Performance*, *3*(4), 547–557. https://doi.org/10.1123/ijspp.3.4.547

Barr, G., & Kantor, B. (2004). A criterion for comparing and selecting batsmen in limited overs cricket. *Journal of the Operational Research Society*, *55*(12), 1266–1274. https://doi.org/10.1057/palgrave.jors.2601800

Bartlett, J., Hatfield, M., Parker, B., Roberts, L., Minahan, C., Morton, J., & Thornton, H. (2020). DXA-derived estimates of energy balance and its relationship with changes in body composition across a season in team sport athletes. *European Journal of Sport Science*, *20*(7), 859–867. https://doi.org/10.1080/17461391.2019.1669718

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Batterham, A., & Hopkins, W. (2006). Making meaningful inferences about magnitudes. *International Journal of Sports Physiology and Performance*, *1*(1), 50–57. https://doi.org/10.1123/ijspp.1.1.50

Bauder, D., Bodnar, T., Parolya, N., & Schmid, W. (2021). Bayesian mean-variance analysis: Optimal portfolio selection under parameter uncertainty. *Quantitative Finance*, *21*(2), 221–242. https://doi.org/10.1080/14697688.2020.1748214

Bellinger, P., Ferguson, C., Newans, T., & Minahan, C. (2020). No influence of prematch subjective wellness ratings on external load during elite Australian Football match play. *International Journal of Sports Physiology and Performance*, *15*(6), 801–807. https://doi.org/10.1123/ijspp.2019-0395

Bellinger, P., & Minahan, C. (2016). The effect of -alanine supplementation on cycling time trials of different length. *European Journal of Sport Science*, *16*(7), 829–836. https://doi.org/10.1080/17461391.2015.1120782

Bernards, J., Sato, K., Haff, G., & Bazyler, C. (2017). Current research and statistical practices in sport science and a need for change. *Sports*, *5*(4), 87. https://doi.org/10.3390/sports5040087

Bishop, D., Spencer, M., Duffield, R., & Lawrence, S. (2001). The validity of a repeated sprint ability test. *Journal of Science and Medicine in Sport*, *4*(1), 19–29. https://doi.org/10.1016/S1440-2440(01)80004-9

Borenstein, M. (2009). *The handbook of research synthesis and meta-analysis* (H. Cooper, L. Hedges, & J. Valentine, Eds.; Vol. 2, pp. 221–235). https://doi.org/10.7758/9781610448864

Borg, D., Barnett, A., Caldwell, A., White, N., & Stewart, I. (2023). The Bias for Statistical Significance in Sport and Exercise Medicine. *Journal of Science and Medicine in Sport*. https://doi.org/10.1016/j.jsams.2023.03.002

Borg, D., Bon, J., Sainani, K., Baguley, B., Tierney, N., & Drovandi, C. (2020). Comment on: 'Moving sport and exercise science forward: A call for the adoption of more transparent research practices'. *Sports Medicine*, *50*(8), 1551–1553. https://doi.org/10.1007/s40279-020-01298-5

Borg, D., Minett, G., Stewart, I., & Drovandi, C. (2018). Bayesian methods might solve the problems with magnitude-based inference. A letter in response to Dr. Sainani. *Medicine and Science in Sports and Exercise*, *50*(12), 2609–2610. https://doi.org/10.1249/mss.0000000000001736

Borg, D., Nguyen, R., & Tierney, N. (2021). Missing data: Current practice in football research and recommendations for improvement. *Science and Medicine in Football*, *0*(ja), null. https://doi.org/10.1080/24733938.2021.1922739

Brown, D., Dwyer, D., Robertson, S., & Gastin, P. (2016). Metabolic power method: Underestimation of energy expenditure in field-sport movements using a global positioning system tracking system. *International Journal of Sports Physiology and Performance*, *11*(8), 1067–1073. https://doi.org/10.1123/ijspp.2016-0021

Buchheit, M., Manouvrier, C., Cassirame, J., & Morin, J.-B. (2015). Monitoring locomotor load in soccer: Is metabolic power, powerful? *International Journal of Sports Medicine*, *36*(14), 1149–1155. https://doi.org/10.1055/s-0035-1555927

Bukiet, B., & Ovens, M. (2006). A mathematical modelling approach to one-day cricket batting orders. *Journal of Sports Science & Medicine*, *5*(4), 495–502. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3861747/

Bürkner, P.-C. (2017). Brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Caldwell, A., Vigotsky, A., Tenan, M., Radel, R., Mellor, D., Kreutzer, A., Lahart, I., Mills, J., Boisgontier, M., Boardley, I., Bouza, B., Cheval, B., Chow, Z. R., Contreras, B., Dieter, B., Halperin, I., Haun, C., Knudson, D., Lahti, J., … Consortium for Transparency in Exercise Science (COTES) Collaborators. (2020). Moving sport and exercise science forward: A call for the adoption of more transparent research practices. *Sports Medicine*, *50*(3), 449–459. https://doi.org/10.1007/s40279-019-01227-1

Castagna, C., Impellizzeri, F., Cecchini, E., Rampinini, E., & Alvarez, J. C. B. (2009). Effects of intermittent-endurance fitness on match performance in young male soccer players. *The Journal of Strength and Conditioning Research*, *23*(7), 1954–1959. https://doi.org/10.1519/JSC.0b013e3181b7f743

Clarke, A., Anson, J., & Pyne, D. (2017). Game movement demands and physical profiles of junior, senior and elite male and female rugby sevens players. *Journal of Sports Sciences*, *35*(8), 727–733. https://doi.org/10.1080/02640414.2016.1186281

Clarke, A., Anson, J., & Pyne, D. (2015). Physiologically based GPS speed zones for evaluating running demands in women's rugby sevens. *Journal of Sports Sciences*, *33*(11), 1101–1108. https://doi.org/10.1080/02640414.2014.988740

Clarke, A., Couvalias, G., Kempton, T., & Dascombe, B. (2019). Comparison of the match running demands of elite and sub-elite women's Australian Football. *Science and Medicine* *in Football*, *3*(1), 70–76. https://doi.org/10.1080/24733938.2018.1479067

Clarke, A., Ryan, S., Couvalias, G., Dascombe, B., Coutts, A., & Kempton, T. (2018). Physical demands and technical performance in Australian Football League Women's (AFLW) competition match-play. *Journal of Science and Medicine in Sport*, *21*(7), 748–752. https://doi.org/10.1016/j.jsams.2017.11.018

Cobb, G., & Moore, D. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly*, *104*(9), 801–823. https://doi.org/10.1080/00029890.1997.11990723

Coutts, A., Quinn, J., Hocking, J., Castagna, C., & Rampinini, E. (2010). Match running performance in elite Australian Rules Football. *Journal of Science and Medicine in Sport*, *13*(5), 543–548. https://doi.org/10.1016/j.jsams.2009.09.004

Crang, Z., Duthie, G., Cole, M., Weakley, J., Hewitt, A., & Johnston, R. (2021). The validity and reliability of wearable microtechnology for intermittent team sports: A systematic review. *Sports Medicine*, *51*(3), 549–565. https://doi.org/10.1007/s40279-020-01399-1

Crielaard, J.-M., & Pirnay, F. (1981). Anaerobic and aerobic power of top athletes. *European Journal of Applied Physiology and Occupational Physiology*, *47*(3), 295–300. https://doi.org/10.1007/BF00422475

Cummins, C., Gray, A., Shorter, K., Halaki, M., & Orr, R. (2018). Energetic demands of interchange and full-match rugby league players. *The Journal of Strength and Conditioning Research*, *32*(12), 3447–3455. https://10.1519/JSC.0000000000001801

Cummins, C., Orr, R., O'Connor, H., & West, C. (2013). Global Positioning Systems (GPS) and microtechnology sensors in team sports: A systematic review. *Sports Medicine*, *43*(10), 1025–1042. https://doi.org/10.1007/s40279-013-0069-2

Cunningham, D., Shearer, D., Carter, N., Drawer, S., Pollard, B., Bennett, M., Eager, R., Cook, C., Farrell, J., Russell, M., & Kilduff, L. (2018). Assessing worst case scenarios in movement demands derived from global positioning systems during international rugby union matches: Rolling averages versus fixed length epochs. *PLoS One*, *13*(4), e0195197. https://doi.org/10.1371/journal.pone.0195197

Dalton-Barron, N., Whitehead, S., Roe, G., Cummins, C., Beggs, C., & Jones, B. (2020). Time to embrace the complexity when analysing GPS data? A systematic review of contextual factors on match running in rugby league. *Journal of Sports Sciences*, *38*(10), 1161–1180. https://doi.org/10.1080/02640414.2020.1745446

Davids, K., Lees, A., & Burwitz, L. (2000). Understanding and measuring coordination and control in kicking skills in soccer: Implications for talent identification and skill acquisition. *Journal of Sports Sciences*, *18*(9), 703–714. https://doi.org/10.1080/02640410050120087

Deb, K., & Datta, R. (2012). Hybrid evolutionary multi-objective optimization and analysis of machining operations. *Engineering Optimization*, *44*(6), 685–706. https://doi.org/10.1080/0305215X.2011.604316

Delaney, J., Duthie, G., Thornton, H., Scott, T., Gay, D., & Dascombe, B. (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., Scott, T., Thornton, H., Bennett, K., Gay, D., Duthie, G., & Dascombe, B. (2015). Establishing duration-specific running intensities from match-play analysis in rugby league. *International Journal of Sports Physiology and Performance*, *10*(6), 725–731.

