# Proceedings Book
# 17th Australasian Conference on Mathematics and Computers in Sport
# (ANZIAM Mathsport 2024)

La Trobe University

Melbourne and Bundoora Campuses

3-5 July 2024

Edited by Ray Stefani and Adrian Schembri

ISBN 978-0-646-86238-5

# Keynote Speakers

## Steven Stern

Steven Stern has conducted teaching and research as a Professor of Data Science at Bond University since 2017, as an ABS Chair of Statistics at Queensland University of Technology from 2013 to 2016 and as a faculty member of the Australian National University for two decades. His teaching excellence has been recognized by Stanford University in 1993, by ANU in 2002, by Bond University in 2019, and by the Australian Teaching and Learning Council and Universities Australia for his work in statistical and data science education. He has also conducted research as a data analyst and consultant for numerous government departments and industry bodies, covering data analytics and statistical methodology. He is probably best known for his research into sports statistics, including developing both team and player ranking methods and as the co-creator, and official custodian since 2014, of the Duckworth-Lewis-Stern method, the official approach to handling interruptions in international and most high-profile domestic limited-over cricket matches and competitions. He presented his research in person at ANZIAM MathSport conferences from 2008 to 2016 and online in 2020. He was a keynote speaker at the Inaugural MathSport Asia Conference in 2018, held at Jamshedpur, India.

## Joshua Kadlec-Cavanaugh

After completing a Bachelor of Science at RMIT in 2014, Joshua Kadlec-Cavanaugh worked as a quantitative analyst at BETIA Racing & Wagering Technologies, modelling thoroughbred & greyhound racing markets. In 2017, Josh was brought on in a casual capacity by the Sunshine Coast Lighting, during their inaugural season, to work in their performance analysis team. He was involved with coding game-day statistics, live-match reporting, and post-match reporting, so that players & coaching staff would be better informed. Josh has worked with Tennis Australia during the Australian Open, assisting in analysis of the shot-by-shot data, compiled during the tournament as well as dashboard visualisations to present to coaching staff for opponent analysis. Currently, Josh is Head of Data & Performance at Ciaron Maher Racing, Australia's leading horse racing stable, where data is used throughout the operation to better prepare horses for race-day, identifying quality bloodstock through biomechanics and pedigree analysis, optimising race-day placement and tactics, as well as assisting in large scale operations of the business which trains over 600 horses nationally.

## Darren O'Shaugnessy

Darren O'Shaughnessy is Head of Analytics and Strategy at St Kilda Football Club. 2024 is his tenth MathSport conference, and it has always been a hugely enjoyable forum with terrific people and ideas. Darren's degree was in theoretical and statistical physics (ANU) before working in parallel computing and at the Australian Artificial Intelligence Institute (1990s style!). With the late great Ted Hopkins, he built Champion Data's analysis suite, then consulted to Hawthorn through its premiership dynasty from 2012 onwards. This is his sixth season at St Kilda, who have finally been convinced to give the nerd a go at running strategy. He has worked with broadcasters, coaches, and analysts in twenty sports and always emphasises the importance of mathematics and randomness in analysis.

# Mathsport 2024 Proceedings Table of Contents

Momentum in Australian Football

Dan B. Dwyer [a,b], Ryan Woods [a], Christopher M. Young [a]

[a] *Centre for Sport Research, Deakin University*
[b] *Corresponding author: dan.dwyer@deakin.edu.au*

**ABSTRACT**

Anecdotally, the term "momentum" is in common use when a team's performance is discussed by spectators and coaches. While momentum, Adler & Adler (1978), has been explored scientifically in some sports (Fry & Shukairy, 2012; Pill, 2013; and Weimer et al., 2023), it has not been defined or analysed in Australian Football (AF). The purpose of this study was to compare the utility of adapted published definitions, and a novel definition of momentum in AF. Four definitions of momentum were used including (1) possession share, (2) effective disposal count, (3) forward ball movement and (4) forward ball movement weighted by score. The value of each definition of momentum was calculated for each of 3,456 quarters from four seasons (2017-2021, not 2020) of the AFL. The usefulness of each definition was compared with respect to its association (Spearman) with score margin and the outcome of the quarter (Wilcoxon). All measures of momentum had a positive association with score margin. The momentum definition that had the strongest association ($r = 0.886$, $p < 0.05$) with score margin was forward ball movement weighted by scoring. Teams with more of this type of momentum won quarters 85.4% of the time. The other definitions of momentum were not as strongly associated with score margin, but momentum was different ($p < 0.001$, ES = 0.32-0.53)) depending on whether a team won or lost each quarter. This is the first study to provide empirical evidence of momentum in AF. The current study supports the notion that teams with more momentum have an increased likelihood of winning. The best performing definition of momentum reported here, provides an opportunity for future exploration of how teams can change or maintain momentum within a match.

*Keywords*: Performance analysis, consistent, dominance

## 1. INTRODUCTION

The concept of momentum in sport can refer to either a physical or mental state that can affect a team's performance or the outcome of the game. Momentum in sport is a complex phenomenon that has been widely discussed by the media, fans, athletes and coaches for decades. (1–5) Analysts within the media have reported and discussed the role of momentum on performance on a regular basis. When observing matches, listeners are also exposed to commentators often mentioning momentum being with one team or that the momentum within a match has shifted.

Currently, there are various opinions in the scientific community about the relationship between momentum and performance and the concept has remained elusive and unresolved. (6–10) Several researchers maintain that momentum is a fundamental factor that can influence performance, (1–

3,11–13) while others contend it is a perception created by the media and fans that might exist but is too difficult or impossible to identify quantitatively. (4,5,14–16)

An extensive literature search for peer-reviewed journal articles related to momentum and performance in Australian Football (AF) resulted in no current published articles discussing the concept in detail. Expanding this search to include non-peer-reviewed articles returned two conference papers from 2013 (19) and 2020, (20) which briefly mention momentum in AF. According to Pill, (19) momentum can be considered as a team's ability to control the interaction of the football in terms of mass of possession and ball movement. Josman et al. (20) attempted to understand the inherent causes of momentum and its potential effect on performance and match outcome; however, they do not clearly define momentum in the study. The lack of peer-reviewed journal articles was surprising given that momentum is so often discussed amongst several sportswriters, broadcasters and fans.

Due to the lack of research on momentum in AF, there is no definitive answer to what momentum is and whether it has any association with the outcome of a match. It is evident that this phenomenon needs to be explored in AF to provide coaches, players and performance staff with a working definition of momentum and an understanding of its importance. An understanding of momentum in AF might permit the development of strategies and tactics that assist in the establishment and or maintenance of momentum.

The aim of the current research is to create (or adapt from previous studies) definitions of momentum that can be functionally applied to AF and to determine which definition (if any) is most applicable in its association with performance and quarter outcome. The research questions include: 1) What is the best method of quantifying momentum? Is it associated with score margin? 2) Is there a difference in total team momentum between winning and losing quarter outcomes? 3) Can the methods that are addressed be implemented in real time?

## 2. METHODS

The Deakin University Human Research Ethics Committee (DUHREC) granted approval for an ethics modification (project number 2017-079).

### Data

Play-by-play data (i.e., the transactional list of events that occur during a match) from the Australian Football League (AFL) was provided by Champion Data™, the official statistical provider for the AFL. The data were collected across multiple seasons in the AFL (from 2017-2021). The 2020 season was excluded from the analysis due to the extensive adjustments in response to the COVID-19 pandemic (i.e., reduction in match duration, increase in match frequency, and match location), making it incomparable to previous years. There were 1.34M player actions and a total of 3,456 quarters across the four seasons of AFL. Each play/technical action in the dataset included attributes for which players were involved, the field location, and the outcome of the action.

The dataset contained metadata variables, including season, match ID number, match location, and period (quarter) of the match. The technical variables within the dataset that were used for this analysis included: team name, opposition, possession duration (measured in seconds), game events (effective disposals, goals and behinds), the direction in which the team is kicking towards, and the standard x and y location on the field. Forward ball movement (total metres gained) was calculated by the change in x location of the ball on the field:

$$Forward\ Ball\ Movement = \sum_{t=0-max\ time}^{t=time} x_{location} of\ ball_t - x_{location} of\ ball_{t-1}$$

Data cleaning and processing was performed in R Studio™ to simplify the dataset, removing variables that were irrelevant to the analysis. A total of 2.98M rows of raw data were removed from the data set. The data was then split into seasons, then into matches, and finally organised at the quarter level.

**Definitions of Momentum**

There were four different measures of momentum that were used for comparison in this study. The four that were created/adapted for use in this study were:

1) Possession share (time in possession) (18): Measured as a percentage, based on the time a team has possession of the ball compared to the other (e.g., team A had possession of the ball 60% in the first quarter compared to 50% for team B).
2) Effective disposal count: This definition is simply the total effective disposals a team has in a quarter. The player actions at each moment in a game is recorded by Champion Data, thus allowing the sum of those effective disposals to be calculated for each team.
3) Forward ball movement (metres gained) (16): Measured using the standard x location of the ball on the field as a team moves towards their goal line. The value represents the net movement of the ball toward the goal that the team is kicking toward. The location of the ball on the y-axis of the coordinate system is disregarded.
4) Forward ball movement weighted by scoring: Similar to the previous definition of momentum, however, it expands upon simple metres gained measurement by multiplying the distance (metres gained) by 1.1 for each behind and 1.6 for each goal.

$$Forward\ ball\ movement\ (weighted) = forward\ ball\ movement\ \times (behinds\ x\ 1.1) \times (goals\ x\ 1.6)$$

**Statistical Analysis**

Normality of distribution was performed using the Kolmogrov-Smirnov (K-S) test. This test indicated that the data was not normally distributed ($p<0.05$). Therefore, non-parametric tests were used for statistical analysis of all measures of momentum throughout the study. A comparison of medians (Wilcoxon paired sample test) was used to compare teams that won and lost for each definition of momentum. The correlation between each measure of momentum according to the score margin was analysed using the Spearman correlation coefficient (r). A Fisher's R-Z transformation was performed to compare the correlation for each definition of momentum. Finally, the proportion of teams that had more momentum when they outscored their opponent for the quarter was calculated (as a %).

## 3. RESULTS

A comparison of the median values of momentum between winning and losing teams for all four measures is outlined in Table 1. Momentum, as measured by possession share, between winning and losing teams was 0.531 (0.079) and 0.528 (0.078), respectively. When momentum is measured by effective disposal count, the median value for winning teams was 102 (19.0) compared to 91 (18.0) for losing teams. Forward ball movement resulted in a median of 991.2 (254.9) for winning teams and 858.2 (250.8) for losing teams. Whereas momentum as a measure of forward ball movement weighted by scoring, the median values for winning and losing teams were 8790.14 (3361.59) and 4354.63 (4531.38), respectively. The Wilcoxon test indicates that there was a significant difference between the median of winning and losing teams of each quarter for all four measures of momentum.

Table 1. Normality test and a comparison of medians for winning and losing teams for each measure of momentum

| | **Momentum Definition** | | | |
|---|---|---|---|---|
| | **Possession share** | **Effective disposals** | **Forward ball movement** | **Forward ball movement – weighted by score** |
| **Normality of distribution (K-S)** | D = 0.625, $p < 0.001$ | D = 1.0, $p < 0.001$ | D = 1.0, $p < 0.001$ | D = 0.994, $p <0.001$ |
| **Win – Loss Median (IQR)** | 0.531 (0.079) – 0.528 (0.078) | 102 (19) – 91 (18) | 991.2 (254.9) – 858.2 (250.8) | 8790.14 (3361.59) - 4354.63 (4531.38) |
| **Wilcoxon paired sample test (win-loss)** | Z = -26.808, $p <0.001$ | Z = -30.722, $p < 0.001$ | Z = -27.93, $p < 0.001$ | Z = -44.56, $p < 0.001$ |
| **Effect Size (Wilcoxon)** | 0.32 | 0.36 | 0.33 | 0.53 |

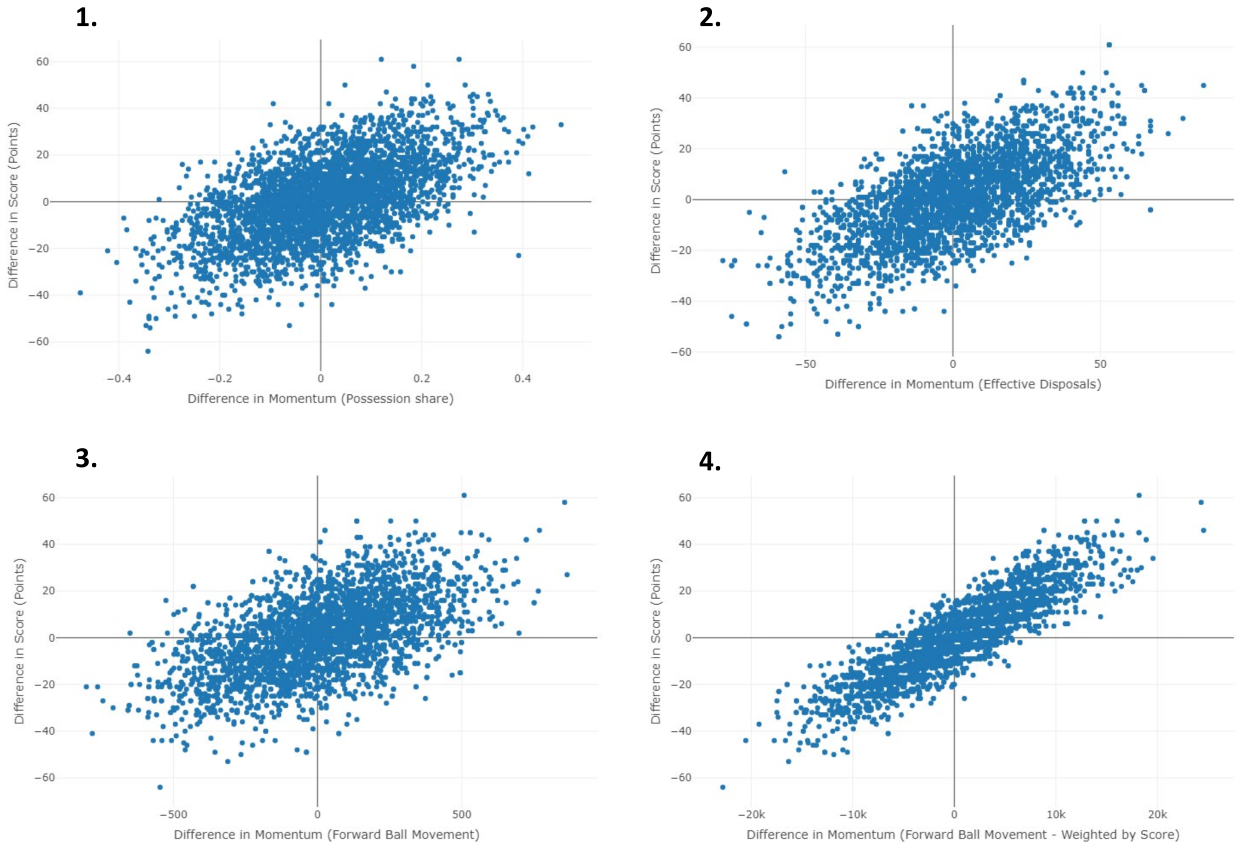
When momentum is analysed as the share of possession in a quarter, there is a medium correlation with score margin ($r = 0.544$, Table 2.). The team with more momentum (i.e., a higher share of possession) at the end of each quarter won that quarter 70.4% of the time. Results show that effective disposals as a measure of momentum in a quarter had a medium correlation with score margin ($r = 0.630$). Teams that had more effective disposals in a quarter than their opposition won that quarter 73.4% of the time. Forward ball movement (measured as metres gained) also proved to have a medium correlation with score margin ($r = 0.562$). The team with more momentum, using forward ball movement as the definition, won the quarter 70.1% of the time. The definition of momentum with the highest correlation with score margin for each quarter was shown to be forward ball movement, weighted by scoring. This measure had a correlation coefficient of $r = 0.886$, and the team that had more momentum in this regard won the quarter 85.4% of the time.

Table 2. Comparison of each definition of momentum in relation to quarter outcome.

| | Correlation coefficient with score margin (r) | Comparison with Possession share correlation (Fisher's r-to-z transformation) | Proportion of teams that had more momentum when they won on points (%) |
|---|---|---|---|
| **Possession share** | 0.544 | NA | 70.4 |
| **Effective disposals** | 0.630* | z= -5.468, $p<0.05$ | 73.4 |
| **Forward ball movement** | 0.562 | z= -1.078, $p>0.05$ | 70.1 |
| **Forward ball movement – weighted by score using simple multiplication** | 0.886* | z=-32.958, $p<0.05$ | 85.4 |

*= statistically significant compared to possession share

Figures 1-4 illustrate each measure of momentum as the net difference, compared to the difference in scores for every game across the four seasons. As evident in Figure 4, forward ball movement weighted by score has a better 'line of best fit' (values are closer together) compared to the other measures of momentum. This also supports the correlation coefficients displayed in Table 3.




**Figures 1-4.** Comparisons of the difference in momentum for all definitions with the difference in points for each quarter across the four seasons.

The difference in momentum (net) between teams over 4-minute time intervals (Figure 5) is shown to coincide with score margin over the course of a game. The best-performing measurement of momentum within an AFL match was considered to be, forward ball movement weighted by score. As evident in Figure 5, there are fluctuations of momentum for each team throughout the game, illustrated by the multiple peaks and troughs.

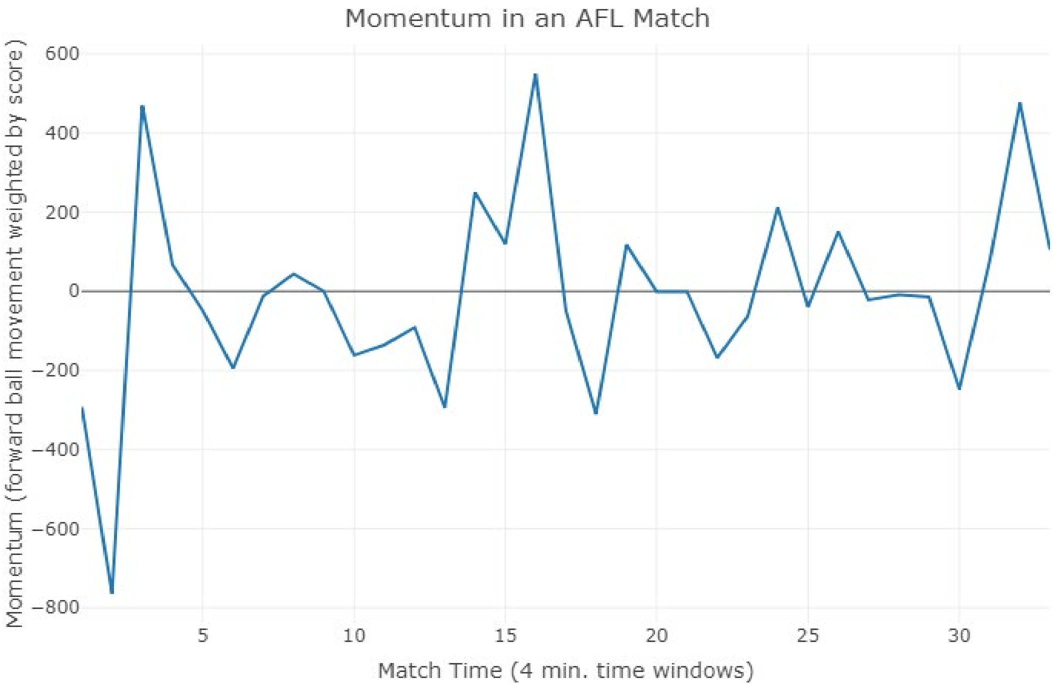



**Figure 5.** Net momentum between two teams across an AFL match using 4-minute time intervals. Momentum is based on the best-performing definition in results – forward ball movement weighted by score.

## 4. DISCUSSION

The main purpose of this study was to design measures of momentum created (or adapted) from past research on momentum in other sports, of which the most applicable definition of momentum should be associated with performance or quarter outcome.

All four measures of momentum had a positive association with score margin and quarter outcome. Teams that experience more momentum tend to outscore their opponents and, thus, win more quarters within a match. There are studies on individual (17) and team sports (5) that suggest that establishing momentum increases the probability of winning a match. However, there is minimal research to suggest an association between momentum and score margin in a period (i.e., quarter). A study on nine-ball billiards (17) discovered that a competitor with more momentum, measured as the player that wins a game (or two) within a match, has an increased probability of winning subsequent games and the match. Although there is no evidence of enhanced performance, research into basketball, football, and wrestling discovered that the team that first establishes momentum within a match has an increased probability of winning. (5) Therefore, the current study supports the notion that players (or teams) with more momentum have an enhanced possibility of winning; however, there is no existing research exploring momentum during a period within a match.

The definition that had the strongest association with score margin was, forward ball movement weighted by scoring. Teams that had more of this type of momentum were more likely to outscore their opponent and, thus, win the quarter. The present study contradicts current research on the measures of momentum that are used. The study conducted by Fry and Shukairy (16) used ‘yards gained’ as a momentum measure, occurring after a ‘momentum changing play’ (successful fourth-down conversion, unsuccessful fourth-down conversion, turnovers, allowing scores). Their findings indicated that there was no significant increase in yards gained after a momentum-changing play had occurred. However, the findings from their study may have been in contrast to the current study due to momentum being measured as a discrete variable that only occurs after a specific event has occurred.

Generally, research on momentum in sport employs a single measure of momentum to investigate its existence. However, similar to the present study, there are researchers who compared multiple definitions of momentum. McCutcheon et al. (5) and Fry & Shukairy (16) used panels of experts (players, coaches and analysts) to try to define momentum and determine whether it existed. Unlike the present study, the findings from these researchers suggested that there was little to no evidence of momentum influencing performance. Although they reported evidence of momentum, it may not mean that it does not exist. (16) When comparing multiple measures of momentum in other sports, it may be more appropriate to formulate definitions based on previous research and key performance indicators (KPIs) instead of expert opinion.

There are some limitations in this study that could be addressed in future study designs and by researchers exploring the concept of momentum in AF. The first limitation is that the best-performing method of measuring momentum is inherently biased because it includes information about scoring in its definition. Forward ball movement weighted by scoring has a better opportunity to have a stronger association with score margin because both variables are score-based. This study also did not analyse intervals throughout the quarter that compare short (in distance) possession chains and long possession chains. It is possible that teams that do not have as much forward ball movement as their opposition still win a quarter.

## 5. CONCLUSIONS

This is the first study to provide empirical evidence of momentum in AF. Momentum is strongly associated with score margin and quarter outcome when it is measured as forward ball movement weighted by score. Three other definitions of momentum were evaluated, and none achieved a stronger association with either measure of performance. While it is possible to create alternative definitions of momentum in AF, they may not be considered useful unless they have a stronger association than the best-performing measure of momentum reported here. This paper contributes to the current understanding of momentum in AF, but also more broadly to sporting contexts. It provides an opportunity for future exploration into how teams can change or maintain momentum within a match.

**Acknowledgements**
We would like to thank Champion Data for licensing access to the data used in this study.

**References**


1. Adler P, Adler PA. The role of momentum in sport. Urban Life. 1978;7(2):153–75.
2. Vallerand RJ, Colavecchio PG, Pelletier LG. Psychological momentum and performance inferences: A preliminary test of the antecedents-consequences psychological momentum model. Journal of Sport and Exercise Psychology. 1988;10(1):92–108.
3. Taylor J, Demick A. A multidimensional model of momentum in sports. Journal of Applied Sport Psychology. 1994;6(1):51–70.
4. Cornelius A, Silva III JM, Conroy DE, Petersen G. The projected performance model: Relating cognitive and performance antecedents of psychological momentum. Perceptual and Motor Skills. 1997;84(2):475–85.
5. McCutcheon LE. Does the establishment of momentum lead to athletic improvement? Perceptual and Motor Skills. 1997;85(1):195–203.
6. Bar-Eli M, Avugos S, Raab M. Twenty years of “hot hand” research: Review and critique. Psychology of Sport and Exercise. 2006;7(6):525–53.
7. Crust L, Nesti M. A review of psychological momentum in sports: Why qualitative research is needed. Athletic Insight. 2006;8(1):1–15.
8. Avugos S, Köppen J, Czienskowski U, Raab M, Bar-Eli M. The “hot hand” reconsidered: A meta-analytic approach. Psychology of Sport and Exercise. 2013;14(1):21–7.
9. Morgulev E, Avugos S. Beyond heuristics, biases and misperceptions: the biological foundations of momentum (hot hand). International Review of Sport and Exercise Psychology. 2020;1–21.
10. Morgulev E. Success breeds success: Physiological, psychological, and economic perspectives of momentum (hot hand). Asian Journal of Sport and Exercise Psychology. 2023;3(1):3–7.
11. Arkes J, Martinez J. Finally, evidence for a momentum effect in the NBA. Journal of Quantitative Analysis in Sports. 2011;7(3).
12. Mace FC, Lalli JS, Shea MC, Nevin JA. Behavioral momentum in college basketball. Journal of Applied Behavior Analysis. 1992;25(3):657–63.
13. Perreault S, Vallerand RJ, Montgomery D, Provencher P. Coming from behind: On the effect of psychological momentum on sport performance. Journal of Sport and Exercise Psychology. 1998;20(4):421–36.
14. Gilovich T, Vallone R, Tversky A. The hot hand in basketball: On the misperception of random sequences. Cognitive Psychology. 1985;17(3):295–314.
15. Moesch K, Bäckström M, Granér S, Apitzsch E. Hard fact or illusion? An investigation on momentum in female elite handball from a team perspective. International Journal of Sport and Exercise Psychology. 2014;12(2):106–20.
16. Fry MJ, Shukairy FA. Searching for momentum in the NFL. Journal of Quantitative Analysis in Sports [Internet]. 2012;8(1). Available from: https://doi.org/10.1515/1559-0410.1362
17. Adams RM. Momentum in the performance of professional tournament pocket billiards players. International Journal of Sport Psychology. 1995;26(4):580–7.
18. Reed D, Hughes M. An Exploration of Team Sport as a Dynamical System. International Journal of Performance Analysis in Sport [Internet]. 2006 Nov 1;6(2):114–25. Available from: https://doi.org/10.1080/24748668.2006.11868377
19. Pill S. Developing theoretically informed practice: the forward press in Australian football as an example of the dynamics of a complex system. In: Proceedings of the 28th ACHPER International Conference, Melbourne. 2013. p. 102.

20. Josman C, Gupta R, Robertson S. Markov chain models for the near real-time forecasting of Australian football league match outcomes. In: Soft Computing for Problem Solving 2019: Proceedings of SocProS 2019, Volume 2. Springer; 2020. p. 111–25.

# ESTIMATING SKILL IN ELITE AND SUB ELITE AUSTRALIAN FOOTBALLERS

Dave Matteo [a,c] , David Carey [a] , Joshua Ruddy [b], Matthew Varley [a]
*[a] La Trobe University, [b] Essendon Football Club, c Corresponding author: dmatteo30@gmail.com*

**Submitted for consideration as (select one):**

- ☒ Oral presentation
- ☐ Poster

**Abstract**

Sporting organisations can benefit from the ability to accurately and objectively compare player skill [1-3] to inform decisions around player retention and trading. Within team, or within league player comparisons can be made directly with appropriate performance metrics [4-6]. However, comparisons become more difficult when they are made between different leagues, as player performances are impacted by the strength of the opposition, ability of teammates, style of play, and playing conditions [7-10]. In Australian Football, there is a need to compare players from elite junior teams as they get drafted into the Australian Football League (AFL), and organisations in the AFL need the ability to compare skill between players playing in the AFL and the Australian Football 2nd tier leagues. This study aimed to create ratings for each player that can be interpreted as Player's Skill Rating (PSR) which is comparable between leagues. The second aim was to provide an Age Included Player Skill Rating (AIPSR), this was because PSR represents a player's skill with age controlled for. Including age effects into PSR to create AIPSR will allow for these metrics to have more flexibility around the types of decision making an Australian Football club may use them for. This study provides a methodological framework for how player skill can be objectively measured without the impact of differences between AFL and 2nd tier leagues. It expanded on previous research by refining the Linear Mixed Model previously used, applying tighter constraints on the model as well as comparing frequentist and Bayesian approaches. This study also differed by the interpretation of the player and age coefficients as the measure of skill instead of using the Linear Mixed Model to normalise Fantasy Football Points. *Keywords*: *Australian Football, Performance, Player Skill*

**Acknowledgements**

Thank you to Steven Azzopardi for his expertise in Australian Football Data and to Essendon Football Club for access to player data from Champion Data.

**References**

1. Bonney, N., et al., *Australian Football Skill-Based Assessments: A Proposed Model for Future Research.* Front Psychol, 2019. **10**: p. 429-429.
2. Carling, C., T. Reilly, and A.M. Williams, *Performance assessment for field sports*. 2008: Routledge.
3. D'Isanto, T., et al., *Assessment of sport performance: Theoretical aspects and practical indications.* Sport Mont, 2019. **17**: p. 79-82.
4. Wolf, S., M. Schmitt, and B. Schuller, *A football player rating system.* Journal of Sports Analytics, 2020. **6**(4): p. 243-257.
5. McIntosh, S., S. Kovalchik, and S. Robertson, *Validation of the Australian Football League Player Ratings.* International journal of sports science & coaching, 2018. **13**(6): p. 1064-1071.
6. Thomas, A., et al., *Competing process hazard function models for player ratings in ice hockey.* The Annals of Applied Statistics, 2013: p. 1497-1524.
7. Coates, D. and B. Oguntimein, *The length and success of NBA careers: Does college production predict professional outcomes.* International Journal of Sport Finance, 2010. **5**(1): p. 4-26.
8. Kharrat, T., I.G. McHale, and J.L. Peña, *Plus–minus player ratings for soccer.* European Journal of Operational Research, 2020. **283**(2): p. 726-736.
9. Woods, C.T., J. Jarvis, and I. McKeown, *Differences between Elite and Semi-Elite Australian Football Conceptualised through the Lens of Ecological Dynamics.* Sports (Basel), 2019. **7**(7).
10. Johnston, R.J., et al., *Movement Demands and Metabolic Power Comparisons Between Elite and Subelite Australian Footballers.* The Journal of Strength & Conditioning Research, 2015. **29**(10): p. 2738-2744.

# EXPECTED SCORES AND GOAL KICKING SKILL IN AUSTRALIAN FOOTBALL

Hayden Pfingst[a], Prasanna Sritharan[a], Steven Azzopardi[a,b], David Carey[a*]

[a] *La Trobe University, Sport and Exercise Science, School of Allied Health, Human Services & Sport*
[b] *Essendon Football Club*
[*] *Corresponding author: d.carey@latrobe.edu.au*

## Abstract

Expected score (xScore) is a tool that has been developed in invasion sports to provide an understanding of the goal scoring opportunities generated by teams and the finishing performance of players. xScore allows for the calculation of expected difference (xDiff) and expected pass benefit (xPB) to analyse different aspects of the sport. xDiff is the difference between a player's xScore from a shot and the actual outcome of the shot. xPB assess whether a player within a scoring opportunity passed the ball to someone in a better or worse scoring position, which provides insight into players decision making abilities when in goal scoring situations. Play-by-play data from the 2021 and 2022 AFL seasons that included location and context information on every kick instance within an official AFL match was analysed. Explanatory variables considered included distance, angle, pressure, and dominant foot. A random intercept for player was included and interpreted as a player-skill variable within models. Models were compared using a drop in deviance test. Results from the player-adjusted xScore model will be presented and contrasted with raw scores and non-adjusted xScore approaches.

**Keywords**: Expected score, Australian Football, player skill

## 1. INTRODUCTION

In Australian Rules Football (ARF) goal kicking skill has the ability to make or break team's chances of winning. Similar to other invasion sports, not all shots have the same level of difficulty even though they are all worth the same on the scoreboard. The final score, and accuracy percentages paint a simple picture of a team's shooting performance and can fail to reflect the context of the shots and quality of scoring oppurtunities generated that contributed to deciding the game. One aspect of context that is often excluded is the identity of the player taking the shot. Within ARF players are recruited based on different physical and skill capabilities to play certain roles within teams. Simply, not all players primary role is to kick goals. Evaluating all players the same creates for a bias towards players whose role revolves around scoring goals and players who have a higher number of opportunities.

Expected goals (xG) is an tool that has been utilized in other invasion sports such as soccer [1, 2, 3, 4], ice hockey [5, 7] and field hockey [6] that has the capabilities of being integrated into ARF as xScore. xG is a method for evaluating a shot quality based on shot context factors such as *distance* and *angle* from goal. xG has been implemented in invasion sports to gain greater understanding of the quality of shot opportunities players are attempting. It has allowed for players and teams to be analysed in more context reflective way when looking at shooting performance. ARF has attempted to integrate xG into the sport, with the xG model re-named as an "expected score" (xScore), as there is more than one way to score in the sport (goals and behinds) [8, 9, 10, 11]. Sport specific variables that can be explored include whether a player kicked with their *dominant foot*, how much *pressure* is being applied to the kicker and what *phase of play* they took the shot. Excluding player as an input to the model creates an unrealistic assumption that all players have the same goal kicking ability. The introduction of player as a variable provides an additional layer of information that would assist in making the model more precise. xScore could adjust the probabilities of scoring based on the historical performance of an individual or historically how well a player of a certain position shoots.

xScore is a useful concept that lays the foundation for addition tools that can be used to help gain deeper understanding of various components of the game. An additional application of xScore could be to measure whether a player is making the correct decision to pass when they were in an attacking position. Using xScore to compare the quality of shot opportunity different players have within a chain could allow for an analysis of the decision-making process.

The overall objective of this study is to design and implement xScore into Australian Rules Football with three specific aims:

1. Use expected score and expected difference to evaluate shooting performance in the Australian Football League during the 2023 season.
2. Create shooting zone for analysis to compare different position groups and find the strongest and weakest zones for each position type.
3. Create a new evaluation model to understand whether players are making appropriate decisions to pass the ball when in a goal scoring position.

## 2. METHODS

### 2.1 Data

Match play-by-play (transactional) data collected by Champion Data from season 2021 to 2023 of the AFL was used within the study. The data beginning in 2021 is due to the introduction of the stand rule in the 2021 season. The play-by-play data collected by Champion Data contain kick context variables such as pressure level, kick location given as 2D coordinates, and the intended target of the kick (whether the player was going for the goals or not). When assessing the shot at goal, the data was filtered for the variable `KICK_INTENT=goal` meaning the player was having a shot at goal no matter whether they score a goal, behind or miss. Each kick is recorded with relevant information within the transactional data including the on-field locations as spatial coordinates $x$ and $y$ with origin at the centre of the ground. The analysis was conducted using R.

### 2.2 Expected Score (xScore) model creation and application.

The xScore model was built using shots at goal from the 2021 and 2022 AFL seasons. Mixed effect models were created using points as the dependant variable with other relative variables included as fixed effects. A number of mixed effect models were created and trialled with player and position being interchanged as random intercept and slope variables which can all be seen in table 2.1. The models were created using the *lme4* package.

*Table 2.1 Expected score models tested ("ns" denotes natural splines applied to a variable).*

| Model | Formula |
|---|---|
| Model 1(xSp) | POINTS ~ ns(dist_goal, 4) + ns(angle, 4) + SHOT_DOM_FOOT + PHASE_OF_PLAY + PRESSURE_LEVEL + (1 \| FULLNAME) |
| Model 2 | POINTS ~ dist_goal + angle + SHOT_DOM_FOOT + PHASE_OF_PLAY + PRESSURE_LEVEL + (position \| FULLNAME) |
| Model 3 (xS) | POINTS ~ ns(dist_goal, 4) + ns(angle, 4) + SHOT_DOM_FOOT + PHASE_OF_PLAY + PRESSURE_LEVEL |
| Model 4 | POINTS ~ dist_goal + angle + SHOT_DOM_FOOT + PHASE_OF_PLAY + PRESSURE_LEVEL + homevsaway + (1 \| FULLNAME) |
| Model 5 | POINTS ~ dist_goal + angle + SHOT_DOM_FOOT + PHASE_OF_PLAY + PRESSURE_LEVEL + (1 \| FULLNAME) + (1 \| position) |

The variables that were included in the models include *Distance from goal, angle from goal, pressure level, phase of play* and whether they kicked on their *dominant foot*. The *angle* is the angle at the centre point of goal from where the shot is being taken relative to directly in front. The *distance* and *angle* variables were entered into the model as spline terms so that non-linear relationships could be fit for those variables. The purpose of the non- linear relationships for these two variables is because a metre change in distance under 10 metres doesn't have the same impact on the shot as a metre change 40 metres away from goal. *Pressure* is another variable that was important in creating the xScore mixed-effect models. *Pressure* is a statistic that is recorded for each disposal in the play-by-play data. Different *pressure levels* can significantly impact the quality of shot a player is taking. Finally *phase of play* and *dominant foot* were trialled within the models. *Phase of play* is determined with the `PRESSURE_LEVEL=set` indicating it was a set shot and `PRESSURE_LEVEL!=set` indicating it was general play. Thus *Phase of play* is a binary variable with only two outcomes, set shot or general play. *Dominant foot* is a variable that was not included in the raw transactional data but was derived from it by counting how many kicks every player performed on each foot throughout the year. The foot that had the higher number of kicks was determined to be their dominant foot. Thus, *dominant foot* was also a binary variable with the outcomes being 1 (kick was taken on dominant foot) and 0 (kick was not taken on dominant foot). A drop in deviance test was used to compare the models with the best being selected using $R^2$ and *Chi-Squared* comparisons. The best model that was determined to be the best was Model 1, and was named xScore Player Adjusted (xSp). The xSp model is the primary model for this study and was used to do the further analysis.