Delaney, J., Thornton, H., Duthie, G., & Dascombe, B. (2016). Factors that influence running intensity in interchange players in professional rugby league. *International Journal of Sports Physiology and Performance*, *11*(8), 1047–1052. https://doi.org/10.1123/ijspp.2015-0559

Delaney, J., Wileman, T., Perry, N., Thornton, H., Moresi, M., & Duthie, G. (2019). The validity of a global navigation satellite system for quantifying small-area team-sport movements. *The Journal of Strength and Conditioning Research*, *33*(6), 1463. https://doi.org/10.1519/JSC.0000000000003157

Deutsch, M., Kearney, G., & Rehrer, N. (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

Dodd, K., & Newans, T. (2018). Talent identification for soccer: Physiological aspects. *Journal of Science and Medicine in Sport*, *21*(10), 1073–1078. https://doi.org/10.1016/j.jsams.2018.01.009

Doncaster, G., Page, R., White, P., Svenson, R., & Twist, C. (2020). Analysis of physical demands during youth soccer match-play: Considerations of sampling method and epoch length. *Research Quarterly for Exercise and Sport*, *91*(2), 326–334. https://doi.org/10.1080/02701367.2019.1669766

Duthie, G., Robertson, S., & Thornton, H. (2021). A GNSS-based method to define athlete manoeuvrability in field-based team sports. *PLoS One*, *16*(11), e0260363. https://doi.org/10.1371/journal.pone.0260363

Dutka, T., Lamboley, C., McKenna, M., Murphy, R., & Lamb, G. (2012). Effects of carnosine on contractile apparatus Ca2+ sensitivity and sarcoplasmic reticulum Ca2+ release in human skeletal muscle fibers. *Journal of Applied Physiology*, *112*(5), 728–736. https://doi.org/10.1152/japplphysiol.01331.2011

Emmonds, S., Heyward, O., & Jones, B. (2019). The challenge of applying and undertaking research in female sport. *Sports Medicine-Open*, *5*(1), 1–4. https://doi.org/10.1186/s40798-019-0224-x

Evans, M., & Moshonov, H. (2006). Checking for prior-data conflict. *Bayesian Analysis*, *1*(4), 893–914. https://doi.org/10.1214/06-BA129

Everaert, I., Stegen, S., Vanheel, B., Taes, Y., & Derave, W. (2013). Effect of beta-alanine and carnosine supplementation on muscle contractility in mice. *Medicine and Science in Sports and Exercise*, *45*(1), 43–51. https://doi.org/10.1249/mss.0b013e31826cdb68

Exercise and Sports Science Australia. (2019). *The professional standards documents*. https://www.essa.org.au/Public/Professional_Standards/The_professional_standards.aspx

Falk, B., Lidor, R., Lander, Y., & Lang, B. (2004). Talent identification and early development of elite water-polo players: A 2-year follow-up study. *Journal of Sports Sciences*, *22*(4), 347–355. https://doi.org/10.1080/02640410310001641566

Fereday, K., Hills, S., Russell, M., Smith, J., Cunningham, D., Shearer, D., McNarry, M., & Kilduff, L. (2020). A comparison of rolling averages versus discrete time epochs for assessing the worst-case scenario locomotor demands of professional soccer match-play. *Journal of Science and Medicine in Sport*, *23*(8), 764–769. https://doi.org/10.1016/j.jsams.2020.01.002

Ferris, D., Gabbett, T., McLellan, C., & Minahan, C. (2018). Basal markers of inflammation, muscle damage, and performance during five weeks of pre-season training in elite youth rugby league players. *Journal of Athletic Enhancement*, *7*(2). https://doi.org/10.4172/2324-9080.1000286

Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, *345*(6203), 1502–1505. https://doi.org/10.1126/science.1255484

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). Guidelines for assessment and instruction in statistics education (GAISE) report. In *Alexandria: American Statistical Association*.

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review*, *70*(1), 1–25. https://doi.org/10.1111/j.1751-5823.2002.tb00336.x

Gallo, T., Cormack, S., Gabbett, T., & Lorenzen, C. (2017). Self-reported wellness profiles of professional Australian Football players during the competition phase of the season. *The Journal of Strength and Conditioning Research*, *31*(2), 495–502. https://doi.org/10.1519/jsc.0000000000001515

Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. Springer Science & Business Media. https://doi.org/10.1007/978-1-4020-8383-9

Gillies, D. (2000). *Philosophical theories of probability*. Psychology Press. https://doi.org/10.4324/9780203132241

Glassbrook, D., Doyle, T., Alderson, J., & Fuller, J. (2019). The demands of professional rugby league match-play: A meta-analysis. *Sports Medicine - Open*, *5*(1), 24. https:

//doi.org/10.1186/s40798-019-0197-9

Govus, A., Coutts, A., Duffield, R., Murray, A., & Fullagar, H. (2018). Relationship between pretraining subjective wellness measures, player load, and rating-of-perceived-exertion training load in american college football. *International Journal of Sports Physiology and Performance*, *13*(1), 95–101. https://doi.org/10.1123/ijspp.2016-0714

Griffin, J., Larsen, B., Horan, S., Keogh, J., Dodd, K., Andreatta, M., & Minahan, C. (2020). Women's football: An examination of factors that influence movement patterns. *The Journal of Strength & Conditioning Research*, *34*(8), 2384–2393. https://doi.org/10.1519/jsc.0000000000003638

Griffin, J., Newans, T., Horan, S., Keogh, J., Andreatta, M., & Minahan, C. (2021). Acceleration and high-speed running profiles of women's international and domestic football matches. *Frontiers in Sports and Active Living*, *3*, 71. https://doi.org/10.3389/fspor.2021.604605

Gunantara, N. (2018). A review of multi-objective optimization: Methods and its applications. *Cogent Engineering*, *5*(1), 1502242. https://doi.org/10.1080/23311916.2018.1502242

Halson, S. (2014). Monitoring training load to understand fatigue in athletes. *Sports Medicine*, *44*(2), 139–147. https://doi.org/10.1007/s40279-014-0253-z

Hannon, M., Coleman, N., Parker, L. J., McKeown, J., Unnithan, V., Close, G., Drust, B., & Morton, J. (2021). Seasonal training and match load and micro-cycle periodization in male premier league academy soccer players. *Journal of Sports Sciences*, *39*(16), 1838–1849. https://doi.org/10.1080/02640414.2021.1899610

Harrison, X., Donaldson, L., Correa-Cano, M. E., Evans, J., Fisher, D., Goodwin, C., Robinson, B., Hodgson, D., & Inger, R. (2018). A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ*, *6*, e4794. https://doi.org/10.7717/peerj.4794

Haugnes, P., Torvik, P.-Ø., Ettema, G., Kocbach, J., & Sandbakk, Ø. (2019). The effect of maximal speed ability, pacing strategy, and technique on the finish sprint of a sprint cross-country skiing competition. *International Journal of Sports Physiology and Performance*, *14*(6), 788–795. https://doi.org/10.1123/ijspp.2018-0507

Hennessy, L., & Jeffreys, I. (2018). The current use of GPS, its potential, and limitations in soccer. *Strength & Conditioning Journal*, *40*(3), 83–94. https://doi.org/10.1519/SSC.0000000000000386

Hermassi, S., Aouadi, R., Khalifa, R., Tillaar, R. van den, Shephard, R., & Chelly, M. S. (2015). Relationships between the yo-yo intermittent recovery test and anaerobic performance tests in adolescent handball players. *Journal of Human Kinetics*, *45*, 197–205. https://doi.org/10.1515/hukin-2015-0020

Hobson, R., Saunders, B., Ball, G., Harris, R., & Sale, C. (2012). Effects of -alanine supplementation on exercise performance: A meta-analysis. *Amino Acids*, *43*(1), 25–37. https://doi.org/10.1007/s00726-011-1200-z

Hodun, M., Clarke, R., De Ste Croix, M., & Hughes, J. (2016). Global positioning system analysis of running performance in female field sports: A review of the literature. *Strength & Conditioning Journal*, *38*(2), 49–56. https://doi.org/10.1519/SSC.0000000000000200

Hopkins, W., Marshall, S., Batterham, A., & Hanin, J. (2009). Progressive statistics for studies in sports medicine and exercise science. *Medicine & Science in Sports & Exercise*, *41*(1), 3–12. https://doi.org/10.1249/MSS.0b013e31818cb278

Horn, J., Nafpliotis, N., & Goldberg, D. (1994). *A niched Pareto genetic algorithm for multiobjective optimization*. 82–87. https://doi.org/https://doi.org/10.1109/ICEC.1994.350037

Hulin, B., Gabbett, T., 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

Ingebrigtsen, J., Brochmann, M., Castagna, C., Bradley, P., Ade, J., Krustrup, P., & Holtermann, A. (2014). Relationships between field performance tests in high-level soccer players. *The Journal of Strength and Conditioning Research*, *28*(4), 942–949. https://doi.org/10.1519/JSC.0b013e3182a1f861

John, L., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*(5), 524–532. https://doi.org/10.1177/0956797611430953

Johnston, K., Wattie, N., Schorer, J., & Baker, J. (2018). Talent identification in sport: A systematic review. *Sports Medicine*, *48*(1), 97–109. https://doi.org/10.1007/s40279-017-0803-2

Johnston, R., Gibson, N., Twist, C., Gabbett, T., MacNay, S., & MacFarlane, N. (2013). Physiological responses to an intensified period of rugby league competition. *The Journal of Strength and Conditioning Research*, *27*(3), 643–654. https://doi.org/10.1519/JSC.0b013e31825bb469

Jones, A., Kirby, B., Clark, I., Rice, H., Fulkerson, E., Wylie, L., Wilkerson, D., Vanhatalo, A., & Wilkins, B. (2021). Physiological demands of running at 2-hour marathon race pace.

Journal of Applied Physiology, *130*(2), 369–379. https://doi.org/10.1152/japplphysiol.00647.2020

Jones, M., West, D., Crewther, B., Cook, C., & Kilduff, L. (2015). Quantifying positional and temporal movement patterns in professional rugby union using global positioning system. *European Journal of Sport Science*, *15*(6), 488–496. https://doi.org/10.1080/17461391.2015.1010106

Kelly, A., & Williams, C. (2020). Physical characteristics and the talent identification and development processes in male youth soccer: A narrative review. *Strength & Conditioning Journal*, *42*(6), 15–34. https://doi.org/10.1519/ssc.0000000000000576

Kempton, T., Sirotic, A., & Coutts, A. (2017). A comparison of physical and technical performance profiles between successful and less-successful professional rugby league teams. *International Journal of Sports Physiology and Performance*, *12*(4), 520–526. https://doi.org/10.1123/ijspp.2016-0003

Kenny, D., & Judd, C. (1986). Consequences of violating the independence assumption in analysis of variance. *Psychological Bulletin*, *99*(3), 422–431. https://doi.org/10.1037/0033-2909.99.3.422

Kruschke, J. (2014). *Doing bayesian data analysis: A tutorial with r, JAGS, and stan*. Academic Press. https://doi.org/10.1016/B978-0-12-405888-0.00001-5

Kwon, S.-S., Lee, K. M., Chung, C. Y., Lee, S. Y., & Park, M. S. (2014). An introduction to the linear mixed model for orthopaedic research. *JBJS Reviews*, *2*(12). https://doi.org/10.2106/JBJS.RVW.N.00009

Lenth, R. (2020). *Emmeans: Estimated marginal means, aka least-squares means*. https:

//CRAN.R-project.org/package=emmeans

Lievens, E., Bellinger, P., Van Vossel, K., Vancompernolle, J., Bex, T., Minahan, C., & Derave, W. (2021). Muscle typology of world-class cyclists across various disciplines and events. *Medicine and Science in Sports and Exercise*, *53*(4), 816–824. https://doi.org/10.1249/MSS.0000000000002518

Ligges, U., & Mächler, M. (2003). Scatterplot3d - an r package for visualizing multivariate data. *Journal of Statistical Software*, *8*(11), 1–20. http://www.jstatsoft.org

Lockie, R., Jalilvand, F., Moreno, M., Orjalo, A., Risso, F., & Nimphius, S. (2017). Yo-yo intermittent recovery test level 2 and its relationship with other typical soccer field tests in female collegiate soccer players. *The Journal of Strength and Conditioning Research*, *31*(10), 2667–2677. https://doi.org/10.1519/JSC.0000000000001734

Lüdecke, D. (2020). *sjPlot: Data visualization for statistics in social science*. https://CRAN.R-project.org/package=sjPlot

Lüdecke, D., Ben-Shachar, M., Patil, I., Waggoner, P., & Makowski, D. (2021). Performance: An r package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, *6*(60), 3139. https://doi.org/10.21105/joss.03139

Lüdecke, D., Patil, I., Ben-Shachar, M., Wiernik, B., Waggoner, P., & Makowski, D. (2021). See: An r package for visualizing statistical models. *Journal of Open Source Software*, *6*(64), 3393. https://doi.org/10.21105/joss.03393

Maquirriain, J., Baglione, R., & Cardey, M. (2016). Male professional tennis players maintain constant serve speed and accuracy over long matches on grass courts. *European Journal of Sport Science*, *16*(7), 845–849. https://doi.org/10.1080/17461391.2016.1156163

Mara, J., Thompson, K., Pumpa, K., & 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

Marler, T., & Arora, J. (2004). Survey of multi-objective optimization methods for engineering. *Structural and Multidisciplinary Optimization*, *26*(6), 369–395. https://doi.org/10.1007/S00158-003-0368-6

Mastroddi, F., & Gemma, S. (2013). Analysis of Pareto frontiers for multidisciplinary design optimization of aircraft. *Aerospace Science and Technology*, *28*(1), 40–55. https://doi.org/10.1016/j.ast.2012.10.003

McCaskie, C., Young, W., Fahrner, B., & Sim, M. (2019). Association between preseason training and performance in elite Australian football. *International Journal of Sports Physiology and Performance*, *14*(1), 68–75. https://doi.org/10.1123/ijspp.2018-0086

McElreath, R. (2018). *Statistical rethinking: A bayesian course with examples in r and stan*. Chapman; Hall/CRC. https://doi.org/10.1201/9781315372495

McKay, A., Stellingwerff, T., Smith, E., Martin, D., Mujika, I., Goosey-Tolfrey, V., Sheppard, J., & Burke, L. (2021). Defining training and performance caliber: A participant classification framework. *International Journal of Sports Physiology and Performance*, *17*(2), 317–331. https://doi.org/10.1123/ijspp.2021-0451

McNeish, D. (2016). On using bayesian methods to address small sample problems. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(5), 750–773. https://doi.org/10.1080/10705511.2016.1186549

Medicine & Science in Sports & Exercise. (2023). *Information for authors*. https://edmgr.ov id.com/msse/accounts/ifauth.htm

Mengersen, K., Drovandi, C., Robert, C., Pyne, D., & Gore, C. (2016). Bayesian estimation of small effects in exercise and sports science. *PLoS One*, *11*(4), e0147311. https://doi.org/10.1371/journal.pone.0147311

Minahan, C., Newans, T., Quinn, K., Parsonage, J., Buxton, S., & Bellinger, P. (2021). Strong, fast, fit, lean, and safe: A positional comparison of physical and physiological qualities within the 2020 Australian Women's Rugby League team. *The Journal of Strength and Conditioning Research*, *35*(Suppl 2), S11–S19. https://doi.org/10.1519/JSC.0000000000004106

Mohr, M., Krustrup, P., & Bangsbo, J. (2003). Match performance of high-standard soccer players with special reference to development of fatigue. *Journal of Sports Sciences*, *21*(7), 519–528. https://doi.org/10.1080/0264041031000071182

Morin, J.-B., Le Mat, Y., Osgnach, C., Barnabò, A., Pilati, A., Samozino, P., & Prampero, P. di. (2021). Individual acceleration-speed profile in-situ: A proof of concept in professional football players. *Journal of Biomechanics*, *123*, 110524. https://doi.org/10.1016/j.jbiomech.2021.110524

Morris, T. (2000). Psychological characteristics and talent identification in soccer. *Journal of Sports Sciences*, *18*(9), 715–726. https://doi.org/10.1080/02640410050120096

Nakai, M., & Ke, W. (2011). Review of the methods for handling missing data in longitudinal data analysis. *International Journal of Mathematical Analysis*, 13.