Model 1 (xSp) was used to provide an xScore for every shot at goal from the 2023 season. Using the *predict()* function in R, the model was applied to every shot within the dataset. Within the *predict()* function the parameter *allow.new.levels = TRUE* was used to allow new players who didn't play in 2021 and 2022 to be modelled with a random effect of 0. The *predict()* function provided each shot with an xScore value.

### 2.3 Expected Difference (xDiff)

Expected difference (xDiff) is an analysis tool that compares players xScore with the outcome of the shot and is described by the equation:

$$xDiff = Points\ scored\ from\ shot - xScore$$

Once an xDiff was generated for each shot it was used to analyse player, team, position, and location shooting performances. Mean xDiff was used to complete multiple aspects of the analysis as it shows the average xDiff and provides an insight into particularly player performances over time.

### 2.4 Player Positions and Shooting Zones

In order to achieve Aim 2, new shot zones were defined and each player was assigned a playing position. The player position labels used within this study are outlined in Table 2.2 and were assigned to each player by Champion Data. Zones were created by applying unique identifiers to each distance zone crossed with each angle zone. 35 new zones were created which can be seen in Figure 2.1.

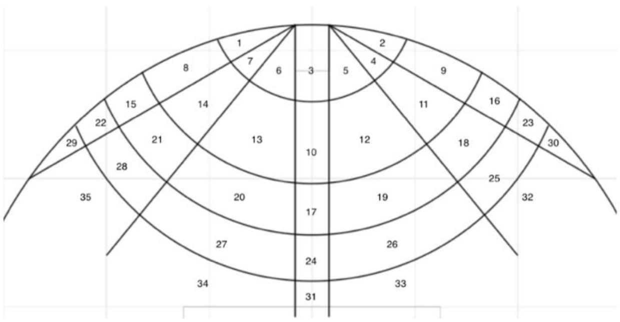



*Figure 2.1 Shot zones*

Each position group was investigated by looking at their mean xDiff in each newly created shooting zone. Each position type was analysed to identify their best and worst zones when shooting at goal. The position and zone analysis are the foundation for providing an outcome of sub aim 2.

### 2.5 Expected Pass Benefit (xPB)

Expected Pass Benefit (xPB) is a value that is assigned to the last player to pass the ball before a player takes a shot at goal. xPB was calculated to evaluate whether players who deliver the final pass would have a better chance of scoring if they had a shot at goal instead. Using xScore generating an xPB for the passer is possible.

xPB is the difference between the xScore of the player who makes the final pass and the xScore of the player who takes the shot.

$$xPB = xScore\ shot\ taker - xScore\ passer$$

It is proposed as a tool to aid the evaluation of the decision-making process during an assist (i.e., passing to a player with a greater or lesser potential to score) and therefore does not consider the final outcome. For example, players who pass the ball to a teammate in a better scoring position will generate a positive xPB as the teammates xScore is higher than theirs. Oppositely if a player passes to a teammate in a worse position, they will receive a negative xPB indicating they may have been better to attempt the shot at goal themselves. The player adjustments within the xScores play a significant role within the metric as it means that better goal kickers will be negatively impacted by passing the ball as they are generally the player who should be attempting the shot. A higher xPB indicates a better decision to pass the ball and players with high mean xPB will be considered as good decision makings when electing to pass the ball inside forward 50.

The type of possession chain that was included was kick to kick chains. Goal chains that come from 50 metre penalties, dispute play, stoppages, and handball to kick chains were excluded from the analysis as they didn't have the appropriate data to generate an xScore. Goal and behind assists were used to find the passer within the play-by-play data and were filtered to find only assists that were kicks. After xPB was calculated for every assist across the 2023 season, instances where the player passed the ball from further than 60 metres away from goal as they are considered to not be in a realistic scoring position. Player analysis required the number of passes to be filtered to players who had more than 6 assists throughout the season as that was the top 25% threshold of total number of assists by players within 60 metres. Mean xPB was extracted for all teams and players and was analysed to find trends.

## 3. RESULTS

### 3.1 xScore model

The selection process to find the best model resulted in the model with player as the random intercept being the best model (xSp). Table 3.1 display the results from the drop in deviance test that was conducted. The model included fixed variables such as distance and angle, whether it was kicked with the dominant foot, the phase of play and what type of pressure the kicker was under. The model included a random intercept variable of player which will adjust each prediction based on the historical performance of each player. Table 3.2 shows that all of the fixed effects except for Pressure Level [None] are statistically significant with P values $<0.05$. Figure 3.1 highlights 9 of the top 10 goal scorers from the 2022 AFL season and shows how their intercepts will be adjusted each time they have a shot. The figure indicates that the number of points scored by a shot is increased when 6 of the 9 chosen players have a shot. Three of the 9 players have their number of expected points lowered when they have a shot at goal.

The expected score non player adjusted (xS) model was generated to show the performance of the xSp model and how important the player adjustments are to the model. Figure 3.2 shows the individual player adjustments that are applied to each player xScore when having a shot at goal along the x axis. The y axis shows each players mean xDiff overall from the 2023 season using the xS model. The majority of players who had a significant number of shots are centred around the mean xDiff of 0 which means players are shooting around how the model expects. Players with only a small sample of shots at goal are able to have large xDiff values, however the player random effect estimates remain small for those players, highlighting the effect of shrinkage in mixed effect regression models.

*Table 3.1 Drop in deviance test results.*

| Model | Model 3 | Model 4 | Model 5 | Model 1 | Model 2 |
|---|---|---|---|---|---|
| npar | 10 | 12 | 12 | 17 | 38 |
| AIC | 93715 | 93678 | 93674 | 93566 | 93722 |
| BIC | 93794 | 93773 | 93769 | 93700 | 94022 |
| logLik | -46847 | -46827 | -46825 | -46766 | -46823 |
| deviance | 93695 | 93654 | 93650 | 93532 | 93646 |
| Chisq | | 40.6069 | 4.1254 | 118.3504 | 0 |
| PR(>Chisq) | | 1.52E-09 | | < 2.2e-16 | 1 |
| P Value | 1 | <0.001 | 1 | <0.001 | 1 |

*Table 3.2 Model results for Expected Score Player Adjusted Model (xSp)*

| | POINTS | | |
|---|---|---|---|
| *Predictors* | *Estimates* | *CI* | *p* |
| (Intercept) | 7.28 | 7.04 – 7.52 | **<0.001** |
| dist goal [1st degree] | -2.93 | -3.14 – -2.71 | **<0.001** |
| dist goal [2nd degree] | -3.89 | -4.18 – -3.59 | **<0.001** |
| dist goal [3rd degree] | -6.84 | -7.69 – -5.99 | **<0.001** |
| dist goal [4th degree] | -4.29 | -5.87 – -2.71 | **<0.001** |
| angle [1st degree] | -0.43 | -0.59 – -0.28 | **<0.001** |
| angle [2nd degree] | -0.69 | -0.87 – -0.52 | **<0.001** |
| angle [3rd degree] | -2.00 | -2.32 – -1.68 | **<0.001** |
| angle [4th degree] | -2.51 | -2.78 – -2.24 | **<0.001** |
| SHOT DOM FOOT [1] | -0.50 | -0.63 – -0.37 | **<0.001** |
| PHASE OF PLAY [0] | -0.56 | -0.75 – -0.37 | **<0.001** |
| PRESSURE LEVEL [Closing] | -0.90 | -1.10 – -0.69 | **<0.001** |
| PRESSURE LEVEL [Corralling] | -0.21 | -0.40 – -0.01 | **0.035** |
| PRESSURE LEVEL [None] | 0.09 | -0.15 – 0.32 | 0.474 |
| PRESSURE LEVEL [Physical] | -2.21 | -2.46 – -1.96 | **<0.001** |
| **Random Effects** | | | |
| $\sigma^2$ | 5.89 | | |
| $\tau_{00}$ FULLNAME | 0.05 | | |
| ICC | 0.01 | | |
| N FULLNAME | 693 | | |
| Observations | 20260 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.173 / 0.180 | | |

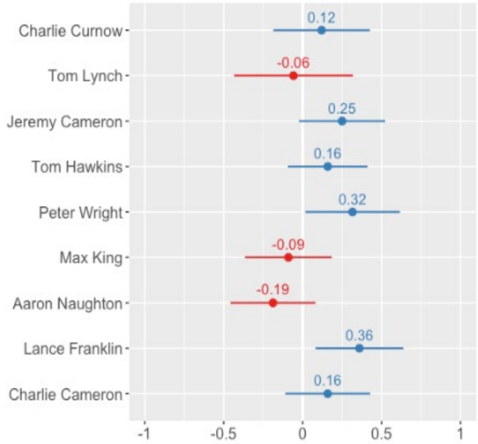



*Figure 3.1. Random effect estimates for top goal kickers in the 2022 season*

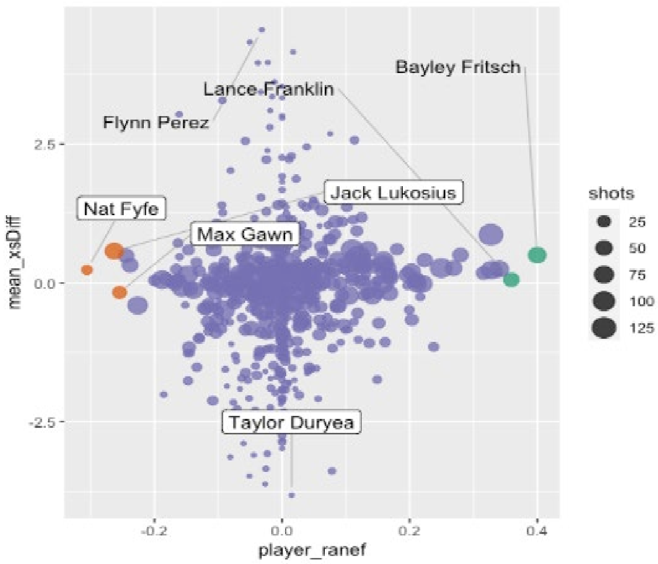



*Figure 3.2. Mean xS Diff compared with player random effect.*

### 3.2 Team Performance xSp vs Actual

Figure 3.3 shows how many points teams actually scored against how many points they were expected to score. 9 out of 18 teams scored more points than they were expected throughout the 2023 season. Out of the top 8 most successful teams throughout the 2023 season only 4 of those teams finished the regular season with more actual points than what was expected. Table 3.3 shows how the ladder would have looked at the end of the 2023 season if teams scored exactly how many points were expected based on the model. 5 out of the 18 teams would have finished higher on the ladder if they scored their expected points. Opposingly 8 out of 18 teams finished higher than expected.

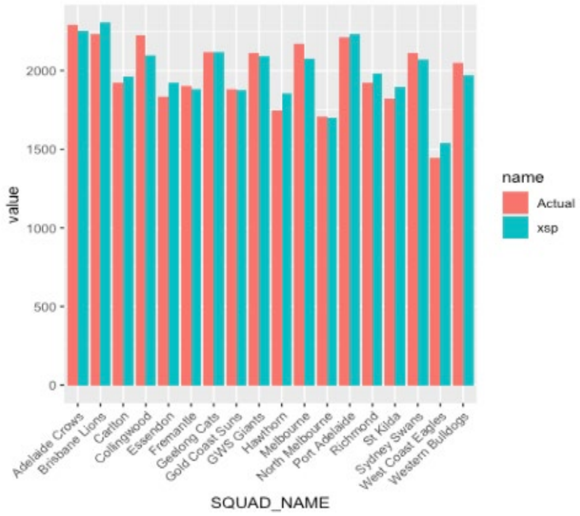



*Figure 3.3. 2023 teams Expected vs Actual score*

*Table 3.3. 2023 Season Ladder using xSp Points*

| Predicted Finish | Actual Finish | Team | Wins | For | Against | Percentage | Points |
|---|---|---|---|---|---|---|---|
| 1st | 2nd | Brisbane Lions | 17 | 2205 | 1665 | 132 | 68 |
| 2nd | 1st | Collingwood | 17 | 1981 | 1735 | 114 | 68 |
| 3rd | 3rd | Port Adelaide | 16 | 2173 | 1813 | 120 | 64 |
| 4th | 10th | Adelaide Crows | 15 | 2160 | 1826 | 118 | 60 |
| 5th | 4th | Melbourne | 14 | 1965 | 1665 | 118 | 56 |
| 6th | 5th | Carlton | 14 | 1900 | 1730 | 110 | 56 |
| 7th | 6th | St Kilda | 13 | 1821 | 1698 | 107 | 52 |
| 8th | 7th | GWS Giants | 13 | 1990 | 1892 | 105 | 52 |
| 9th | 12th | Geelong Cats | 12 | 2049 | 1866 | 110 | 48 |
| 10th | 11th | Essendon | 12 | 1861 | 1980 | 94 | 48 |
| 11th | 9th | Western Bulldogs | 11 | 1857 | 1771 | 105 | 44 |
| 12th | 8th | Sydney Swans | 10 | 1987 | 1846 | 108 | 40 |
| 13th | 14th | Fremantle | 10 | 1802 | 1862 | 97 | 40 |
| 14th | 13th | Richmond | 10 | 1911 | 2020 | 95 | 40 |
| 15th | 15th | Gold Coast Suns | 9 | 1815 | 2027 | 90 | 36 |
| 16th | 16th | Hawthorn | 8 | 1748 | 2050 | 85 | 32 |
| 17th | 17th | North Melbourne | 3 | 1616 | 2249 | 72 | 12 |
| 18th | 18th | West Coast Eagles | 3 | 1482 | 2628 | 56 | 12 |

### 3.3 Expected Difference Zones and Positions

The positional breakdown of shooting attempts per zone provided a mean xDiff for every newly created shot zone which varied between position type. Figure 3.4 provides a summary of which position type had the largest positive mean xDiff in the various shot zones. The zones that are blank indicate that either all positions have a negative mean xDiff within the zone or each position had fewer than 9 shot attempts from the zone. Each position can be analysed using xDiff for each new shot zone.

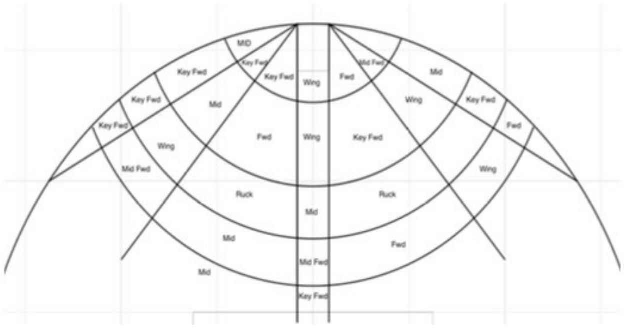


*Figure 3.4 Zones with position that has the largest positive mean xDiff.*

*Figure 3.5 Zones with position that has the largest negative mean xDiff.*

### 3.4 Expected Pass Benefit

Once the xPB was generated for every pass that led to a score every player was generated a mean xPB using those results. All players were found to have a positive mean xPB. Figure 3.6 showcases the players with the largest 3 mean xPB and the lowest 3 mean xPB and shows the spread of xPB for each of their passes. Out of the 6 players 5 have a small spread meaning they are consistent with their xPB each time they elect to pass. One of the six players produced a wide spread of results.

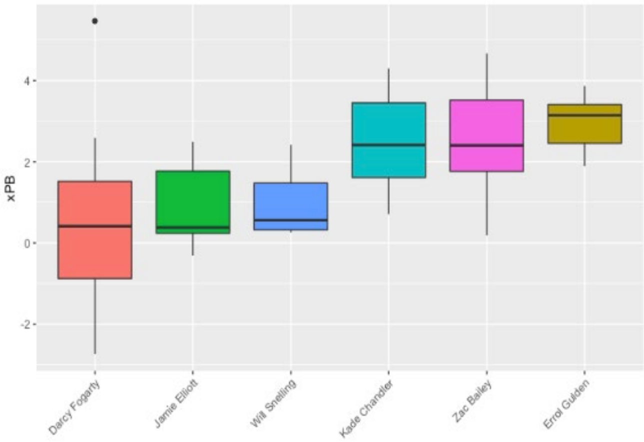



*Figure 3.6 Three players with the highest and lowest mean xPB spread*

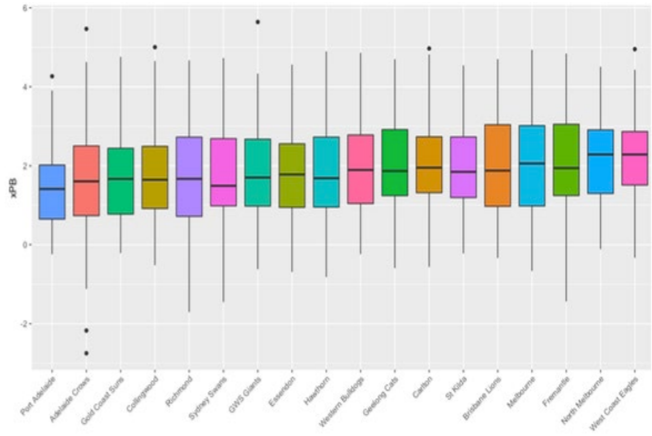



*Figure 3.7 xPB spread for every team within the AFL*

The teams with the highest xPB were West Coast Eagles (2.26), North Melbourne (2.16) and Fremantle (2.09) whereas Port Adelaide (1.51), Adelaide Crows (1.6) and the Gold Coast Suns (1.66) had the lowest mean xPB. Figure 3.7 shows the spread of every xPB players from every team generated throughout the 2023 AFL season. The Adelaide Crows and Richmond had the largest spread between the minimum and maximum xPB and the most consistent teams with the smallest spread were Port Adelaide and North Melbourne.

## 4. DISCUSSION

xScore has the capability to be implemented into Australian Rules Football and be a useful tool to analyse multiple aspects of the game at both a team and player level. xScore is useful to analyse team performance and understand whether they are generating good quality scoring opportunities. A higher team xScore is positive as it indicates the team's attacking method is providing better quality scoring opportunities. The introduction of xDiff to work with xScore is a tool that is best used to evaluate goal shooting performance rather than assessing the quality of scoring opportunity. Using both tools can allow teams to evaluate a wide range of aspects of the games from how well teams are generating good quality scoring opportunities to evaluating player and team shooting performances. When intergrading xScore into ARF using a player adjusted model provides an additional layer of information into the model. Using player as a random effect within the model provides a calculation that not all players have the same ability when shooting at goal. The addition of player adjustments adds expectation on typically classified 'good goal kickers' as they will generate a higher xScore than other players for the same opportunity. The player adjustments will also impact the xDiff that is generated from shots as better goal kickers will be additionally punished by missed opportunities and traditionally bad goal kickers will receive an additional boost if they score. The player adjustments are reflected when analysing teams xScore as they will be higher when getting the ball to the better goal kickers. Teams xScore is enhanced when the player who is shooting at goal has a history of success. Player adjusted xScore models should be utilised within ARF as they can be a multi-faceted tool for analysing player and team performance and provide greater insights when used to also implement xDiff and xPB into the analysis.

### 4.1 xScore Player Adjusted Model

The xSp model was shown to be the best model by the drop in deviance test results. In the drop in deviance test 5 models were compared including a simple linear regression model that had no random effects. Only two models provided significance which were models 1 and 4. The only difference between the models were model

4 included whether the player attempting the shot was from the home or away team. Model 1 (xSp) was the model that provided the best results from the drop in deviance test and used for the analysis in the paper. Fixed effect variables (distance, angle, pressure) had the strongest influence on model predictions of expected score, the magnitude of player-level effects were smaller bit still contributed to the model. Figure 3.1 is an example of what impact the random effects can have on individual player predictions. The figure shows how the top 10 goal kickers from 2022 would have their xScore adjusted whenever they attempt a shot. Assumptions are traditionally made that because players kick a lot of goals, they have good accuracy in front of goal, which is shown to be wrong by their random effects. The random effects show that 3 out of the top 10 goal kickers have negative player adjustments indicating throughout the 2021 and 2022 seasons they kicked under the number of points that were expected. The player adjustments appear to be a useful inclusion to xScore as it provides a skill-based variable that differentiates each player based on the skill of goal kicking that they have produced throughout previous years.

Figure 3.2 has mean xDiff based on the xS model and the player adjustments along the different axis. The majority of the points of the figure should be close to the 0 line on the y axis but Flynn Perez and Taylor Duryea are prime examples of outliers that would be created. Limited shots are the reason these players are on different extreme sides of the figure, but player adjustments would assist in bringing those players closer to 0. Player adjustment assist in removing the outliers via shrinkage in mixed models. Applying player adjustments to the xScore model was an important addition which provides a deeper understanding and an additional layer of analysis that can assist in the analysis of goal kicking.

### 4.1.1 Application of xSp vs Actual Results

In the 2023 season 9 of the 18 teams scored above what was expected. Out of the 9 teams that scored above expected only 4 finished inside the top 8 at the end of the season. Although half of the teams finished above expected looking, those numbers shouldn't be looked at in isolation and should instead be assessed per game to provide better context. Table 3.3 is a representation of the ladder at the end of the home and away season if teams score what was expected by the xSp model. The table shows that some teams capitalised on their opportunities and other teams wasted crucial opportunities which cost them at the end of the season. Two team's standout for those particular reasons as the Sydney Swans capitalised on their opportunities and climbed the ladder whereas the Adelaide Crows wasted their opportunities and slid out of finals contention. The Adelaide crows finished the season with more actual points than xSp points, but it was certain games that they shot poorly in that provided them with a $10^{th}$ place season instead of an expected $4^{th}$ place finish. The model expected Adelaide crows would have won games against Richmond (Round 2), Collingwood (Round 7), Melbourne (Round 19), Brisbane (Round 22), and Sydney (Round 23) which they instead lost all those games. The actual vs xSp ladder highlights the importance of goal kicking on team success.

### 4.2 Expected Difference

Expected difference (xDiff) is a tool related xScore that can be used to analyse goal kicking performance of players and teams. The benefit of using xDiff over other traditional goal kicking evaluation tools is it considers the difficulty of the shot that is being attempted. xDiff allows for analysis of how effectively players are finishing different quality opportunities. xDiff rewards players for capitalizing on hard quality shot opportunities and continually capitalize on easy opportunities. xDiff also punishes players for missing easy quality shot opportunities that should be scored.

Sub aim 2 involved breaking down the forward 60 into 35 new zones which were a combination of the angle and distance zones that currently exist within the data and evaluating each position type in those newly created zones. Figure 3.4 and figure 3.5 reflect the best and worst position types from each of the new zones. Key forwards performed the best in the most zones within figure 3.4 with eight out of twenty-eight zones showing them with the highest mean xDiff. Other positions that had a large presence within the figure include midfielders (6 out of 28 zones) and wingers (5 out of 28 zones) meaning those three positions represent 68% of the figure. Oppositely figure 3.5 shows which positional type has the largest negative xDiff in each zone and is left blank if all positions had a positive mean xDiff or did not have more than 9 attempts within the zone. The Rucks have the most zones within figure 3.5 with seven out of twenty-seven having rucks as the worst shooters. Forwards (6 out of 27) and midfielder- forwards (5 out of 27) were the other positions that had a significant number of zones in which they were the worst performing. The implementation of xDiff for the different zones and positions can help clubs with training and recruitment. xDiff can help with training as visualisations such as figures 3.5 can help identify which positions require more training from certain areas of the ground to improve their goal kicking. Furthermore visualisations can be created for clubs most frequent goal kickers to identify their areas that need improvement and areas that they are good in to try and get them the ball in those areas. Recruiters can use this information to identify areas of concern in relation to goal kicking and target players who excel in those areas to recruit.

### 4.3 Expected Pass Benefit

Quantitative evaluation of on-field player decision making is difficult. xPB may provide a review tool that can assess players on whether they were better off having a shot at goal rather than passing the ball. Based on the xScores that are generated for players having a shot at goal and players who provide an assist to them xPBs are generated. A negative xPB indicates the player who passed the ball should have considered taking the shot themselves rather than pass the ball as they had an easier shot at goal. Player adjustments within the xSp model also impact xPB as players with higher player adjustments are negatively impacted when passing the ball to worse goal kickers. Figure 3.6 shows the players with the highest and lowest mean xPB, and the player adjustment impacts are evident with Darcy Fogarty (Player Adjustment = +0.34) having the worst result. Positively all the players with more than 9 assists have positive mean xPBs. In Figure 3.6 it shows Darcy Fogarty did have assists that had negative xPBs indicating he made the potentially incorrect decision to pass the ball. 3 of his 9 assists provided a negative xPB indicating 1 out of 3 times he reordered an assist he should have attempted a shot. xPB allows to identify players such as Darcy Fogarty and allow coaches to correctly assess the player and understand why he is passing to players in worse positions and possibly encourage the player to attempt the shots more as it may be beneficial to the team.

Teams can use xPB to evaluate whether as a team they are creating better quality shot opportunities for each other and can identify whether the positions they are getting their player to pass into are good areas to shoot from. Surprisingly the two teams that finished last on the ladder at the end of the 2023 season had the highest mean xPB of all teams. West Coast and North Melbourne throughout the season continually got the ball to their better goal kickers in better positions to achieve these results. Alternatively they also had the lowest number of assists throughout the season which could contribute to their high means. xPB is a tool that can be used at both a team and player level to assist coaches to continue to develop and educate players on what decisions they should be making when in a scoring position.

### 4.4 Limitations and Future Study

Gaps within the data limit the depth of analysis that can occur in regards to certain contextual variables. In the xScore model building process the direction the player is facing is not recorded which means it is unsure whether the player is facing the goal making it an easier shot or if the player is facing a different direction making it harder. The limitation of this is that if a player attempts shots with the same exact variables but they are facing different directions they are recorded at the same difficulty. This is a limitation as it impacts the difficulty of the shot and there is no variable within the data to address the issue. Future research should look at what other variables could be included into the model to make it more precise. Variables such as weather and type of kick should be considered as they can impact the probability of scoring. Additionally adding a variable that considers the results from the shot zone and position type analysis within the model could be explored.

Within the xPB calculations only assists are utilised as the data makes it difficult to find a previous possession to evaluate all previous possession to shots on goal. Using only assists neglects shots on goal that completely missed which would traditionally be difficult shots to attempts. Finding a way to include all previous possessions would allow for the tool to be more impactful as it would provide more data to assess and possibly lower xPBs. Also only using assists that came from kicks skews the eventual shots towards being set shots naturally giving them a better xScore. The limitations that have appeared within the xPB data should be assessed and rectified in future research.

## 5. CONCLUSIONS

Expected score models that include player-level effects have the potential to provide insight into multiple aspects of ARF. xScore permits other evaluation tools such as xDiff and xPB which can provide an in depth analysis of various aspects of ARF from the quality of scoring opportunities teams are generating to how well players perform when shooting at goal and whether players are making a smart decision to pass the ball when in scoring opportunities themselves. xScore provides a wholistic look at shots that are being attempted by factoring in different constraints that players are confronted with. Future work should look at trialling different variables within the model such as type of kick and weather conditions.

**Acknowledgements**
We wish to thank Essendon Football Club for their assistance in data access.

**References**

1. Anzer G, Bauer P. A Goal Scoring Probability Model for Shots Based on Synchronized Positional and Event Data in Football (Soccer). Frontiers in Sports and Active Living. 2021;3.
2. Brechot M, Flepp R. Dealing With Randomness in Match Outcomes: How to Rethink Performance Evaluation in European Club Football Using Expected Goals. Journal of Sports Economics. 2020;21(4):335-62.
3. Cavus M, Biecek P, editors. Explainable expected goal models for performance analysis in football analytics. Proceedings - 2022 IEEE 9th International Conference on Data Science and Advanced Analytics, DSAA 2022; 2022.
4. Kumar S, Jagannath V, Visalakshi P, editors. Quantifying Shot Quality and Predicting the Goal Probability for Football Shots. 2022 2nd Asian Conference on Innovation in Technology, ASIANCON 2022; 2022.
5. McCorriston J, Reed C. Expect the Expected: Approximating the Caliber of Possession Using Shot Quality.
6. Myers BR, Daly AM. Expanding Expected Goals Methodology in Field Hockey. Sport Journal. 2022.
7. Paerels T. Play for the Point or Go for the Win: Expected Goals in the National Hockey League. 2020.
8. theafllab. 2021. [cited 2023]. Available from: https://theafllab.wordpress.com/2021/03/24/introducing-aflxscore/.
9. Murray, H. S., Drovandi, C., Carr, E. J., & Corry, P. (2022). Statistical modelling of goalkicking performance in the Australian Football League. Journal of Science and Medicine in Sport, 25(8), 690-695.
10. Anderson, D., Breed, R., Spittle, M., & Larkin, P. (2018). Factors affecting set shot goal-kicking performance in the Australian football league. Perceptual and Motor Skills, 125(4), 817-833.
11. Dwyer, D. B., & Young, C. M. (2024). Shots at goal in Australian football: historical trends, determinants of accuracy and common strategies. Journal of Science and Medicine in Sport.

# ESTIMATING THE CAUSAL EFFECT OF DIFFERENT INSIDE 50 DECISIONS ON SCORING IN AUSTRALIAN FOOTBALL

Michael Allen[a], Matthew C. Varley[a], Demis Maslo[b], David Carey[a*]

[a] *La Trobe University, Sport and Exercise Science, School of Allied Health, Human Services & Sport*
[b] *Victorian Institute of Sport*
[*] *Corresponding author: d.carey@latrobe.edu.au*

## Abstract

This study aimed to analyse the impact of different decisions on scoring in Australian rules football. Specifically, it explored the outcomes of passing versus taking a shot when players were within a reachable distance of goal. Using a dataset comprising 6,081 Disposals of Interest (DOI) sourced from play-by-play data, the study examined players' likelihood of taking a shot vs. passing to a teammate within 35 and 60 meters from the goal, considering various contextual factors influencing the player in possession. Causal inference was employed, utilizing propensity score matching to compare similar situations resulting in different decisions. A total of 1,045 pairs meeting an intra-pair variation limit of 5% were grouped for analysis. Additionally, video footage was manually reviewed to tag additional variables, which were then applied to 665 DOI to understand their impact on shot propensity. New propensity scores were computed for these instances, resulting in 109 matched pairs. The study used score and equity as outcome measures. Analysis across four linear models consistently indicated a positive causal effect favouring taking a shot at goal, with equity showing a greater advantage over chain points. Comparing matched controls under similar game contexts revealed that players opting to take a shot were, on average, 0.83 points better off than those choosing to pass (95% CI [0.15 – 1.51]). Propensity score matching demonstrated a high level of adaptability, suggesting its utility for future research exploring different seasons or decision-making processes in Australian Rules Football. The findings offer valuable insights for coaches and support staff seeking to enhance scoring efficiency through tactical development.

**Keywords**: AFL, causal inference, decision making

## 1. INTRODUCTION

Australian footballers face many situations throughout a match that require split-second decisions to be made. One such decision is whether to pass the ball to a teammate or take a shot at goal when within a reachable distance of goal. Over the past decade, Australian Football League (AFL) scoring averages have fallen, with teams scoring approximately ten points less per match than they did ten years ago (1). This is despite the total number of inside 50s (number of times the ball is moved into the forward zone) remaining stable. Whilst factors such as rule changes, interchange caps, and improved defensive tactics may explain this decline, in this paper, decision making of the player with the ball will be investigated as a way for teams to increase scoring.

There has been limited research assessing decision making in Australian football (AF). Passing decisions have been analysed using expected outcomes (the probability of the ball reaching a teammate or opposition player) (2), where short kicks with lower decision values (the expected outcome of the pass, divided by the maximum outcome in the player's kicking range) have been favoured over longer passes with higher decision values. This is likely due to easier identification and execution of these options. In practical terms, a lower decision value translates to a lower expected advantage for the team as a result of executing a pass. Gaining field position may give the team a greater chance of scoring, however, these options may increase the risk of a turnover, thus, players may prefer to maintain possession of the ball with lower risk passes. Other research that modelled decision making in AF also found a trend towards shorter pass options (3). This previous analysis displays a focus on spatiotemporal data obtained from global positioning system (GPS) and local positioning system (LPS) technology for model creation. The decision making analysed in this research observed the success of passing, but not in the context of scoring or decisions to pass or take a shot.

The effectiveness of different decisions in sport is difficult to measure using data collected from matches, as only one action is taken in each situation. The complexity of behaviour in AF, in terms of ball movement and player positioning, means that exact replications of situations with different decisions are rare, making it difficult to make direct comparisons between the two. For example, a player may be faced with a situation where they have two possible decisions, A (take a shot at goal) or B (pass to a teammate). Here, only the outcome following the chosen decision can be observed, while the outcome if the other decision had been chosen is unknown. The difference in outcomes for the different decisions is known as the causal effect (4). Randomised controlled trials are considered the gold standard of testing for causal relationships, whereby results from control and placebo groups allow for an understanding of the effect that a treatment has on an outcome (5). These, however, are inappropriate for implementation in AF in-game decision making research due to the impracticality of enforcing decisions that are made throughout a match. For example, allowing certain players to only take a shot at goal and others to only pass the ball to observe the outcomes of each decision is impossible to administer during competitive match-play, as players naturally have a choice for the decisions they make on field. To overcome this, observational causal inference may be suitable for calculating the causal effects of decisions (4). One approach to observational causal inference is propensity score matching (6). Propensity scores estimate the probability of receiving a treatment (making a decision), based on the pre-treatment (pre-decision) characteristics (6). These characteristics provide an understanding as to why a decision may have been made. In this paper, these characteristics relate to contextual variables surrounding a player in possession of the ball. These inform the likelihood of a player taking a shot, so that the effectiveness of different decisions can be assessed in different circumstances. Propensity score matching then allows situations with a similar propensity, but where a different decision was made, to be paired together. This reduces bias between observations of difference decisions, allowing for a more balanced calculation of causal effects (7).

Propensity score matching has previously been implemented in other sports' research, such as soccer (8). Examining the effect that crossing the ball had on shot generation, it was found that possession chains involving this action led to a shot 5% more often than those that did not. Propensity score matching was also implemented in ice hockey research to analyse zone entry methods in the National Hockey League (9). Identifying the causal effect of carry-in and dump-in methods, the authors estimated that a team could score approximately one extra goal every 50 carry-ins. Such results may be useful for developing decision making processes, with this project looking to apply a similar method for analysis.

This research will investigate the scoring impact of deciding to pass or take a shot at goal in AF when faced with similar contextual factors. The findings may provide a starting point for coaches and support staff to develop tactics that improve future scoring efficiency.

## 2. METHODS

### Data

The data used in this project was collected by Champion Data, the principal statistics provider of the AFL and the video footage was collected by Essendon Football Club (EFC), an AF team in the Australian Football League (AFL). In game event data from every match of the 2020 AFL season (162 matches) was included in this data set, which contained 667 players and 80 variables related to each statistic collected throughout the season. These included variables related to disposal context and effectiveness, match context, and information about possession chains and locations on the field. Match footage only included 17 EFC matches. Data analysis was conducted using R and Hudl Sportscode version 12.2.14.

### Disposal of Interest

A disposal of interest (DOI) was used to assess the decision (either passing or taking a shot at goal) made by the players and the effectiveness of each decision. This was defined as the first kick in a possession chain within 35 to 60 metres from the attacking goal and had to be unimpeded (i.e., was not smothered upon kicking the ball) (Figure 1). This distance was chosen due to an expected even distribution between the decision to pass or take a shot and was thought to be a key area for creating scoring opportunities. In total, 6,081 DOI were identified for analysis. The data contained a kick intent label that allowed for shots at goal that did not register a score to be identified.

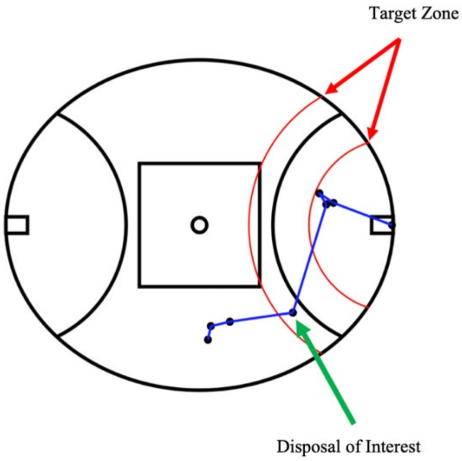



***Figure 1:*** A typical Australian rules football ground with the red lines used to demonstrate the target area that was used to qualify a disposal of interest. The blue markings demonstrate the movement within an example possession chain, and the first kick within the zone labelled as the DOI.

**Decision context variables**

Contextual variables were included for each DOI to identify similar situations and provide an understanding of why the player made a certain decision. The description of these contextual variables is shown in Table 1. Distance and angle to goal were calculated from the manually annotated x- and y-coordinates provided by Champion Data. Pressure points were also assigned by Champion Data (Table 2) based on the proximity of the defender closest to the player with the ball.

***Table 1:*** Contextual variables that surround a disposal of interest.

| Variable | Description |
|---|---|
| Period | Quarter number |
| Period Seconds | Time in quarter |
| Pressure level | Level of pressure applied to the player disposing of the ball |
| Margin | Score differential for the team in possession of the ball |
| Distance | Distance from goal (metres) |
| Angle | Angle from the centre of goal |

***Table 2:*** Classification of pressure points variable.

| Level | Description |
|---|---|
| Set | After mark or free kick without playing on |
| No Pressure | Not close enough to rush the ball carrier |
| Corralling | Guarding space but not approaching |
| Chasing | Gaining from behind or forcing rushed disposal |
| Closing | In area and approaching ball carrier |
| Physical | Making contact before disposal |

A limitation of the context variables listed in Table 1 is that they do not contain any information about off-the-ball players, in particular passing options and their defensive coverage. As spatiotemporal data was not available, video tagging was used to record player positioning, which allowed for greater contextual detail of situations identified as a DOI. These variables were only used for analysis of EFC matches, with EFC and their opponents being tagged because of time constraints. Hudl Sportscode was used to manually tag match footage to measure additional contextual variables that could not be obtained directly from technical statistics, as described in Table 3. A total of 665 DOI were tagged with additional video context variables.