National Rugby League. (2020). *Tiffany Slater honoured at NSW Her Sport Her Way awards*. https://www.nrl.com/news/2020/03/05/tiffany-slater-honoured-at-nsw-her-sport-her-way-awards/

National Rugby League. (2022). *Statement on NRLW expansion*. https://www.nrl.com/news/2022/06/15/statement-on-nrlw-expansion/

Newans, T., Bellinger, P., Buxton, S., Quinn, K., & Minahan, C. (2021). Movement patterns and match statistics in the national rugby league women's (NRLW) premiership. *Frontiers in Sports and Active Living*, *3*. https://doi.org/10.3389/fspor.2021.618913

Newans, T., Bellinger, P., Dodd, K., & Minahan, C. (2019). Modelling the acceleration and deceleration profile of elite-level soccer players. *International Journal of Sports Medicine*, *40*(5), 331–335. https://doi.org/10.1055/a-0853-7676

Newans, T., Bellinger, P., Drovandi, C., Buxton, S., & Minahan, C. (2022). The utility of mixed models in sport science: A call for further adoption in longitudinal data sets. *International Journal of Sports Physiology and Performance*, *17*(8), 1289–1295. https://doi.org/10.1123/ijspp.2021-0496

Newans, T., Bellinger, P., & Minahan, C. (2022). The balancing act: Identifying multivariate sports performance using Pareto frontiers. *Frontiers in Sports and Active Living*, *4*. https://doi.org/10.3389/fspor.2022.918946

Newell, J., Aitchison, T., & Grant, S. (2014). *Statistics for sports and exercise science: A practical approach*. Routledge. https://doi.org/10.4324/9781315847542

Nomura, S., Ogata, Y., Komaki, F., & Toda, S. (2011). Bayesian forecasting of recurrent earthquakes and predictive performance for a small sample size. *Journal of Geophysical Research: Solid Earth*, *116*(B4). https://ui.adsabs.harvard.edu/link_gateway/2011JGRB.116.4315N/doi:10.1029/2010JB007917

Novak, A., Impellizzeri, F., Trivedi, A., Coutts, A., & McCall, A. (2021). Analysis of the worst-case scenarios in an elite football team: Towards a better understanding and application. *Journal of Sports Sciences*, *39*(16), 1850–1859. https://doi.org/10.1080/02640414.2021.1902138

NRL.com. (2021). *The broader game: NRLW games up to 70 minutes to promote fatigue factor*. https://www.nrl.com/news/2021/07/08/the-broader-game-nrlw-games-up-to-70-minutes-to-promote-fatigue-factor/

Ottosson, R., Engström, P., Sjöström, D., Behrens, C., Karlsson, A., Knöös, T., & Ceberg, C. (2009). The feasibility of using pareto fronts for comparison of treatment planning systems and delivery techniques. *Acta Oncologica*, *48*(2), 233–237. https://doi.org/10.1080/02841860802251559

Patel, A., Bracewell, P., Gazley, A., & Bracewell, B. (2017). Identifying fast bowlers likely to play test cricket based on age-group performances. *International Journal of Sports Science & Coaching*, *12*(3), 328–338. https://doi.org/10.1177/1747954117710514

Pérez-Toledano, M. Á., Rodriguez, F., García-Rubio, J., & Ibañez, S. J. (2019). Players' selection for basketball teams, through performance index rating, using multiobjective evolutionary algorithms. *PLoS One*, *14*(9), e0221258. https://doi.org/10.1371/journal.pone.0221258

Pion, J., Lenoir, M., Vandorpe, B., & Segers, V. (2015). Talent in female gymnastics: A survival analysis based upon performance characteristics. *International Journal of Sports Medicine*, *94*(11), 935–940. https://doi.org/10.1055/s-0035-1548887

Ploutz-Snyder, R., Fiedler, J., & Feiveson, A. (2014). Justifying small-n research in scientifically amazing settings: Challenging the notion that only "big-n" studies are worthwhile. *Journal of Applied Physiology*, *116*(9), 1251–1252. https://doi.org/10.1152/japplphysiol.01335.2013

Ponsich, A., Jaimes, A. L., & Coello, C. A. C. (2012). A survey on multiobjective evolutionary algorithms for the solution of the portfolio optimization problem and other finance and economics applications. *IEEE Transactions on Evolutionary Computation*, *17*(3), 321–344. https://doi.org/10.1109/TEVC.2012.2196800

Pyne, D., Gardner, A., Sheehan, K., & Hopkins, W. (2005). Fitness testing and career progression in AFL football. *Journal of Science and Medicine in Sport*, *8*(3), 321–332. https://doi.org/10.1016/s1440-2440(05)80043-x

Queensland Department of Education and Training. (2023). *Literacy and numeracy fact sheet*. https://education.qld.gov.au/parents/Documents/factsheet-l-n.pdf

Quinn, K., Newans, T., Buxton, S., Thomson, T., Tyler, R., & Minahan, C. (2020). Movement patterns of players in the Australian Women's Rugby League team during international competition. *Journal of Science and Medicine in Sport*, *23*(3), 315–319. https://doi.org/10.1016/j.jsams.2019.10.009

Quintana, D., & Williams, D. (2018). Bayesian alternatives for common null-hypothesis significance tests in psychiatry: A non-technical guide using JASP. *BMC Psychiatry*, *18*(1), 178.

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rago, V., Krustrup, P., Martín-Acero, R., Rebelo, A., & Mohr, M. (2020). Training load and submaximal heart rate testing throughout a competitive period in a top-level male football team. *Journal of Sports Sciences*, *38*(11-12), 1408–1415. https://doi.org/10.1080/02640414.2019.1618534

Rampinini, E., Sassi, A., Morelli, A., Mazzoni, S., Fanchini, M., & Coutts, A. (2009). Repeated-sprint ability in professional and amateur soccer players. *Applied Physiology, Nutrition, and Metabolism*, *34*(6), 1048–1054. https://doi.org/10.1139/H09-111

Richter, C., O'Reilly, M., & Delahunt, E. (2021). Machine learning in sports science: Challenges and opportunities. *Sports Biomechanics*, 1–7. https://doi.org/10.1080/14763141.2021.1910334

Rienhoff, R., Hopwood, M., Fischer, L., Strauss, B., Baker, J., & Schorer, J. (2013). Transfer of motor and perceptual skills from basketball to darts. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00593

Roocks, P. (2016). Computing Pareto frontiers and database preferences with the rPref package. *The R Journal*, *8*(2), 393–404. https://doi.org/10.32614/RJ-2016-054

Rudsits, B., Hopkins, W., Hautier, C., & Rouffet, D. (2018). Force-velocity test on a stationary cycle ergometer: Methodological recommendations. *Journal of Applied Physiology*, *124*(4), 831–839. https://doi.org/10.1152/japplphysiol.00719.2017

Russell, M., Sparkes, W., Northeast, J., Cook, C., Love, T., Bracken, R., & Kilduff, L. (2016). Changes in acceleration and deceleration capacity throughout professional soccer match-play. *The Journal of Strength and Conditioning Research*, *30*(10), 2839–2844. https://doi.org/10.1519/jsc.0000000000000805

Sainani, K. (2018). The problem with "magnitude-based inference". *Medicine and Science in*

Sports and Exercise, *50*(10), 2166–2176. https://doi.org/10.1249/MSS.0000000000001645

Sainani, K., Borg, D., Caldwell, A., Butson, M., Tenan, M., Vickers, A., Vigotsky, A., Warmenhoven, J., Nguyen, R., & Lohse, K. (2021). Call to increase statistical collaboration in sports science, sport and exercise medicine and sports physiotherapy. *British Journal of Sports Medicine*, *55*(2), 118–122. http://doi.org/10.1136/bjsports-2020-102607

Sánchez-García, M., Sánchez-Sánchez, J., Rodríguez-Fernández, A., Solano, D., & Castillo, D. (2018). Relationships between sprint ability and endurance capacity in soccer referees. *Sports*, *6*(2), 28. https://doi.org/10.3390/sports6020028

Santos-Fernandez, E., Wu, P., & Mengersen, K. (2019). Bayesian statistics meets sports: A comprehensive review. *Journal of Quantitative Analysis in Sports*, *15*(4), 289–312. https://doi.org/10.1186/s12888-018-1761-4

Scott, J., Hill, S., Barwood, D., & Penney, D. (2021). Physical literacy and policy alignment in sport and education in Australia. *European Physical Education Review*, *27*(2), 328–347. https://doi.org/10.1177/1356336X20947434

Seshadri, D., Thom, M., Harlow, E., Gabbett, T., Geletka, B., Hsu, J., Drummond, C., Phelan, D., & Voos, J. (2021). Wearable technology and analytics as a complementary toolkit to optimize workload and to reduce injury burden. *Frontiers in Sports and Active Living*, *2*. https://doi.org/10.3389/fspor.2020.630576

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. (2020). *Afex: Analysis of factorial experiments*. https://CRAN.R-project.org/package=afex

Speed, H., & Andersen, M. (2000). What exercise and sport scientists don't understand.

Journal of Science and Medicine in Sport, *3*(1), 84–92. https://doi.org/10.1016/S1440-2440(00)80051-1

Stølen, T., Chamari, K., Castagna, C., & Wisløff, U. (2005). Physiology of soccer: An update. *Sports Medicine (Auckland, N.Z.)*, *35*(6), 501–536. https://doi.org/10.2165/00007256-200535060-00004

Suarez-Arrones, L., Portillo, J., Pareja-Blanco, F., Villareal, E. S. de, Sánchez-Medina, L., & Munguía-Izquierdo, D. (2014). Match-play activity profile in elite women's rugby union players. *The Journal of Strength and Conditioning Research*, *28*(2), 452–458. https://doi.org/10.1519/jsc.0b013e3182999e2b

Tapia, M. G. C., & Coello, C. A. C. (2007). *Applications of multi-objective evolutionary algorithms in economics and finance: A survey*. 532–539. https://doi.org/10.1109/CEC.2007.4424516

Thornton, H., Armstrong, C., Rigby, A., Minahan, C., Johnston, R., & Duthie, G. (2020). Preparing for an Australian Football League Women's League season. *Frontiers in Sports and Active Living*, *2*, 216. https://doi.org/10.3389/fspor.2020.608939

Thornton, H., Delaney, J., Duthie, G., & Dascombe, B. (2019). Developing athlete monitoring systems in team sports: Data analysis and visualization. *International Journal of Sports Physiology and Performance*, *14*(6), 698–705. https://doi.org/10.1123/ijspp.2018-0169

Thornton, H., Nelson, A., Delaney, J., Serpiello, F., & Duthie, G. (2019). Interunit reliability and effect of data-processing methods of global positioning systems. *International Journal of Sports Physiology and Performance*, *14*(4), 432–438. https://doi.org/10.1123/ijspp.2018-0273

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

Till, K., Cobley, S., Morley, D., O'hara, J., Chapman, C., & Cooke, C. (2016). The influence of age, playing position, anthropometry and fitness on career attainment outcomes in rugby league. *Journal of Sports Sciences*, *34*(13), 1240–1245. https://doi.org/10.1080/02640414.2015.1105380

Turner, A., Jones, B., Stewart, P., Bishop, C., Parmar, N., Chavda, S., & Read, P. (2019). Total score of athleticism: Holistic athlete profiling to enhance decision-making. *Strength & Conditioning Journal*, *41*(6), 91–101. https://doi.org/10.1519/SSC.0000000000000506

Vaeyens, R., Lenoir, M., Williams, M., & Philippaerts, R. (2008). Talent identification and development programmes in sport. *Sports Medicine*, *38*(9), 703–714. https://doi.org/10.2165/00007256-200838090-00001

Varley, M., Elias, G., & Aughey, R. (2012). Current match-analysis techniques' underestimation of intense periods of high-velocity running. *International Journal of Sports Physiology and Performance*, *7*(2), 183–185. https://doi.org/10.1123/ijspp.7.2.183

Vigh-Larsen, J., Dalgas, U., & Andersen, T. (2018). Position-specific acceleration and deceleration profiles in elite youth and senior soccer players. *The Journal of Strength and Conditioning Research*, *32*(4), 1114–1122. https://doi.org/10.1519/JSC.0000000000001918

Waldron, M., Highton, J., Daniels, M., & Twist, C. (2013). Preliminary evidence of transient fatigue and pacing during interchanges in rugby league. *International Journal of Sports Physiology and Performance*, *8*(2), 157–164. https://doi.org/10.1123/ijspp.8.2.157

Weaving, D., Young, D., Riboli, A., Jones, B., & Coratella, G. (2022). The maximal intensity period: Rationalising its use in team sports practice. *Sports Medicine-Open*, *8*(1), 1–9. https:doi.org/10.1186/s40798-022-00519-7

Wells, G., Elmi, M., & Thomas, S. (2009). Physiological correlates of golf performance. *The Journal of Strength and Conditioning Research*, *23*(3), 741–750. https://doi.org/10.1519/JSC.0b013e3181a07970

Welsh, A., & Knight, E. (2015). "Magnitude-based inference": A statistical review. *Medicine and Science in Sports and Exercise*, *47*(4), 874–884. https://doi.org/10.1249/MSS.0000000000000451

Whitehead, M. (2010). *Physical literacy* (pp. 8–19). Routledge. https://doi.org/10.4324/9780203881903

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. https://ggplot2.tidyverse.org

Wickham, H. (2021). *Tidyr: Tidy messy data*. https://CRAN.R-project.org/package=tidyr

Wickham, H., François, R., Henry, L., & Müller, K. (2021). *Dplyr: A grammar of data manipulation*. https://CRAN.R-project.org/package=dplyr

Yarkoni, T. (2019). The generalizability crisis. *Behavioral and Brain Sciences*, 1–37. https://doi.org/10.1017/S0140525X20001685

Zondervan-Zwijnenburg, M., Peeters, M., Depaoli, S., & Van de Schoot, R. (2017). Where do priors come from? Applying guidelines to construct informative priors in small sample research. *Research in Human Development*, *14*(4), 305–320. https://doi.org/10.1080/15427609.2017.1370966

A Pareto Frontiers for Multivariate Cricket Performance

A.1 Abstract

In Twenty20 cricket, there is a trade-off relationship between batting average and strike rate as well as bowling strike rate, economy, and average. This study presents Pareto frontiers as a tool to identify athletes who possess an optimal ranking when considering multiple metrics simultaneously. 884 matches of Twenty20 cricket from the Indian Premier League were compiled to determine the best batting and bowling performances, both within a single innings and across each player's career. Pareto frontiers identified nine optimal batting innings and six batting careers. Pareto frontiers also identified three optimal bowling and five optimal bowling careers. Each frontier identified players that were not the highest ranked athlete in any metric when analysed univariately. Pareto frontiers can be used when assessing talent across multiple metrics, especially when these metrics may be conflicting or uncorrelated. Pareto frontiers can

identify athletes that may not have the highest ranking on a given metric but have an optimal balance across multiple metrics that are associated with success in a given sport.

> ℹ The following chapter is a copy of the published manuscript:
>
> **Newans, T.**, Bellinger, P., & Minahan, C. Identifying multivariate cricket performance using Pareto frontiers. MathSport Conference 2022.
>
> As co-author of the paper "Identifying multivariate cricket performance using Pareto frontiers", I confirm that Timothy Newans has made the following contributions:
>
> - Study concept and design
> - Data collection
> - Data analysis and interpretation
> - Manuscript preparation
>
> Name: Clare Minahan
>
> Date: 29/03/2023

A.2 Introduction

The need to identify attributes to quantify optimal performance is evident for every sport (K. Johnston et al., 2018). With the exception of a few single-skill sports (Rienhoff et al., 2013), most athletes require a number of attributes to perform in their given sport. These attributes can encompass physical (Kelly & Williams, 2020), physiological (Dodd & Newans, 2018), mental (Morris, 2000), or skill-based characteristics (Davids et al., 2000), that all can contribute to the performance of a player. Attributes such as speed, endurance, agility, strength, power, and accuracy are common across multiple sports (Davids et al., 2000), and each attribute can have multiple variables seeking to quantify that attribute. As such, coaches and support staff are consistently looking for new variables that could be used to either quantify new attributes of interest or develop more variables to better quantify already-identified attributes with the hope that these new variables can identify previously-hidden talent or interrogate subtle differences between different athletes. However, with the increase in the number of attributes of interest, the likelihood that an athlete excels in every attribute decreases. Consequently, methods are required that can analyse multiple attributes simultaneously, rather than viewing each attribute in isolation.

While traditional research statistical techniques focus around identifying the mean and standard deviation of a population (Hopkins et al., 2009), sports typically are not interested in the mean during talent identification processes, rather, they are looking for outliers. That is, coaches and support staff are looking for athletes that sit the furthest away from the mean in the direction that success is defined. Therefore, when multiple attributes are of interest, selection of athletes is by choosing athletes that sit the further away from the mean within each attribute. While this process can work when variables are positively correlated, this process can miss talent when variables are negatively correlated. For instance, at the elite level, there is a negative correlation between maximal sprint speed and endurance capacity

(Sánchez-García et al., 2018). However, running-based team sports require athletes possess both speed and endurance to play at the elite level and, therefore, players necessarily need to trade off between having optimal speed and optimal endurance. In its simplicity, if both speed and endurance were equally required for success, selecting the top-n sprinters and the top-n endurance runners may not be the optimal athletes for that sport.

Consequently, both attributes need to be viewed in tandem. The process of optimising the balance of multiple attributes is termed 'multi-objective optimisation'. Mathematically, they aim to create the perfect balance of the attributes of interest. If a data point was defined as: $\vec{x_1}\,\varepsilon\, X$, it is, therefore, better than another data point defined by: $\vec{x_2}\,\varepsilon\, X$ if $f_i(\vec{x_1}) \leq f_i(\vec{x_2})$ for all metrics $i\ \varepsilon\ \{1,\ 2,\ ...,\ k\}$ and $f_i(\vec{x_1}) < f_i(\vec{x_2})$ for at least one metric $j\ \varepsilon\ \{1,\ 2,\ ...,\ k\}$. Once these conditions have been met, the remaining points are deemed Pareto-optimal and form what is called the Pareto frontier.

In Twenty20 cricket, there are multiple facets within both batting and bowling that can define success. Unlike Test cricket and, to an extent, One-Day cricket where scoring as many runs as possible regardless of how many deliveries faced is of most importance, Twenty20 crickets requires batters to score faster (i.e., higher strike rate) and for bowlers to concede minimal runs which, in some cases, can come at the expense of preserving their wicket. Therefore, there is a trade-off relationship between batting average and strike rate as well as bowling economy, average, and strike rate within Twenty20 cricket. For example, early on in an innings the risk-return of attempting to hit six runs off a ball is significantly different than in the final over of an innings. Similarly, a bowler needs to balance taking wickets while also conceding as few runs as possible. For instance, when bowling four overs, it is again difficult to determine whether taking three wickets for 50 runs is of more worth than taking no wickets but only conceding eight runs as the three wickets may not have been worth conceding 50 runs. As both attributes within each domain are of interest, Pareto frontiers can be used to determine batters and bowlers that may not record the highest in either variable but display an optimal balance of the two attributes. Therefore, when assessing the quality of players, it is necessary to utilise tools that can analyse these data sets without favouring one metric over another. Therefore, the present study aimed to use Pareto frontiers to identify the best performing Twenty20 batters and bowlers.