*Table 3:* Variables manually annotated from video footage.

| Variable | Description |
|---|---|
| Zone Density | Number of players within 50m of goal that are within a reachable distance from the player in possession of the ball |
| Outnumber | The numerical difference in the identified zone between the attacking and defending teams |
| Nearest Pressure | The direction of the defender closest to the player in possession of the ball |
| Attacker Closest to Goal | The distance of the closest player to goal on the attacking team |
| Defender Closest to Goal | The distance of the closest player to goal on the defending team |

**Propensity Score matching**

Propensity score matching was used to calculate the average effect of choosing to take a shot at goal versus trying to pass to a teammate when between 35-60m from goal and faced with a similar situation context. Propensity scores were used to identify similar situations to calculate the causal effect of different decisions. This involved estimating the probability of a player taking a shot at goal based on the contextual factors presented in Table 1. Logistic regression was used to calculate these scores due to its effectiveness in previous research (10). Attempted shot on goal (binary) was used as the dependent variable and contextual factors as the independent variables (Table 4). For example, a propensity score of 0.65 would translate to a 65% probability that the player would take a shot at goal based on the factors present at the time of disposal. To account for non-linearity, natural splines were used for the time in quarter, margin, distance to goal, and angle from goal variables. Two models were created, one using context variables from the event data and the other using the video tagged additional context variables in addition to the event data. The matchit function in the Matchit R package (11) was used for propensity score matching (Table 4). This involved matching instances with similar propensities, but with different decisions made, into pairs. To improve the quality of matches a caliper was set at 0.05. This established an intra-pair difference limit of 5%, meaning pairs could have a maximum of 5% difference in estimated probability of taking a shot. These pairs were used for estimating the causal effects of taking a shot, with the method producing new data sets that were balanced and unbiased for use in analysis.

*Table 4:* Propensity score matching variables using the Matchit package ("ns(…)" indicates that natural splines with 5 degrees of freedom were used to allow for non-linear relationships in the model for a variable).

| Variable | Description |
|---|---|
| Dependent variable | Action taken (shot or pass) |
| Context variables (event model) | Period + Pressure_level + ns(period_seconds) + ns(margin) + ns(distance) + ns(angle) |
| Context variables (event + video model) | Period + Pressure_level + ns(period_seconds) + ns(margin) + ns(distance) + ns(angle) + zone_density + outnumber + nearest_pressure + closest_att + closest_def |
| Method | Nearest neighbour |
| Caliper Value | 0.05 |

**Estimating the effectiveness of different decisions**

Two ways of evaluating the outcome of a decision were used. The first was to simply use the number of points scored in the possession chain containing the disposal (and decision) of interest. This could be 6 if the chain ended in a goal, 1 if ending in a behind, or 0 for all other endings (including a turnover). The alternative approach was to incorporate the field equity of the end state of the chain (12, 13). This concept estimates the value of having the ball in a particular field location and game state (e.g., a stoppage inside the forward 50). Although possession chains may not result in a score, the position on the field where the chain ends may lend itself to a better chance of scoring next (i.e., positive equity), this was considered important when evaluating the decision of the player. Under the score equity framework chains ending in a goal were still valued at +6, behinds at +1.22, inside 50 ball-ups at +0.38, inside 50 throw-ins at +0.45, and turnovers at -0.33.

The creation of balanced data sets with propensity score matching allowed for a more estimation of causal effects without the bias introduced by the context of different decisions. A linear regression model was used to calculate the average treatment effect (ATE), using the chain points as the dependent variable and the action taken as the independent variable. This was applied to the subset of data remaining after matching had occurred. The treatment in this instance related to taking a shot, and the effect indicating which decision produced a scoreboard advantage, using matched controls. A positive value would indicate an advantage for taking a shot, whereas a negative value would favour passing. This was applied to both the event-only, and event and video data sets to determine the influence that additional context variables had on the ATE. The effect was also calculated based on chain equity to observe whether this produced an increased or decreased advantage. Again, this was applied to both data sets, with equity used as the dependent variable and the action taken as the independent variable.

## 3. RESULTS

### Propensity Score model

The propensity score model results are shown in Figure 2. The figures show the estimated relationship between each contextual variable and the probability that a player will take a shot at goal. The addition of natural splines produced a noticeable effect on variable estimations in the logistic regression model. The probability of taking a shot was lowest at the greatest angles, while highest when the angle was minimised.

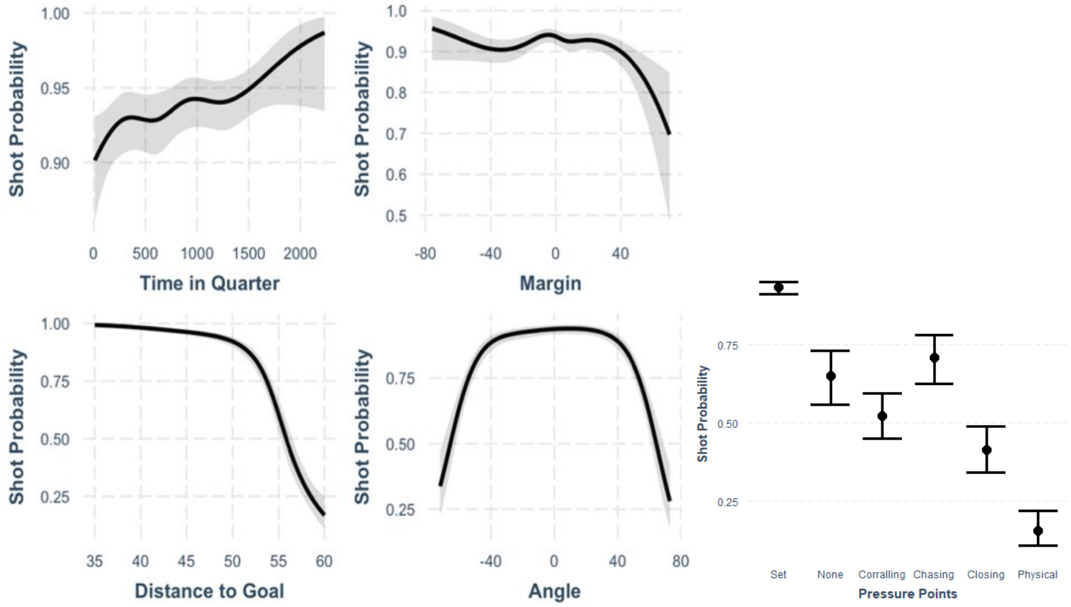



***Figure 2:*** Adjusted predictions of propensity model. In each sub-figure one variable is changed while the others are constant at their mean values (mean time = 767s, margin = 0.8 points, distance = 49m, angle = -2 degrees, pressure = set)

The calibration plot shown in Figure 3 demonstrates the level of agreement between predicted and actual outcomes for the propensity score model using event data only. The close proximity to the dotted line indicates that the model produced results similar to the expected distribution, making it appropriate for use in propensity score matching. For example, a predicted probability of 75% would indicate that 75% of instances with similar propensity scores would result in a shot at goal being taken. Figure 3b presents the agreement from the event and video data set, with this model also demonstrating appropriate approximation of propensity scores. The smaller sample size was responsible for a larger confidence interval, however, still appeared to fit the data well. This model was also used for propensity score matching.

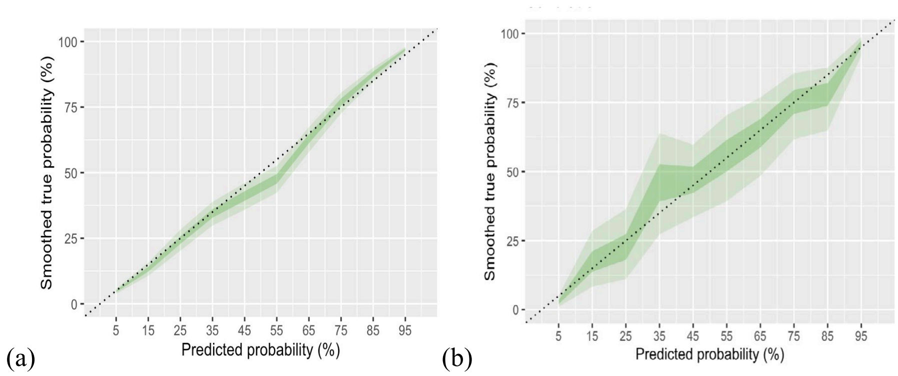



***Figure 3: (a)*** Calibration plot comparing observed outcomes and predicted outcomes based on propensity scores from event data. ***(b)*** Calibration plot comparing observed outcomes and predicted outcomes based on propensity scores from event and video data.

**Propensity Score Matching**

As seen in Figure 4, the initial action distribution of unmatched DOIs skewed in opposite directions, with very little overlap of frequencies at similar propensities. Matching using a 5% intra-pair variability limit produced 1,045 pairs that contained greater similarity in terms of propensity (Figure 5). An example of matching pairs can be seen in Table 5, with each grouping containing one shot and one pass option. 109 pairs were created from video tagging.

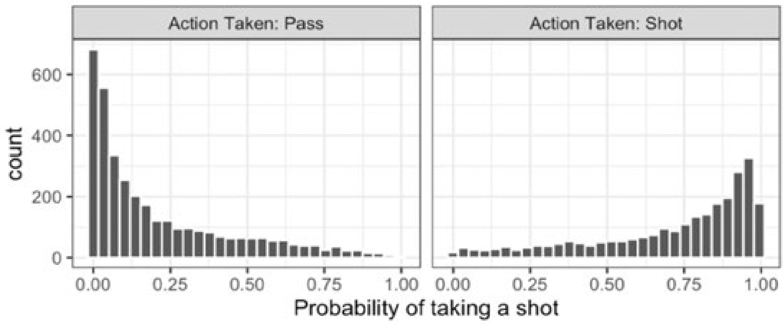



***Figure 4:*** Action distributions based on propensity scores using event data before matching

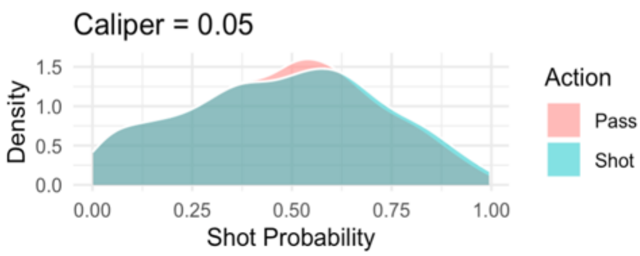



***Figure 5:*** Propensity distributions following matching without a caliper and using a 0.05 caliper limit.

***Table 5:*** Example pairs following propensity score matching using event contextual variables.

| Player | Qtr | Seconds | Pressure | Margin | Distance (m) | Angle | Propensity | Action | Pair |
|---|---|---|---|---|---|---|---|---|---|
| T. Liberatore | 1 | 582 | Closing | 2 | 47 | -21 | 0.4100 | Shot | 1 |
| T. Liberatore | 2 | 1343 | Corralling | 1 | 46 | 47 | 0.4099 | Pass | 1 |
| J. Dawson | 3 | 825 | Chasing | -32 | 60 | -32 | 0.0163 | Shot | 2 |
| J. Harmes | 1 | 450 | Closing | -7 | 57 | -37 | 0.0162 | Pass | 2 |
| R. West | 4 | 45 | Set | 26 | 45 | 13 | 0.9421 | Shot | 3 |
| R. Sloane | 4 | 1510 | Set | 15 | 48 | -37 | 0.9246 | Pass | 3 |

### Effect of extra video context variables on propensity scores

The effect that video variables had on propensity scores was significant in some cases. Using only event data initially, distance and angle variables appeared to have the greatest effect on the likelihood of a player taking a shot at goal. Noticeable changes could be seen when further variables obtained from video footage were included in calculations. For example, an observation of a player 52m from goal, directly in front, in a close game in the second quarter was estimated have a probability of taking a shot of 46% under the event-only propensity score model. When extra context from video was added to the model the probability increased to 100% because there were no teammates within the entire 50m zone (and therefore no passing options).

### Causal Effects

A positive average treatment effect was found for taking a shot in all four linear regression models (Table 6). This is likely due to shots predominantly producing a score, irrespective of propensities. Over time this accumulation of points may be greater than those resulting from passing, which is reliant on finding another link in the possession chain to score. It is also worth noting that although a player may have a low likelihood of taking a shot, this does not mean that a passing option is available or easy to execute. This is highlighted in Table 7, which shows that only 39.81% of passing chains led to a score, compared to 75.21% of shots. With missed shots still gaining 0.22 points of equity, passes that were regathered by the opposition (-0.33 points) further widened the effect between actions. However, when observing only the chains that ended in a score, the average points scored were greater in passing chains (3.80 points) than shots (2.95 points).

***Table 6:*** Average treatment effects of taking a shot over passing.

| Outcome Variable | Context variables | Average Treatment Effect | 95% CI | p-value |
|---|---|---|---|---|
| Chain Points | Event | + 0.46 points | 0.24 – 0.68 | 0.0000006 |
| Chain Equity | Event | + 0.56 points | 0.34 – 0.78 | 0.00004 |
| Chain Points | Event + video | + 0.68 points | 0 – 1.36 | 0.04495 |
| Chain Equity | Event + video | + 0.83 points | 0.15 – 1.51 | 0.01694 |

***Table 7:*** Scoring outcomes of pass and shot decisions.

| | Pass | Shot |
|---|---|---|
| **Score of any kind** | 39.81% | 75.21% |
| **Average score per scoring chain** | 3.80 points | 2.95 points |

## 4. DISCUSSION

The aim of this paper was to estimate the causal effects of taking a shot at goal vs passing when between 35 and 60 metres from goal in AF during the 2020 AFL season. Causal inference, using propensity score matching, was conducted to appropriately analyse this effect in situations that contained a similar likelihood of taking a shot but resulted in a different decision (pass or shot) being made. To identify these action probabilities, contextual factors were gathered from technical statistics and video footage from throughout the season, which were used in logistic regression models for calculating propensity scores. Chain equities were also established to assess the impact of a decision beyond just the points that resulted from a chain of possession. While not included in their study design, Anderson et. al., (2018) also cited the possible advantage of non-scoring outcomes, and the indirect effect that advancing the ball towards the attacking goal may have on scoring (14). Forward 50 stoppages and behinds were found to have a positive effect on the attacking team due to a higher probability of scoring next, whereas forward 50 turnovers produced a negative effect. Using four different linear regression models to estimate the average treatment effect, each model found a significant ($p<0.05$) positive effect for taking a shot, with equity models demonstrating a greater advantage than models with the same variables using chain points as the independent variable.

The causal effect findings of the two decisions are potentially due to the greater likelihood of scoring when taking a shot. When opting to pass to a teammate, the success of the chain is reliant on finding at least one extra link in the possession chain, which may not always be obvious or easy to execute. This would reflect previous research which found that turnovers were the most frequent outcome of inside 50 entries (15). A high proportion (73.79%) of non-scoring chains in this paper ended in a turnover, which produced -0.33 points of equity, demonstrating the trade-off of trying to find a teammate in a better position. Of the 3,514 passes made, 1,399 chains resulted in scores (783 goals and 616 behinds) and 1,593 ended in a turnover, meaning a team was more than twice as likely to give the ball back to the opposition than they were to score a goal. This does not, however, mean that passing should be excluded as a valid decision for players between 35 and 60 metres from goal. In situations where a player was unable to score, creating a forward 50 stoppage demonstrated a positive equity for ball ups (+0.38 points) and throw ins (+0.45 points). These findings are consistent with equity findings by Jackson (2016), with non-error possessions or disposals found to produce positive equity for the attacking team, while forward 50 errors resulted in negative equity (12). The field position of these stoppages is the likely reason for the superior scoring outcomes, despite teams in their defensive 50 winning more first possessions. The scoring efficiency of shots, as seen in Table 12, was nearly double that of passes, which is a likely reason for the positive causal effect seen in Table 11. Despite passes producing a higher average score per scoring chain, the increased likelihood of not scoring at all was detrimental to the causal effect of this decision.

Distance and angle to goal were two variables that demonstrated noticeable variation amongst propensities, whereby the greater these two values were, the lower the chance that a player would take a shot at goal. This is likely due to a player's confidence in scoring a goal from these locations. These findings are supported by previous research, where it was found that players that were less than 30 degrees from goal were approximately 2.6 times more likely to score a goal than those on a greater angle (14). It has also been found that fewer behinds were scored from set shots directly in front of goal (16), which may explain the reasons for passing from greater distances and angles. Identifying teammates in a better position to score may override the confidence a player has in their abilities to score from their location. It has also been found that when taking a set shot in excess of 50 metres, approximately a quarter of attempts landed inside the field of play (14). Referring back to Figure 4, the propensity to take a shot also began to fall sharply around this distance as players began to reach their maximum kicking range. The success from longer distances may also impact a player's willingness to take a shot. Previous research demonstrated changes in kicking technique between 30 and 40 metres from goal in AF (17), which may lead to inconsistent scoring accuracy from certain distances. While this was not conducted under match conditions, kicking techniques and characteristics may also be worth considering to estimate its effect on decision making in AF, if a reliable method of classification could be established.

The inclusion of player positioning variables obtained from video tagging (zonal density, outnumber, nearest pressure, and the attacker and defender closest to goal) produced noticeable changes to propensity scores and a better representation of the circumstances under which the decision was made. While previous studies have found the importance of kicking in producing favourable match outcomes (18), the inclusion of a technical description of the contextual factors would have given greater perspective of the influence of a kick. The additional context obtained from tagging provided greater insight into the factors affecting a decision, why a decision was made, and which variables were most impactful on a decision. The inclusion of valid spatiotemporal data from wearable tracking devices may have provided greater accuracy and reliability of the video-based variables for player positioning at the point of a DOI, due to the interpretation used in tagging. Circumstances did not allow for reliability testing from others to take place, meaning the ability to replicate findings is unclear. This would also allow for more specific variables to be added that couldn't be estimated manually. For example, distances between attacking and defending players may provide further insight into the number of available pass options, as well as the estimated probability of non-scoring outcomes. The effect of handballs as a passing option may also be worth exploring. With this paper only using kicks as a possible DOI, handballing may potentially improve the quality of shots taken, and would need to be assessed to observe whether this would be more beneficial for the team's performance.

The current gaps in AF literature also make it difficult to compare findings and assess the validity of results; however, this paper demonstrated similarities to papers focussing on decision making in other sports. As previously mentioned, propensity score matching has been used to analyse shot generation in soccer when players cross the ball or use other ball movement methods (8). Crossing the ball was shown to have an estimated advantage over possession chains without this action. Despite looking at ways to generate a shot, rather than actual score outcomes, the premise remains similar to this paper, with a player being forced to make a decision that they think will have a positive impact on the team's performance. In a similar manner, Toumi and Lopez's paper used propensity score matching to estimate zone entry decision making in the National Hockey League (9). Again, the causal effect of decisions demonstrated comparable characteristics, whereby a carry-in method was estimated to produce an additional goal for every 50 carry-ins. This method appears to have a high level of adaptability in AF, allowing for implementation across different seasons or areas of observation.

## 5. CONCLUSIONS

This paper aimed to estimate the causal effect of deciding to take a shot at goal instead of passing between 35 and 60 metres from goal in AF. Contextual variables relating to match circumstances (time and current margin), and player positioning with and without possession of the ball, obtained from technical statistics and video footage, allowed for propensity score estimation using logistic regression. These scores were used in propensity score matching to identify the advantage of one decision over the other, using matched controls. Taking a shot at goal was found to have a positive average treatment effect in all four linear regression models when using chain points or chain equity as the outcome measure, and when using technical statistics for context or including video tagged information as well. This method demonstrates adaptability to different seasons and highlights the potential for causal inference to be used as an effective method for analysis of decision making in AF. The findings of this study may provide a potential starting point for future tactical development by coaches and support staff to improve scoring efficiency.

**Acknowledgements**

We wish to thank Essendon Football Club for their assistance in data access.

**References**

1. AFLTables. Yearly Totals and Averages 2021 [Available from: https://afltables.com/afl/stats/yearly.html.
2. Spencer B, Jackson K, Bedin T, Robertson S. Modeling the Quality of Player Passing Decisions in Australian Rules Football Relative to Risk, Reward, and Commitment. Front Psychol. 2019;10(1777).
3. Spencer B, Bedin T, Farrow D, Jackson K. A method for evaluating player decision-making in the Australian Football League. 14th Australasian Conference on Mathematics and Computers in Sport; Sunshine Coast, Australia2018.
4. Spirtes P. Introduction to causal inference. Journal of Machine Learning Research. 2010;11(5).
5. Hariton E, Locascio JJ. Randomised controlled trials - the gold standard for effectiveness research: Study design: randomised controlled trials. BJOG. 2018;125(13):1716.
6. Rosenbaum PR, Rubin DB. Reducing Bias in Observational Studies Using Subclassification on the Propensity Score. Journal of the American Statistical Association. 1984;79(387):516-24.
7. D'Agostino RB, Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Stat Med. 1998;17(19):2265-81.
8. Wu LY, Danielson AJ, Hu XJ, Swartz TB. A contextual analysis of crossing the ball in soccer. Journal of Quantitative Analysis in Sports. 2021;17(1):57-66.
9. Toumi A, Lopez, M. From grapes and prunes to apples and apples: Using matched methods to estimate optimal zone entry decision-making in the National Hockey League 2019 [Available from: https://rpubs.com/atoumi/zone-entries-nhl.
10. Lee J, Little TD. A practical guide to propensity score analysis for applied clinical research. Behav Res Ther. 2017;98:76-90.
11. Greifer N. MatchIt: Getting Started 2021 [Available from: https://cran.r-project.org/web/packages/MatchIt/vignettes/MatchIt.html#ref-stuart2010.
12. Jackson K. Assessing player performance in Australian football using spatial data 2016.
13. O'Shaughnessy DM. Possession versus position: strategic evaluation in AFL. J Sports Sci Med. 2006;5:533.
14. Anderson D, Breed R, Spittle M, Larkin P. Factors Affecting Set Shot Goal-kicking Performance in the Australian Football League. Perceptual and Motor Skills. 2018;125:003151251878126.
15. Aarons MF, Young CM, Bruce L, Dwyer DB. The effect of team formation on defensive performance in Australian football. Journal of Science and Medicine in Sport. 2021.
16. Nicholls MER, Loetscher T, Rademacher M. Miss to the Right: The Effect of Attentional Asymmetries on Goal-Kicking. PLOS ONE. 2010;5(8):e12363.
17. Blair S, Robertson S, Duthie G, Ball K. The effect of altering distance on goal-kicking technique in Australian football ISBS. 2018.
18. Robertson S, Back N, Bartlett J. Explaining match outcome in elite Australian Rules football using team performance indicators. Journal of Sports Sciences. 2015.

# AN ECOLOGICAL DYNAMICS APPROACH TO RUNNING RELATED INJURY

Ben Jones [a,b,c], Jonathan Wheat[b], Kane Middleton[a], David Carey[a], Ben Heller [b]

[a] *School of Allied Health, Human Services and Sport, La Trobe University, Melbourne, VIC 3086, Australia*
[b] *Sport and Physical Activity Research Centre, Sheffield Hallam University, Olympic Legacy Park, 2 Old Hall Rd, Sheffield, S9 3TY, United Kingdom*
[c] *Corresponding author: b9jones@ltu.edu.au*

**Submitted for consideration as (select one):**

- ☒ Oral presentation
- ☐ Poster

**Abstract**

Running is a complex human movement that requires the non-linear interaction of independent biological systems, and the interaction of the runner with their task and environment [1,2]. A high-order signal, such as stride time or COM acceleration time series, that emerges from these non-linear interactions will contain fluctuations that reflect the ability of the human body to utilise its abundant degrees of freedom [2]. These fluctuations can be measured in terms of their magnitude and structure (complexity) [3]. Since changes within these fluctuations may be seen during the development of a running-related injury, longitudinal assessment of variability measures may support injury prediction [2]. However, the between-day reliability of such measures within healthy individuals is largely unknown. This study quantified the reliability of measures frequently used to assess variability in running gait. Biomechanical data was collected using inertial measurement devices whilst injury-free recreational runners completed a series of running trials on a treadmill across two visits. Between-day reliability was quantified using intraclass correlation coefficients, the standard error of measurement and the minimum detectable change. Reliability was found to be dependent on the variability measure and the biomechanical variable to which it was applied. This has implications for studies conducting longitudinal monitoring of variability measures and the potential for injury forecasting.

*Keywords*: Running gait, biomechanics, movement variability, complexity, overuse injury

**References**

[1] K. Davids, P. Glazier, D. Araújo, R. Bartlett, Movement Systems as Dynamical Systems, Sports Med 33 (2003) 245–260. https://doi.org/10.2165/00007256-200333040-00001.

[2] S.T. Fonseca, T.R. Souza, E. Verhagen, R. Van Emmerik, N.F.N. Bittencourt, L.D.M. Mendonça, A.G.P. Andrade, R.A. Resende, J.M. Ocarino, Sports Injury Forecasting and Complexity: A Synergetic Approach, Sports Med 50 (2020) 1757–1770. https://doi.org/10.1007/s40279-020-01326-4.

[3] R.E.A. Van Emmerik, S.W. Ducharme, A.C. Amado, J. Hamill, Comparing dynamical systems concepts and techniques for biomechanical analysis, Journal of Sport and Health Science 5 (2016) 3–13. https://doi.org/10.1016/j.jshs.2016.01.013.

# Mathematical Modelling of Oil Patterns in Tenpin Bowling

Curtis Hooper [a,b] ,
[a] *Sports Technology Institute, Loughborough University, England*

[b] *Corresponding author:* *c.hooper@lboro.ac.uk*

## Abstract

The sport of tenpin bowling has many hidden intricacies that are not immediately obvious. One such feature is the thin film of oil that is applied to the lane by sophisticated lane machines. It is there first and foremost to protect the lane surface from the friction generated with the bowling balls. However, the placement of the oil in leagues and tournaments, known as the oil pattern, dictates how easy it is for bowlers to achieve high scores, where abouts on the lane they should play and with what type of bowling ball they should use. In this paper we shall discuss a mathematical model consisting of two coupled first order differential equations that describes the process of the application of oil to a bowling lane by lane machines manufactured by Kegel. The results of the model show a novel 3-dimensional representation of the oil pattern which can be used to guide the design of oil patterns as well as helping bowlers devise their optimal strategy for playing on them. We will discuss how to make oil patterns more difficult for bowlers to score on without simply increasing the volume of oil applied to the lane which causes problems for the pin setters and lane machinery.

**Keywords**: Tenpin bowling, mathematical modelling, oil patterns

## 1. INTRODUCTION

Tenpin bowling is played by millions of people worldwide each year and remains the number one participation sport in the United States of America (USBC). The aim of the game is to roll a ball down a long, narrow lane such that it knocks down 10 pins that are arranged in an equilateral triangle. The lane is 60 feet long from the foul line to the centre of the front pin, and 3.5 feet wide from the right edge to the left edge of the lane, beyond which is a gutter. The width of the lane is divided into 39 individual boards each just over 1 inch wide. Board 2R is the board closest to the right gutter as one looks from the foul line towards the pins, board 2L is the closest board to the left gutter, and 20 C is in the middle of the lane. From right to left the boards are denoted 2R, 3R, 4R, …, 19R 20C 19L, …, 4L, 3L, 2L.

However, there are many hidden features that need to be considered by amateurs and professionals so that they can maximise their chance to knock down all 10 pins in 1 shot, which is known as a strike. One particular feature is the thin film of lane conditioner, which is more commonly referred to as oil, that is applied to the lane. Historically, applying the oil to a lane was a manual task involving spray guns, but nowadays, it's executed with precision by advanced lane machines. This process mirrors the steps taken in resurfacing sports surfaces like curling or ice hockey, where a fine layer of water is applied post-cleaning.

In total, between 15 mL and 35 mL of oil is typically applied to a lane, with the location of the oil contributing to the area of the lane that a bowler will play in, and how big their target is in order to have a good chance of getting a strike. Typically, most of the oil is concentrated towards the front and centre sections of the bowling lane, gradually decreasing as it extends further down the lane. The thickness of the oil is approximately 10 mm perpendicular to the lane's surface at its maximum. If the bowler imparts rotation on the ball upon release such that it rotates about an axis that is not perpendicular to the velocity vector, it will initially travel in a straight line due to the high concentration of oil in the front part of the lane, causing minimal friction and allowing for skidding. As the volume of oil decreases between the ball and the lane, friction increases, causing the ball to follow the direction of rotation and the overall path of the ball becomes curved.

Equations governing the motion of a sphere rolling on a surface have been extensively studied across various sports (Cross, 1998) (Hubbard & Smith, 1999). In the case of tenpin bowling, where the ball both rolls and rotates, specific studies have been conducted to model the ball's parabolic path (Huston et al, 1979) (Frohlich, 2004). Experimental methods, such as using inertial measurement units within bowling balls, facilitate the extraction of relevant ball properties (King et al, 2011).

To maximise their target area, bowlers typically play in a part of the lane where there is less oil near the edges of the lane and more oil towards the centre to guide the ball toward their intended target line. This strategic placement ensures that even if the ball deviates from its initial trajectory, the rotation coupled with the oil will guide it towards the desired target area at the pins as illustrated in Figure 1. Adjusting the ratio of oil volume between the centre and outer parts of the lane can impact the level of accuracy required for consistent striking. A higher ratio provides a larger margin for error, while a lower ratio demands greater precision due to more uniform friction across the lane surface as shown in Figure 2.

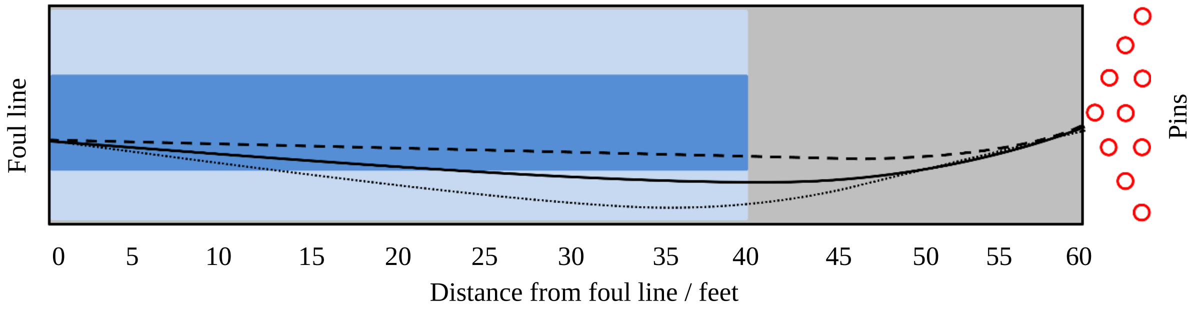



Figure 1: Trajectories of a bowling ball on an easy oil pattern. The intended shot is the solid line, a shot missed to the left is the dashed line, and a shot missed to the right is the dotted line. The oil pattern is 40 feet long, and darker shades of blue correspond to more oil.

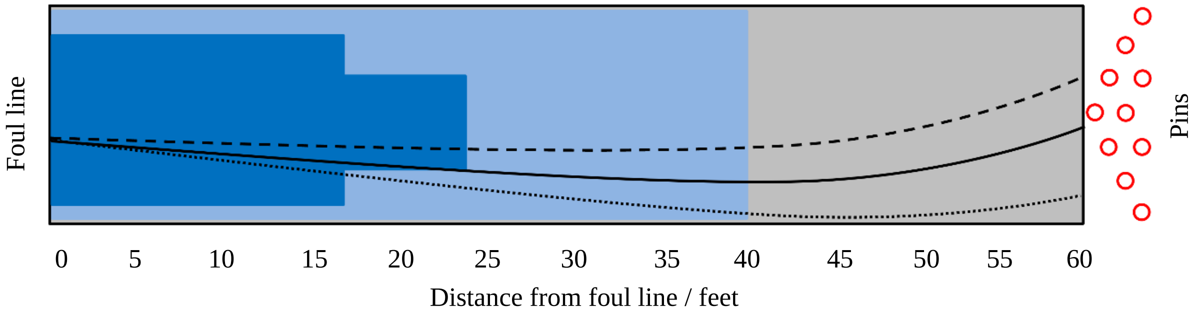



Figure 2: Trajectories of a bowling ball on a difficult oil pattern. The intended shot is the solid line, a shot missed to the left is the dashed line, and a shot missed to the right is the dotted line. The oil pattern is 40 feet long, and darker shades of blue correspond to more oil.

This paper reviews the mathematical model developed by the author (Hooper, 2023) and uses it to plot a range of oil patterns of varying lengths and difficulty. The operational principles of the oiling mechanism are explained first, before deriving the model equations and fitting unknown parameters using experimental data. The results of the model are illustrated for various oil patterns, with both 3D and 2D representations provided.

## 2. APPLICATION OF OIL

The oiling mechanism is located at the back of the machine. At the front of the machine is a cleaning mechanism that removes any oil that is already on the lane as the machine moves towards the pins. This helps to ensure that the oil being applied to the lane as the machine travels towards the pins, known as the *forward* pass, is being applied to a completely dry surface. The cleaning mechanism is disabled as the machine travels back from the pins towards the foul line on the *reverse* pass.

The oiling mechanism studied here (Kegel Sanction Technology 2024) consists of three main components: the oil head (OH), transfer brush (TB), and buffer brush (BB). The OH is fixed to a rail that runs across the width of the lane, parallel to the foul line, once the lane machine is positioned on the lane and ready to apply the oil pattern. It moves along the rail at a constant speed, spanning from the second board in from the right gutter (board 2R) to the second board in from the left gutter (board 2L). Oil is continuously pumped from the OH between selected boards as it traverses between boards 2R and 2L at a predetermined pump setting $v_p$ which is usually in the range 40 µL – 60 µL. The machine offers six preset speeds, denoted by $S_s$ $(s = 1, 2, \dots, 5, 6)$, to travel along the lane at ranging from 0.254 m/s (10”/s) to 0.762 m/s (30”/s), but the first stream of oil is applied to the BB whilst the machine is stationary at the foul line.

Oil discharged from the OH is directly applied to the TB, which is in contact with the cylindrical BB rotating at a constant angular velocity. The BB applies oil to the lane when in contact with it. Upon reaching the pre-programmed oil pattern length $P_L$ on the forward pass, as the machine moves from the foul line toward the pins, the BB is raised from the lane, marking the end of oil application for the forward pass. The lane machine can also apply oil on the reverse pass, from the pins back to the foul line, with the BB reestablishing contact with the lane at a designated distance known as the reverse brush drop distance. Additional loads can be added as per the pattern designer's specifications.

## 3. MATHEMATICAL MODEL

To create a model for the oiling mechanism discussed in Section 2, each of the 37 boards from 2L to 2R are treated separately. The volumes of oil on each board are treated as functions of $x$ which represents the distance down the lane from the foul line.

FORWARD PASS

The rate of change of oil on board $i$ of the TB as the machine travels down lane is given by

$$\frac{dv_1^i}{dx} = P_i(x) - Q_i(x),$$

where $P_i(x)$ is the rate at which oil is applied to the TB, and $Q_i(x)$ is the rate at which oil leaves the TB. The function $P_i(x)$ takes the form of a series of $\text{sech}^2$ pulses to describe when the pump applies a load of oil to TB as

$$P_i(x) = K_s \sum_{k=1}^{F_i} \text{sech}^2\big(A_s(x - x_k)\big),$$

where $F_i$ is the number of loads of oil that board $i$ receives on the forward pass, and $x_k$ is the distance from the foul line at which load $k$ is applied. The parameters $K_s$ and $A_s$ are dependent on the speed setting $s$ and the pump setting $v_p$ of the machine and are calculated later. For $Q_i(x)$, a simple relationship is taken whereby the rate at which oil is removed from the TB is proportional to the difference of oil on the TB and BB as

$$Q_i(x) = k_1(v_1^i(x) - v_2^i(x)),$$

where $v_2^i(x)$ is the volume of oil on the TB at the position corresponding to board $i$. The initial condition for this equation is

$$v_1^i(0) = v_0^i + v_p,$$

where $v_0^i$ is the volume of oil on the TB at the start of the forward pass of the machine. The differential equation for the BB takes the same form as that for the BB, and is given by

$$\frac{dv_2^i}{dx} = k_1\left(v_1^i(x) - v_2^i(x)\right) - k_2\left(v_2^i(x) - v_3^i(x)\right),$$

where $k_2$ is a constant, and $v_3^i(x)$ is the volume of oil on the lane at the start of the forward pass. We assume that the lane is completely dry and does not have any oil on it, thus $v_3^i(x) \equiv 0$. The initial condition for the BB is

$$v_2^i(0) = v_2^i + k_0 v_p,$$

where $v_2^i$ is the volume of oil on the BB at the start of oiling, and $k_0 \in [0,1]$ is a constant. Finally, the volume of oil transferred to the lane on the forward pass is given as some factor of the oil on the BB over the lane as

$$v_F^i = k_3 v_2^i,$$

where $k_3$ is a constant.

REVERSE PASS

The equations for the brushes on the reverse pass take much the same form as their equations in the forward pass above. The main difference is that the oil on the lane given by $v_3^i(x)$ is given by $v_F^i(x)$ as there now will be some oil on the lane resulting from the forward pass of the machine, and that the initial conditions for the brushes are given by the solution they take at the end of the forward pass. The equations for the reverse pass are given by the following system:

$$\frac{d\hat{v}_1^i}{dx} = K_s \sum_{k=1}^{r_i} \operatorname{sech}^2\big(A_s(x - \hat{x}_k)\big) - k_1\left(\hat{v}_1^i(x) - \hat{v}_2^i(x)\right), \qquad \hat{v}_1^i(0) = v_1^i(P_L)$$

$$\frac{d\hat{v}_2^i}{dx} = k_1\left(\hat{v}_1^i(x) - \hat{v}_2^i(x)\right) - k_2\left(\hat{v}_2^i(x) - v_F^i(x)\right), \qquad \hat{v}_2^i(0) = v_2^i(P_L),$$

$$v_R^i = k_3 \hat{v}_2^i,$$

Then, the total volume of oil transferred onto board $i$ of the lane is given by

$$v_T^i = v_F^i + v_R^i.$$

LOAD DISTANCES

The time taken for the OH to travel across the rail is ~ 1.67 seconds. This can be used along with the speed of the machine and the boards where the pump is open to calculate the distances where each board receives a load of oil (Hooper 2023).

## 4. PARAMETER FITTING

The constants $K_s$ and $A_s$ can be found by considering that the area underneath the $\operatorname{sech}^2$ pulses must be equal to $v_p$. Therefore we have

$$V_p = \int_{-\infty}^{\infty} K_s \operatorname{sech}^2(A_s x)dx\,,$$

which gives

$$V_p = \frac{2K_s}{A_s}.$$

The width of a pulse along a board is also dependent on the speed of the machine and is given by $W_s = 1.67S_s/37$. Taking measurements of full width at tenth maximum of a $\operatorname{sech}^2$ curve of the form used in the function $P_I(x)$ gives the equation

$$\frac{K_s}{10} = K_s \operatorname{sech}^2(A_s x),$$

which has solutions

$$x_{1,2} = \pm\frac{1}{A_s}\operatorname{sech}^{-1}\left(\sqrt{\frac{1}{10}}\right),$$

hence the width of a pulse is given as the difference between $x_1$ and $x_2$ as

$$W_s = \frac{2}{A_s}\operatorname{sech}^{-1}\left(\sqrt{\frac{1}{10}}\right).$$

Solving $K_s$ and $A_s$ gives

$$K_s = \frac{V_P}{W_s}\operatorname{sech}^{-1}\left(\sqrt{\frac{1}{10}}\right),$$

$$A_s = \frac{2}{W_s}\operatorname{sech}^{-1}\left(\sqrt{\frac{1}{10}}\right).$$

The remaining parameters were fitted by numerically fitting the model solution to best match measurements taken of an oil pattern that was applied to a bowling lane. The model was then validated with those fitted parameters by using them to plot the volume of oil of a different oil pattern and comparing the solution to measurements taken of that oil pattern when it was applied to a lane. The agreement was close (Hooper, 2023).

Further validation was provided as the model captures a feature of the machines without any constraint being imposed. It is standard practice when lane maintenance is being performed during leagues and tournaments to apply the oil pattern to 2 lanes to get the imprint of the pattern onto the brushes, and then to redo those 2 lanes and continue on and dress the rest of the lanes being used. For example, if lanes 1 – 8 are being used for a league, the order that the lanes would be dressed would be 1, 2, 1, 2, 3, …, 7, 8. Lanes 1 and 2 are referred to as a *burn* pair, and are done twice to allow sufficient oil to build up on the brushes such that the output of oil onto the lanes thereafter is consistent.

## 4 RESULTS

We now present solve the governing equations by using the 4$^{th}$ order Runge-Kutta scheme in MATLAB and show the results of a variety of oil patterns of different levels of difficulty, volumes, and lengths. In the first instance, the progression of the solution is shown from the first lane with complete dry brushes to when the solution becomes stable to illustrate the burn pair requirement. After, the solutions are shown as the oil patterns on the 5$^{th}$ lane. The oil patterns chosen to show are Stone Street (Figure 3), Turnpike (Figure 4), World Bowling London (Figure 5), and World Bowling Melbourne (Figure 6).

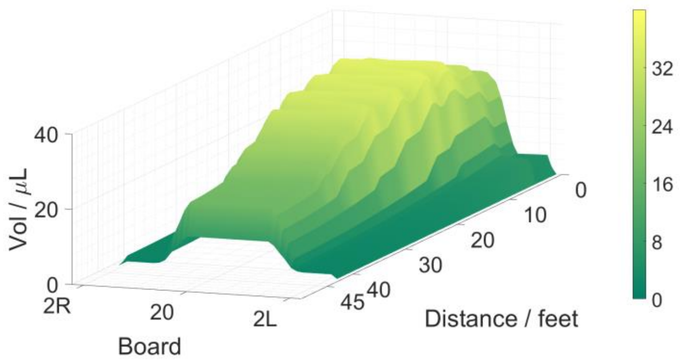



Figure 3: An ‘easy’ medium length oil pattern (Kegel Stone Street)

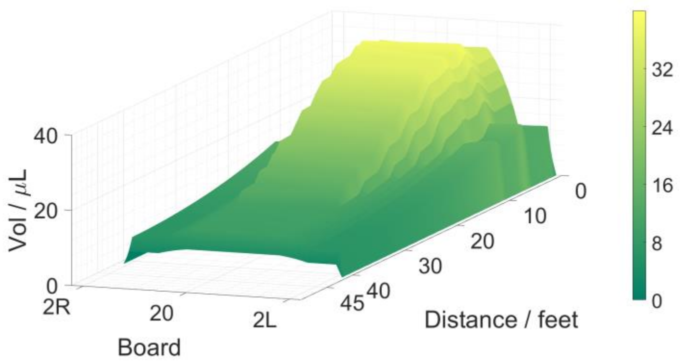



Figure 4: A ‘difficult’ medium length oil pattern (Kegel Turnpike)

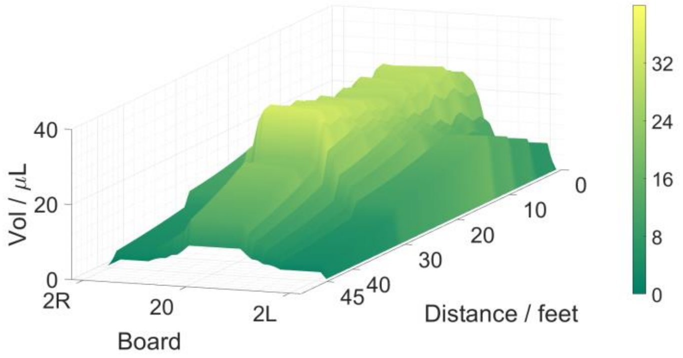



Figure 5: A difficult long length oil pattern (World Bowling London 2020)

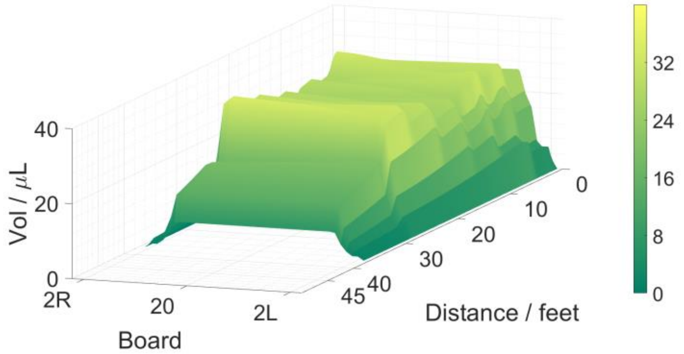



Figure 6: A difficult short length oil pattern (World Bowling Melbourne 2020)

The volume of oil across the lane for each of the 4 patterns are shown in Figure 7 at 15 ft from the foul line, and at 30 ft in Figure 8.

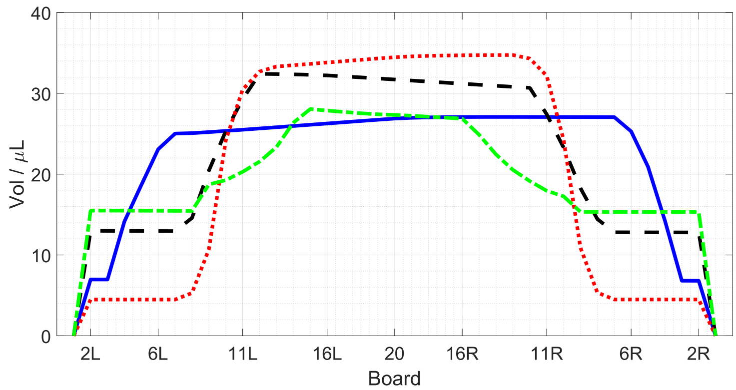



Figure 7: Volume of oil across the width of the lane for Stone Street (red, dot), Turnpike (black, dash), London (green, dash dot), and Melbourne (blue, solid), at 15 feet from the foul line.

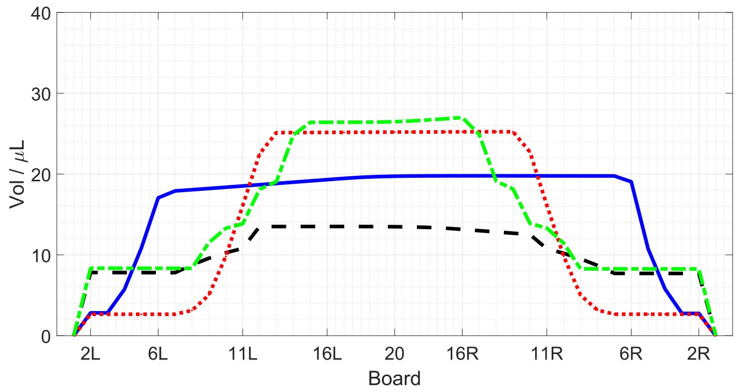



Figure 8: Volume of oil across the width of the lane for Stone Street (red, dot), Turnpike (black, dash), London (green, dash dot), and Melbourne (blue, solid), at 30 feet from the foul line.

## 5. DISCUSSION

On comparison of Figures 3 – 6 it is clear to see the different shapes of the oil patterns presented. For the two medium length patterns in Figure 3 and Figure 4, it can be seen why the pattern presented in Figure 3 presents an easier opportunity for a bowler to strike regularly than the pattern in Figure 4. Indeed, there is more oil in the middle of the lane (between boards 10L and 10R) on the easier pattern than there is on the harder pattern. However, the location of the oil on these boards along the lane is important. There is approximately the same volume of oil in the middle and front of the lane, between boards 10L and 10R and 0 – 15 feet, on both patterns. The additional oil that is there on the easier pattern is located further down lane, from 20 feet onwards. This decrease in oil with distance can be seen by comparing the red dot curve with the black dash curve in Figure 7 and Figure 8, which shows the volume of oil across the lane at 15 feet and 30 feet respectively. At 15 feet, the two curves in question are of a similar height, whereas at 30 feet the black dash curve of the more difficult pattern is significantly lower than the red dot curve of the easier pattern. As discussed previously, this will result in shots that are missed towards the centre of the lane not hooking until the ball gets further down lane on the easier pattern due to the sustained low friction level, which in turn means the ball will not deviate as much in the $x$ direction, thus will finish at the desired part of the headpin giving a high chance of a strike. On the contrary, shots missed towards the centre of the lane on the difficult pattern will hook at around the same distance as it would has the ball travelled along the intended trajectory, therefore the boards covered in the $x$ direction will also be the similar, resulting in the ball not finishing at the desired part of the headpin and a strike not occurring.

The 44 ft oil pattern shown in Figure 5 is 3 ft longer than the difficult pattern of Figure 4, but the volume of oil towards the outer part of the lane are similar at 15 feet and 30 feet which can be seen from the green dash -dot and black dash curves in Figure 7 and Figure 8. Consequently, balls that are too far to the right on the long 44 ft oil pattern will simply not be able to cover enough boards in the $x$ direction to make it back the desired area of the headpin for a strike. Bowlers refer to this area as an out of bounds. The length of the pattern and sustained high volume of oil towards the outer part of the lane will force bowlers to make their shots closer to the headpin, and to not miss towards the outer part of the lane.

The middle part of the lane of the 37 ft oil pattern shown in Figure 6 between boards 6L and 6R is very flat both across the width of the lane and down the length of the lane in the $x$ direction, which means that shots thrown in this area will all react to the friction in a similar fashion due to its uniformity. Recall that this is not what a bowler wants as they would like shots that miss their target to all finish at the same place, which requires shots thrown around the target line to respond differently. The affect this has on strategy is for the bowlers to play towards the edge of the lane near the gutter where there is less oil towards the gutter and more oil towards the centre. Playing in this part of the lane poses its own difficulties for many bowlers as they do not want their ball to go into the gutter. However, for optimal strategy on this oil pattern, it is the best area of the lane to play in.

## 6. CONCLUSIONS

A mathematical model to describe the application of oil to a bowling lane has been reviewed and used to plot a variety of oil patterns. The solutions can clearly illustrate why some oil patterns are more difficult than others. The distribution of oil across the length of the lane plays a pivotal role in strategy along with the distribution of oil across the width of the lane which is more commonly considered. The solutions from the model can illustrate these features in a clear, novel way. They can be used by bowlers and coaches alike, to better understand the surface of the lane. It is also a useful tool for oil pattern designers, as they can get a good approximation of what an oil pattern will look like on a lane without dressing a lane and taking measurements, or by watching bowlers play on it and then making educated guesses as to what the oil pattern looks like.

There are some features of the current generation of Kegel Lane machines that have not been captured by the model presented here, such as the variable buffer brush rotation speed, but further experimental measurements are required in order for the model to capture such features accurately.

**Acknowledgements**

We wish to thank Kegel for their support of this work, as well as Raymond Teece and Rob Thurlby for many useful discussions throughout.

**References**

Cross R. (1998) The trajectory of a ball in lawn bowls. American Journal of Physics; 66(8), 735–738.

Frohlich C. (2004) What makes bowling balls hook. American Journal of Physics, 72(9), 1170–1177.

Hooper, C.G. (2023) Mathematical modelling of the application of lane conditioner to a Tenpin bowling lane, Journal of Sports Engineering and Technology, 0(0), https://doi.org/10.1177/17543371231217021

Hubbard M and Smith T. (1999) Dynamics of golf ball-hole interactions: rolling around the rim. Journal of Dynamic Systems, Measurement, and Control, 121(1), 88–95.

Huston, R.L., Passerello, C., Winget J.M., et al., (1979) On the dynamics of a weighted bowling ball. Journal of Applied Mechanics, 46(4), 937–943.

Kegel Sanction Technology (2024) https://www.kegel.net/sanction-technology-overview

Kegel (2024) Stone Street Oil Pattern, https://patternlibrary.kegel.net/pattern/1acfe9af-cf52-ec11-8c62-000d3a5afff3

Kegel (2024) Turnpike Oil Pattern, https://patternlibrary.kegel.net/pattern/7faecdb2-cf52-ec11-8c62-000d3a5afd36

Kegel (2024) 2020 World Bowling London Oil Pattern, https://patternlibrary.kegel.net/pattern/235b22e5-2754-ec11-8c62-000d3a5afd36

Kegel (2024) 2020 World Bowling Melbourne Oil Pattern, https://patternlibrary.kegel.net/pattern/d060c54d-2854-ec11-8c62-000d3a5afd36

King L., Perkins N.C., Churchill H., et al., (2011) Bowling ball dynamics revealed by miniature wireless MEMS inertial measurement unit, Sports Engineering, 13, 95–104.

USBC – United States Bowling Congress Press Room, https://bowl.com/press-room

# AN ANALYSIS OF A TEST CRICKET SERIES

Graham Pollard [a,c] , Tristan Barnett [b]
[a] *University of Canberra*
[b] *Macquarie University*
[c] *Corresponding author: grahamhpollard@gmail.com*

**Abstract**

Test cricket data was analysed to identify the presence of a 'home advantage' effect. A model was devised for a 4-tests series. It was used to evaluate the importance of a win versus a draw and the importance of a draw versus a loss at each possible scoreline. A method for determining the excitement at each possible scoreline was described and it was related to the importance of a win versus a draw and the importance of a draw versus a loss. A method for using these importances to evaluate the return a team achieves by increasing the probability of a win versus a draw, and/or the probability of a draw versus a loss was given. An alternative approach to measuring excitement was also given, producing similar results. A comparison of some characteristics of n-tests series (n = 2, 3, 4 and 5) was given. A 'biformat' of two 4-tests series, one held in each country, was proposed.

**Keywords**: Test cricket, test series, importance of a win versus a draw, importance of a draw versus a loss, excitement of a test match within a test series, efficiency of a test series.

## 1. INTRODUCTION

A test series consists of playing an agreed number of test matches (2, 3, 4 or 5) in the one country. In a test series between two more established teams (such as England, Australia, India, South Africa), this number of matches is typically 3 or 4, although in the case of England/Australia, it has been 5. What are the typical characteristics of a test series? For example, what is the likelihood of a draw in a test match? Is there a home advantage? Are some tests more important and more exciting? In which situations within a test series does a team get a greater reward when 'lifting its play'? Is there a 'best' number of matches within a series? These questions are considered in this paper.

The excitement in a series is often related to its closeness. Lynch (2021) noted that 'there have so far been 506 series comprising 3 or more tests. The closest of all in respect of average runs per wicket scored by each team was in the India Vs Australia series in 2000-01, which included India's famous victory in Calcutta after following on. India, who ended up winning the series 2-1, averaged 34.164 runs per wicket, while Australia averaged 34.160.' Drawn test series can be very exciting as well, and Tripathi (2021) describes 'the three greatest drawn Test series over the last 3 decades'.

There are two broad aspects to excitement within test cricket…the excitement within a particular test match as that match unfolds, and the excitement that a particular match has within the broader test series. This paper outlines a mathematical approach to the second of these two excitement situations. It is noted that the methodology of Vecer, Ichiba and Laudanovic (2007) might well provide a useful approach to measuring the excitement within any particular test match.

The objectives of this study were several.

1. To analyse recent test cricket data for the likelihood of a test match draw (versus a win or a loss), and to identify any home advantage effect.
2. To develop a mathematical analysis of a 'best of 4' tests series.
3. To use this analysis to evaluate the importance of a win versus a draw, and the importance of a draw versus a loss, for each possible 'score-line' within the series.
4. To evaluate the excitement of a test match at a given 'score-line' within the test series.
5. To develop a relationship between excitement and importances at a given 'score-line'.
6. To use the importances to evaluate the increased likelihood of winning a test series when the likelihood of winning a particular test or tests is increased.
7. To note a second measure of the excitement at a given 'score-line'.
8. To compare the main scoring system features of a 2-tests, 3-tests, 4-tests and 5-tests series.

## 2. METHODS

Recent test cricket data was analysed to identify relevant values for the probability of a win, a draw and a loss in a test match. The data was also analysed for any 'home-advantage' effect. A branching diagram and associated recurrence methods were used to analyse a '4-tests' series. Some characteristics such as the importance of a win versus a draw, and the importance of a draw versus a loss were evaluated at each possible state within a 4-tests series. Making use of this branching diagram, a method for evaluating the excitement of the various scoring situations that can arise within a 4-tests series was demonstrated. The relationship between excitement and importances at a score-line was developed. A method for determining the 'return' a team obtains by 'lifting play' at a score-line was derived. An alternative approach to the analysis of excitement was given. A comparison of some characteristics of a 2-tests, 3-tests, 4-tests, and 5-tests series was carried out.

## 3. RESULTS

AN ANALYSIS OF RECENT TEST SERIES DATA

Yearly test cricket data used in this paper was obtained from the website of the International Cricket Council. A link to the data for the calendar year 2016 is given. Table 1 gives a summary of drawn test matches in the calendar years from 2003 to 2022.

| Years | # Drawn tests | # tests | Percentage of tests drawn |
|---|---|---|---|
| 2003-2006 | 47 | 190 | 24.7 |
| 2007-2010 | 46 | 162 | 28.4 |
| 2011-2014 | 42 | 167 | 25.1 |
| 2015-2018 | 28 | 185 | 15.1 |
| 2019-2022 | 21 | 148 | 14.2 |
| Total | 184 | 852 | 21.6 |

Table 1. The number of drawn test matches, 2003-2022.

The percentage of tests that resulted in a draw decreased in about 2015 to about 15%. There could be many reasons for this decrease. A new style of more aggressive test cricket could well be an important factor. Thus, it was decided to focus on the test cricket series in the period 2015-2022. The home advantage in test cricket can be seen in Table 2.

| Year | # Home wins | # Draws | #Away wins | Total |
|---|---|---|---|---|
| 2015 | 19 | 10 | 15 | 44 |
| 2016 | 27 | 5 | 14 | 46 |
| 2017 | 24 | 7 | 14 | 45 |
| 2018 | 29 | 6 | 15 | 50 |
| 2019 | 24 | 3 | 13 | 40 |
| 2020 | 14 | 4 | 4 | 22 |
| 2021 | 18 | 7 | 19 | 44 |
| 2022 | 21 | 8 | 12 | 41 |
| 2015-2022 | 176 (53.0%) | 50 (15.1%) | 106 (31.9%) | 332 |

Table 2. Home wins, draws and 'away' wins in test cricket.

To study this home-advantage effect, it was decided to pair countries, working backwards from 2022. For example, Australia played England in 5-tests series in Australia in 2021 and in 2017, and Australia played England in 5-tests series in England in 2019 and 2015. The results are in Table 3.

| Location/Number of Wins and Draws | Australia | Draw | England | Total |
|---|---|---|---|---|
| In Australia 2021 and 2017 | 8 | 2 | 0 | 10 |
| In England 2019 and 2015 | 4 | 1 | 5 | 10 |
| Total | 12 | 3 | 5 | 20 |

Table 3. Australia versus England, 2015-2021.

The frequency of a draw in Table 3 is $3/20 = 0.15$. If we denote a (common) home advantage factor by H and the marginally more successful Australian factor by A, then $(0.425+A+H)$ can be estimated by $8/10$, and

(0.425+A-H) by 4/10, giving estimates H = 0.2 and A = 0.175. Estimates such as H and A have large standard errors. As Australia had a better performance in this table, A turned out positive.
Secondly, India played England in 8 tests in 2021. See Table 4. India was more successful overall, and both teams performed better ‘at home’ than ‘away’. As in the above paragraph, if we denote a (common) home advantage factor by H and the marginally more successful Indian factor by I, then (0.4375+I+H) can be estimated by 3/4, and (0.4375+I-H) by 2/4, giving H = 0.125 and I = 0.1875.

| Location/Number of Wins and Draws | India | Draw | England | Total |
|---|---|---|---|---|
| In England, August, 2021 | 2 | 1 | 1 | 4 |
| In India, February, 2021 | 3 | 0 | 1 | 4 |
| Total | 5 | 1 | 2 | 8 |

Table 4. India versus England, 2021.

This process of selecting pairs of countries that had played two test series (‘home’ and ‘away’) was continued, working ‘backwards’ from 2022. Test series of 4-tests was considered firstly, and then the case of test series of 3-tests was studied. See Table 5 which includes the above two cases.

In 6 of the paired test series in Table 5 the home advantage was positive, 3 times it was zero and 3 times it was negative. By virtue of the ordering of the first team mentioned in column 1, the country effect could not be negative. On 10 occasions it was positive and twice it was zero.

| Pair of countries (Location Years) | # Tests | Country | Home |
|---|---|---|---|
| 1.Australia Vs England (Aust 21, 17; Eng 19, 15) | 20 | A=0.175 | H = 0.2 |
| 2.India Vs England (England, August 21; India February 21) | 8 | I=0.188 | H = 0.125 |
| 3. India Vs Australia (Australia 2020, India 2017) | 8 | I=0.125 | H = 0 |
| 4.England Vs South Africa (S Africa 2019, England 2017) | 8 | E=0.25 | H = 0 |
| 5.England Vs West Indies (W Indies 2022, England 2020 | 6 | E=0 | H = 0.333 |
| 6.Australia Vs Pakistan (Pakistan 2022, Australia, 2016) | 6 | A=0.333 | H = 0.333 |
| 7.England Vs Pakistan (Pakistan 2022, England 2020) | 6 | E=0.333 | H = -0.333 |
| 8.India Vs South Africa (S Africa 2021, India 2019) | 6 | I=0.166 | H = 0.333 |
| 9.England Vs West Indies (W Indies 2019, England 2017) | 6 | E=0 | H = 0.167 |
| 10.England Vs Sri Lanka (Sri Lanka 2018, England 2016) | 6 | E=0.417 | H = -0.167 |
| 11.Pakistan Vs West Indies (W Indies 2017, Pakistan 2016) | 6 | P=0.166 | H = 0 |
| 12.India Vs Sri Lanka (Sri Lanka July 2017, India Nov 2017 | 6 | I=0.333 | H = -0.333 |
| Weighted averages | 92 | 0.201 | 0.076 |

Table 5. A ‘paired’ analysis of test series of 5-, 4- and 3- tests, 2015-2022.

There were four cases where one of the above pairings played an (additional) ‘unpaired’ 4-tests series in the period 2015-2022. This provided an opportunity to see how well the estimates in Table 5 fitted these ‘new’ observations. The four cases were:

1. Australia Vs India in Australia in December 2018.
2. England Vs Pakistan in England in July 2016.
3. South Africa Vs England in South Africa in December 2015.
4. India Vs South Africa in India in November 2015.

Consider the first case of Australia Vs India in Australia, where there was 1 draw, India won 2 tests and Australia won 1 test. Noting in Table 5 that I = 0.125 and H = 0, and conditioning on 1 draw in this series in 2018, then using these statistics and the above methodology, India would expect to have won (4-1)/2 + 3*0.125 = 1.875 tests and Australia 1.125 tests. These expected values are quite similar to the observed, despite the large standard errors of the estimates.

The analysis in the above paragraph was repeated for the other 3 cases above. Predicting these ‘unpaired’ test series results using the conclusions of the corresponding ‘paired’ series produced somewhat surprisingly accurate results, despite the large standard errors of the estimates.

## A MATHEMATICAL ANALYSIS OF A ‘BEST OF 4-TESTS’ CRICKET SERIES

Here we consider an analysis of a ‘best of 4-tests’ cricket series between two strong cricket teams such as India and Australia. In a matching of strong teams the probability of a drawn match is likely to be a bit higher than on average. (Note that in the 12 test matches played between India and Australia from February 2017 to January 2021, there were 3 draws (25%)) In the following mathematical analysis we assume the probability of a drawn match is 0.2, the probability that Team A (the better team) wins a test match is 0.5, and the probability Team B wins a test match is 0.3. Thus, Team A's probabilities of (win, draw, loss) are assumed to be (pw = 0.5, pd = 0.2, pl = 0.3). Also, we suppose a coin is tossed to determine the series-winner if the series is drawn after the 4 tests are played. This is simply a useful devise for the following analysis.

An analysis of this 4-tests series with the above parameters is given in Table 6. The first test is called state 1, the second test is either state 2, 3 or 4 depending on whether the first test was won, drawn or lost by Team A. Continuing in this systematic manner, state 5 corresponds to where Team A has won the first two tests, state 7 corresponds to where the series is evenly balanced after 2 tests, and state 9 corresponds to where Team A has lost the first 2 tests. State 10 corresponds to where Team A is ahead by 1 test match after 3 tests have been played, state 11 corresponds to where the series is evenly balanced after 3 tests, and state 12 corresponds to where Team A is behind by 1 test match after 3 tests have been played. There are thus 12 'live states' from where Team A can win or draw or from where Team B can win or draw. The 'coin-toss state' if the series is drawn at 2-2 is called 'state 13'. Note that the $4^{th}$ test is typically played even if the series is not 'alive' at that stage.

| Test No. | A's lead | State | P(A wins series) | Imp(W vs D) | Imp(D vs L) | Excitement |
|---|---|---|---|---|---|---|
| 1 | 0 | 1 | 0.6728 | 0.1795 | 0.2365 | 0.16055 |
| 2 | 1 | 2 | 0.8335 | 0.105 | 0.23 | 0.1281 |
| 2 | 0 | 3 | 0.654 | 0.23 | 0.27 | 0.196 |
| 2 | -1 | 4 | 0.4175 | 0.27 | 0.225 | 0.2025 |
| 3 | 2 | 5 | 0.955 | 0 | 0.15 | 0.063 |
| 3 | 1 | 6 | 0.85 | 0.15 | 0.25 | 0.15 |
| 3 | 0 | 7 | 0.62 | 0.25 | 0.35 | 0.23 |
| 3 | -1 | 8 | 0.35 | 0.35 | 0.25 | 0.25 |
| 3 | -2 | 9 | 0.125 | 0.25 | 0 | 0.125 |
| 4 | 1 | 10 | 0.85 | 0.0 | 0.5 | 0.21 |
| 4 | 0 | 11 | 0.6 | 0.5 | 0.5 | 0.4 |
| 4 | -1 | 12 | 0.25 | 0.5 | 0 | 0.25 |

Table 6. The analysis of a '4-tests' series when (W/D/L) probabilities for Team A are (0.5, 0.2, 0.3).

## IMPORTANCE OF A TEST MATCH WITHIN A TEST SERIES

The probability that Team A wins the series from each state can be calculated using backwards recurrence methods. These probabilities are given in column 4 of Table 6. In a similar way to that of Morris (1977)*, the importance of a win (versus a draw)* in a particular state is defined as the probability of A winning the series given A wins in that state minus the probability of A winning the series given a draw in that state. These values are given in column 5 and can be seen to differ markedly from state to state. Correspondingly, *the importance of a draw (versus a loss)* in a particular state is defined as the probability of A winning the series given a draw in that state minus the probability of A winning given a loss in that state. These values are given in column 6. It is noted that whenever Team A (the better team) is equal or ahead in the series, the importance of a draw (versus a loss) is equal to or greater than the importance of a win (versus a draw).

For these parameters, the probability Team A wins the 4-tests series is 0.6728 (see the first row of Table 6). Thus, the probability Team A loses the 4-test series is 0.3272. As it can be shown using forwards recurrence methods that state 13 occurs with probability 0.209, the W/D/L probability breakdown (for the series) for Team A is 0.6728-(0.209/2) / 0.209 / 0.3272-(0.209/2), that is, 0.5683 / 0.209 / 0.2227.

Interestingly, and as an aside, the probability of a drawn 4-tests series is very similar to that of a single test match for these parameters.

## EXCITEMENT OF A TEST MATCH WITHIN A TEST SERIES

As in the definition of excitement given by Pollard (2017), *the excitement of a particular state within the test series is equal to the expected value of the absolute size of the change in a team's probability of an overall win as a result of that test being played.* For example, noting that state 1 is followed by state 2, 3 or 4 if the first test is won, drawn or lost respectively by Team A, the excitement in state 1 is equal to 0.5*(0.8335-0.6728) + 0.2*(0.6728-0.654) + 0.3*(0.6728-0.418) = 0.16055. The excitements of the other states can be calculated correspondingly and are given in column 7 of Table 6. Not surprisingly, the most exciting state is state 11 when the series is tied with only one test left to play, and the least exciting state is state 5 when Team A is ahead by 2 test wins after just 2 tests have been played. Further, note that the excitement in state 8 is 0.25, whilst it is only 0.15 in state 6. This is not surprising as the better team is '1 behind' in state 8, whilst it is '1 ahead' in state 6.

MATHEMATICAL EXPRESSIONS FOR EXCITEMENT

Firstly, we consider those states where the effect of a *drawn* match is to *not decrease* the probability of Team A winning the series. (In Table 6 these states are states 2, 5, 6, 10 and 11.) For one such state. the excitement of that state as defined in the above section is given by the equation

$$Excitement = (P(Won) - P)) * pw + (P(Drawn) - P)) * pd + (P - P(Lost)) * pl,$$

where P is the probability of winning the series given the present state, P(Won) is the probability of winning the series having won the present test, P(Drawn) is the probability of winning the series having drawn the present test, and P(Lost) is the probability of winning the series having lost the present test. Adding and subtracting pl*P(Lost) from the right-hand side of this equation, and noting that P = pw*P(Won)+pd*P(Drawn)+pl*P(Lost), it follows that

$$Excitement = 2pl * \left(P - Prob(Lost)\right)$$

e.g. for state 2, P = 0.834, Prob(Lost) =0.62, and pl = 0.3, and this equation can be verified. Note that the second to last equation can be rewritten as

$$Excitement = (Iw + d) * pw + (d) * pd + (Id - d) * pl,$$

where Iw is the importance of a win versus a draw, Id is the importance of a draw versus a loss and d = (P(Drawn)-P) is not negative. *This equation gives the relationship of excitement to importances for the cases when d = (P(Drawn)-P) is not negative*.

Secondly, for the states where the effect of a drawn test match is to not increase the probability of Team A winning the series, the corresponding three equations are

$$Excitement = (P(Won) - P)) * pw + \left(P - P(Drawn)\right)) * pd + (P - P(Lost)) * pl,$$

$$Excitement = 2pw * (Prob(Won) - P), \text{ and}$$

$$Excitement = (Iw - d) * pw + (d) * pd + (Id + d) * pl,$$

where d= P-P(Drawn) is not negative.

For example, in state 8 we have P(Won) = 0.6, P = 0.35, P(Drawn) = 0.25, P(Lost) = 0, Iw = 0.35, d = 0.1, Id = 0.25 and above equations give an excitement of 0.25. *This last equation gives the relationship of excitement to importances for the cases when d = (P - P(Drawn)) is not negative*.

THE EFFECT OF INCREASED PROBABILTIES IN PARTICULAR TEST MATCHES

In this section we show how importances can be used to find the increased probability of winning the series given an increased probability of winning rather than drawing (or drawing rather than losing) one or more test matches. For example, suppose Team A plays better at the location of the second test and its probabilities for W/D/L at *each* of states 2, 3 and 4 are 0.6/0.2/0.2. That is, its probability of a loss is decreased by 0.1 and its probability of a win is increased by 0.1 in this second test. We call this type of lift which depends only on the number of the test (test number 1, 2, 3 or 4), and not on the relative scores of the teams, a 'type 1' lift. It can be seen from Table 6 that the importance of a win versus a loss (obtained by adding the importance of a win versus a draw and the importance of a draw versus a loss) is 0.335, 0.5 and 0.495 in states 2, 3 and 4 respectively.

Further, we assume that, if it so wishes, Team A has the capacity to convert some potential losses into draws by playing 'very conservatively'. We assume Team A will do this if the series is tied or they are one test down at the beginning of the third test (ie. states 7 and 8), or if they are one test up at the beginning of the fourth test (ie. state 10). This is equivalent to changing its W/D/L probabilities from 0.5/0.2/0.3 to 0.5/0.3/0.2 in states 7, 8 and 10. This type of lift, which depends on both the number of the test and the 'score sheet', is called a 'type 2' lift. Note in Table 6 that the importance of a draw versus a loss is 0.35, 0.25 and 0.5 in states 7, 8 and 10 respectively.

The increased probability of winning the test series for Team A by 'lifting its game' as described in the above two paragraphs can be evaluated using the following equation (Pollard and Pollard, 2010)

$$Increase = \sum Ni * Ii * di$$

where the summation is over the states where the lifting occurs, $Ii$ is the (relevant) importance of state $i$ prior to the lifting, $di$ is the size of the probability lift in state $i$ and $Ni$ is the expected number of times state $i$ is visited *after the lifting has occurred*. This equation is demonstrated in Table 7. The importances of a win versus a loss for states 2, 3 and 4 in Table 7 are derived very simply from Table 6, and the importances of a draw versus a loss for states 7, 8 and 10 come from Table 6. By 'lifting' in these (two different) ways, Team A increases its probability of winning the series by 0.07 from 0.673 to 0.743 (see Table 7). Note that this value of 0.743 can be verified by using the 'new' W/D/L probability values for each state with (backwards) recurrence methods.

| State | Relevant Importance, Ii | Change in probability, di | Ni | Ni*Ii*di |
|---|---|---|---|---|
| 2 | (0.105+0.23) = 0.335 | 0.1 | 0.5 | 0.01675 |
| 3 | (0.23+0.27) = 0.5 | 0.1 | 0.2 | 0.01 |
| 4 | (0.27+0.225) = 0.495 | 0.1 | 0.3 | 0.01485 |
| 7 | 0.35 | 0.1 | 0.32 | 0.0112 |
| 8 | 0.25 | 0.1 | 0.1 | 0.0025 |
| 10 | 0.5 | 0.1 | 0.294 | 0.0147 |
| | | | | Sum=0.07 |

Table 7. Calculating the increase in the probability of winning the series when 'lifting' play.

When 'lifting' in (say) two states, the total effect on the probability of winning the series 'interacts positively' when the total probability lift for the two states is greater than the sum of the two individual probability lifts. Consider, for example, states 3 and 6. Lifting the probability of a win and decreasing the probability of a loss by 0.1 in state 3 (parameters become 0.6/0.2/0.2 for state 3) gives a series probability increase of winning of 0.01. Also, lifting the probability of a draw and decreasing the probability of a loss in state 6 by 0.1 (parameters become 0.5/0.3/0.2 for state 6) gives a series probability increase of winning of 0.005, whilst lifting in the above manner in both states 3 and 6 gives a series probability increase of 0.016 (greater than 0.01 + 0.005). This may not be surprising as state 6 is more likely to occur given there is lifting in state 3.

When 'lifting' in two states, the total effect on the probability of winning the series can also 'interact negatively'. Using corresponding lifts to those in the above paragraph, it can be shown that state 2 alone (with parameters 0.6/0.2/0.2) has a probability lift of 0.017 and state 7 alone (with parameters 0.5/0.3/0.2) has a probability lift of 0.012, but the states interact negatively giving a total lifting of only 0.027 (less than 0.017 + 0.012). Again, this may not be surprising as state 7 is less likely to occur given the lifting in state 2.