A.3 Methods

The present study comprised all 884 matches of the first 14 editions of the men's Indian Premier League (IPL), India's domestic T20 cricket competition. The data set contained 566 batters and 467 bowlers. Collectively, there were 13,357 individual batting innings with observations ranging from 1-208 innings per batter, while there were 10,925 individual bowling innings with observations ranging from 1-180 innings per bowler.

```
library(tidyverse)
library(rPref)
library(patchwork)
library(scatterplot3d)
```

```
options(scipen = 999)

iplbat <- read_csv('www/data/Study_A_pareto_iplbat.csv') # Men's batting scorecards

iplbowl <- read_csv('www/data/Study_A_pareto_iplbowl.csv') # Men's bowling scorecards
```

To summarise the data, two summary statistics were generated for batting and three summary statistics were generated for bowling. The summary statistics were as follows:

- Batting Average: runs scored divided by frequency of dismissal
- Batting Strike Rate: runs scored divided by balls faced
- Bowling Average: runs conceded divided by wickets taken
- Bowling Strike Rate: balls bowled divided by wickets taken
- Bowling Economy: runs conceded divided by overs (i.e., 6 balls) bowled

```
dismissals <- iplbat %>%
  group_by(id) %>%
  filter(Dismissed == T) %>%
  summarise(Dismissals = n()) ## Calculate number of dismissals


notouts <- iplbat %>%
  group_by(id) %>%
  filter(Dismissed == F) %>%
  summarise(NotOuts = n()) ## Calculate number of not outs
```

```
sumBat <- iplbat %>%
  group_by(id, Batter, LastName) %>%
  summarise(
    TotalRuns = sum(R),
    TotalBalls = sum(B),
    Innings = n()
  ) %>%
  ungroup() %>%
  left_join(dismissals) %>%
  left_join(notouts) %>%
  mutate(
    Dismissals = case_when(is.na(Dismissals) ~ as.integer(0),
                           T ~ Dismissals),
    NotOuts = case_when(is.na(NotOuts) ~ as.integer(0),
                        T ~ NotOuts),
    Average = TotalRuns / Dismissals,
    StrikeRate = TotalRuns / TotalBalls * 100) ## Calculate career batting average and str

filtBat <- sumBat %>%
  filter(Innings >= 20) ## Filter only those with 20 or more batting innings
```

```r
sumBowl <- iplbowl %>%
  group_by(id, Bowler, LastName) %>%
  summarise(
    Innings = n(),
    Balls = sum(Balls),
    Wickets = sum(W),
    Runs = sum(R)
  ) %>%
  mutate(
    Average = Runs / Wickets,
    Economy = Runs / Balls * 6,
    StrikeRate = Balls / Wickets
  ) ## Calculate bowling average, economy, and strike rate

filtBowl <- sumBowl %>%
  filter(Innings > 20) %>%
  ungroup() ## Filter only those with more than 20 bowling innings
```

To understand both the batting and bowling attributes within cricket, four Pareto frontiers for were established within the data set:

i. *Pareto-optimal Batting Innings*

This analysis outlined the highest runs scored within an innings at the highest strike rate.

```
BatInnPareto <- psel(iplbat %>% filter(R > 0),high(R)*high(SR),top_level = 999)
```

ii. *Pareto-optimal Batting Career*

This analysis outlined the highest batting average across a career at the highest strike rate. To provide a more accurate career report, batters required to have played a minimum of 20 innings which left 163 eligible batters.

```
BatCarPareto <- psel(filtBat,high(Average)*high(StrikeRate), top_level = 999) %>%
      filter(Average > 20 | StrikeRate > 100)
```

iii. *Pareto-optimal Bowling Innings*

This analysis outlined the most wickets taken in an innings at the lowest economy.

```
BowlInnPareto <- psel(iplbowl,high(W)*low(Econ),top_level = 999)
```

iv. *Pareto-optimal Bowling Career*

This analysis outlined the lowest bowling average across a career at the lowest economy and lowest strike rate. To provide a more accurate career report, bowlers required to have bowled in more than 20 matches, which left 145 eligible bowlers.

```
BowlCarPareto <- psel(filtBowl,low(Average)*low(StrikeRate)*low(Economy),top_level = 999)
      filter(Average < 50)
```

The *rPref* package (Roocks, 2016) was used in R v 4.1.0 (R Core Team, 2019) to determine the Pareto frontiers using the psel function with the 'top_level' argument set to 999 to ensure every athlete was assigned to a frontier.

A.4 Results

Pareto-optimal Batting Innings

Nine Pareto-optimal innings were identified with extremities ranging from 6 runs off 1 ball (i.e., strike rate = 600) to 175 off 66 balls (i.e., strike rate = 265.15). Additionally, the solution of 6 runs off 1 ball has been attained eight times. The IPL batting innings Pareto frontier is displayed in Figure A.1 and the batters are listed in Table A.1.

```
ggplot(BatInnPareto, aes(x = R, y = SR)) +
    geom_point(alpha=0.1, size = 3) +
    geom_text(data = BatInnPareto %>% filter(.level == 1 & R != 6 & !(LastName %in% c("de
    geom_text(data = BatInnPareto %>% filter(.level == 1 & LastName == "de Villiers"), aes
    geom_text(data = BatInnPareto %>% filter(.level == 1 & LastName == "Pollard"), aes(lab
    geom_text(data = BatInnPareto %>% filter(.level == 1 & LastName == "Russell"), aes(lab
    geom_text(data = BatInnPareto %>% filter(.level == 1 & LastName == "Pathan"), aes(labe
```

```
geom_text(data = BatInnPareto %>% filter(.level == 1 & LastName == "Miller"), aes(labe
geom_text(data = BatInnPareto %>% filter(.level == 1 & LastName == "Gayle"), aes(label
geom_line(data = BatInnPareto %>% filter(.level == 1), alpha = 0.5, colour = "darkgree
geom_line(data = BatInnPareto %>% filter(.level == 1), alpha = 0.5, colour = "red")+
theme_minimal() +
annotate(geom = "text",x = 18,y = 600, label = "8 players",color = "red")+
scale_y_continuous(breaks = seq(100,600,100))+
coord_cartesian(xlim = c(0,180))+
theme_minimal() +
labs(x = "Runs Scored in an Innings",
     y = "Innings Batting Strike Rate") +
theme(axis.title = element_text(size = 16),
      panel.grid.minor.y = element_blank(),
      axis.text = element_text(size = 16, color = "black"))
```

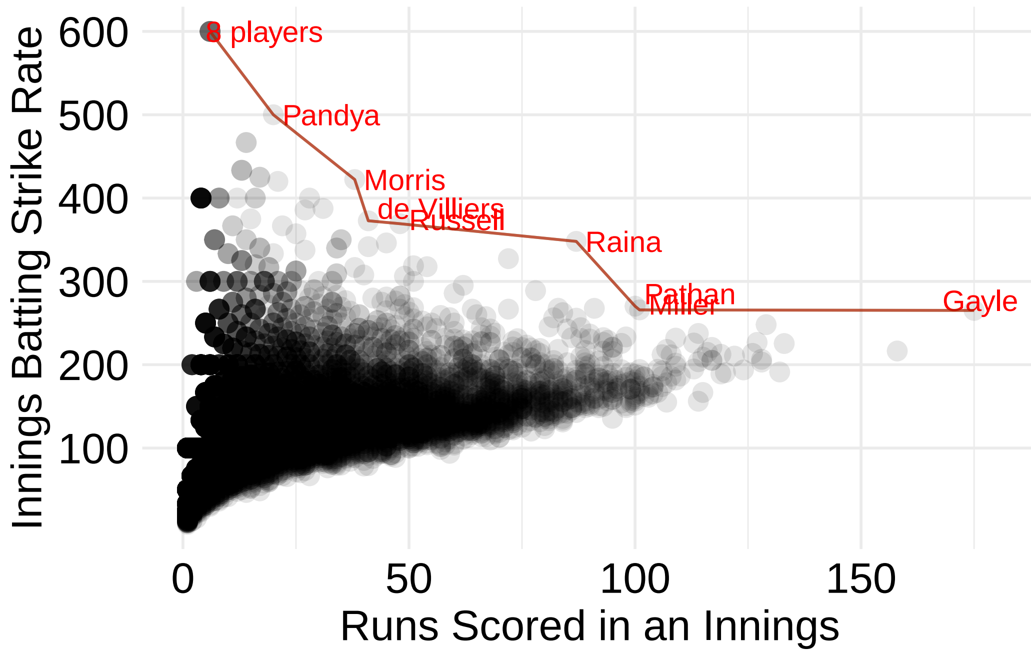

Figure A.1: Pareto-optimal batting within an innings with the Pareto frontier highlighted in red. N.B. For illustrative purposes, points were filtered out if both their runs scored was below 50 and their strike rate was below 100.

```
BatInnPareto %>%
  filter(.level == 1) %>%
  mutate(SR = round(SR,2)) %>%
  select(Batter,R,B,SR,Season,Match = Season.Match.No) %>%
  arrange(-R)
```

Table A.1: List of all Pareto-optimal IPL batting within an innings.