It can be shown that there are '*no paired interactions and no triple interaction*' between the three states 2, 3 and 4. Using the 'basic parameters' 0.5/0.2/0.3 (and 0.6/0.2/0.2 for a state where there is a lift in play), the increase in the probability of winning the series by lifting in state 2 only is 0.017, by lifting in state 3 only is 0.01, and by lifting in state 4 only is 0.015. Thus, given no paired or triple interactions, the return by lifting in *whichever of these three states actually occurs* is given by the sum of these values (0.042). These zero interactions are not surprising as the test series enters only one of these 3 states, and thus might be called '*structural zero interactions*'. Other 'zero-paired interactions' can of course occur. For example, for the parameter values used, the lift (to parameter values 0.6/0.2/0.2) in state 2 can be shown to have a zero interaction with a 'type 2' lift to values 0.5/0.3/0.2 in state 6, although these two states can occur in the one realization of a test series.

AN ALTERNATIVE APPROACH TO THE MEASUREMENT OF EXCITEMENT

In the analysis of excitement given in section above the final outcome was assumed to be a win to Team A or a win to Team B, as it was assumed that there would be a coin-toss to determine the winner if the series was tied after 4 tests. In that analysis the size of Team A's increase in probability of winning the series given a win in a particular 'state' of the series is equal to the decrease in Team B's probability of winning the series. Thus, if the measure of excitement of a test match is defined as the modulus of the change in Team A's probability of winning the series *plus* the modulus of the change in Team B's probability of winning as a result of that test being played, *this measure of excitement would be simply twice the measure given in the above section.*

The approach to excitement taken in this section was by Vecer, Ichiba and Laudanovic (ibid). The final outcome of the series is now assumed to be a win to Team A, a draw, or a win to Team B. It is noted that a draw in a test match can also be exciting just as can be a win to Team A or a win to Team B. At each stage and state within the test series the excitement of a draw is defined as the expected value of the modulus of the change in the probability of a draw as a result of that test being played. This corresponds to the definition of the excitement of a test win by Team A being equal to the expected value of the modulus of the change in probability of Team A winning the series as a result of that test being played, and the excitement of a win by Team B being equal to the expected value of the modulus of the change in Team B's probability of wining the series as a result of that test being played. *The total excitement of a test match is defined as the sum of the two largest of these three excitements.*

Table 8 lists 15 possible states (St) in column 1. States 1 to 12 were described earlier in Table 6. State 13 now represents the 'dead' state where Team A has won prior to the 4th test being played, and state 14 represents the 'dead' state where Team B has correspondingly won. State 15 is the state where the test series is drawn after all 4 tests have been played. Column 2 gives the frequency of occurrence for each of these states. Columns 3, 4 and 5 give the probability that the series is won by Team A, is drawn, or is won by Team B starting in the state for that row. Note that the values in columns 3, 4 and 5 of row 1 agree with earlier values These values in columns 3, 4 and 5 for states 1 to 12 have been calculated using backwards recursion methods, starting with Test 4 in states 10, 11 and 12. Making use of the values in columns 3, 4 and 5, columns 6, 7 and 8 give the excitement (Ex) for a series win by Team A, a draw, and a win by Team B starting in that state. Thus, for example, in state 3, noting that the next state is state 6 with probability 0.5, state 7 with probability 0.2, and state 8 with probability 0.3, the excitement for a series win by Team A is equal to 0.5*mod(0.56-0.79)+0.2*mod(0.56-0.45)+0.3*mod(0.56-0.25) = 0.23. Correspondingly, the excitement for a series draw in state 3 is equal to 0.5*mod(0.188-0.12) +0.2*mod(0.188-0.34) +0.3*mod(0.188-0.2) =0.068. Also, the excitement for a series win by Team B is equal to 0.5*mod(0.252-0.09)+0.2*mod(0.252-0.21)+0.3*mod(0.252-0.55) = 0.179.

| St | Freq | P(A) | P(D) | P(B) | Ex A | Ex D | Ex B | Ex A,B | T Ex2 | T Ex1 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0.5685 | 0.2086 | 0.2229 | 0.1795 | .04584 | 0.1419 | 0.3214 | 0.3214 | 0.3214 |
| 2 | 0.5 | 0.748 | 0.171 | 0.081 | 0.179 | 0.101 | 0.081 | 0.260 | 0.280 | 0.2562 |
| 3 | 0.2 | 0.56 | 0.188 | 0.252 | 0.23 | 0.068 | 0.179 | 0.409 | 0.409 | 0.392 |
| 4 | 0.3 | 0.275 | 0.285 | 0.44 | 0.175 | 0.055 | 0.23 | 0.405 | 0.405 | 0.405 |
| 5 | 0.25 | 0.91 | 0.09 | 0 | 0.126 | 0.126 | 0 | 0.126 | 0.252 | 0.126 |
| 6 | 0.2 | 0.79 | 0.12 | 0.09 | 0.21 | 0.12 | 0.126 | 0.336 | 0.336 | 0.3 |
| 7 | 0.34 | 0.45 | 0.34 | 0.21 | 0.27 | 0.096 | 0.21 | 0.48 | 0.48 | 0.46 |
| 8 | 0.12 | 0.25 | 0.2 | 0.55 | 0.25 | 0.12 | 0.27 | 0.52 | 0.52 | 0.5 |
| 9 | 0.09 | 0 | 0.25 | 0.75 | 0 | 0.25 | 0.25 | 0.25 | 0.5 | 0.25 |
| 10 | 0.285 | 0.7 | 0.3 | 0 | 0.42 | 0.42 | 0 | 0.42 | 0.84 | 0.42 |
| 11 | 0.188 | 0.5 | 0.2 | 0.3 | 0.5 | 0.32 | 0.42 | 0.92 | 0.92 | 0.8 |
| 12 | 0.171 | 0 | 0.5 | 0.5 | 0 | 0.5 | 0.5 | 0.5 | 1.0 | 0.5 |
| 13 | 0.275 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 14 | 0.081 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15 | 0.2086 | n.a. | n.a. | n.a. | n.a. | n.a. | n.a. | n.a. | n.a. | n.a. |

Table 8. The frequency of entering states 1 to 15 in one realization of a test series, the probability of team A (finally) winning the series, (finally) drawing it, and (finally) losing it from states 1 to 12 respectively, the expected change in the absolute value of the probability of Team A winning the series, of drawing the series, and of losing the series as a result of the test match being played, for each possible state or test-series situation.

Column 9 gives the excitement of a win by Team A or a win by Team B (Ex A,B) obtained by adding the corresponding values in columns 6 and 8. Column 10 gives *the total excitement, T Ex2, defined as the sum of the two largest elements of Ex A, Ex D and Ex B in that row*. Note that to add all three of these excitements would involve an element of 'double counting', as P(A), P(D) and P(B) must sum to 1 for each row. Thus, for example, the total excitement, T Ex2, of being in state 1 is equal to the sum 0.1795+0.1419 = 0.3214. Column 11 (last column, T Ex1) gives twice the excitement values in Table 6 for the reason described above. Thus, the total excitement (T Ex1) of being in state 1 is equal to 0.3214. As another example, the total excitement (TEx2) in the sequence of test results Win by B/Draw/Win by A/Win by A (with the sequence of states being 1/4/8/11) is equal to 0.3214 + 0.405 + 0.52 + 0.92 = 2.1664 using this second approach to excitement which can include the excitement of a draw. Note that the corresponding total excitement using the first definition of excitement is equal to 0.3214 + 0.405+ 0.5 + 0.8 = 2.0264 (refer to Table 8).

Several observations can be made in Table 8.

1. The values in column 9 (Ex A,B) and column 11 (T Ex1) are quite similar. Six times they are equal, and on the other 6 occasions the value in column 11 is a little smaller than that in column 9. Thus, the

size of the total excitement at each stage of the series, as given by the earlier analysis, is quite similar to its value (excluding draws) using the second approach.
2. The excitement of a draw, Ex D, is typically somewhat small relative to the sum of the other excitements, Ex A,B = Ex A + Ex B, early in the test series, and particularly when the series is also evenly matched. However, Ex D does increase substantially in the 4th test.

It can be seen in Table 8 that the first 3 tests must be 'live tests', in the sense that either team A or team B has the possibility of winning or drawing the series. However, the probability that the 4th test is 'live' is only 0.644 (= 0.285+0.188+0.171 in Table 8), and 'dead' tests have zero measures of excitement (by definition). The weighted averages (for each of the four tests) of the excitement measures T Ex2, T Ex1 and Ex D in Table 8 are given in Table 9. The substantial increase in the draw excitement measure, Average Ex D, as the test series progresses can be seen in Table 9.

| Test Number | P(Live test) | Average T Ex2 | Average T Ex1 | Average Ex D |
|---|---|---|---|---|
| 1st | 1 | 0.3214 | 0.3214 | 0.04584 |
| 2nd | 1 | 0.3433 | 0.328 | 0.0806 |
| 3rd | 1 | 0.4008 | 0.3304 | 0.12504 |
| 4th | 0.644 | 0.90584 | 0.55217 | 0.41205 |

Table 9 Average excitements per 'live' test.

A BRIEF COMPARISON OF 1-TEST, 2-TESTS, 3-TESTS, 4-TESTS AND 5-TESTS SERIES

We have considered a test series consisting of four test matches. Using the same parameters, we note characteristics of each of the 1-test, 2-tests, …, 5-tests series. The probability that the series is drawn is much bigger for the 2-tests case than for the other cases (see Table 10). This is a major weakness of the 2-tests system. The ratio P/Q, where P is the probability that the better team wins the series and Q is the probability that the better team loses the series, increases as the number of tests played increases. This is not surprising. The mean number of 'dead tests' is 0.34 for the 3-tests case, 0.356 for the 4-tests case and 0.630 for the 5-tests case. Thus, the proportion of 'dead' tests is 0.113, 0.089, and 0.126 respectively. This smallest value for the proportion of 'dead' tests might be seen as a minor strength of the 4-tests case, relative to the 3-tests and 5-tests cases.

| Number tests, N | 1-test | 2-tests | 3-tests | 4-tests | 5-tests |
|---|---|---|---|---|---|
| P(win), P | 0.5 | 0.45 | 0.56 | 0.569 | 0.614 |
| P(loss), Q | 0.3 | 0.21 | 0.252 | 0.223 | 0.227 |
| P(Draw) | 0.2 | 0.34 | 0.188 | 0.209 | 0.159 |
| P/Q | 1.667 | 2.143 | 2.222 | 2.550 | 2.709 |
| Mean ('live' tests) | 1 | 2 | 2.66 | 3.644 | 4.370 |
| Mean number of 'Dead' tests | 0 | 0 | 0.34 | 0.356 | 0.630 |
| RE1 = ((P-Q)/Mean)Ln(P/Q) | 0.102 | 0.091 | 0.092 | 0.089 | 0.088 |
| RE2 = ((P-Q)/N)Ln(P/Q) | 0.102 | 0.091 | 0.082 | 0.081 | 0.077 |

Table 10. Characteristics of the n-tests series (n = 1, 2, 3, 4 and 5.).

Miles (1984) studied the efficiency of sports scoring systems, paying particular attention to tennis (where a match cannot result in a draw). Given two scoring systems with the same likelihood of correctly identifying the better competitor, the scoring system with the smaller expected duration is the more efficient one. He called the approach 'symmetric sequential analysis' and made use of the optimal efficiency results of Wald and Wolfowitz (1948) in statistical hypothesis testing. He showed that when comparing the efficiency of two tennis scoring systems, one only needs to compare the values of ((P-Q)/Mean)Ln(P/Q), where P is the probability that the better player wins, Q = 1-P and Mean is the mean number of points played. He called this expression the 'relative efficiency' and it is denoted by RE1 in Table 10. Pollard and Pollard (2012) extended this efficiency concept to the case of some other scoring systems, including ones where the scoring system could result in a win, draw or loss to each team. This result is used in Table 10. The 'Mean' in Table 10 is the average number of 'live' tests, as in many sporting competitions 'play ceases' as soon as one player or team must win the contest. This is not the case in a test series as the series continues even though the winner of the series might have already been determined. Table 10 includes an *approximation* to the above relative efficiency expression. It is called RE2 and involves replacing the mean number of 'live tests' by the total number of tests played. This would appear to be reasonable for test cricket series. The ratio of any two of these relative efficiencies gives the efficiency of one test series relative to another. Thus, the efficiency of 4-tests versus 3-tests is given by the ratio (0.081/0.082)*100 (98.7%).

The probability of a drawn series in the 2-tests case might be considered too high (0.34) for practical purposes. Also, the higher value of P/Q for the 4-tests case versus the 3-tests case might make the 4-tests system more appealing. As noted above, in terms of the number of 'dead' tests, the 4-tests case might be considered somewhat preferable to the 5-tests case.

## 5. CONCLUSIONS

The percentage of test matches resulting in a draw has decreased in recent years to about 15%. The home team has won about 53% of tests and the other team has won about 32% of tests.

A method for determining the importance of a win versus a draw and the importance of a draw versus a loss for each possible 'score-to-date' situation has been described. This method was used to evaluate the return a team achieves by increasing the probability of a win versus a draw, and/or the probability of a draw versus a loss at each possible score. A relationship between the excitement and the importances at a scoreline was determined. Two methods for determining the excitement were given. The first is based on the modulus of the change in a team's likelihood of winning the series as a result of that particular test being played. The second method includes an additional element that measures the excitement of a 'drawn test match' at each possible score. The two measures of excitement were comparable. Other measures of the 4-tests series such as the relative likelihood of each team winning, the expected number of 'dead' tests, and the efficiency of the series were compared with the same measures for the 3-tests and 5-tests series. These measures suggest that the 4-tests series might be considered somewhat preferable to the 3-tests and 5-tests cases.

Home-advantage is certainly a factor in test cricket. Often it would appear to be a bigger factor than the advantage of winning the toss at the commencement of a test match (and often choosing to bat first). Test series appear to be planned and played in pairs. The first 4-tests series in Country A and the second 4-tests series in country B a year or so later. This second 4-tests series in country B forms the *second half* of a biformat of 8-tests matches in the two countries. It also forms the *first part* of another biformat of 8-tests matches in which the next 4-tests are played back in country A. Every test match held within such a 4-tests series must have a positive excitement measure with respect to the 8-tests biformat series of which the current 4 tests are the first four tests. Thus, there would never be a test match with zero excitement (in such a 'rolling biformat' structure). Note that such a biformat is completed after every 4-tests series.

In such a biformat structure, there would be the excitement within each test match, the excitement within each 4-tests series, and the excitement of the test matches within the 'dual' series. Further, at the end of each 'dual' test series (eg 8 test matches), there would *more often* (than at present) be an overall winner (with series draws being rarer than at present). Note that the winner of the biformat of 8-tests may not be the winner of the last 4-tests series played. The authors are not aware of a comparable competition format, but this structure would appear to have some merit in test cricket where there is typically a home advantage. Further, completing a 'cycle of competition' at the end of each and every 4-tests series would appear to be an attractive and novel characteristic. It could also increase incentives for the teams where, at times, without it, some incentives might dissipate.

It is clear that winning the toss in a test match is better than losing it. The advantage of playing a test match at home might be reduced by giving the visiting team the choice of batting or bowling first in each of the test matches.

The methods of this paper could be applied directly to other sports that have a contest consisting of a series of 'matches' each of which can result in a win, draw or loss.

**References**

Lynch, S. 2021. *What is the closest Test series ever played, by average runs per wicked scored by each team?* Available at: https://What is the closest Test series ever played, by average runs per wicket scored by each team? - ESPN (Accessed: 21 May 2023).

Miles, R. E. 1984. 'Symmetric Sequential Analysis: The efficiencies of Sports Scoring Systems (with Particular Reference to those of Tennis)', *Journal of the Royal Statistical Society. B*, *46(1)*, 93-108.

Morris, C. 1977. 'The most important points in tennis', In *Optimal Strategies in Sport*, eds S.P. Ladany and R.E. Machol, 131-140. Amsterdam: North-Holland. (Vol 5 in Studies in Management Sciences and Systems).

Pollard, G. H. 2017 'Measuring Excitement in Sport', *Journal of Sports Analytics*, *3(1)*, 37-43.

Pollard, G. H. and Pollard, G. N. 2010. 'The efficiency of scoring systems with outcome dependent points", Proceedings of the Tenth Australasian Conference on Mathematics and Computers in Sport, Darwin, Australia, edited by A. Bedford and M. Ovens, MathSport (ANZIAM). 35-42.

Pollard, G.H. and Pollard, G.N. 2012. A general measure for the relative efficiency of any two scoring systems. *Proceedings of the Eleventh Australasian Conference on Mathematics and Computers in Sport,* Melbourne, Australia, edited by A. Bedford and A Schembri, MathSport (ANZIAM), 162-169.
Roy, S. 2019. *5 greatest test series in the last 40 years.* Available at: https://5 Greatest Test series in the last 40 years (sportskeeda.com) (Accessed: 8 June 2023).
The International Cricket Council, available at: https:// www.icc-cricket.com/.(Accessed: 8 June 2023).
Test Match Results in 2016 and Statistics (crictotal.com) (Accessed: 23 May 2023)
Tripathi, D. 2021. *3 greatest drawn Test series over the last 3 decades.* Available at: https://www.sportskeeda.com/cricket/test-7567f. (Accessed: 21 May 2023).
Vecer, J. Ichiba, T. and Laudanovic, M. 2007. On probabilistic excitement in sports games, *Journal of Quantitative Analysis in Sports,* vol 3, no. 3, article 6.
Wald, A. and Wolfowitz, J. 1948. Optimum character of the sequential probability ratio test, *Ann. Math. Statist.,* vol. 19, 326-339.

This page is intentionally blank.

# MATH _ SPORT: The Preposition Proposition

Steven E Stern [a,b]

[a] *Bond University, Centre for Data Analytics, Gold Coast, Australia*
[b] *Corresponding author: sstern@bond.edu.au*

**Submitted for consideration as (select one):**

- ☒ Oral presentation
- ☐ Poster

**Abstract**

In this talk, I will discuss the journey of mathematical contributions to, for and in various sports. From its beginnings, sports participants, fans, and administrators have been focused on numbers. Numbers to quantify their game, to determine its outcomes and enshrine its best participants. I will make some observations about the early forays of mathematical thinking into sports and end by discussing my own beliefs of how mathematics is now entering the realm of not only measuring sport, but also becoming part of its rules and structure. To this end, I will discuss the importance of holistic and principled thinking, so as to avoid notable unintended consequences. In particular, I will focus on cricket as a key example of an early adopter of many mathematical components to its rules and regulations (as evidenced by DLS, DRS and even the Kendix Ranking System). Further, I will begin a wider discussion of developing appropriate principles to guide further mathematical contributions in the area of ranking criteria, from the grand scale of international rankings to the small scale of round robin tournament secondary criteria for separating teams tied on primary competition points.

*Keywords*: Cricket, Principles-based Reasoning, Ranking Criteria, Round Robin Tournaments, Team Ratings

**Acknowledgements**

I wish, as always, to acknowledge the amazing support, collaboration and friendship of Frank Duckworth and the late Tony Lewis. Their generosity of spirit gave me the most wonderful opportunity of my professional life, and I hope they are proud of my ongoing attempts to pay that generosity forward.

**References**

James, B. (2010). The New Bill James Historical Baseball Abstract. Simon & Schuster.

Duckworth, FC & Lewis, AJ (2004). A successful Operational Research intervention in one-day cricket. *The Journal of the Operational Research Society*. *55*, 749–759.

Stern, S. E. (2009). An Adjusted Duckworth-Lewis Target in Shortened Limited Overs Cricket Matches. *The Journal of the Operational Research Society*. *60*, 236–251.

# SPEEDS: SPORT AND EXERCISE SCIENCE EXCELLENCE THROUGH DATA SCIENCE

Eugene Sachkou[a], David L. Carey[a*], Minh Huynh[a], Haresh Suppiah[a], Richard Little[b], Matthew Varley[a], Andrew Govus[a]

[a] *La Trobe University, Sport and Exercise Science, School of Allied Health, Human Services & Sport*
*b Victorian Institute of Sport*
[*]*Corresponding author: d.carey@latrobe.edu.au*

**Submitted for consideration as (select one):**

- ☒ Oral presentation
- ☐ Poster

**Abstract**

A recent survey of Australian high performance sport practitioners identified common deficiencies in statistical knowledge within sport and exercise science, and proposed recommendations for improvement, as well as a pathway for sport and exercise science practitioners to increase their capability. The survey found that although practitioners may have a basic knowledge of data, they face a significant obstacle in enhancing their understanding and application of statistics due to the lack of accessible, contextualised, and industry-relevant statistical education materials that use open-source datasets related to high performance sport.

This project aims to curate a repository of contextualised datasets for use within sport and exercise science as well as statistical education in the form of online learning materials. The platform created will be open to contributions from the academic and sport communities with the goal of creating a lasting and valuable resources for data science and statistics capability uplift in sport science.

*Keywords*: Education, open data, sport science.

**References**

Grohmann, A. and Kauffeld, S. (2013). Evaluating Training Programs: Development and Correlates of the Questionnaire for Professional Training Evaluation. *International Journal of Training and Development*, Vol. 17, Issue 2, pp. 135-155, 2013.

Sainani, K., & Chamari, K. (2022). Wish list for improving the quality of statistics in sport science. *International Journal of Sports Physiology and Performance*, *17*(5), 673-674.

Sainani, K. L., Borg, D. N., Caldwell, A. R., Butson, M. L., Tenan, M. S., Vickers, A. J., ... & Bargary, N. (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.

# ESPORTS AS A FORM OF ACTIVE AGEING: AN EXAMINATION OF LONGEVITY IN STARCRAFT 2'S TOP RANKINGS

Indeesha Senarath [a] , Minh Huynh [a,d]

[a] *LaTrobe University*
[d] *Corresponding author: m.huynh@latrobe.edu.au*

## Abstract
Esports has rapidly emerged as a dominant form of entertainment and competitive play in the digital era. While the realm of esports is often associated with younger players, a rising trend of older participants has prompted intriguing questions regarding age, performance, and longevity in competitive gaming. This study sets out to explore the relationship between age and sustained success in one of the most iconic and skill-intensive esports titles, Starcraft 2. Specifically, we investigate whether players who ascend to the top 10 rankings at an older age exhibit a different trajectory in terms of their tenure within the elite echelons compared to their younger counterparts. Drawing from a comprehensive dataset spanning several years, we chart the ranking positions of the top Starcraft 2 players, alongside variables such as their ages and the duration of their stay within the top rankings. Our preliminary analysis suggests a potential trend: players who enter the top 10 at a younger age may, on average, enjoy a more extended presence within these prestigious ranks than those who achieve this feat later in life. If substantiated, these findings could offer valuable insights into the dynamics of ageing within the esports domain. They underscore the concept that esports, far from being a young person's pursuit exclusively, offers avenues for active ageing where older individuals can engage, compete, and achieve at high levels. Nonetheless, the interplay between age, neuroplasticity, reflexes, and gaming strategies remains a multifaceted tapestry that warrants further exploration.

**Keywords**: Esport, Active Aging, Starcraft 2, Longevity, Top Rankings

## 1. INTRODUCTION
The digital arena of esports has surged in popularity and complexity, challenging traditional narratives surrounding competition, professional trajectories, and age-related limitations. While traditional sports often present an age-dependent curve of performance, esports occupies a unique space where cognitive agility, strategic depth, and split-second decision-making reign supreme. Starcraft 2, with its intricate blend of strategy and frenetic gameplay, stands as an emblematic representative of this new age of competition (Chan et al., 2022; Ward and Harmon, 2019). Often associated with young prodigies, the esports realm has prompted a pertinent inquiry: How does age impact one's competitive journey, especially at the zenith of performance rankings?

Historical data from conventional sports reveal discernible patterns of athletes peaking in their mid to late twenties, underpinned by factors such as physical endurance, recovery rates, and muscular prowess. In stark contrast, esports primarily homes in on cognitive faculties, ranging from tactical planning to hand-eye coordination. The widespread assumption posits that younger individuals, blessed with heightened neuroplasticity and swift reflex responses, hold an innate advantage. Yet, this hypothesis remains relatively untested, especially when considering the duration of a player's stay at the top of competitive leaderboards.

The illustrious game of Starcraft 2 serves as a fertile ground for this exploration. As a real-time strategy titan, it has cultivated a rich history and a dedicated competitive scene (O'Camb, 2022). Given its demanding nature, Starcraft 2 offers invaluable insights into the interplay between age and sustained success. Before delving into the mechanics of this relationship, it's crucial to recognize the broader context: the notion of active ageing in a digital world. As global demographics shift towards older populations, the quest for meaningful, age-inclusive engagements intensifies. Can esports serve as a beacon, offering a platform where age does not confine competitive spirit? And if so, do inherent age-related factors dictate the trajectory and duration of a player's competitive pinnacle?

In light of these questions, our investigation seeks to determine the relationship between a player's age upon entering Starcraft 2's top 10 rankings and the duration of their presence within this elite tier. The central hypothesis posits that as the age at which a player enters the top 10 increases, the number of weeks they maintain this position will decrease. In addition to validating or challenging this hypothesis, our study aims to contribute to the broader discourse surrounding esports, age dynamics, and the potential for digital platforms to foster active ageing across all age brackets. The subsequent sections will provide a comprehensive breakdown of our methodology, data analysis, and findings, each elucidating the nuances of age and excellence in the world of Starcraft 2.

## 2. METHODS

### DATA COLLECTION

Data for this study was meticulously sourced from the online repository http://aligulac.com/. Aligulac is renowned for its comprehensive database, documenting match statistics for Starcraft 2 professional matches. For the scope of this investigation, we focused on data spanning from March 10, 2010, to the present.

### PLAYER RATING SYSTEMS

Central to our analysis was the Glicko rating system—a model employed by the website to rank players. The Glicko system, a brainchild of Professor Mark E. Glickman, is a refined version of the Elo rating system. Initially conceptualized for board games like chess, the Glicko model introduces a variable, called the rating deviation (RD), which signifies the confidence in a player's rating. The lower the RD, the more certain the rating is, and vice-versa.

In the realm of traditional sports, the Glicko model has been applied to ascertain player or team rankings, especially in games where outcomes are binary (win/loss). It captures not only a player's skill but also the certainty about that skill, offering a dynamic and more nuanced understanding of player performance over time. Starcraft 2, with its intensely competitive nature and binary outcomes, lends itself well to the Glicko model. The game's intricate strategies, rapid decision-making, and varying gameplay styles mean that player performance can fluctuate. The modified Glicko model employed by Aligulac accounts for these nuances, ensuring that the ratings are not just a reflection of wins and losses but also a testament to a player's consistency and adaptability (Tekofsky et al., 2015; Ćirović and Ćirović, 2019).

### ANALYSIS

For the core of our analysis, we zoomed in on the top 10 players, as determined by their Glicko ratings, for each week from 2010 to the present. This approach offered a snapshot of the elite player pool's composition and dynamics over more than a decade. Given the count nature of our dependent variable (weeks in the top 10), we employed a Poisson regression to model the data. However, to account for potential overdispersion, we also considered the suitability of a Negative Binomial Regression. Preliminary analyses were undertaken to check for overdispersion in the Poisson model by examining the ratio of the deviance to the degrees of freedom and assessing the Pearson chi-squared statistic.

## 3. RESULTS

### DESCRIPTIVE STATISTICS

The descriptive statistics of the primary variables are presented in Table 1. On average, players spent approximately 36.26 weeks in the top 10, with a large standard deviation of 48.48 weeks. This suggests considerable variability in the duration players remained within the top 10 rankings, ranging from as short as a single week to as long as 215 weeks.

Table 1: Descriptive statistics of top Starcraft 2 players (2010 to present)

| | Mean | Standard Deviation | Minimum | Maximum |
|---|---|---|---|---|
| Weeks (n) | 36.26 | 48.48 | 1 | 215 |
| Age top10 (years) | 21 | 2.61 | 15.43 | 30.15 |
| Age (years) | 30 | 3.77 | 19 | 42 |
| Earnings ($) | 276,221 | 26,5792.4 | 2,278 | 1,246,333 |
| Matches Played (n) | 1,401 | 1,085 | 151 | 4,631 |
| Matches Offline (%) | 33.01 | 17.17 | 0 | 80.2 |

The average age at which players entered the top 10 was approximately 21.24 years, with a relatively modest standard deviation of 2.61 years. The age of players during the time of data collection averaged around 30.76 years, indicating that many of these players had maintained a professional career spanning nearly a decade or longer. In terms of earnings, players amassed an average of $276,221 throughout their careers. However, there was a substantial range in earnings, with some players earning as little as $2,278 and others securing as much as

$1,246,333. The players, on average, participated in about 1401.77 matches. Some had played as few as 151 matches, while others boasted a staggering 4631 matches, highlighting the diversity in match exposure and experience. It's also worth noting that approximately 33% of these matches were played offline, emphasizing the importance of face-to-face competitions in a predominantly online esports landscape.

NEGATIVE BINOMIAL REGRESSION

Upon fitting the Poisson regression model, we observed evidence of overdispersion, which suggested that the variance was greater than the mean for our dependent variable (weeks in the top 10). Given this overdispersion, we proceeded with a Negative Binomial regression, which is better equipped to handle such variability in count data. From the Negative Binomial regression results presented in Table 2, both the age when a player entered the top 10 (Age_top10) and the number of matches played are statistically significant predictors of the duration a player spends in the top 10 rankings.

Table 2: Results for Duration in Top 10 Rankings Relative to Age of Entry and Matches Played

| Predictors | Incidence Rate Ratios | 95% CI | p-value |
|---|---|---|---|
| (Intercept) | 278.92 | 38.99 to 2076.22 | < 0.001 |
| Age_top10 | 0.87 | 0.80 to 0.95 | 0.003 |
| Matches Played | 1.02 | 1.01 to 1.03 | < 0.001 |

Observations: 91
$R^2$ Nagelkerke: 0.337

The incidence rate ratio for Age_top10 is 0.87, suggesting that for each additional year in age when entering the top 10, players are expected to spend about 13% less time in the top 10 rankings, holding all else constant. Similarly, the incidence rate ratio for Matches Played is approximately 1.02, indicating a very slight (but statistically significant) increase in the expected duration in the top 10 rankings for each additional match played. The $R^2$ (Nagelkerke) of 0.337 indicates that approximately 33.7% of the variability in the duration spent in the top 10 rankings is explained by the model.

## 4. DISCUSSION

As hypothesised, there is an inverse relationship between the age at which a player enters the top 10 and the duration of their stay within this elite circle. Drawing parallels with traditional sports where physical prowess tends to wane with age, in Starcraft 2, older players entering the top tier may find it challenging to maintain their elite status for prolonged periods. This observation is in line with the broader discourse around cognitive faculties and neuroplasticity. Younger minds, often more agile and adaptive, could have an edge in rapidly evolving games like Starcraft 2, which demand swift decision-making, real-time strategy adaptations, and intricate gameplay mechanics.

The results, as delineated in Table 2, elucidate that age, when a player enters the top 10 rankings, significantly impacts the duration spent in these coveted spots. Specifically, for each additional year in age upon entering the top 10, a player's expected time in this ranking reduces by approximately 13%. This finding corroborates our initial hypothesis, suggesting that younger players tend to have a longer stint in the top 10 rankings compared to their older counterparts. This could be attributed to various factors, including possibly better adaptability, quicker reflexes, or longer potential career longevity for younger players. However, it's also essential to recognize that entering at a younger age might provide players with more extended periods of potential professional play, allowing them longer durations to remain competitive. Furthermore, the variable 'Matches Played' was also a significant predictor, although the effect size was relatively small. This might signify that with increased exposure and experience, players can marginally extend their presence in the top 10, suggesting the importance of consistent practice and engagement in competitive matches.

The implications of our findings are manifold. From a training perspective, aspiring players might be motivated to start their professional journey early, thereby optimizing their cognitive agility alongside gaining match experience. Esports organizations and coaches might tailor training regimes, recognizing the differential strengths of younger and more experienced players. Our study also touches upon the broader theme of esports as a form of active ageing. The average age of players in our dataset was approximately 30 years, a testament to the longevity of esports careers. While the competitive edge might be sharper in younger players, the world of esports remains inclusive, allowing older players to remain engaged, competitive, and relevant. This is a compelling narrative for the digital age, where traditional notions of age and competitiveness are continuously being redefined (Ashinoff, 2014).

However, our study isn't without its limitations. The data sourced from Aligulac, while comprehensive, might not encapsulate all nuances of a player's journey. For instance, external factors like changes in training environments, personal life events, and shifting game meta can all influence a player's performance, aspects not captured in our dataset. Additionally, the Glicko rating system, though robust, has its inherent biases and might not always reflect the real-time dynamics of player performance. Another consideration is the potential for survivorship bias. By focusing on top players, we may have overlooked the journey of many who started young but couldn't sustain their performance or those older entrants who might have remained just outside the top 10. Such players might offer additional insights into the age-performance paradigm.

In conclusion, the relationship between age, experience, and sustained excellence in Starcraft 2 is intricate and layered. While our study sheds light on some aspects of this dynamic, it also underscores the need for more granular, longitudinal research. As esports continues to burgeon, intersecting with themes of active ageing and redefining competitive trajectories, it remains a fertile ground for further academic exploration and practical insights. Through such studies, we hope to further enrich our understanding, not just of the game but also of the evolving narrative of age and excellence in the digital era.

**Acknowledgements**
We wish to thank http://aligulac.com/ for continuously providing StarCraft II match and rating data for the general public.

**References**

Chan G, Huo Y, Kelly S, Leung J, Tisdale C, Gullo M. (2022). The impact of eSports and online video gaming on lifestyle behaviours in youth: A systematic review. Computers in Human Behavior, 126:106974.

O'Camb J. (2022). Strategy in StarCraft 2 learning curves. Honours thesis, Simon Frase University.

Ward MR, Harmon AD. ESport Superstars. (2019). Journal of Sports Economics, 20(8), 987-1013.

Tekofsky S, Spronck P, Goudbeek M, Plaat A, Herik Jvd. (2015). Past Our Prime: A Study of Age and Play Style Development in Battlefield 3. IEEE Transactions on Computational Intelligence and AI in Games, 7(3):292-303.

Ćirović Z, Ćirović N (2019). A StarCraft 2 Player Skill Modeling. y-BIS 2019 Conference Book: Recent Advances in Data Science and Business Analytics.

Ashinoff BK. (2014). The potential of video games as a pedagogical tool. Frontiers in Psychology, 5.

This page is intentionally blank.

# CRACKING THE LOTTERY CODE: TOOLS TO UNCOVER SYSTEMATIC DATA TAMPERING

Erica Mealy [a], Anthony Bedford [a,d]
[a] *University of the Sunshine Coast*
[d] *Corresponding author:* *abedford@usc.edu.au*

**Submitted for consideration as (select one):**

- ☒ Oral presentation - Mealy
- ☐ Poster

**Abstract**

Online games of chance, including lotteries, rely on large random number generation systems. Despite stringent requirements from regulators to ensure statistically proven randomness, as computation power has increased, so too has the capacity for subtle manipulation of datasets. In this research, we look at an array of processes used to find, and detect, any hint of seeded (a.k.a. doctored or fixed) data in both hypothetical and actual sets of lottery data. The question is: Are these systems able to pick up even the *slightest* fixing of data, which would benefit a designer manipulating the random number generation process post generation? We develop, using Python, a suite of tools with embedded, established statistical methods to test this hypothesis. Current state-of-the-art processes and procedures used for the detection of systematic data are outlined, and the methodology to find potential sets of non-randomness is covered. Expanding upon this we explore processes of creating random data, seeding random data, and using externally generated random data to then subtly doctor, in the ever so slightest way, to determine if we can systematically win. We evaluate how one would hide the non-systematic information and propose processes in segmenting the data to find potential systematic non-randomness in a set of data for lotteries.

*Keywords*: random numbers, randomness, statistical tests

**References**

Jun, B., & Kocher, P. (1999). The Intel random number generator. Cryptography Research Inc. white paper, 27, 1-8.

Lott, S. F. (2005). Empirical Tests of Random Number Generators. Retrieved from https://www.itmaybeahack.com/homepage/_static/rngtest/rngdoc.html

Niederreiter, Harald (1992). Random Number Generation and Quasi-Monte Carlo Methods., Society for Industrial and Applied Mathematics, https://doi.org/10.1137/1.9781611970081

Office of Liquor and Gaming Regulation. (2020). Random number generator minimum technical requirements v1.5. Retrieved from https://www.publications.qld.gov.au/ckan-publications-attachments-prod/resources/15ec1054-2ebc-46fe-80c7-569eb733c58d/random_number_generator_minimum_technical_requirements_v1_5_clean.pdf

Office of Liquor and Gaming Regulation. (2016). Random number generator minimum technical requirements v1.4.1 Retrieved from https://www.publications.qld.gov.au/dataset/2117b434-a133-48ab-acef-96a3feb2f905/resource/8cef84d5-10cb-421a-beee-064eb60196ae/download/randomnumbergeneratorsminimumtechnicalrequirementsv141.pdf

Tasmanian Liquor and Gaming Commission. (2023, July). Fully Automated Table Game Technical Standards. Liquor and Gaming, General Gaming Standards. Retrieved from https://www.treasury.tas.gov.au/Documents/TLGC%20Fully%20Automated%20Table%20Game%20Technical%20Standards%20-%201%20July%202023.PDF

Victorian Commission for Gambling and Liquor Regulation (VCGLR). (2020). 3.20 Random Number Generator and Symbol Selection. In The Technical Requirements Document for Melbourne Casino v4.06 (pp. 24-25). Victoria State Government.

# TOWARDS COMPUTER VISION SOLUTIONS IN HORSERACING

Anthony Bedford [a,d] , Erica Mealy [a] , Abigail Koay [a], Peter Embleton [a]

[a] *University of the Sunshine Coast*
[d] *Corresponding author: abedford@usc.edu.au*

**Submitted for consideration as (select one):**
- ☒ Oral presentation – Bedford to present
- ☐ Poster

**Abstract**
Horse racing is one of the world’s oldest sports. In Australia, it has evolved into a huge entertainment industry, large employer, a source of tax and revenue, and over five-billion-dollar contributor to the economy, with over $9.1 billion of value-added impact. In 2016/17 there were 2,175 TAB meetings and 461 non-TAB meetings, and there are an estimated 19,000 thoroughbred races per year. From this, in Australia it is estimated wagering turnover (2022-23) was $29.1 billion, so there is of course money at stake. One big opportunity exists for technology to assist with the sport – in protest hearings. The current reason for a protest is when a jockey, owner, trainer, or steward alleges interference by one party against another during a race that may have affected the outcome of a race. If a protest is upheld by officials, the runner that caused the interference is placed directly after the horse interfered with. If a protest is dismissed by officials, the original result of the race stands. In the hearing, only verbal testimony and replays of vision are used which can be bias or incomplete. Other features such as time data, or any inference used, save for the final margin between horses (as opposed to say that used in cricket for dismissals) would provide additional evidence and support to decision making and lower the risk of protest. In this work, we explore the emerging area of computer vision in horseracing, and note problems faced with existing technology, such as variable frame, variable field of vision, speed of camera, and multi-object tracking. We then develop a framework for this process using computer vision and define how it would be set-up to assist greatly in the decision-making utilising inferred and affected horse speeds.

*Keywords*: Computer Vision, Horseracing, framework

**References**
Ng, WWY, Liu, X, Yan, X, Tian, X, Zhong, C, & Kwong, S (2023). Multi-object tracking for horse racing. Information Sciences, Elsevier, https://www.sciencedirect.com/science/article/pii/S0020025523005364

Colle, P (2022). What AI can do for horse-racing?. arXiv preprint arXiv:2207.04981, arxiv.org, https://arxiv.org/abs/2207.04981

Yoda, I (2012). Horse position information analyzing and displaying method. US Patent 8,107,679, Google Patents, https://patents.google.com/patent/US8107679B2/en

G. N. Iyer, V. S. Bala, B. Sohan, R. Dharmesh and V. Raman, "Automated Third Umpire Decision Making in Cricket Using Machine Learning Techniques," 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2020, pp. 1216-1221, doi: 10.1109/ICICCS48265.2020.9121078.

# THE IMPACT OF WINNING ON THE NUMBER OF ARTICLES PUBLISHED ABOUT HOST TEAMS DURING TOURNAMENTS

Grant Maxwell [a], Paul Bracewell [b,c,d]

[a] *EightOne*
[b] *Footmarx,* [c] *Endorsco*
[d] *Corresponding author: paul@endorsco.com*

**Submitted for consideration as (select one):**

- ☒ Oral presentation
- ☐ Poster

**Abstract**

The media plays a pivotal role in drawing public attention to events of national and international significance, shaping perceptions and narratives around these occurrences. In the context of New Zealand, this influence was particularly evident as the country hosted three major international sporting events between 2021 and 2023. These events included the Women's Rugby World Cup in 2021, the Women's Cricket World Cup in 2022, and the FIFA Women's World Cup in 2023, which was jointly hosted by Australia and New Zealand.

To understand the media's impact on public engagement with these events, a comprehensive analysis was conducted. This analysis tracked the number of mentions of the host team and each tournament in mainstream media outlets, overlaying these findings with key events and performances throughout each tournament's duration.

Interestingly, the results revealed a correlation between the prevalence of media reporting and the performance of the host team. When the New Zealand teams performed well, there was a noticeable increase in media coverage, suggesting that success on the field directly influences the level of media attention received.

This pattern of media coverage has significant implications for brand exposure and the activation of events during major tournaments. It suggests that the success of the host team can amplify the reach and impact of associated marketing and promotional activities, enhancing the visibility of sponsors and partners. Furthermore, it underscores the importance of media strategy in leveraging sports events for national branding and international recognition, highlighting the potential for strategic partnerships between sports organisations, media, and commercial entities to maximize the benefits of hosting such events.

*Keywords*: Machine Learning, Media Analysis

# A FRAMEWORK FOR USING IMPORTANT EVENTS IN SPORT FOR STRATEGIC BRAND VISIBILITY

Tram Chau[a] , Paul Bracewell [a,b,c,d]
[a] *Victoria University of Wellington, New Zealand*
[b] *Footmarx,* [c] *Endorsco*
[d] *Corresponding author: paul@endorsco.com*

## Abstract

Organizations strategically place logos at sports venues and on team kits for marketing and branding, leveraging sports sponsorship and advertising to achieve objectives like Brand Visibility, Positive Attribute Association, Target Audience Engagement, and Community and Fan Connection. When events are broadcasted, the reach and impact of these logos are significantly extended, engaging audiences across diverse geographies.

Here, we introduce a framework utilising machine vision to quantify the frequency and visibility of logos during critical moments of sporting events, building upon concepts previously discussed at Mathsport conferences. We have expanded the capabilities of a Python-based tool, originally developed for tracking player movements, to include the identification and capture of brand logos. The primary objective is to accurately detect and assess the prominence of these logos in the dynamic environment of sports. A secondary aim is to link the visibility of logos to significant game moments, thereby evaluating their impact on fan engagement and the likelihood of inclusion in highlight reels.

This approach allows us to explore the relationship between logo visibility at key game instances and the propensity for these moments to be featured in replays or highlight segments. Our goal is to uncover insights into how strategic logo placement influences fan attention and memory, potentially enhancing the effectiveness of sports sponsorships.

Initial findings suggest that this framework can provide valuable insights into the strategic placement of logos and its impact on marketing within sports. By examining logo frequency and context, sponsors and advertisers can better gauge the effectiveness of their placements, optimising strategies to maximise visibility during pivotal moments. This research contributes to a deeper understanding of sports marketing dynamics, presenting a novel methodology for assessing the value of sponsorship in live sports events and subsequent highlight coverage.

**Keywords**: machine vision, sports marketing, deep learning, pre-train model, logo detection.

## 1. INTRODUCTION

Machine learning techniques have significantly advanced sports analysis, enhancing outcomes by providing deeper insights into athlete movement (Ivanovsky et al., 2022; Lee et al., 2020), ball trajectory (Kamble et al., 2019), and event detection (Hong et al., 2018). These applications have been pivotal in refining sports coaching strategies. Beyond these player-centric analyses, tracking the banners of sports sponsors in videos has emerged as a crucial area, yet it remains underexplored. This tracking offers valuable data on brand exposure during broadcasts, but challenges such as varying viewpoints, changing illumination, occlusions, and complex backgrounds have historically hindered accurate detection (Hou et al., 2023).

The availability of large logo image datasets has facilitated the development of detection algorithms. Techniques like tracking-through-detection, which do not require video annotations for each specific logo, leverage pre-trained models through transfer learning to enhance logo detection (Liao et al., 2017). Despite the capability to process videos at a frame rate of 25 to 30 frames per second, the processing time remains a challenge; for instance, Liao et al. (2017) reported processing times of approximately 0.59 seconds per frame, indicating significant delays in real-time applications.

Research has also extended to other objects in sports videos, such as players, soccer balls, and referees (Naik et al., 2022), and even non-sports contexts like agriculture (Ge et al., 2022). For logos, customized detection models such as YOLOv3 and YOLOv4 have been adapted for specific datasets like FlickrLogos-32, which contains images from 32 brands (Wang et al., 2022; Paleček & Chaloupka, 2021). These models are trained on a closed set of known brands, yet the dynamic nature of branding, with new logos continuously emerging, poses ongoing challenges. The need for a detection method capable of identifying logos outside the training dataset is clear.

We introduce a novel deep learning approach designed to efficiently track generic logos in sports videos, adapting to the fast-paced nature of live sports without compromising on accuracy or speed. The subsequent sections are structured to outline the methodology (Section 2), present our findings (Section 3), and discuss the implications of these results along with conclusions (Sections 4 and 5).

## 2. METHODS

The methodology depicted in Figure 1 is structured into four main stages, designed to optimize the efficiency and accuracy of our logo tracking system:

1. Frame Comparison: This initial stage involves comparing consecutive frames to identify and skip those with minimal changes, thereby reducing unnecessary detections and enhancing processing speed.
2. Logo Detection: Each frame, selected based on the criteria established in the previous stage, undergoes logo detection. This involves identifying logos using a deep learning model that scans for known brand symbols.
3. Logo Tracking: After detection, logos are tracked across successive frames. This stage ensures continuity and captures the trajectory of logos throughout the sequence, facilitating an analysis of their visibility and movement.
4. Post-Processing: The final stage groups similar tracked logos, using clustering algorithms to consolidate detections into a concise list. This step simplifies the output and focuses on distinct logo appearances, minimizing redundancy and improving the clarity of results.

These stages collectively form a robust methodology for tracking logos in sports videos, as illustrated in Figure 1.

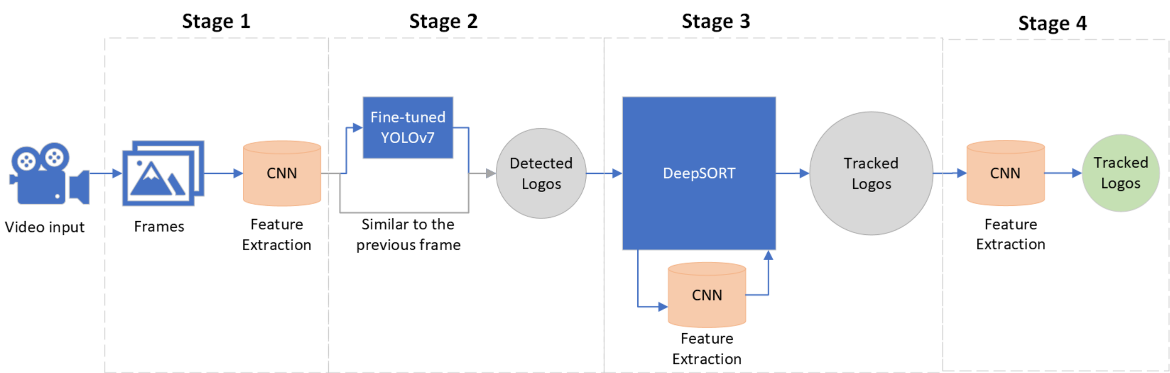



Figure 1: Method to track logo from video by detecting logo in each frame.

DATA

Video for inference is downloaded from YouTube — the highlight rugby match between Wellington and Waikato (All Blacks YouTube channel, 2023a). This is a 12-minute clip with 25 frames per second or 18000 frames in total.

PRE-PROCESSING

In practical applications, it is crucial that our method delivers results within a reasonable timeframe. Our preliminary experiments revealed that detection typically consumes more time than tracking. This is primarily due to the inherent properties of videos, where frames per second (fps) dictate the smoothness of playback. We observed that many consecutive frames exhibit high similarity, making repeated detections on such frames inefficient and resource-intensive.

To address this, we implemented a frame comparison strategy during the pre-processing phase, which leverages deep appearance features extracted from a pre-trained model. This process involves the following steps:

1. Frame Comparison: Each frame is analyzed for similarity to its predecessor. The similarity score is computed and compared against a predetermined threshold of 0.998.
2. Decision Making: If the similarity score exceeds the threshold, indicating minimal change from the previous frame, the detection from the preceding frame is reused. This detected data is directly passed to the tracking module, bypassing the need for re-detection.

3. Detection and Tracking: For frames where the similarity score falls below the threshold, the frame undergoes fresh detection to identify any new or changed logos. These logos are then passed on to the tracking stage.
4. Visual Illustration: Figure 2 visually details this process, showing the frame comparison across 10 consecutive frames along with their respective similarity scores.

This methodology ensures that our system remains efficient by minimizing unnecessary detections, thus saving time and computational resources while maintaining high accuracy in dynamic environments.


Figure 2: Illustratration of the application of our frame comparison methodology across a sequence of ten consecutive frames, from frame 1516 to 1525, using similarity scores based on deep feature appearances between consecutive frames.

Frames 1516 and 1517: Both frames register similarity scores of 0.996 and 0.995, respectively. These scores fall below the threshold of 0.998, indicating noticeable changes between these and their preceding frames. Consequently, they are forwarded to the detection stage for logo identification. Frames 1519 and 1520: These frames exhibit similarity scores above the threshold, with scores sufficiently high to assume no significant visual changes have occurred. Thus, they are skipped in the detection process to conserve computational resources. Frames 1521 to 1523: Similar to frames 1516 and 1517, these frames have scores below the threshold, suggesting new or altered content that requires detection and subsequent tracking. Frame 1524: This frame shows a marked decrease in similarity, scoring only 0.572, indicative of significant visual differences from its predecessor, necessitating a return to the detection stage. Frame 1525: With a similarity score of 0.955, this frame also falls below the threshold, confirming the need for further detection efforts.

This step-by-step analysis not only conserves resources by avoiding redundant detections but also ensures that no significant visual changes go undetected, maintaining the accuracy and reliability of our tracking system.

Moreover, the inference time is defined by other factors such as the frame's resolution, the detected confidence score (the lower the threshold the more detected logos to be tracked and the more time the process is), and other parameters. Regarding pre-processing, the video's resolution will be rescaled to 384x640 to be compatible with the detection model without affecting the model performance.

DETECTION

In the detection stage, we utilize the YOLOv7 model, pre-trained on the Logodet3k dataset, to identify logos within video frames (Wang et al., 2022a; Nathanjc, 2023). The YOLOv7 model has been chosen for its robust performance in detecting small and detailed objects across diverse backgrounds. The configuration of key parameters is crucial for optimizing detection accuracy:

1. Image Size: Tailored to balance between resolution and processing speed, ensuring that logos are detectable even in complex scenes.
2. Detection Confidence Score: Set with a threshold of 0.5 to maintain a balance between accuracy and the number of detections. This means that only logos detected with a confidence score higher than 0.5 are considered reliable enough to be passed onto the tracking stage.
3. IoU (Intersection over Union) Threshold for Non-Max Suppression: This parameter helps reduce redundant logo detections by ensuring that each logo is counted only once, improving the precision of our tracking.

Figure 3 illustrates this process in action, displaying logos detected in frames 1518 and 4749. For instance, in frame 1518, even though three logos are detected, only the logo with a confidence score of 0.75 surpasses our confidence threshold and is forwarded to the tracking algorithm. This selective process ensures that our tracking system focuses on high-probability detections, enhancing both efficiency and accuracy.

Adjusting the confidence score threshold can potentially increase the number of logos tracked by including detections with lower confidence. However, this must be balanced against the risk of introducing false positives, which could lead to erroneous tracking results.


Figure 3: Example of detected logo with their confidence score. Only those with confidence scores higher than the threshold will be transferred to the tracking stage.

TRACKING

Once logos are detected, they advance to the tracking stage, where the DeepSORT algorithm plays a crucial role in maintaining continuity across frames (Wojke et al., 2017). This stage involves several key processes:

1. Deep Feature Extraction: Similar to the pre-processing stage, deep features of each detected logo are extracted using the same pre-trained model. These features capture intricate details necessary for accurate tracking over time.
2. Position Prediction: The Kalman filter is employed to predict the future positions of each logo based on their current states. This prediction assists in maintaining tracking continuity even when logos move or temporarily disappear from view.
3. Association Metric: The combination of deep features and position predictions forms an association metric. This metric is crucial for assigning a consistent identity to each logo across consecutive frames.
4. Continuity of Identity: If a logo maintains its visibility without interruption for up to 50 frames (approximately 2 seconds), it retains its assigned identity.
5. Identity Reassignment: If a logo reappears after a disappearance of more than 50 frames, it is considered a new instance and is assigned a new identity. This helps in distinguishing between continuous and recurring logo appearances.

Figure 4 provides a visual demonstration of this process. It displays a logo tracked across several frames, where slight changes in appearance and position occur. Despite these changes, the logo is consistently recognized and retains the same identity, demonstrating the effectiveness of the DeepSORT algorithm in tracking dynamic objects in video streams.

The tracking algorithm successfully maintains a consistent identity for logos detected in mostly consecutive frames, specifically in frames 4745, 4757-4760, and 4763-4773. Despite slight variations in appearance and intermittent absences spanning less than 50 frames, the system effectively assigns these logos a single identity. This demonstrates the robustness of our tracking method in handling gaps and minor visual changes without losing track of the logo's continuity across the sequence.


Figure 4: Tracking result. Detected logos in mostly consecutive frames (4745, 4757-4760, 4763-4773) with slightly different appearance and missing frames

POST-PROCESSING

The tracking stage is susceptible to identity switch issues, where a single logo might be incorrectly assigned multiple identities. This problem arises due to several factors, including the thresholds set for intermittent logo disappearance and for matching logo appearances. To mitigate these challenges:

1. Feature Extraction: We utilize a pre-trained model for extracting deep features from the first instance of each tracked logo. This method ensures that the most representative features of each logo are used for subsequent comparisons.
2. Grouping Criteria: We employ both disparity distance and cosine similarity scores to determine the similarity between logo instances. Logos that exhibit low disparity distances and high cosine similarities to each other are considered to be the same entity and are merged into a single identity.

Figure 5 illustrates the effectiveness of this post-processing technique. It shows examples of logos that, despite potential variations through the video sequence, have been successfully grouped together as a single identity based on their deep feature similarities. This grouping helps maintain consistency in logo tracking and reduces errors associated with identity switches.


Figure 5: Example of grouping tracked logos. The first logo in each row is the anchor, the other logos in the same row are its neighbours. The neighbours will be grouped into the anchor logo for the final result.

## 3. RESULTS

In an analysis of a 12-minute highlights rugby match, which included 18,000 frames, our pre-processing stage was able to identify that approximately 55% of the frames were highly similar. This significant detection streamlined our processing strategy, reducing the overall inference time to just 5 minutes and 50 seconds, markedly faster than real-time processing.

The efficiency of our detection and tracking algorithms is highlighted by the average detection time of approximately 13 milliseconds per frame, and an even quicker tracking time of about 2 milliseconds per frame. These metrics demonstrate our system's capability to handle large volumes of data swiftly and effectively.

Figure 6 displays the list of the 39 logos that were successfully tracked throughout the video, showcasing the effectiveness of our methodology. Figure 7 provides detailed metrics for the selected logos from Figure 6, offering insights into the frequency and duration of logo appearances, which are critical for assessing brand visibility and impact.


Figure 6: The final list of tracked logos from the video.


Figure 7: The frequencies in the entire video and duration at each occurrence of selected logos from Figure 6.

Furthermore, Figure 8 presents the inference results from a different video—the highlights of a rugby match between Auckland and Waikato, as shown on the All Blacks YouTube channel (2023b). This analysis showcases a distinct set of tracked logos. Despite the video's total length of 18,750 frames, our method efficiently skipped 25% of these frames, completing the inference in just 8 minutes and 50 seconds.

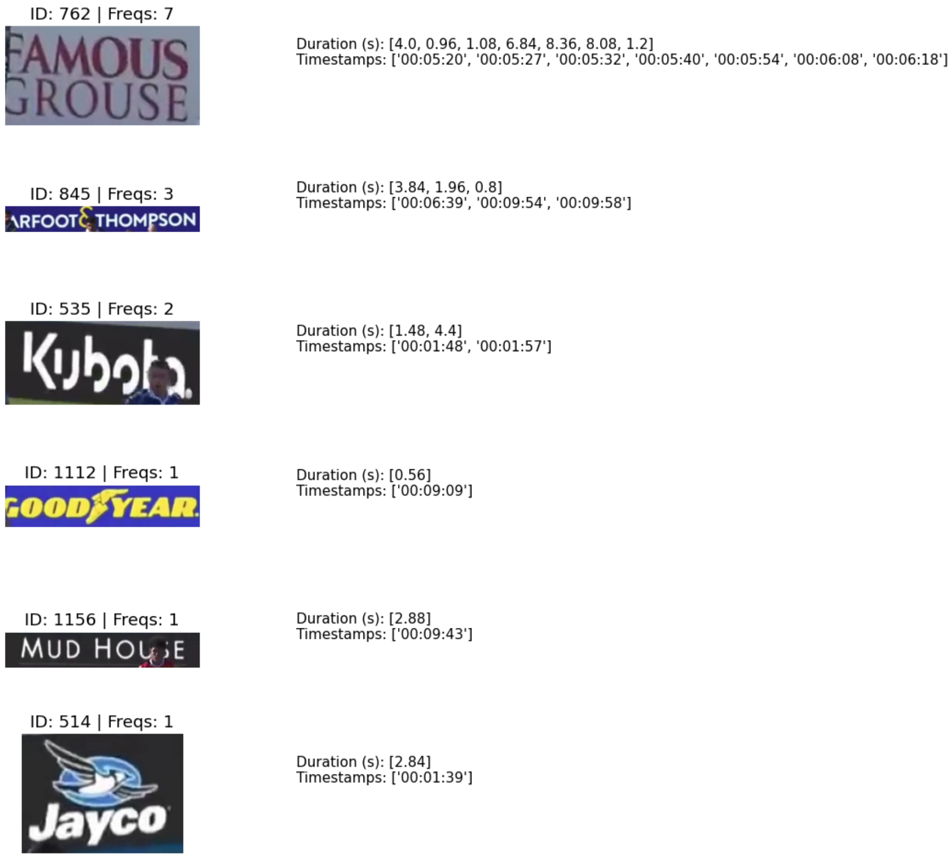



Figure 8: Several detected logos with metrics when running inference for another video.

## 4. DISCUSSION

Deep learning has significantly advanced the detection of generic logos from images, improving tracking methods and the effectiveness of pre- and post-processing stages. In our method, logos are detected in each frame with a minimum confidence score of 0.5. This threshold is adjustable, allowing for the capture of more logos with lower confidence scores and providing flexibility to address specific challenges. The pre-processing stage effectively minimizes redundant detections, while the post-processing stage groups similar logos to reduce the number of tracked logos by a third. However, issues arise when some logos are not fully resolved in post-processing, indicating a need for enhancements to the pre-trained feature extraction model or the integration of an image similarity deep learning model (Chen & He, 2020).

Our approach utilizes a pre-trained model trained on a vast dataset, capable of detecting logos absent from the training set. For unique logos, one-shot or few-shot learning methods are applied (Liu et al., 2023), showcasing the minimal presence of false positives and demonstrating the system’s efficiency in rapidly compiling logo data. This efficiency extends to the inference time, which is optimized to be faster than the video's duration, thus making this method highly applicable in real-world scenarios. Additionally, extracting logo crops provides data on logo sizes, a valuable metric as larger logos are more noticeable to viewers.

Despite these advancements, our method encounters limitations, particularly with logos that are complex or partially obscured, potentially leading to inaccuracies in brand exposure metrics critical for marketing analyses. A comparative analysis with traditional logo detection methods reveals that while our approach generally offers faster processing times and higher accuracy, the scalability and resource utilization, especially in systems with limited computational capabilities, need further optimization. Future research should explore the trade-offs between accuracy and resource efficiency to develop more streamlined models suitable for broader applications.

The implications of our method significantly influence sports marketing and sponsorship strategies by providing real-time, accurate data on logo visibility. This allows sports sponsors to make more informed decisions about logo placement for maximum exposure and opens up new possibilities for dynamic advertising strategies that adapt to audience engagement in real-time. However, the enhanced tracking capabilities also bring ethical considerations, particularly concerning privacy and the pervasive nature of advertising. As these technologies become more integrated into both public and private spaces, establishing guidelines to protect individual privacy while fostering innovative marketing practices is crucial.

## 5. CONCLUSIONS

Pre-trained deep learning models trained on extensive datasets are utilized for generic logo detection. We have developed a method that not only detects and tracks logos in sports videos efficiently but also compiles a comprehensive list of logos, calculating their frequency and duration of occurrence. This streamlined approach enables sports sponsors to swiftly assess brand exposure across individual or multiple videos throughout a season, thereby enhancing their decision-making processes in advertising strategies.

While our method demonstrates promising results in enhancing logo tracking in sports videos, addressing its technical limitations and broader impacts through continuous research and ethical practices will be key to realizing the full potential of deep learning in sports marketing. Future guidelines and regulations should address these concerns, ensuring that advancements in tracking technology are implemented responsibly and ethically.

## References


All Blacks YouTube channel. RD 6 HIGHLIGHTS: Waikato v Wellington (Bunnings NPC 2023). Sep. 2023a. URL https://www.youtube.com/watch?v=h3DA0EQ1_Gw.

All Blacks YouTube channel. RD 7 HIGHLIGHTS: Auckland v Waikato (Bunnings NPC 2023). Sep. 2023b. URL https://www.youtube.com/watch?v=BGiafYxNzIY.

Yuhao Ge, Sen Lin, Yunhe Zhang, Zuolin Li, Hongtai Cheng, Jing Dong, Shanshan Shao, Jin Zhang, Xiangyu Qi, and Zedong Wu. Tracking and counting of tomato at different growth period using an improving yolo-deepsort network for inspection robot. Machines, 10 (6), 2022. ISSN 2075-1702. doi: 10.3390/machines10060489. URL https://www.mdpi.com/2075-1702/10/6/489.

Yuxi Hong, Chen Ling, and Zuochang Ye. End-to-end soccer video scene and event classification with deep transfer learning. In 2018 International Conference on Intelligent Systems and Computer Vision (ISCV), pages 1–4, 2018. doi: 10.1109/ISACV.2018.8369043.

Sujuan Hou, Jiacheng Li, Weiqing Min, Qiang Hou, Yanna Zhao, Yuanjie Zheng, and Shuqiang Jiang. Deep learning for logo detection: A survey. ACM Trans. Multimedia Comput. Commun. Appl., 20(3), oct 2023. ISSN 1551-6857. doi: 10.1145/3611309. URL https://doi.org/10.1145/3611309.

Leonid Ivanovsky, Dmitry Matveev, Vladimir Khryashchev, and Alexander Semenov. Detection and tracking of sport players on videodata using deep learning methods. In 2022 International Siberian Conference on Control and Communications (SIBCON), pages 1–5, 2022. doi: 10.1109/SIB- CON56144.2022.10002998.

P.R. Kamble, A.G. Keskar, and K.M. Bhurchandi. A deep learning ball tracking system in soccer videos. Opto-Electronics Review, 27(1):58–69, 2019. ISSN 1230-3402. doi: https://doi.org/10.1016/j.opelre.2019.02.003. URL https://www.sciencedirect.com/science/article/pii/S123034021830146X.

JungSoo Lee, Sungwon Moon, Do-Won Nam, Jiwon Lee, Ah Reum Oh, and Wonyoung Yoo. A study on sports player tracking based on video using deep learning. In 2020 International Conference on Information and Communication Technology Convergence (ICTC), pages 1161–1163, 2020. doi: 10.1109/ICTC49870.2020.9289223.

Yuan Liao, Xiaoqing Lu, Chengcui Zhang, Yongtao Wang, and Zhi Tang. Mutual enhancement for detection of multiple logos in sports videos. In 2017 IEEE International Conference on Computer Vision (ICCV), pages 4856– 4865, 2017. doi: 10.1109/ICCV.2017.519.

Banoth Thulasya Naik, Mohammad Farukh Hashmi, Zong Woo Geem, and Neeraj Dhanraj Bokde. Deepplayer-track: Player and referee tracking with jersey color recognition in soccer. IEEE Access, 10:32494–32509, 2022. doi: 10.1109/ACCESS.2022.3161441.

Karel Paleˇcek and Josef Chaloupka. Logo detection and identification in system for audio-visual broadcast transcription. In 2021 44th International Conference on Telecommunications and Signal Processing (TSP), pages 357–360, 2021. doi: 10.1109/TSP52935.2021.9522636.

Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv preprint arXiv:2207.02696, 2022a

Jing Wang, Weiqing Min, Sujuan Hou, Shengnan Ma, Yuanjie Zheng, and Shuqiang Jiang. Logodet-3k: A large-scale image dataset for logo detection. ACM Trans. Multimedia Comput. Commun. Appl., 18 (1), jan 2022. ISSN 1551-6857. doi: 10.1145/3466780. URL https://doi.org/10.1145/3466780.
Nathanjc. Logo detection yolov7. 2023. URL https://huggingface.co/spaces/nathanjc/LogodetectionYoloV7/tree/main.
Tianying Liu, Lu Zhang, Yang Wang, Jihong Guan, Yanwei Fu, Jiajia Zhao, and Shuigeng Zhou. Recent few-shot object detection algorithms: A survey with performance comparison. ACM Trans. Intell. Syst. Technol., 14(4), jun 2023. ISSN 2157-6904. doi: 10.1145/3593588. URL https://doi.org/10.1145/3593588.
Nicolai Wojke, Alex Bewley, and Dietrich Paulus. Simple online and realtime tracking with a deep association metric. In 2017 IEEE International Conference on Image Processing (ICIP), pages 3645–3649, 2017. doi: 10.1109/ICIP.2017.8296962.
Xinlei Chen and Kaiming He. Exploring simple siamese representation learning. CoRR, abs/2011.10566, 2020. URL https://arxiv.org/abs/2011.10566.

# THE USE OF MATHEMATICAL MODELS TO PREDICT ATHLETIC PERFORMANCE IN MEN'S AND WOMEN'S 100M, 200M, 400M, LONG JUMP, HIGH JUMP, SHOT PUT AND DISCUS AT THE 2024 PARIS OLYMPIC GAMES

Ian T. Heazlewood

*Sport Science Institute, Sydney, NSW, Australia*
*Corresponding author:* edulearnsport2020@gmail.com

## Abstract

Prediction of future athletic performances is a common practice and frequently based on plotting the changes in world records, however changes in world records do not predict Olympic athletic performances. Data from previous Olympic final performances are a more realistic concept of changing performances at the Olympic Games and provide goals for athletes to achieve in Olympic finals. Previous research has used finalist scores averaging methods and linear and non-linear curve estimation mathematical modelling to predict performances in athletics at the 1996, 2000, 2004, 2012 and 2016 Olympic Games with a high degree of success by accurate prediction of event performance, high $R^2$ values and low residuals in the men's and women's 100m, 400m, long jump, high jump and throwing events. The aim of the research was to replicate the preceding methods to predict averaged performance based on first six finalists in events of men's and women's 100m, 400m, long jump, high jump, shot put and discus. The regression functions used were linear, logarithmic, inverse, quadratic, cubic, compound, power, sigmoidal, growth, exponential and logistic. The results indicated men 100m was multiple solutions with $R^2$ =.844, women 100m cubic $R^2$=.923, men 200m cubic $R^2$=.922, women 200m sigmoidal $R^2$=.696, men 400m cubic $R^2$=.922, women 400m multiple solutions $R^2$=.923 $R^2$=.578, men long jump cubic $R^2$=.821, women long jump inverse $R^2$=.750, men high jump cubic $R^2$=.911, women high jump cubic $R^2$=.936, men shot put cubic $R^2$=.931, women shot put inverse $R^2$=.511, men discus cubic $R^2$=.954 and women discus cubic $R^2$=.913. General trends predict the following performances at the 2024 Paris Olympics compared to 2020 Tokyo performances. Men 100m improvement of 0.1s, 100m women improvement 0.06s, men 400m decrease of 0.13s, women 400m improvement by 0.62s, men long jump improvement 9cm, women long jump improvement 31cm, men high jump improvement 1cm, women high jump improvement by 4cm, men shot put decrease by 66cm, women hot put 1.75m, men discus improvement by 52cm and women discus decrease by 1.22m.

**Keywords:** Predicting athletic event performance, Olympic Games, mathematical modelling, linear and non-linear regression, curve estimation, men and women athletes.

## 1. INTRODUCTION

MATHEMATICAL MODELLING AND PREDICTING FUTURE ATHLETIC PERFORMANCE

The prediction of future athletic performance by athletes at the Olympic Games is a recurring theme, as well as forming the basis for some stimulating discussions on the limits of human performance. Mathematics in sport and exercise and sports science should be based upon substantial evidence, rather they are based on the belief that records are made to be broken and that performances based on past experiences must continue to improve over time. Obviously this concept has limits and at some point the limits of human performance based on human biology and physiology will be achieved. The accessibility of data in the form of results from Olympic Games, world records and world best performances in a specific year allows the analysis of performances in any number of events. From these analyses, changes in performance over time can be observed and predictions of future performance can be made utilising the process of mathematical extrapolation and interpolation. A number of researchers have attempted to predict future performances by deriving and applying a number of mathematical statistical models based on past performances in athletics (Heazlewood and Lackey, 1996, 1998; Heazlewood, 2006a, 2006b, 2011a, 2011b, 2013a, 2013b, Heazlewood & Walsh, 2014a, 2014b). Prendergast (1990) applied the average speeds of world record times to determine a mathematical model for world records. The records or data used in the analysis spanned a 10 year period. Following his analysis, Prendergast (1990) raised the question of whether any further improvements can be expected or if the limits of human performance have been reached. The sports of athletics (Heazlewood and Lackey, 1996, 1998; Heazlewood, 2006a, 2006b, 2008, 2010, 2011a, 2011b, 2013a, 2013b, Heazlewood & Walsh, 2014a, 2014b; Walsh & Heazlewood, 2014) and swimming (Heazlewood & Lackey, 1996; Lackey and Heazlewood, 1998; Heazlewood, 2006a, 2006b) have been addressed in this manner. Historically, the knowledge of future levels of sporting performance has been identified by Banister and Calvert (1980) as beneficial in the areas of talent identification, both long and short term goal setting, and training program development based on the next level of expected future performance. In addition, expected levels of future performance are often used in the selection of national representative teams where performance criteria are explicitly stated in terms of athletics times and distances for example as required entry or qualification standards that are set at each Olympic Games and by World Athletics for entry at the Olympic Games (International Olympic Committee (IOC), 2024). A set of qualification standards and criteria are set by World Athletics prior to the Olympic Games (McAlister, 2022).

Péronnet and Thibault (1989) postulate that some performances, such as the men's 100m sprint is limited to the low 9 seconds, whereas, Seiler (referred to by Hopkins, 2000) envisages no limits on improvements based on data reflecting progression of records over the last 50 years. According to Seiler improvements per decade have been approximately 1% for sprinting, 1.5% for distance running, 2-3% for jumping, 5% for pole vault, 5% for swimming and 10% for skiing for male athletes, whereas female sprint times may have already peaked. The differences for males and females it is thought to reflect the impact of successful drugs in sport testing on females. Previous derived curve estimations that significantly fit the data have also displayed interesting findings as no one curve fits all the data sets.

Over the long term 1924 to 2020, Olympic event performances have improved incrementally sprint events in and significantly in jump and throwing events in the men's and women's 100m, 200m, 400m long jump, high jump, shot put and discus, which will be the Olympic events evaluated in this research (Olympic Results, International Olympic Committee (2024). The significant improvements in the jump and throwing events can be attributed to technical innovations as the hitch kick in long jump, Fosbury flop in high jump and rotational throwing technique in the shot put and discus, combined with increased strength and power training and increased mass and height of throwers adding to technique.

For men from 2004 to 2024 Olympics actual performances by the mean score of the top six finalist in each event (to represent overall athlete performance change) indicates improvements of 8.5 % from 10.82s to 9.90s men 100m (1924-2020), 9.5% from 21.92s to 19.83s men 200m (1924-2020), 8.5% from 48.35s to 44.24s men 400m (1924-2020), 16.22% from 7.09m to 8.24m men long jump (1924-2020), 24.87% from 1.89m to 2.36m men high jump (1924-2020), 34% from 14.59m to 22.12m men shot put (1924-2020), and 33.57% from 44.38m to 66.81m men discus (1924-2020).

For the women 11.95% from 12.30s to 10.83s women 100m (1928-2020), 12.96% from 25.22s to 21.95s women 200m (1948-2020), 8.65% from 54.09s to 49.41s women 400m (1964-2020), 25.32% from 5.53m to 6.93m women long jump (1948-2020), 25.16% from 1.59m to 1.99m women high jump (1928-2020), 33.40% from 13.06m to 19.61m women shot put (1948-2020), and 44.8% from 36.25m to 65.70m women discus (1928-2020).

## TRENDS IN CURVE ESTIMATION AND MATHEMATICAL FUNCTIONS

The curve estimation regression functions derived by Heazlewood and Lackey (1996) to predict performance in the men's and women's produced mix results across the sprint and jump events. In athletics for the men's events the mathematical functions (Heazlewood and Lackey, 1996) that displayed the best goodness of fit were 100m inverse, 400m sigmoidal, long jump cubic and the high jump displayed four functions (compound, logistic, exponential and growth). In the women's events the regression functions that displayed the best goodness of fit were 100m cubic, 400m sigmoidal, long jump inverse and the high jump displayed four functions (compound, logistic, exponential and growth). The curves that fit the data have displayed interesting findings as no one curve fits all the data sets and men and women displayed some different functions linked to gender as men's 100m was inverse whereas women's was cubic, men's long jump cubic and women's long jump inverse. The function for the 400m and high jump were the same for men and women. This may indicate that different events and gender responses are dependent upon different factors that are being trained differently or factors underpinning performance evolving in differences in men and women exercise physiology and biomechanics. This has resulted in different curves or mathematical functions that reflect these improvements in training or phylogenetic changes over time. However, at some point in the future when time catches-up with the actual performance, then how accurately the predictive models reflect reality can be assessed.