Batter	R (B)	Strike Rate	Match
Chris Gayle	175 (66)	265.15	IPL06 Match 31

Batter	R (B)	Strike Rate	Match
David Miller	101 (38)	265.78	IPL06 Match 51
Yusuf Pathan	100 (37)	270.27	IPL03 Match 2
Suresh Raina	87 (25)	348.00	IPL07 Match 59
Andre Russell	48 (13)	369.23	IPL12 Match 17
AB de Villiers	41 (11)	372.72	IPL08 Match 16
Chris Morris	38 (9)	422.22	IPL10 Match 9
Krunal Pandya	20 (4)	500.00	IPL13 Match 17
Numerous	6 (1)	600.00	IPL04 Match 74 [1st occurrence]

Pareto-optimal Batting Career

Six Pareto-optimal batting careers innings were identified. Andre Russell recorded the highest career batting strike rate with 178.57 runs per 100 balls, while KL Rahul recorded the highest batting average with 47.43 runs per dismissal. The IPL batting career Pareto frontier is displayed in Figure A.2 and the batters are listed in Table A.2.

```
ggplot(BatCarPareto, aes(x = Average, y = StrikeRate)) +
    geom_point(data = BatCarPareto %>% filter(.level != 1), alpha=0.3, size = 3) +
    geom_point(data = BatCarPareto %>% filter(.level == 1), color = "red", size = 3)+
    geom_line(data = BatCarPareto %>% filter(.level == 1), color = "red")+
    geom_text(data = BatCarPareto %>% filter(.level == 1 & !(LastName %in% c("de Villiers"
```

```
geom_text(data = BatCarPareto %>% filter(.level == 1 & LastName == "Warner"), aes(labe
geom_text(data = BatCarPareto %>% filter(.level == 1 & LastName == "Bairstow"), aes(la
geom_text(data = BatCarPareto %>% filter(.level == 1 & LastName == "de Villiers"), aes
theme_minimal()  +
scale_x_continuous(limits = c(2,49),breaks = seq(from = 0,to = 50,by = 10))+
labs(x = "Career Batting Average",
     y = "Career Batting Strike Rate") +
theme(axis.title = element_text(size = 16),
      legend.position = "none",
      axis.text = element_text(size = 16, color = "black"))
```

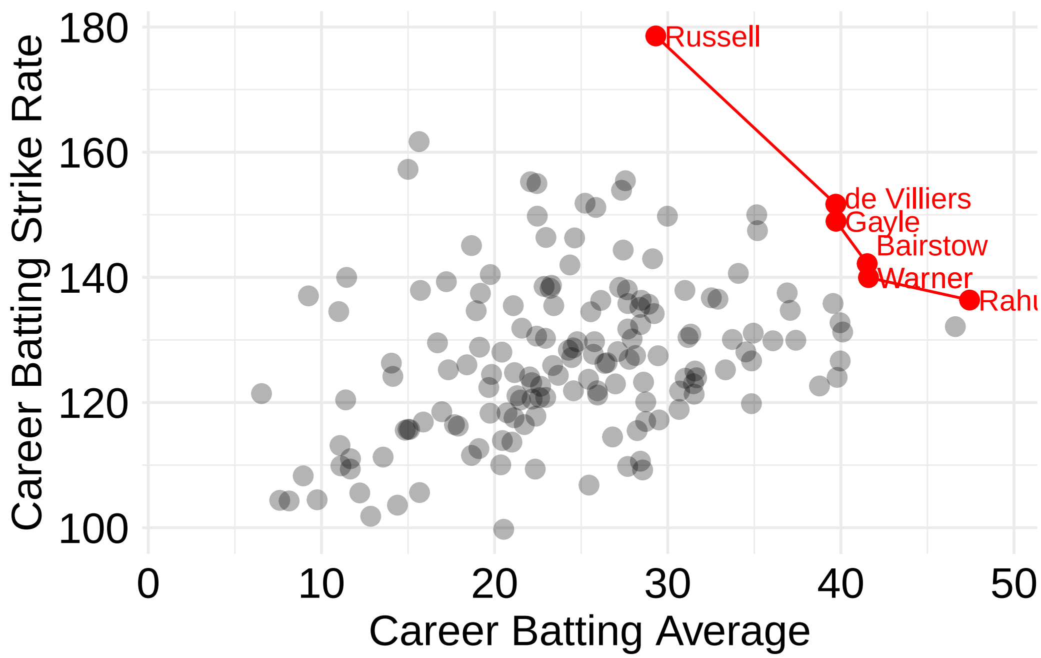

Figure A.2: Pareto-optimal batting across a career with the Pareto frontier highlighted in red. N.B. For illustrative purposes, points were filtered out if both their average was below 20 and their strike rate was below 100.

```
BatCarPareto %>%
  filter(.level == 1) %>%
  select(Batter,Innings,Average,`Strike Rate` = StrikeRate) %>%
  mutate(across(c(Average,`Strike Rate`),round,2)) %>%
  arrange(-Average)
```

Table A.2: List of all Pareto-optimal IPL batting careers.

Batter	Innings	Average	Strike Rate
KL Rahul	85	47.43	136.38

Batter	Innings	Average	Strike Rate
David Warner	150	41.60	139.97
Jonny Bairstow	28	41.52	142.19
Chris Gayle	141	39.72	148.96
AB de Villiers	170	39.71	151.69
Andre Russell	70	29.31	178.57

Pareto-optimal Bowling Innings

Three Pareto-optimal bowling innings were identified: 2/0 by Suresh Raina, 5/5 by Anil Kumble, and 6/12 achieved by Alzarri Joseph. The IPL bowling innings Pareto frontier is displayed in Figure A.3 and the bowlers are listed in Table A.3.

```
ggplot(BowlInnPareto, aes(x = W, y = Econ)) +
  geom_jitter(data = BowlInnPareto %>% filter(.level != 1),aes(x = W, y = Econ), alpha=0.1
  geom_point(data = BowlInnPareto %>% filter(.level == 1), aes(x = W, y = Econ), alpha = 0
  geom_text(data = BowlInnPareto %>% filter(.level == 1), aes(x = W, y = Econ, label = Las
  geom_line(data = BowlInnPareto %>% filter(.level == 1), aes(x = W, y = Econ), alpha = 0.
  theme_minimal() +
  labs(x = "Wickets in an innings",
       y = "Innings Bowling Economy") +
    coord_cartesian(xlim = c(0,6.3))+
```

```r
  theme(axis.title = element_text(size = 16),
        legend.position = "none",
        panel.grid.minor.y = element_blank(),
        axis.text = element_text(size = 16, color = "black"))
```

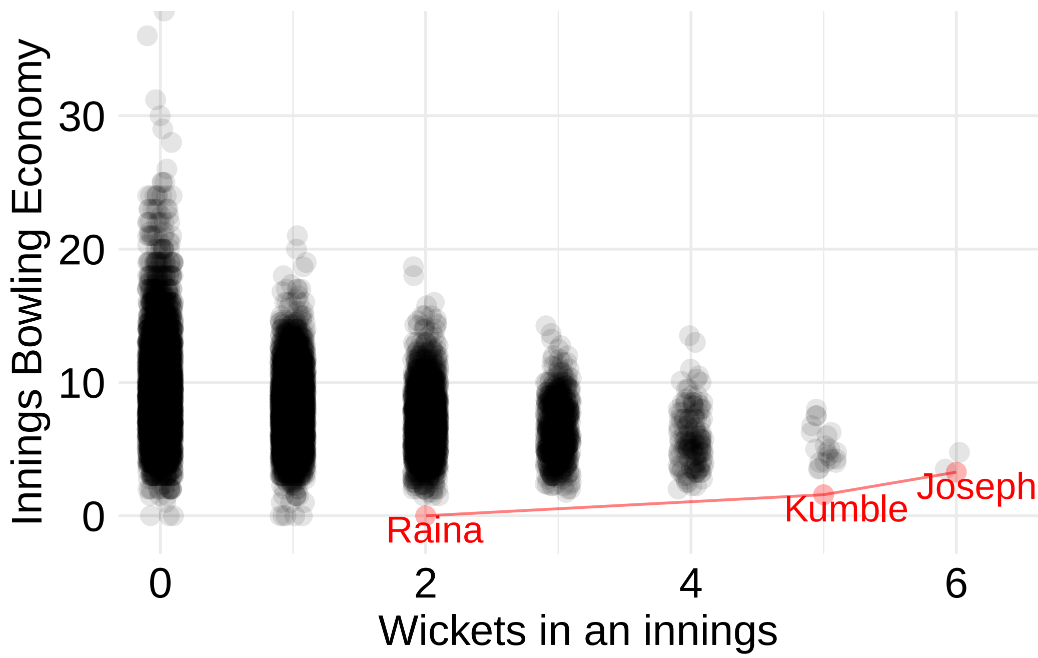

Figure A.3: Pareto-optimal bowling within an innings with the points on the Pareto frontier highlighted in red.

```r
BowlInnPareto %>%
  filter(.level == 1) %>%
  select(Bowler, Overs = O, Wickets = W, Runs = R, Season, Match = Season.Match.No) %>%
  arrange(Wickets)
```

Table A.3: List of all Pareto-optimal IPL bowling innings.