The Heazlewood (2006a, 2006b) study to a degree replicated the findings in the 1996 study (Heazlewood & Lackey, 1996), where the functions and trends for men and women in the 100m, 400m, long jump and high jump were very similar. The study was evaluating the predictions of the 2000 Sydney and 2004 Athens Olympic

Games. It is interesting to observe the models derived in the 2006a and 2006b research moderately overestimated performance for men's and women's 100m although actual times did improve from 2000 to 2004, overestimated performance in the men's and women's 400m, moderately overestimated performance in the men's long jump and significantly in the women's and moderately overestimated performance in the men's and women's high jump. Table 1 indicates the athletic events, type of mathematical function, equations and $R^2$ values from the Heazlewood (2006a, p.85).

| Event | Regression Type | Equation * | $R^{2**}$ |
|---|---|---|---|
| Men's 100m | Inverse | $Y = b_0 + (b_1/t)$ | 0.65890 |
| Women's 100m | Cubic | $Y = b_0 + b_3t^3$ | 0.90175 |
| Men's 400m | S | $Y = e^{(bo + b1/t)}$ | 0.90676 |
| Women's 400m | S | $Y = e^{(bo + b1/t)}$ | 0.84251 |
| Men's Long Jump | Cubic | $Y = b_0 + b_3t^3$ | 0.77735 |
| Women's Long Jump | Inverse | $Y = b_0 + (b_1/t)$ | 0.89389 |
| Men's High Jump | Compound<br>Logistic<br>Exponential<br>Growth | $Y = b_0(b_1)^t$<br><br>$Y = b_0e^{b1t}$<br>$Y = e^{bob1t}$ | 0.94390<br>0.94390<br>0.94390<br>0.94390 |
| Women's High Jump | Compound<br>Logistic<br>Exponential<br>Growth | $Y = b_0(b_1)^t$<br><br>$Y = b_0e^{b1t}$<br>$Y = e^{bob1t}$ | 0.94665<br>0.94665<br>0.94665<br>0.94665 |

Table 1: Athletic events, type of mathematical function, equations and $R^2$ values.

** All $R^2$ values in table 1 significant at $p<0.05$.
* Where $b_0$ = a constant
$b_1$, $b_3$ = regression coefficients
t = year
y = mean result for each event

However, the ability to predict performances at the 2000, 2004, 2008 and 2012 Olympic Games for the men's 100m, 400m, long jump and high jump, based on the 1924 to 2012 data, were moderately accurate with low to moderate residual error.

The prediction performance in the throwing events of shot put and discus for men and women (Heazlewood & Walsh, 2014a, 2014b) produced results different from the earlier sprint and jump findings. Note the 2014a and 2014b studies evaluated all throwing events, the shot, discus, javelin and hammer, however the focus will be the findings for the shot put and discus to be consistent with this current research. The International Olympic Committee (IOC) athletic results was the source for the data set (IOC, 2014) when this study was completed and produced descriptive data and descriptive graphical analysis to indicate trends in World records for the men's throwing events, however the analyses were not based on mathematical predictive modelling. The data set in this research for men was 1936 to 2012 for shot put and discus, and for women was 1948-2012 for shot put and 1928-2012 for discus. The curve estimation regression functions for the men both for shot put and discus were cubic functions with high explained variance (shot put $R^2= .965$ and discus $R^2= .972$) and low residual error with inflexion points both suggesting performance decreases in 2016 and 2020 and subsequently to confirmed for the discus in 2016 and 2020, however not for the shot put as the 2016 prediction was accurate, however the 2020 result displayed a significant improvement in performance and not predicted in the cubic model.

The women's shot put and discus conformed to a cubic function as the best fit to the data and similarly displayed high explained variance (shot put $R^2= .918$ and discus $R^2= .925$) and low residual error. Once again, inflexion points were suggesting performance declines in these two events in the near future and this is what occurred in both events at the 2016 and 2020 Olympic. What is interesting based the results are shot put and discus events for both men and women competing at the Olympics displayed cubic functions that predicted performance plateaux and then performance declines and suggesting in these events the men and women shot put and discus throwers are approaching their performance limits.

The dominant and important research question is, can mathematical models based on nonlinear curve estimation, which have proven to be very successful in fitting and predicting past, current and future Olympic performances for event finals in athletics and swimming for men display equal effectiveness in predicting athletic performances in the men's throwing events of the shot put, discus, hammer and javelin at the Olympic Games, which are past, current and future performances?

HOW TO QUALIFY FOR ATHLETICS AT PARIS 2024

The review of literature has focussed on past research linked to evaluating performance change from actual performance to derive mathematical models to extrapolate and then predict future performance using previous Olympic finalists' performances in sprint, jump and throws athletic events from 1924-2016. This analysis overlaps with other methods that are utilised to set qualification standards for future Olympics, such as 2024 Paris Olympics. In this context selecting performance standards or methods to control the number of athletes that will qualify for Olympic athletic events. The number of athletes selected to compete is limited and based on a quota system. The actual standards change based on previous years performances in certified World Athletic competitions to achieve the quotas per event. If overall athletic performance standards improve then the qualification standards are increased to maintain the quota. Table 2 indicates the entry qualification standard, number of athletes permitted per event, number of athletes qualified by 9/5/2024 and number of Australian athletes qualified by 9/5/2024 for Paris 2024 based on actual event performances of men and women for 100m, 200m, 400m, high jump, long jump, shot put and discus. Table 3 indicates the changing qualification standard for the men' and women's 100m from 2000 to 2024 Olympics and notice the variation in time for the women athletes nowhere the time might increase or decrease, where time for the men is constantly decreasing to maintain the allocated quota of athletes. The quota system linked to qualification standard is a method of predicting performance standards for athletes although not sated in any mathematical model just based on descriptive pragmatism to maintain the quotas.

| Event | Men | Women | Number per event | Number qualified by entry standard 9/5/2024 | Australian qualified athletes 9/5/2024 |
|---|---|---|---|---|---|
| 100m | 10.00s | 11.07s | 56 | 19 men<br>20 women | none<br>none |
| 200m | 20.16s | 22.57s | 48 | 22 men<br>19 women | none<br>none |
| 400m | 45.00s | 50.95s | 48 | 34 men<br>24 women | 1<br>none |
| High Jump | 2.33m | 1.97m | 32 | 9 men<br>11 women | none<br>2 |
| Long Jump | 8.27m | 6.86m | 32 | 11 men<br>14 women | 1<br>none |
| Shot Put | 21.50m | 18.80m | 32 | 15 men<br>14 women | none |
| Discus | 67.20m | 64.50m | 32 | 17 men<br>14 women | none |

Table 2: Entry qualification standard, number of athletes permitted per event, number of athletes qualified by 9/5/2024 and number of Australian athletes qualified by 9/5/2024 for Paris 2024 based on actual event performances of men and women for 100m, 200m, 400m, high jump, long jump, shot put and discus.

Note

1. Entry standards for each event are set by World Athletics who control athletics at the Summer Olympic Games prior to games.
2. 78 days to Paris Olympic Games 26 July - 11 August 2024 from 9 May 2024.
3. The qualification and ranking period for all individual events - other than the 10,000m, marathon, combined events and race walks - will be between 1 July 2023 and 30 June 2024.
4. Qualification entry standard period ends 30 June 2024, which gives athletes 52 days.

| Year | Men | Women |
|---|---|---|
| 2000 | 10.27 | 11.40 |
| 2004 | 10.21 | 11.30 |
| 2008 | 10.21 | 11.32 |
| 2012 | 10.18 | 11.29 |
| 2016 | 10.16 | 11.32 |
| 2020 | 10.05 | 11.12 |
| 2024 | 10.00 | 11.07 |

Table 3: The qualification standard for men's and women's 100m for each Olympic year 2000-2024

Finally, the complete selection system is as follows. Athletes will be able to qualify in two ways for the Paris 2024 Games, with approximately 50% of qualification places based on achieving the entry standard for an event within the qualification period, and the other approximate 50% based on the World Athletics Ranking within the ranking period. The qualification and ranking period for all individual events - other than the 10,000m, marathon, combined events and race walks - will be between 1 July 2023 and 30 June 2024 (McAlister, 2022). The exact numbers based on the qualification standard and qualification by ranking is explained in detail for the 100m, 200m, 400m, long jump, high jump, shot put and discus and the detail is provided by World Athletics (2024) "Rode to Paris" guidelines. The following example provides a detailed example for the women's 100m. So making the qualification standard does not guarantee a berth in the women's 100m quota of 56. For example the USA is only allowed three athletes where they may have more than three times that number who have qualified. In this USA context to qualify for the USA team they usually select the first three athletes from the finals at the USA Olympics Trials. As a consequence athletes would require a place at the USA trials as well as achieving the qualification standard. The predictive models for future competitions as the Olympics combined with the pragmatism of World Athletics criteria to be selected provides performance goals to train and aspire to. The criteria for women sprinters to qualify for 100m for 2024 Paris.

Entry number: 56 - Qualification period for entry standard: 01 July 2023- 30 June 2024.
Entry standard: 11.07
World rankings period: 01 July 2023 - 30 June 2024.
Maximum quota per country is 3 athletes per event.
Number of athletes qualified by:

- By entry standard: 20
- By finishing position at designated competitions: 0
- By world rankings position: to complete the required entry number: 36
- By top list: 0
- By universality places: 0

QP: quota place in event (counts max 3 per country) CP: country place in event -in bold first 3** per country (World Athletics, 2024).

According to the World Athletics Annual Report 2022 (released August 2023) there were 180,000 elite athletes active in 2022 and these athletes will be competing first to qualify for the Olympics and then competing to reach the final and then competing for the Olympic medals. To achieve these outcomes the standards the athletes will have to pursue are the standards suggested by the predictive mathematical models for their respective events, and these standards are higher than the qualifying standards for the 2024 Paris Olympics by World Athletics. Predictive modelling for future Olympic athletic events can provide performance goals to aspire to.

## 2. RESEARCH PROBLEM, QUESTIONS AND HYPOTHESES

RESEARCH PROBLEM

The previous research that attempted to predict future athletic performance was based on linear regression and non-linear models in in the sport of athletics most curve estimation regression techniques found to be non-linear (Heazlewood and Lackey 1996; Heazlewood and Lackey, 1998; Heazlewood, 2006a, 2006b; Heazlewood &

Walsh, 2011a; Heazlewood & Walsh, 2011b; Heazlewood & Walsh, 2014a; Heazlewood & Walsh, 2014b). The research problem is the models developed in earlier research over 10 years ago still apply to current athletic events and enable prediction of future Olympic athletic events such as Paris 2024, Los Angeles 2028 and Brisbane 2032, which will enable athletes to set realistic training and performance goals based on future performance standards to qualify for athletic events and aspire to achieving Olympic finals, which is where the medals are contested? The higher quality of the predictive models will enable higher quality predictions and the principle of sport science and mathematics and computers in sport is to understand and predict sports performance.

RESEARCH QUESTIONS

A number of research questions can be generated that will address the predominant theoretical issues. These are:

1. What is the best mathematical-statistical model that will best fit the data for each event?
2. Does one model fit all events?
3. Will the mean score based on the top six (finalists) better reflect actual changes in human performance that have occurred in the past 100 years in men and 96 years in women?
4. Will the models be identical for the different genders for each event?
5. Will the models be gender specific as well as event specific?
6. Will the mathematical-statistical models enable the accurate prediction of future performances?
7. How much variance will be explained in the different models for each events for each gender?
8. Will the derived mathematical-statistical models result in realistic or absurd predictions that have occurred in previous research of this type?

HYPOTHESES

While prior research has used many physiological factors to predict performance, the aim of this paper is to predict performances primarily at the 2024 Paris Olympic Games, as well as the 2028 Los Angeles Olympic Games and 2032 Brisbane Olympic Games using only easily accessible data, that is, the results of past Olympic finalists. The task of this descriptive research is to determine if there are significant mathematical relationships between the mean of the finalists of different Olympic events over the years. Further to this, it is generally hypothesised that the significant relationship can be used in equations to accurately predict the average times sprint events, lengths and heights of jumps and distances in throwing events in specific Olympic events. The specific events that will be predicted are the men's and women's 100m, 200m, 400m, long jump, high jump, shot put and discus.

The following hypotheses will be formally stated as the positive or alternate hypotheses. The null hypotheses are assumed to be true only for the tests of statistical significance, as statistical tests essentially test the validity of the null hypotheses.

1. The different events will have different models of best fit as they will be dependent on different human factors such as speed, power, endurance, speed endurance, strength and motor co-ordination that are represented by different mathematical-statistical functions.
2. The different genders will have different models of best fit as a result of the different number of years that females have been competing in the different events and as a result of females only recently adopting more stressful physical training programs that will influence the response curves (training and competition performance) over time.
3. The mathematical-statistical models will be both event and gender specific.
4. The mean score for the first six Olympic finalist will more realistically reflect human performance and inferred ability in each event than previous predictive models.
5. The different mathematical models derived will provide a more accurate prediction of future performances both for short time and long time periods.
6. More of the variance will be explained and a better model fit will occur based on nonlinear mathematical functions.
7. The likelihood of absurd mathematical predictions will be reduced or non-existent.

## 3. METHODS

DATA SOURCES

The data sources and data sets were the International Olympic Committee (Olympics, 2024) website, which provides Olympic Games results from all the modern Summer and Winter Olympic Games for all events including athletic results for men and women. An additional data resource is the Wikipedia (2024) website Athletics at the Summer Olympics. This website links to individual Olympic athletic results competed at all Modern Olympic Games from heats and qualifying rounds through to Olympic finals in the events utilised in this research. The events selected for mathematical modelling and prediction were the men's and women's

100m, 200m, 400m, long jump, high jump, shot put and discus as representing a traditional mix of sprinting, jumping and throwing events. The performances, in times or distance as appropriate, for athletes finishing first to sixth were utilised for curve estimation analyses. The sprint times were entered to the hundredth of seconds and distances for jumping and throwing events to the centimetre as provided by the data sources. It is important to note the 1952 Helsinki Olympics were the first Games to employ photo-finish cameras that could record athletes' times simultaneously (Olympics, 2024), as traditional timing techniques had proved unreliable. From 1977 world records in track events will only be accepted by World Athletics if calibrated photo finish-timing apparatus is applied (World Athletics, 2024).

The events selected did not have the same timeframes as women' track and field events were added after the men's events. The data base from the men's events were from 1924-1936 and 1948-2020. Women 100m data was 1928-1936 and 1948-2020; 200m was 1948-2020; 400m was 1964-2020; high jump was 1928-1936 and 1948-2020; long jump was 1948-2020; shot put was 1948-2020 and discus was 1928-1936 and 1948-2020.

The actual 2020 Tokyo Olympics were postponed to 2021 due to Covid19. No Summer Olympics were conducted in 1940 and 1944 due to World War II. There were Olympic boycott of the Moscow in 1980 where some western nations, such as USA, refused to compete and Los Angeles 1984 where eastern bloc nations, such as USSR, refused to compete. This has resulted in disjointed performances at the Olympics as well as the women's 200m, 400m, long jump and shot put having far fewer Olympic competitive years. These disruptions and later inclusions of events for women may influence some historical trends of performance changes over time. The influence of sports doping in the 1970' and 1980's as mentioned in the review of literature may mask the actual trends that occurred without doping.

In athletics assistive and resistive winds are thought to influence performance in such events as the 100m and long jump and the wind variable can be corrected to assess performance in still air conditions. These wind correction calculations are not presented in this paper just the times and distances reported for the athletic events, however correcting for the influence of wind may result in slightly different values for the original data.

Prior to 1952 non-electronic and less accurate methods, such as hand timing, was applied and differences in sprint times can be approximately 0.2 seconds different when comparing hand timing to electronic timing as hand timing has lower. An example is an electronic time of 10.00s for the 100m would be 9.80s using hand time and this error has not been corrected for with the hand timed performances prior to 1952. This may result in minor errors in the mathematical model

MODEL SELECTION AND DATA FIT CRIERIA

The data set was analysed utilising IBM SPSS Statistics-Version 26 (IBM, 2024). The following eleven curve estimation regression mathematical models were applied. An identical approach utilised in the evaluation of mathematical models that evaluate performances changes over time with Olympic Games (Heazlewood & Lackey, 1996; Lackey & Heazlewood, 1998; Heazlewood, 2006a, 2006b; Heazlewood, 2013a; Heazlewood & Walsh, 2014a; Walsh & Heazlewood; 2014b), World Athletic Championships (Heazlewood; 2013b) and decline in Masters Athletic performance with age in track and field events (Heazlewood & Lackey, 1998).

1. Linear - equation is Y = b0 + (b1 * t). The series values are modelled as a linear function of time. (1)

2. Logarithmic - equation is Y = b0 + (b1 * ln(t)). (2)

3. Inverse - equation is Y = b0 + (b1 / t). (3)

4. Quadratic - equation is Y = b0 + (b1 * t) + (b2 * t**2). The quadratic model can be used to model a series that "takes off" or a series that dampens. (4)

5. Cubic - is defined by the equation Y = b0 + (b1 * t) + (b2 * t**2) + (b3 * t**3). (5)

6. Power - equation is Y = b0 * (t**b1) or ln(Y) = ln(b0) + (b1 * ln(t)). (6)

7. Compound - equation is Y = b0 * (b1**t) or ln(Y) = ln(b0) + (ln(b1) * t). (7)

8. S-curve - equation is Y = e**(b0 + (b1/t)) or ln(Y) = b0 + (b1/t). (8)

9. Logistic - equation is Y = 1 / (1/u + (b0 * (b1**t))) or ln(1/y-1/u) = ln (b0) + (ln(b1) * t) where u is the upper boundary value. After selecting Logistic, specify the upper boundary value to use in the

regression equation. The value must be a positive number that is greater than the largest dependent variable value. (9)

10. Growth - equation is Y = e**(b0 + (b1 * t)) or ln(Y) = b0 + (b1 * t). (10)

11. Exponential - is Y = b0 * (e**(b1 * t)) or ln(Y) = ln(b0) + (b1 * t). (11)

The following criteria for model selection were applied to select the best fitted model with the data sets.

GENERAL METHOD OF DETERMINING THE APPROPRIATE REGRESSIONS MODELS
To investigate the hypotheses of model fit and prediction, the eleven regression models were individually applied to each of the athletic and swimming events. The regression equation that produced the best fit for each event, that is, produced the highest coefficient of determination (abbreviated as $R^2$), was then determined from these eleven equations. The specific criteria to select the regression equation of best were the magnitude of $R^2$, the significance of the analysis of variance alpha or p-value and the residuals.

The Coefficient of Determination

The coefficient of determination ($R^2$) is a measure of accuracy of the model used. A coefficient of determination of 1.00 indicates a perfectly fitting model where the predicted values match the actual values for each independent variable (Norušis, 1993; Garson, 2010; Hair et al., 2006). Where more than one model was able to be selected due to an equal $R^2$, the simplest model was used under the principle of parsimony, that is, the avoidance of waste and following the simplest explanatory model.

Residuals

The residuals are the difference between the actual value and the predicted value for each case, using the regression equation (Norušis, 1993; Garson, 2010; Hair et al., 2006) and the smaller the residual, the better the fit of the model. For each model the residuals were generated by the SPSS program. A large number of positive residuals indicate that the prediction is an over estimation (faster than the actual performance) and a large number of negative residuals indicates an underestimation (slower time than the actual performance).

Level of Significance

The level of significance, or p value, is a representation of the relationship between the model and the data. The smaller the p value, the higher the level of significance and the greater the relationship where a small p value indicates a small possibility that the closeness of the predicted values to the actual values due to chance is small.

Logical Acceptance Based on Extrapolations

The ability of the model to generate extrapolations that appear to be reasonable when compared to previous means was also taken into consideration. When a model generated extrapolations that appear to be inconsistent with the actual results this model was discarded and the model with the next highest coefficient of determination was selected. The models and mathematical function curve estimations finally selected were based on the preceding model evaluation criteria.

Applying the Model of Best Fit

After selection of the model to be used, according to the criteria previously stated, the equation of best fit was determined by applying the derived constants and coefficients to the generic formula for that model. Using this equation, a prediction of the mean result for the event at each Olympiad was calculated. At this stage, graphs representing the means of past and future performances for each event in each Olympiad were also generated in addition to predicted means using the appropriate regression equation. Final predictions have included prediction for Paris 2024, Los Angeles 2028 and Brisbane 2032 to extend performance predictions into the Olympic Games in near future. To predict the level of performance in the year 2024, 2028 and 2032 the model fit that provided the greatest accuracy was chosen. A series of regressions were made using the best fitting model and data set for each event. Using the constants and coefficients generated by regression models the future predictions were then calculated.

## 4. RESULTS

The results for the curve estimation models for the different events are displayed in table 4. The men's model of best was for the cubic function for all events, which are the 100m, 200m, 400m, long jump, high jump, shot put and discus. This indicates a consistent mathematical trend across all these events in terms of performance change over time. The high $R^2$ values range from 0.821 for the men's long jump to 0.954 for the men's discus.

The women's events indicate a cubic function for 100m, high jump, and discus; a sigmoidal function for 200m, 400m and shot put and an inverse function for long jump. The $R^2$ values for the women's events are lower ranging from 0.517 for the women's shot put, 0.578 for 400m, 0.691 for 200m, 0.751 for long jump, 0.913 for discus 0.923 for 100m to 0.936 for the women's high jump. The higher $R^2$ values for the women's events are for events that have been contested over a longer time frame at the Olympic Games. The women's events with the lower $R^2$ values were introduced after 1948 and the 400m as late as 1964.

| Event | Regression Type | Equation * | R2** |
|---|---|---|---|
| Men's 100m | Cubic | $Y = b_0 + b_3t^3$ | 0.842 |
| Women's 100m | Cubic | $Y = b_0 + b_3t^3$ | 0.923 |
| Men's 200m | Cubic | $Y = b_0 + b_3t^3$ | 0.942 |
| Women's 200m | Sigmoidal | $Y = e^{(bo + b1/t)}$ | 0.696 |
| Men's 400m | Cubic | $Y = b_0 + b_3t^3$ | 0.922 |
| Women's 400m | Sigmoidal | $Y = e^{(bo + b1/t)}$ | 0.578 |
| Men's Long Jump | Cubic | $Y = b_0 + b_3t^3$ | 0.821 |
| Women's Long Jump | Inverse | $Y = b_0 + (b_1/t)$ | 0.751 |
| Men's High Jump | Cubic | $Y = b_0 + b_3t^3$ | 0.911 |
| Women's High Jump | Cubic | $Y = b_0 + b_3t^3$ | 0.936 |
| Men's Shot Put | Cubic | $Y = b_0 + b_3t^3$ | 0.931 |
| Women's Shot Put | Sigmoidal | $Y = e^{(bo + b1/t)}$ | 0.517 |
| Men's Discus | Cubic | $Y = b_0 + b_3t^3$ | 0.954 |
| Women's Discus | Cubic | $Y = b_0 + b_3t^3$ | 0.913 |

Table 4: Events, Derived Mathematical Functions, Equations and $R^2$ Values.

** All $R^2$ values in table 1 significant at $p<0.001$.
* Where b0 = a constant
b1, b3 = regression coefficients
t = year
Y = mean result for each event

Table 5 indicates actual and predicted times for male and female 100m sprint for 1924 based on known data to predicting by extrapolation to expected mean time for the top six finalists for Paris - 2024, Los Angeles - 2028 and Brisbane 2032 Olympics. The high $R^2$=0.842 for men and $R^2$=0.923 for women indicates the curve estimation based on the cubic model is accurate by comparing the actual time to predicted time and the small values for the residuals. The women's model is more accurate and the goodness of fit is excellent as represented by the close fit of the actual and predicted performance times. The predicted time for men in Paris is 9.82s a marginal improvement on the 9.90s in Tokyo and predicted time for women is 10.80s a marginal improvement on the 10.83s in Tokyo. Marginal improvements in 100m time for both men and women are also predicted for Los Angeles - 2028 and Brisbane 2032 Olympics.

| Year | Men 100m Actual Time(s) Cubic $R^2$=0.842 p<.001 | Men 100m Predicted Time(s) | Men 100m Error Time(s) | Women 100m Actual Time(s) Cubic $R^2$=0.923 p<.001 | Women 100m Predicted Time(s) | Women 100m Error Time(s) |
|---|---|---|---|---|---|---|
| 1924 | 10.82 | 10.76 | .06 | na | na | na |
| 1928 | 10.93 | 10.72 | .21 | 12.30 | 12.27 | .03 |
| 1932 | 10.48 | 10.68 | -.20 | 12.06 | 12.17 | -.11 |
| 1936 | 10.57 | 10.63 | -.06 | 11.93 | 12.08 | -.15 |
| 1948 | 10.43 | 10.51 | -.08 | 12.10 | 11.82 | .28 |
| 1952 | 10.43 | 10.47 | -.04 | 11.87 | 11.74 | .13 |
| 1956 | 10.60 | 10.43 | .17 | 11.77 | 11.66 | .11 |
| 1960 | 10.28 | 10.39 | -.11 | 11.42 | 11.59 | -.17 |
| 1964 | 10.31 | 10.35 | -.04 | 11.66 | 11.51 | .15 |
| 1968 | 10.07 | 10.31 | -.24 | 11.26 | 11.44 | -.18 |
| 1972 | 10.33 | 10.27 | .06 | 11.45 | 11.38 | .07 |
| 1976 | 10.21 | 10.23 | -.02 | 11.23 | 11.32 | -.09 |
| 1980 | 10.39 | 10.20 | .19 | 11.19 | 11.26 | -.07 |
| 1984 | 10.24 | 10.16 | .08 | 11.29 | 11.20 | .09 |
| 1988 | 10.34 | 10.12 | .22 | 10.99 | 11.14 | -.15 |
| 1992 | 10.10 | 10.09 | .01 | 10.96 | 11.09 | -.13 |
| 1996 | 10.00 | 10.05 | -.05 | 11.00 | 11.05 | -.05 |
| 2000 | 10.03 | 10.02 | .01 | 11.18 | 11.00 | .18 |
| 2004 | 9.90 | 9.99 | -.08 | 10.99 | 10.96 | .03 |
| 2008 | 9.89 | 9.95 | -.06 | 10.99 | 10.92 | .07 |
| 2012 | 9.83 | 9.92 | -.09 | 10.84 | 10.88 | -.04 |
| 2016 | 9.91 | 9.89 | .02 | 10.85 | 10.85 | .00 |
| 2020 | 9.90 | 9.85 | .04 | 10.83 | 10.82 | .01 |
| 2024 | | 9.82 | | | 10.80 | |
| 2028 | | 9.79 | | | 10.77 | |
| 2032 | | 9.76 | | | 10.75 | |

Table 5: Actual and predicted times for male and female 100m sprint for 1924 to 2032 Olympics.

Note: 2024 - Paris, 2028 - Los Angeles, 2032 - Brisbane

Table 6 indicates actual and predicted times for male and female 200m sprint for 1924 based on known data to predicting by extrapolation to expected mean time for the top six finalists for Paris - 2024, Los Angeles – 2028 and Brisbane 2032 Olympics. The men's 200m conforms to a cubic model with high $R^2$=0.942 and the women's 200m conforms to a sigmoidal model with a lower $R^2$=0.696. One again, the high explained variance for the men 200m indicates a close between the actual and predicted 200m times and the residuals are small. The goodness of fit is not as good for the women's predictive function, which was sigmoidal and with higher residual values. The prediction for men is a marginally slower time where Tokyo 2020 was 19.86s, whereas Paris 2024 is predicted to be 19.92s. For the women a significant improvement is predicted where Tokyo 2020 was 21.95s, whereas Paris 2024 is predicted to be 21.43s. The difference in quality of the models might be attributed to the women's 200m was introduced in 1948, therefore fewer data points and the influence of sports doping by eastern bloc nations producing greater prediction error.

| Year | Men 200m Actual Time(s) Cubic $R^2$=0.942 p<.001 | Men 200m Predicted Time(s) | Men 200m Error Time(s) | Women 200m Actual Time(s) Sigmoidal $R^2$=0.696 p<.001 | Women 200m Predicted Time(s) | Women 200m Error Time(s) |
|---|---|---|---|---|---|---|
| 1924 | 21.92 | 21.98 | -.06 | na | na | na |
| 1928 | 21.97 | 21.82 | .15 | na | na | na |
| 1932 | 21.52 | 21.67 | -.15 | na | na | na |
| 1936 | 21.37 | 21.52 | -.15 | na | na | na |
| 1948 | 21.30 | 21.12 | .18 | 25.22 | 24.15 | 1.07 |
| 1952 | 21.17 | 21.00 | .17 | 24.25 | 23.99 | .26 |
| 1956 | 21.18 | 20.89 | .29 | 23.92 | 23.83 | .09 |
| 1960 | 20.71 | 20.78 | -.07 | 24.73 | 23.68 | 1.05 |
| 1964 | 20.62 | 20.68 | -.06 | 23.35 | 23.53 | -.18 |
| 1968 | 20.24 | 20.58 | -.34 | 22.82 | 23.38 | -.56 |
| 1972 | 20.33 | 20.49 | -.16 | 22.67 | 23.23 | -.56 |
| 1976 | 20.54 | 20.41 | .13 | 22.60 | 23.08 | -.48 |
| 1980 | 20.36 | 20.33 | .03 | 22.37 | 22.94 | -.57 |
| 1984 | 20.23 | 20.26 | -.03 | 22.10 | 22.79 | -.69 |
| 1988 | 20.08 | 20.20 | -.12 | 21.88 | 22.65 | -.77 |
| 1992 | 20.34 | 20.14 | .20 | 22.17 | 22.51 | -.34 |
| 1996 | 19.89 | 20.09 | -.20 | 22.36 | 22.37 | -.01 |
| 2000 | 20.19 | 20.05 | .14 | 22.37 | 22.23 | .14 |
| 2004 | 20.06 | 20.01 | .05 | 22.40 | 22.09 | .31 |
| 2008 | 20.08 | 19.98 | .10 | 22.06 | 21.96 | .10 |
| 2012 | 19.78 | 19.95 | -.17 | 22.24 | 21.82 | .42 |
| 2016 | 20.06 | 19.94 | .12 | 22.11 | 21.69 | .42 |
| 2020 | 19.86 | 19.92 | -.06 | 21.95 | 21.56 | .39 |
| 2024 | | 19.92 | | | 21.43 | |
| 2028 | | 19.92 | | | 21.30 | |
| 2032 | | 19.93 | | | 21.17 | |

Table 6. Actual and predicted times for male and female 200m sprint for 1924 to 2032 Olympics.

Table 7 indicates actual and predicted times for male and female 400m sprint for 1924 to 2032 Olympics. In this event the men had a good model fit between the data and the derived predictive cubic function $R^2$=0.922 and residual error very small. The predicted performance over the 8 years indicate performance stability with minimal change in performance from 2020 to 2032 and maybe a slight performance decrease. Where Tokyo 2020 was 44.24s for men and predicted to be 44.37s for Paris 2024. Whereas the women's model was once again a sigmoidal function with significantly lower $R^2$=0.578 and larger residual error. The 400m event is a late addition to the track program as the event was not competed from 1964 onwards and the first 400m in Tokyo 1964 was won by Betty Cuthbert from Australia. It is interesting to note that in women's 400m Paris 2024 a significant increase in performance is predicted has where Tokyo was 49.41s Paris 2024 is expected to be 48.79s with this improvement trend continuing to Brisbane 2032.

Table 8 indicates actual and predicted height for male and female high jump for 1924 to 2032 Olympics. The event was first competed in by women 1928 so data points over time are comparable with men. Both men ($R^2$=0.911) and women ($R^2$=0.936) displayed the best fit with a cubic function with high explained variance as well as the small residual error. In the context of men high jumpers performance stability is predicted over the next 8 years with only a 1cm improvement where Tokyo 2020 was 2.36m Paris 2024 is expected to be 2.37m. The women are also expected to display a marginal increase in jump height from Tokyo 2020 1.99m to Paris 2024 to 2.02m.

Table 9 actual and predicted distance for men and women long jump for 1924 to 2032 Olympics. Once again the curve estimation function was a cubic function with high explained variance $R^2$=0.821 with small residual error, like all other functions derived for the men events in this study. The women's long jump was an inverse $R^2$=0.751 with lower percentage of explained variance and larger residual error and once again, the data points to derive the relationship were fewer for women as the event was first contested in 1948. For men the prediction for Paris 2024 of 8.32m compared to Tokyo of 8.24m. From 2024 to 2032 the prediction is stable at approximately 8.32m-8.33m. The predictions for women for Paris 2024 are more significant at 7.18m with steady increase to Brisbane 2032 at 7.36m, whereas Tokyo 2020 was 6.93.

| Year | Men 400m Actual Time(s) Cubic $R^2$=0.922 p<.001 | Men 400m Predicted Time(s) | Men 400m Error Time(s) | Women 400m Actual Time(s) Sigmoidal $R^2$=0.578 p<.001 | Women 400m Predicted Time(s) | Women 400m Error Time(s) |
|---|---|---|---|---|---|---|
| 1924 | 48.35 | 48.52 | -.17 | na | na | na |
| 1928 | 48.37 | 48.18 | .19 | na | na | na |
| 1932 | 47.53 | 47.85 | -.32 | na | na | na |
| 1936 | 47.13 | 47.54 | -.41 | na | na | na |
| 1948 | 47.47 | 46.69 | .78 | na | na | na |
| 1952 | 46.62 | 46.44 | .18 | na | na | na |
| 1956 | 47.32 | 46.20 | 1.12 | na | na | na |
| 1960 | 45.45 | 45.97 | -.52 | na | na | na |
| 1964 | 45.84 | 45.76 | .08 | 54.09 | 52.07 | 2.02 |
| 1968 | 45.11 | 45.57 | -.46 | 52.50 | 51.84 | .66 |
| 1972 | 45.16 | 45.39 | -.23 | 51.79 | 51.61 | .18 |
| 1976 | 45.06 | 45.22 | -.16 | 50.64 | 51.38 | -.74 |
| 1980 | 45.24 | 45.07 | .17 | 50.42 | 51.16 | -.74 |
| 1984 | 44.76 | 44.93 | -.17 | 49.98 | 50.93 | -.95 |
| 1988 | 44.51 | 44.81 | -.30 | 50.30 | 50.71 | -.41 |
| 1992 | 44.47 | 44.70 | -.23 | 49.79 | 50.49 | -.70 |
| 1996 | 44.46 | 44.61 | -.15 | 49.09 | 50.27 | -1.18 |
| 2000 | 44.73 | 44.53 | .20 | 49.70 | 50.05 | -.35 |
| 2004 | 44.47 | 44.46 | .01 | 49.84 | 49.84 | .00 |
| 2008 | 44.73 | 44.42 | .31 | 50.19 | 49.62 | .57 |
| 2012 | 44.56 | 44.38 | .18 | 49.86 | 49.41 | .45 |
| 2016 | 44.38 | 44.36 | .02 | 50.09 | 49.20 | .89 |
| 2020 | 44.24 | 44.36 | -.12 | 49.41 | 49.00 | .41 |
| 2024 | | 44.37 | | | 48.79 | |
| 2028 | | 44.40 | | | 48.58 | |
| 2032 | | 44.44 | | | 48.38 | |

Table 7. Actual and predicted times for male and female 400m sprint for 1924 to 2032 Olympics.

| Year | Men High Jump Actual Distance(m) Cubic $R^2$=0.911 p<.001 | Men High Jump Predicted Distance(m) | Men High Jump Distance(m) Error | Women High Jump Actual Distance(m) Cubic $R^2$=0.936 p<.001 | Women High Jump Predicted Distance(m) | Women High Jump Distance(m) Error |
|---|---|---|---|---|---|---|
| 1924 | 1.89 | 1.85 | .04 | na | na | na |
| 1928 | 1.90 | 1.88 | .02 | 1.52 | 1.48 | .04 |
| 1932 | 1.95 | 1.92 | .03 | 1.59 | 1.52 | .07 |
| 1936 | 1.97 | 1.95 | .02 | 1.58 | 1.56 | .02 |
| 1948 | 1.94 | 2.04 | -.10 | 1.60 | 1.66 | -.06 |
| 1952 | 1.97 | 2.07 | -.10 | 1.61 | 1.70 | -.09 |
| 1956 | 2.09 | 2.10 | -.01 | 1.67 | 1.73 | -.06 |
| 1960 | 2.09 | 2.12 | -.03 | 1.70 | 1.76 | -.06 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 1964 | 2.13 | 2.15 | -.02 | 1.76 | 1.78 | -.02 |
| 1968 | 2.17 | 2.17 | .00 | 1.79 | 1.81 | -.02 |
| 1972 | 2.18 | 2.19 | -.01 | 1.86 | 1.83 | .03 |
| 1976 | 2.20 | 2.21 | -.01 | 1.89 | 1.86 | .03 |
| 1980 | 2.27 | 2.23 | .04 | 1.92 | 1.88 | .04 |
| 1984 | 2.30 | 2.25 | .05 | 1.95 | 1.90 | .05 |
| 1988 | 2.35 | 2.27 | .08 | 1.96 | 1.92 | .04 |
| 1992 | 2.33 | 2.28 | .05 | 1.95 | 1.94 | .01 |
| 1996 | 2.35 | 2.30 | .05 | 1.99 | 1.95 | .04 |
| 2000 | 2.33 | 2.31 | .02 | 2.00 | 1.97 | .03 |
| 2004 | 2.34 | 2.32 | .02 | 1.99 | 1.98 | .01 |
| 2008 | 2.33 | 2.34 | -.01 | 2.00 | 1.99 | .01 |
| 2012 | 2.23 | 2.34 | -.11 | 2.00 | 2.01 | -.01 |
| 2016 | 2.34 | 2.35 | -.01 | 1.96 | 2.01 | -.05 |
| 2020 | 2.36 | 2.36 | .00 | 1.99 | 2.02 | -.03 |
| 2024 | | 2.37 | | | 2.03 | |
| 2028 | | 2.37 | | | 2.03 | |
| 2032 | | 2.37 | | | 2.04 | |

Table 8. Actual and predicted height for male and female high jump for 1924 to 2032 Olympics.

| Year | Men Long Jump Actual Distance(m) Cubic $R^2$=0.821 p<.001 | Men Long Jump Predicted Distance(m) | Men Long Jump Distance(m) Error | Women Long Jump Actual Distance(m) Inverse $R^2$=0.751 p<.001 | Women Long Jump Predicted Distance(m) | Women Long Jump Error Distance(m) |
|---|---|---|---|---|---|---|
| 1924 | 7.09 | 7.12 | -.03 | na | na | na |
| 1928 | 7.42 | 7.21 | .21 | na | na | na |
| 1932 | 7.22 | 7.29 | -.07 | na | na | na |
| 1936 | 7.69 | 7.37 | .32 | na | na | na |
| 1948 | 7.34 | 7.59 | -.25 | 5.53 | 6.01 | -.48 |
| 1952 | 7.26 | 7.66 | -.40 | 5.91 | 6.08 | -.17 |
| 1956 | 7.46 | 7.72 | -.26 | 5.98 | 6.15 | -.17 |
| 1960 | 7.86 | 7.78 | .08 | 6.18 | 6.21 | -.03 |
| 1964 | 7.63 | 7.84 | -.21 | 6.41 | 6.28 | .13 |
| 1968 | 8.17 | 7.89 | .28 | 6.54 | 6.34 | .20 |
| 1972 | 8.01 | 7.95 | .06 | 6.59 | 6.41 | .18 |
| 1976 | 8.01 | 7.99 | .02 | 6.56 | 6.48 | .08 |
| 1980 | 8.18 | 8.04 | .14 | 6.90 | 6.54 | .36 |
| 1984 | 8.10 | 8.08 | .02 | 6.72 | 6.60 | .12 |
| 1988 | 8.18 | 8.12 | .06 | 6.91 | 6.67 | .24 |
| 1992 | 8.22 | 8.16 | .06 | 6.85 | 6.73 | .12 |
| 1996 | 8.25 | 8.19 | .06 | 7.00 | 6.80 | .20 |
| 2000 | 8.33 | 8.22 | .11 | 6.81 | 6.86 | -.05 |
| 2004 | 8.36 | 8.24 | .12 | 6.96 | 6.93 | .03 |
| 2008 | 8.20 | 8.27 | -.07 | 6.81 | 6.99 | -.18 |
| 2012 | 8.15 | 8.29 | -.14 | 6.84 | 7.05 | -.21 |
| 2016 | 8.26 | 8.30 | -.04 | 6.99 | 7.11 | -.12 |
| 2020 | 8.24 | 8.32 | -.08 | 6.93 | 7.18 | -.25 |
| 2024 | | 8.33 | | | 7.24 | |
| 2028 | | 8.33 | | | 7.30 | |
| 2032 | | 8.34 | | | 7.36 | |

Table 9. Actual and predicted distance for male and female long jump for 1924 to 2032 Olympics.