Bowler	Overs	Wickets	Runs	Match
Suresh Raina	0.3	2	0	IPL04 Match 52
Anil Kumble	3.1	5	5	IPL02 Match 2
Alzarri Joseph	3.4	6	12	IPL12 Match 19

Pareto-optimal Bowling Career

Five Pareto-optimal bowling careers were identified, with Doug Bollinger achieving the lowest average, Rashid Khan achieving the lowest economy, while Kagiso Rabada recorded the lowest strike rate. The IPL bowling career Pareto frontier is displayed in Figure A.4 and the bowlers are listed in Table A.4.

```
BowlCarPareto$color <- case_when(BowlCarPareto$.level == 1 ~ 2,
                                 BowlCarPareto$.level > 1 ~ 1)
BowlCarPareto$Label[BowlCarPareto$.level == 1] <- BowlCarPareto$LastName[BowlCarPareto$.le
BowlCarParetoPlot <-scatterplot3d(BowlCarPareto[c("Economy","Average","StrikeRate")], type
                              xlab="Career Bowling Economy",
                              ylab="Career Bowling Strike Rate",
                              zlab="Career Bowling Average")
zz.coords <- BowlCarParetoPlot$xyz.convert(BowlCarPareto$Economy, BowlCarPareto$Average, B
text(zz.coords$x,
```

```
    zz.coords$y,
    labels = BowlCarPareto$Label,
    cex = .8,
    pos = 2,
    col = "red")
```

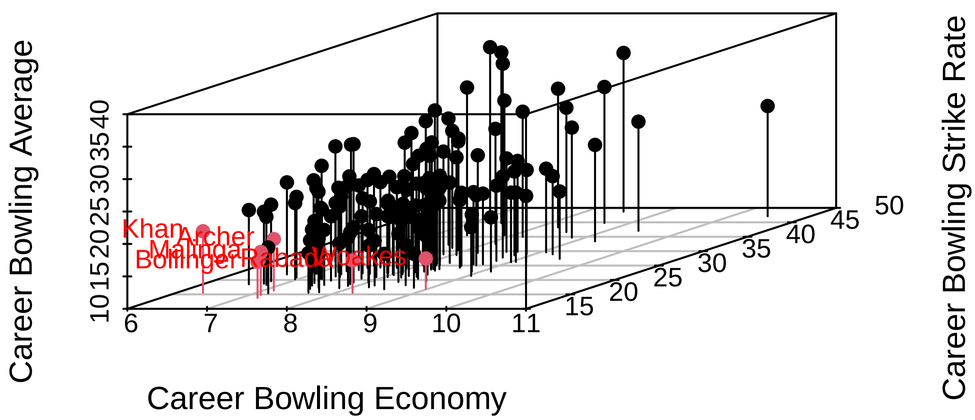

Figure A.4: Pareto-optimal bowling across a career with the points on the Pareto frontier highlighted in red.

```
BowlCarPareto %>%
  filter(.level == 1) %>%
  select(Bowler, Innings, Average, Economy, `Strike Rate` = StrikeRate) %>%
  mutate(across(c(Average:`Strike Rate`),round,2)) %>%
  arrange(Average)
```

Table A.4: List of all Pareto-optimal IPL bowling careers.

Bowler	Innings	Average	Economy	Strike Rate
Doug Bollinger	27	18.73	7.22	15.57
Kagiso Rabada	59	19.71	8.22	14.39
Lasith Malinga	122	19.79	7.14	16.63
Jofra Archer	35	21.33	7.13	17.93
Rashid Khan	86	21.46	6.40	20.12

A.5 Discussion

This study sought to use Pareto frontiers to visualise optimal Twenty20 cricket batting and bowling performances, both within an innings as well as across a career. By analysing performance multivariately, rather than simply analysing multiple variables univariately, players can be deemed optimal despite not being objectively highest in a single variable. When conflicting attributes are of equal interest, Pareto frontiers can view these variables in tandem as the expectations of an individual to attain the highest level in both attributes univariately may be unfeasible. All four Pareto frontiers contained at least one athlete that was not the highest ranked athlete in any metric when analysed univariately, and yet was deemed Pareto-optimal due to their balance in the metrics of interest.

The main advantage of Pareto frontiers highlighted in the present study is identifying athletes who are optimal across multiple metrics even when they are not the highest ranked in any metric. This was most evident where Chris Gayle, when viewed univariately, has the 9th-highest career batting average (39.72), which is 6.71 runs per innings lower than the highest (Figure 2). Similarly, he has the 14th-highest strike rate, striking at 148.96 which is 29.61 runs per 100 balls lower than the highest. However, when considering both metrics simultaneously and visualising these metrics, he is one of the best batsmen across the 14 seasons of the IPL.

The present study also illustrated how Pareto frontiers can be used to visualise talent in more than 2 dimensions. For example, while Jofra Archer has the sixth-lowest bowling average, 14th-lowest economy, and the 19th-lowest strike rate (see Figure 4), he can be deemed a Pareto-optimal bowler as there are no other bowlers who supersede him across all three metrics. While there will be some correlations between the three bowling metrics (i.e., average, economy, and strike rate) as the metrics are related (e.g., wickets taken is the denominator of average and numerator of strike rate), visualising the third dimension is still necessary as the reader would still need to multiply the x and y values to understand where they would sit in the third dimension.

In the present study we chose to observe batting and bowling as purely independent roles within cricket; however, there are also avenues for Pareto frontiers to be established for all-rounders within cricket (i.e., players that are picked for both their batting and bowling ability). However, it should be noted that if an all-rounder Pareto frontier were to be established with both batting average and strike rate as well as bowling average, economy, and strike rate, the resulting five-dimensional outputs, while valid and executable, become increasingly difficult to interpret and visualise. To do such an analysis, a factor-reduction technique such as principal components analysis should be considered and the Pareto frontier could be built from the extracted components (e.g., batting and bowling).

While the present study is designed to be an introduction for sports scientists to the concept of Pareto frontiers, it should also be considered that there is some level of uncertainty surrounding each observation in the career Pareto frontiers due to the differing number of observations. For example, Jonny Bairstow is deemed Pareto-optimal as he is currently striking at 142.19 at an average of 41.52 after 28 innings; however, it is right to assume that it is more uncertain that he lies on the frontier than AB de Villiers who has 170 observations. Therefore, future research could consider providing confidence or credible intervals around the probability that an individual lies on the Pareto frontier. Consequently, it is then feasible that a probability that an individual sits on the first, second, or third frontier could be calculated.

While the present study used Twenty20 cricket to illustrate the power and usefulness of Pareto frontiers, the concept can be widely applied within sports science data sets, especially when the variables of interest are uncorrelated or negatively correlated. Pareto frontiers can still be established between two positively correlated metrics; however, it is likely that there will be less 'hidden' athletes on this frontier as naturally the athletes who are high in one metric will be high in the other metric. Future research should apply Pareto frontiers across different avenues within sports performance analysis which have multi-faceted determinants as there are many other possibilities within sports whereby Pareto frontiers can reveal athletes who possess the optimal balance of the metrics of interest.

A.6 Conclusion

With the proliferation of various physiological, mechanical, and skill-related attributes associated with performance, Pareto frontiers should be used within sports science to visualise multiple performance metrics. By analysing opposing data in tandem, more feasible expectations and benchmarks can be established to reveal talent that may have been missed when analysing multiple metrics univariately.

B Downloadable Data Sets

Study 1:

- NRLW mean speed mixed model data set

Study 2:

- Men's BBL batting data set
- Women's BBL batting data set
- Men's BBL bowling data set
- Women's BBL bowling data set

Study 3:

- -Alanine Bayesian data set

Study 4:

- NRLW Full Match data set

- NRLW Half-By-Half data set

Study 5:

- Bayesian Mixed Model Pareto data set

Appendix A:

- IPL batting data set
- IPL bowling data set

C Downloadable R Scripts

Literature Review: Literature review code

Study 1: Study 1 code

Study 2: Study 2 code

Study 3: Study 3 code

Study 4: Study 4 code

Study 5: Study 5 code

Discussion: Discussion code

Appendix A: Appendix A code