Table 10. Actual and predicted distance for male and female shot put for 1924 to 2032 Olympics indicates a very good function fit for a cubic model $R^2$=0.931 and for men and a relatively poor predictive model based on a sigmoidal function $R^2$=0.517 and large residual error for women and this data set has fewer data points as the event for women was introduced at the London 1948 Olympics. The model for men appears realistic and the prediction for Paris 2024 is a slight decrease in performance to 21.46m when compared to Tokyo 2020 at 22.12m and stable performance from 2024 to 2032. The women's model does not appear to be realistic in terms of future predictions as Paris 2024 is predicted to be 21.74m, whereas Tokyo performance was 19.96m and predictions for Los Angeles 2028 is 22.09m and Brisbane 2032 is 22.44m. At his point in time the top four women throwers have thrown between 20.03m to 20.68m in 2024 rankings in the women's shot put and there are only 48 days for qualification in athletics for Paris 2024 and well below a mean score of 21.74m.

| Year | Men Shot Put Actual Distance(m) Cubic $R^2$=0.931 $p<.001$ | Men Shot Put Predicted Distance(m) | Men Shot Put Distance(m) Error | Women Shot Put Actual Distance(m) Sigmoidal $R^2$=0.517 $p<.001$ | Women Shot Put Predicted Distance(m) | Women Shot Put Error Distance(m) |
|---|---|---|---|---|---|---|
| 1924 | 14.59 | 14.14 | .45 | na | na | na |
| 1928 | 15.29 | 14.76 | .53 | na | na | na |
| 1932 | 15.56 | 15.35 | .21 | na | na | na |
| 1936 | 15.69 | 15.92 | -.23 | na | na | na |
| 1948 | 16.01 | 17.46 | -1.45 | 13.06 | 15.86 | -2.80 |
| 1952 | 16.90 | 17.92 | -1.02 | 14.33 | 16.14 | -1.81 |
| 1956 | 17.58 | 18.35 | -.77 | 15.79 | 16.42 | -.63 |
| 1960 | 18.51 | 18.75 | -.24 | 16.47 | 16.70 | -.23 |
| 1964 | 19.48 | 19.13 | .35 | 17.35 | 16.99 | .36 |
| 1968 | 19.91 | 19.48 | .43 | 18.39 | 17.28 | 1.11 |
| 1972 | 21.09 | 19.80 | 1.29 | 19.61 | 17.57 | 2.04 |
| 1976 | 20.74 | 20.09 | .65 | 20.61 | 17.87 | 2.74 |
| 1980 | 20.86 | 20.36 | .50 | 21.19 | 18.17 | 3.02 |
| 1984 | 20.75 | 20.60 | .15 | 19.02 | 18.47 | .55 |
| 1988 | 21.53 | 20.81 | .72 | 20.87 | 18.78 | 2.09 |
| 1992 | 20.91 | 20.99 | -.08 | 19.92 | 19.09 | .83 |
| 1996 | 20.81 | 21.15 | -.34 | 19.56 | 19.41 | .15 |
| 2000 | 20.99 | 21.28 | -.29 | 19.56 | 19.73 | -.17 |
| 2004 | 20.72 | 21.38 | -.66 | 19.06 | 20.05 | -.99 |
| 2008 | 20.87 | 21.45 | -.58 | 19.39 | 20.38 | -.99 |
| 2012 | 21.32 | 21.50 | -.18 | 19.70 | 20.72 | -1.02 |
| 2016 | 21.43 | 21.51 | -.08 | 19.78 | 21.05 | -1.27 |
| 2020 | 22.12 | 21.50 | .62 | 19.61 | 21.39 | -1.78 |
| 2024 | | 21.46 | | | 21.74 | |
| 2028 | | 21.39 | | | 22.09 | |
| 2032 | | 21.30 | | | 22.44 | |

Table 10. Actual and predicted distance for male and female shot put for 1924 to 2032 Olympics.

Table 11 displaying actual and predicted distance for male and female discus for 1924 to 2032 Olympics indicates very good cubic predictive functions for both men ($R^2$=0.954 and low residual error for 1996 to 2020) and women ($R^2$=0.913 and low residual error for 2012 to 2020) discus throwers. For men the prediction is marginal increase in performance from 66.81m Tokyo 2020 to 67.33m Paris 2024 with performance stability from 2024 to 2032. For women the prediction for Paris 2024 is a marginal decrease 64.48m from Tokyo 2020 which was 65.70. Both these prediction for men and women discus throwers, based on the evidence, appear highly plausible.

| Year | Men Discus Actual Distance(m) Cubic $R^2$=0.954 p<.001 | Men Discus Predicted Distance(m) | Men Discus Distance(m) Error | Women Discus Actual Distance(m) Cubic $R^2$=0.913 p<.001 | Women Discus Predicted Distance(m) | Women Discus Error Distance(m) |
|---|---|---|---|---|---|---|
| 1924 | 44.38 | 42.85 | 1.53 | na | na | na |
| 1928 | 46.08 | 44.91 | 1.17 | 36.25 | 32.47 | 3.78 |
| 1932 | 47.86 | 46.89 | .97 | 37.30 | 35.87 | 1.43 |
| 1936 | 48.92 | 48.78 | .14 | 40.86 | 39.09 | 1.77 |
| 1948 | 50.01 | 53.91 | -3.90 | 40.07 | 47.69 | -7.62 |
| 1952 | 52.77 | 55.44 | -2.67 | 45.79 | 50.20 | -4.41 |
| 1956 | 54.07 | 56.87 | -2.80 | 50.76 | 52.53 | -1.77 |
| 1960 | 56.72 | 58.22 | -1.50 | 52.37 | 54.68 | -2.31 |
| 1964 | 59.09 | 59.48 | -.39 | 56.37 | 56.65 | -.28 |
| 1968 | 62.36 | 60.65 | 1.71 | 55.29 | 58.45 | -3.16 |
| 1972 | 63.19 | 61.72 | 1.47 | 63.80 | 60.06 | 3.74 |
| 1976 | 65.18 | 62.71 | 2.47 | 66.80 | 61.49 | 5.31 |
| 1980 | 65.77 | 63.60 | 2.17 | 67.14 | 62.74 | 4.40 |
| 1984 | 65.56 | 64.40 | 1.16 | 62.97 | 63.82 | -.85 |
| 1988 | 66.97 | 65.11 | 1.86 | 69.67 | 64.70 | 4.97 |
| 1992 | 63.75 | 65.73 | -1.98 | 66.24 | 65.41 | .83 |
| 1996 | 66.19 | 66.25 | -.06 | 66.11 | 65.94 | .17 |
| 2000 | 67.57 | 66.69 | .88 | 65.26 | 66.28 | -1.02 |
| 2004 | 66.23 | 67.03 | -.80 | 65.65 | 66.44 | -.79 |
| 2008 | 67.32 | 67.27 | .05 | 62.13 | 66.41 | -4.28 |
| 2012 | 67.48 | 67.43 | .05 | 66.04 | 66.21 | -.17 |
| 2016 | 66.60 | 67.49 | -.89 | 65.61 | 65.81 | -.20 |
| 2020 | 66.81 | 67.46 | -.65 | 65.70 | 65.24 | .46 |
| 2024 | | 67.33 | | | 64.48 | |
| 2028 | | 67.12 | | | 63.53 | |
| 2032 | | 66.80 | | | 62.40 | |

Table 11. Actual and predicted distance for male and female discus for 1924 to 2032 Olympics.

## 5. DISCUSSION

All men's' events 100m, 200m, 400m, long jump, high jump, shot put and discus displayed cubic functions that best fitted the data from 1924-2020 with high explained variance ($R^2$ values 0.821- 0.954), with small residual error when actual compared to predicted performance and significant p-values where p< .001 in all events) and no unrealistic predictions were presented for near future Olympic Games as 2024 Paris, 2028 Los Angeles and to 2032 Brisbane. This finding suggest the incremental changes in men's performances are consisted across years, however the cubic functions in many events indicate a law of diminishing returns and in events as men's 100m, 200m, 400m, long jump, high jump, shot put and discus we may have achieved a performance plateau as the 2016 and 2020 past Olympics and 2024, 2028 and 2032 indicate static performances or no significant changes in these events. Possibly most high performance athletes have reached the limits provided by training and most are aware of the specific training requirements for their specific events. If evolutionary phylogenetic trends are occurring in human motor performance we will have observe many generations and future Olympic Games to re-evaluate performance changes in athletics. The men's 100m and 400m sprint cubic function trends of this research are not consistent with the research of Heazlewood and Lackey (1996) and Heazlewood (2006a, 2006b) where the 100m was a poorer fit inverse function and the 400m was a sigmoidal function, however the goodness of fit for the current cubic function was higher than the Heazlewood and Lackey (1996) and Heazlewood (2006a, 2006b) research and based on a larger number of data point.

The men's long jump cubic function in this research is consistent with the findings of the Heazlewood and Lackey (1996) and Heazlewood (2006a, 2006b) research with high explained variance, whereas the high jump results are not. In the Heazlewood and Lackey (1996) and Heazlewood (2006a, 2006b) research the best fit functions were compound, logistic, exponential and growth functions which all displayed high explained variance. In this research based on a larger set of data points the best fit function was the cubic function although with high explained variance and low residual error. In both the men's high jump and long jump the Heazlewood and Lackey (1996) and Heazlewood (2006a, 2006b) the mathematical functions over predicted performances increases in these events at the 1996, 2000, 2004 and 2008 Olympic Games.

The research of Heazlewood and Walsh (2014a) on the trends in men's throws at the Olympic Games indicated an excellent cubic function solution based on data from the 1936 to 2012 Olympics for the shot put and discus in predicting 2016 performance in these events at Rio de Janeiro 2016 Olympics, as well as a future performance plateaux in these events occurring in 2016 to 2024 and these earlier predictions have been corroborated by this research.

The women's events were varied in terms of what curve estimation functions were the best fits. The women's 100m ($R^2$ = .923), high jump ($R^2$ = .936), and discus ($R^2$ = .954) were cubic functions and displayed high explained variance and small residual errors. The models predicted marginal and realistic increase in performances for 100m and high jump and marginal realistic decline in discus performance for the 2024, 2028 and 2032 Olympic Games. The cubic function for the women's 100m is consistent with the Heazlewood and Lackey (1996) and Heazlewood (2006a, 2006b) research and the Heazlewood and Walsh (2014b) predicting the woman's discus trends at the Olympic Games including the plateau in performance.

The 200m ($R^2$ = .696), 400m ($R^2$ = .578), and shot put ($R^2$ = .517) were sigmoidal functions and displayed lower explained variance and larger residual errors. The sigmoidal functions all indicated significant but plausible improvements in the 200m and 400m performances for 2024, 2028 and 2032, however the improvements predicted in the shot put model, the lowest fitting model, were implausible based on current performances by women in this event where the predicted distances were for 2024 21.74m, 2028 22.09m and 2032 22.44m. The actual score in 2020 Tokyo was 19.61m and the sigmoidal model predicted Tokyo to be 21.39m representing a significant difference between actual and predicted distance. The 400m sigmoidal function is consistent with earlier research of Heazlewood and Lackey (1996) and Heazlewood (2006a, 2006b), however the sigmoidal function for the women's shot put in the current research is inconsistent with the Heazlewood and Walsh (2014b) which indicated a high goodness of fit to a cubic function and a more realistic prediction of performance for 2016 and 2020 Olympic Games and for the 2024, 2028 and 2032 Games.

The long jump ($R^2$ = .751) best fit was an inverse function with moderately high explained variance and moderate residual error. The actual 2020 Tokyo long jump distance was 6.93m, however in the predictive model 2020 was 7.18m, 2024 was 7.24m, 2028 was 7.30m and 2032 was 7.36m, which represent significant improvements and which are probably unrealistic predictions. However, we will see the truth of these prediction in the next three Olympic Games. The inverse function for the women's long was consistent with the research of Heazlewood and Lackey (1996) and Heazlewood (2006a, 2006b), however the explained variance $R^2$ = .751 in this research was lower than the Heazlewood (2006a, 2006b) research where it was explained variance $R^2$ = .894.

The different events for men did not have different models of best fit which would be dependent on different human factors such as speed, power, endurance, speed endurance, strength and motor co-ordination that are represented by different mathematical-statistical functions. This finding refutes hypothesis one, as all events in the study were best fitted to a cubic function and this suggests sprints, jumps and throws are dependent on strength, speed and power and which have been developed to their potential as the data and models suggest men have achieved performance plateaux in these events.

Women did generate some specific predictive functions linked to specific events where 100m, high jump and discus were accurate good fit cubic functions; 200m, 400m and shot put were poor fit sigmoidal functions and long jump was a moderate fit inverse function. The newer Olympic events added to the athletic program for women, as the 200m, 400m, shot put and long jump might still be in the developing stage in terms of training as the suggestion via the predictive models are suggesting significant performance improvements in these events. So at this stage in the research there is corroboration of different models for different athletic events for women and contrary to the results displayed by the men. These results for women corroborate hypotheses one and two.

In terms of hypothesis three the mathematical-statistical models will be both event and gender specific was not supported for the men's events in this research as all the best fitted models were cubic functions, unlike previous research (Heazlewood and Lackey, 1996: Heazlewood, 2006a, 2006b), where the functions for the 100m was inverse, 400m was sigmoidal, long jump was cubic, long jump was compound, logistic, exponential and growth, and high jump was compound, logistic, exponential and growth.

In the women's events the current research partially supported this hypothesis as the 100m, high jump and discus displayed cubic functions, 200m, 400m and shot put sigmoidal and long jump inverse. The previous research (Heazlewood and Lackey, 1996: Heazlewood, 2006) also supported hypothesis three where 100m was inverse, 400m was sigmoidal, long jump was inverse and high jump was compound, logistic, exponential and growth. However the grouping of the events within the different mathematical does not present a conceptual consistency and may represent the longevity of the events as the longest running events in the Olympics are the 100m, high jump and discus. Another reason might be the influence of sports doping on women's performances in the women's and highlighting when specific women's world records were established which resulted in the natural progression of women's being interrupted or masked by sports doping. The 100m and 200m records were set in 1988, the 400m in 1985, long jump in 1988, high jump in 1987, shot put in 1987 and discus in 1988 in the majority of cases by East German and USSR (World Athletics, 2024). At this point time World Athletics and the World Anti-Doping Agency (WADA, 2024) are believed to be successful in policing drugs in sport (World Athletics Doping Policy, 2024; WADA, 2024).

Hypothesis four the mean score for the Olympic finalist will more realistically reflect human performance and inferred ability in each event than previous predictive models is a more substantive approach to understanding performances changes in the events in this research. If World records were utilised the number of data points will be very small as world records are infrequently broken as indicated previously the world record in the men's 100m and 200m were set in 2009 or 15 years ago, the 400m 2016 or 8 years ago, the long jump in 1991 or 33 years ago, high jump 1993 or 31 years ago, shot put 2023 and discus 2024, which would result in irregular non-comparable timelines and minimal data points and non-substantive curve estimation solutions over time.

In the majority of events the different mathematical models derived will provided a more accurate prediction of future performances both for short time and long time periods corroborates hypothesis five and hypothesis seven that the likelihood of absurd mathematical predictions will be reduced or non-existent.

In most events the prediction for Paris 2024, Los Angeles 2028 and Brisbane 2032 were realistic and many events performance plateaux suggesting performance improvement limits have been achieved based on the top six Olympic finalists in each event. It important to highlight the predictions were unrealistic for the women's 200m, 400m, and shot put, which were sigmoidal functions with lower explained variance values ($R^2$ = .517 - .696) with higher residual error, which resulted less predictive functions and which appear to overestimate performance improvements, which based on current performance levels in these events actual performances are significantly lower than predicted performances.

Finally, hypothesis six states, more of the variance will be explained and a better model fit will occur based on nonlinear mathematical functions. This has been substantiated in this research, as all the men's events the best goodness of fits for the data and predictive equations were cubic functions with high explained variance. In the case of the women's events three functions were cubic with high explained variance, three were sigmoidal with lower explained variance and one was inverse.

## 6. CONCLUSIONS

Mathematical modelling to understand trends in the athletic events of 100m, 200m, 400, long, jump, high jump, shot put and discus can provide insights to the changes over time for men and women high performance athletes competing at the Olympic Games. Such as, what standard of performance will be required to make the top six finalists in sprint, jump and throw events in future Olympic Games for both men and women athletes?

The derived models can provide plausible answers as to will athletic performances in specific events at the Olympic Games continue to improve? According to this research for 2024 Paris Olympics the answer might be no in the men's 100m, 200m, 400m, high jump and marginal improvements in the long jump, discus and shot put (although a significant improvement in 2020 occurred the predictive model predicts performance stability in the future or even decline).

For 2024 Paris Olympics women athletes performance improvements may continue in the 100m, 200m, 400m, long jump, and marginally in the high jump, significantly in the shot (but not likely based on actual performance as of May 2024 world ranking via top performances) and for performance decline in the discus.

Specifically, predictive models comparing the 2024 Paris Olympics to the 2020 Tokyo performances indicate men 100m improvement of 0.1s, 100m women improvement 0.06s, men 400m decrease of 0.13s, women 400m improvement by 0.62s, men long jump improvement 9cm, women long jump improvement 31cm, men high jump improvement 1cm, women high jump improvement by 4cm, men shot put decrease by 66cm, women hot put 1.75m, men discus improvement by 52cm and women discus decrease by 1.22m.

The mathematical models that appeared both event and gender specific in earlier cited research was not corroborated in this research as all men's events were described by cubic functions as the best goodness of model fit, which may suggest a common model of factors improving performance in men athletes in these events on the long term (1924 to 2024) and performance stasis based on recent short term responses (2016 to 2024).

The women athletes did display some mathematical models specific to events, such as cubic functions for 100m, high jump and discus, sigmoidal function for 200m, 400m and shot put and an inverse function for the long jump, which may indicate different exercise physiological and biomechanical factors determining performance change over time. It was noted the events with longer Olympic participation times 1928 to 2020) displayed the cubic functions for the women and similar to the men who have been competing in these events from 1924 to 2020 in the data set for this research. The sigmoidal function events of 200m, 400m and shot put for women were added well after these events were introduced for men and not have transitioned to a cubic function that is reflective of longer performance timeframes.

**Acknowledgements**
I acknowledge all the men and women Olympic athletes who have competed in the previous Olympic Games in the 100m, 200m, 400m, long jump, high jump, shot put and discus who enabled the development of the data set utilised in this research.

**References**

Banister, E.W. & Calvert, T.W. (1980) Planning for future performance: Implications for long term training. Canadian Journal of Applied Sports Science 5, 170176.

Edwards D. and Hopkins, W., Targets for the future. Modern Athlete and Coach. 17 (3), 34-36 (1979).

Garson, D.G. (2010) Curve estimation. Statnotes: Topics in Multivariate Analysis. Retrieved 1/12/2010 from http://faculty.chass.ncsu.edu/garson/pa765/statnote.htm.

Hair, J.E., Block, W., Babin, B.J., Anderson, R.E. & Tatham, R.L. (2006) Multivariate Data Analysis. (6th Ed.). Pearson - Prentice Hall.

Heazlewood, I. & Lackey, G. (1996). The use of mathematical models to predict elite athletic performance at the Olympic Games. Conference proceedings 3rd Conference on Mathematics and Computers in Sport. Bond University (30 September - 2 October, 1996) Queensland.

Heazlewood, I. & Lackey, G. (1998). Mathematical models that predict performance decline in elite veteran athletes 30-90 years in the sprint distance and jump events. Conference proceedings 4th Conference on Mathematics and Computers in Sport. Bond University (13-15 July, 1998), Queensland. Pp.51-78.

Heazlewood, T. (2006a). Prediction versus reality: The use of mathematical models to predict elite performance in swimming and athletics at the Olympic Games. Conference Proceedings of the Eighth Australasian Conference on Mathematics and Computers in Sport. 3-5 July 2006, pp. 2-11. Coolangatta, Queensland, Australia.

Heazlewood, T. (2006b). Prediction versus reality: The use of mathematical models to predict elite performance in swimming and athletics at the Olympic Games. Journal of Sports Science and Medicine. 5, pp. 541-547.

Heazlewood, I. (2008) Prediction versus reality: The use of mathematical models to predict elite performance in swimming and athletics at the Olympic Games. Oral presentation to Australian Statistical Society (NSW Branch).University of Sydney. 24/9/2008.

Heazlewood, I. (2010). Prediction versus reality: The use of mathematical models to predict elite performance in swimming and athletics at the Olympic Games. Model Predicting 2012 Performances. 2010 High Performance Sports Research Workshop, Australian Institute of Sport. Canberra, Australia 12-14 May, 2010. Invited Poster Presentation.

Heazlewood, I.T. & Walsh, J. (2011a). Mathematical models that predict 2012 London Olympic Games swimming and athletic performances. In Proc. 3rd International Conference on Mathematics in Sport. Pp.71-78. Salford, UK.

Heazlewood, I.T. & Walsh, J. (2011b). Mathematical Models that predict athletic performance in sprint and jump events for current and future world champions. In Proc. 3rd International Conference on Mathematics in Sport. Pp.64-70. Salford, UK.

Heazlewood, I. (2013a). The congruence between mathematical models predicting 2012 London Olympic Games swimming and athletic Performances and Actual Performance. 9th International Symposium on Computer Science in Sport (IACSS2013). June 19-22, 2013. Istanbul, Turkey. Oral presentation

Heazlewood, I. (2013b). The congruence between mathematical models predicting 2011 Athletic World Championship performances and actual performance in the 100m, 400m, long jump and high jump. 9th International Symposium on Computer Science in Sport (IACSS2013). June 19-22, 2013. Istanbul, Turkey. Oral presentation.

Heazlewood, I.T. & Walsh, J. (2014a). Mathematical models that predict athlete performance in the men's throwing events at the Olympic Games. Conference proceedings 12th Australian Conference on Mathematics and Computers in Sport. Bond University (25-27 June, 2014), Darwin, Australia. Pp. 48-53.

Heazlewood, I.T. & Walsh, J. (2014b). Mathematical models that predict athlete performance in the women's throwing events at the Olympic Games. Conference proceedings 12th Australian Conference on Mathematics and Computers in Sport. Bond University (25-27 June, 2014), Darwin, Australia. Pp. 54-59.

Hopkins, W. (2000) Limits to performance. SportScience 4, Retrieved 3/4/2006 from http://www.sportsci.org/jour/0002/inbrief.html#limits. IAAF. (2008)

International Olympic Committee (2014). Olympic Results: Your source for event results and medalists from every Olympic Games. International Olympic Committee.

International Olympic Committee (2024). Olympic Results: Your source for event results and medalists from every Olympic Games. International Olympic Committee. https://olympics.com/en/olympic-games/olympic-results.
Lackey, G. & Heazlewood, I. (1998). Use of Mathematical Models to Predict Elite Swimming Performance. Conference proceedings 4th Conference on Mathematics and Computers in Sport. Bond University (13-15 July, 1998), Queensland. Pp. 79-110.
Linthorne, N., The effect of wind on 100m sprint times. Journal of Applied Biomechanics, 10 (2) 110 - 131 (1993).
McAlister, S. (2022). How to qualify for athletics at Paris 2024. The Olympics qualification system explained. https://olympics.com/en/news/how-to-qualify-paris-2024-athletics-qualification-system-explained.
Morton, R. H., The supreme runner: A theory of running and some of his physiological attributes. The Australian Journal of Science and Medicine in Sport. 16 (2) 26-28 (1984).
Norušis, M., SPSS for windows: Base systems user's guide release 6.0, (Chicago, SPSS Inc, 1993).
Peronnet F. and Thibault G., Mathematical analysis of running performance and world running records. Journal of Applied Physiology, 6 (7) 453-465, (1989).
Prendergast, K., What do world running records tell us? Modern Athlete and Coach. 28 (1) 33 -36 (1990).
Walsh, J. & Heazlewood, I. (2014). Performance variability in the throwing events. Proceedings of the International Association of Computer Science in Sport Conference (IACSS2014). Ed. Assoc. Prof. Ian Heazlewood, Assoc. Prof. Anthony Bedford, Darwin, Australia, June 22-24. 2014. Pp. 109-115.
Wikipedia (2024). Athletics at the Summer Olympics. https://en.wikipedia.org/wiki/Athletics_at_the_Summer_Olympics.
Wikipedia (2024). Men's 100 metres world record progression. Wikipedia. https://en.wikipedia.org/wiki/Men%27s_100_metres_world_record_progression.
World Athletics. (2021). Anti-Doping: Anti-Doping Rules. World Athletics. https://worldathletics.org/about-iaaf/documents/anti-doping.
World Anti-Doping Agency (WADA). (2024). Raising the game for clean sport. https://www.wada-ama.org/en.
World Athletics (2024). Competition Rules: Rules, Polices & Resources. https://www.athletics.com.au/competition-rules/.
World Athletics (2024). World Record progression of 100 Metres. World Athletics. https://worldathletics.org/records/by-progression/16647.

# FROM THE ANCIENT OLYMPICS TO BRISBANE 2032: METHODS FOR ADDING RELEVANT AND ENTERTAINING OLYMPIC SPORTS

Raymond Stefani
*California State University Long Beach, USA*
*Corresponding author: Raystefani@aol.com*

## Abstract
Olympic hosts have selected sports having relevant, widely followed competition, and sports providing a high level of entertainment. The 861 events of the Ancient Olympics included athletics (49%), combat sports like boxing and wrestling (32%) and chariot racing (11%). For entertainment, 4% involved equestrian competition, and 4% included dramatic acting, trumpet playing, and lyre playing. In 1896, including the marathon brilliantly linked modern with ancient, creating worldwide interest. From 1896-1936, host nations, with IOC approval, selected sports that maintained relevance and entertainment value at their locations. After WW2, non-medal demonstration sports were contested from 1948-1992 to fit local interest. Six of those sports were so popular they became continuing core sports: badminton, baseball, handball, judo for women, taekwondo, and tennis. Six sports were added directly to continuing status: archery, judo for men, softball, table tennis, triathlon, and volleyball. After 1992, the Olympics were too large for demonstration sports. From 2004 to 2016, trying to limit sports federations to 28 proved too cumbersome: baseball and softball were dropped and golf and rugby 7s relaced them. Great success followed the introduction of Olympic Agenda 2020. Tokyo 2000 was allowed to add up to 500 athletes in recognised, non-core Olympic sports, with a 10,500-athlete core sports limit. Added were skateboarding, sport climbing, surfing, karate, baseball, and softball. Paris 2024 had to include new sports athletes in the 10,500-athlete limit; hence they included just skateboarding, sport climbing and surfing, which the IOC then made continuing core sports. Eclectic breakdancing was also included. The Los Angels 2028 Olympics will include the 3 new core sports plus baseball, softball, lacrosse, cricket, squash, and flag football. Brisbane 2032 will inherit the schedule from LA 2028, probably continuing lacrosse, cricket, and squash and maybe baseball and softball. Netball may well be included given its popularity in Australia.

*Keywords*: Ancient Olympics, Modern Olympics, Olympic Agenda 2020, selecting sports, women's sports

## 1. INTRODUCTION
Every four years from 776 BC to 1000 years later, and from 1896 to the present, the world's sports enthusiasts have turned their attention to the Olympic Games because of the well-deserved reputation for adjusting the plethora of offered sports to keep them relevant and entertaining for all to see. We examine the sports offered in Ancient Greece to reveal trends in sports selection. For the modern Olympic Games, we identify four eras with differing methods for sports section, the most recent of which, Olympic Agenda 2020, has been very successful at adding popular new sports to the Olympic program.

## 2. ANCIENT OLYMPICS
Thanks to the Perseus Project of Tufts University (Perseus Project, 2024) as sponsored by The Melon Foundation, The Annenberg Project and many others, ancient texts were scoured to create a list of contested events from 776 BC through 277 AD. These are summarized in Wikipedia List of Ancient Greek Victors (2024). Each dated event contains the winner and winner's city. No winners were recorded after 277 AD. We placed that data base into an Excel spread sheet and deleted events that had ambiguous or no event definitions. The remaining 861 events contested from 776 BC to 269 AD were sorted into the five categories in Tables 1 and 2.

| Sport | Events | First Year | Last Year | Total Events | % of 861 |
|---|---|---|---|---|---|
| Athletics | 7 | 776 BC | 269 AD | 419 | 49% |
| Combat | 6 | 708 BC | 221 AD | 279 | 32% |
| Chariot Racing | 10 | 680 BC | 241 AD | 94 | 11% |
| Equestrian Racing | 3 | 648 BC | 197 AD | 36 | 4% |
| Artistic Performance | 4 | 396 BC | 261 AD | 33 | 4% |
| Total | 30 | | | 861 | |

Table 1: The 5 Sports Contested at the Ancient Olympics from 776 BC though 269 AD

| Athletics (7) | Combat (6) | Chariot Racing (10) | Equestrian Racing (3) | Artistic Performance (4) |
|---|---|---|---|---|
| Stadion | Boxing | Apene | Foals' Race | Herald Competition |
| Stadion-Boys | Boxing-Boys | **Chariot Race** | Horse Race | **Lyre Playing** |
| Diaulos | Pankration | **Chariot-Foals** | Mares' Race | **Tragedy Competition** |
| Diaulos in Armor | Pankration-Boys | **10 Horse Chariot** | | Trumpeter Competition |
| Dolichos | Wrestling | Synoris | | |
| Pentathlon | Wrestling-Boys | Synoris-Foals | | |
| Pentathlon-Boys | | Synoris-Colts | | |
| | | Tethrippon | | |
| | | Tethrippon-Foals | | |
| | | Tethrippon-Colts | | |

Table 2: The 30 Events from Table 1 for the 5 Sports (The five events in red were only contested in 65 AD)

The Ancient Olympics began with just one athletics running event, the Stadion, which is the origin of the word stadium, Romano (1993). The Stadion was a straight 600-foot sprint, very popular in Greece. The athletics events of Table 2 involved 49% of all the 861 events contested over the 1000 years of the Ancient Olympics. The Games expanded scope to include the hand-to-hand combat events of Table 2, starting in 708 BC. Combat sports comprised 32% of all events contested.

A velodrome was completed near the stadium in Olympia to accommodate chariot racing in 680 BC. The chariot was an important part of warfare, so peacetime competition was very popular. The chariot racing events in Table 2 comprised 11% of all the 861 events analysed. Equestrian-racing was first added for entertainment purposes in 648 BC. The three equestrian events of Table 2 added another dimension to the Games, totalling 4% of all events contested.

The accumulated relevance and entertainment provided by the Olympic Games that had started with one athletics event in 776 BC and then was significantly augmented by Olympic Games organizers who added three additional sports by 648 BC, incentivized the building of three more stadia and the beginning of three more Games in Ancient Greece. The Pythian (at Delphi) Games began in 582 BC, the Isthmian Games began in 581 BC, and the Nemean Games began in 573 BC. The Nemean and Isthmian Games were held every two years while the Olympic and Phythian Games were held every four years. Collectively there were six competitions over every four-year period, owing their origin to a lone running event in 776 BC at Olympia. For information about those other Games see Wikipedia Heraean Games (2024), Wikipedia Isthmian Games (2024), and Wikipedia Pythian Games (2024).

The remaining 4% of all events contested were artistic performance events, beginning in 396 BC. Lacking loudspeakers, trumpeters and herald yellers provided crowd information and control, as well as wartime information transmission. Competition in those areas became popular and were added appropriately.

Another facet of competition is identified in Table 2 by five events shown in red, which were only contested in 65 AD. Also, in 65 AD, Herald competition was restored after a 420-year absence. At a time with Rome dominating Greece, all six of those 65 AD events were won by Nero. It was probably a bad life choice to try to defeat Nero. It is thus clear that competition in the past was also be influenced by a strong individual and a country's dominance, much as the modern world experienced during the Cold War.

The Olympic Games were dedicated to Zeus, a male god, and thus only men could compete, Miller (2004) and Were Women Allowed in the Olympics (2015, December). Unmarried women could and did attend. Married women were not to attend, Miller (2004). However, at least six women beat the system at the Olympic Games, owning and training the horses that won eight chariot and equestrian events. They realized that the person who owned the horses, regardless of gender, was deemed official champion and not the jockey or chariot driver who was paid by the owner and thus not eligible to be an official winner. Also in ancient Greece, Games dedicated to Zeus' mythical wife Hera, the Heraean Games, were contested only by women. The Ancient Greek historian and geographer, Pausanias, (Miller, 2004), published information about the Heraean Games, but most publicity dealt with the men's Olympic Games. Women ran multiples of 500 feet on the same track at Olympia, but at a different time than when the men ran multiples of 600 feet. Thus, women were considered about 5/6 or 83% as fast as men, on a crude scale with only 100-foot increments. In today's Olympics, the Olympic champion women in running, speed skating, swimming, and rowing are 90% as fast as their male counterparts, Stefani (2014).

## 3. THE MODERN OLYMPICS BEGIN: 1896-1936 (DOMINANCE BY HOST NATIONS)

The Modern Olympics began in 1896, appropriately in Athens, Greece. Table 3 summarizes the first 10 Games, five before World War 1, followed by five more before World War 2. The information in Table 3 was taken from Wallechinsky and Loucky (2012 and earlier editions). Before the 1896 games began, much was publicized about the marathon run which linked ancient with modern. That popular theme was made even more pronounced when a Greek, Spyridon Louis won, successfully launching the new edition of the Olympic Games. The host nation (below in red) won the most medals in seven of those 10 Games, including the first five. The fraction of medals won by the 10 top nations in Table 3 averages 34%, compared 10% won by the USA, the highest national medal winner, at Tokyo 2020. Competition has tightened significantly since the first 10 Games.

| | | | Most Medals | | | Second Most | | |
|---|---|---|---|---|---|---|---|---|
| **Olympics** | **Gold Medals** | **Medals Awarded** | **Nation** | **Medals** | **% Won** | **Nation** | **Medals** | **% Won** |
| **1896 Athens** | 44 | 122 | **GRE** | **47** | **39%** | USA | 19 | 16% |
| **1900 Paris** | 96 | 276 | **FRA** | **102** | **37%** | USA | 53 | 19% |
| **1904 St. Louis** | 110 | 284 | **USA** | **238** | **84%** | GBR | 15 | 5% |
| **1908 London** | 110 | 323 | **GBR** | **145** | **45%** | USA | 47 | 15% |
| **1912 Stockholm** | 101 | 309 | **SWE** | **65** | **21%** | USA | 61 | 20% |
| | | | | | | | | |
| **1920 Antwerp** | 156 | 435 | USA | 96 | 22% | SWE | 63 | 14% |
| **1924 Paris** | 124 | 366 | USA | 99 | 27% | **FRA** | **38** | **10%** |
| **1928 Amsterdam** | 110 | 327 | USA | 56 | 17% | GER | 31 | 9% |
| **1932 Los Angeles** | 115 | 346 | **USA** | **104** | **30%** | ITA | 36 | 10% |
| **1936 Berlin** | 130 | 388 | **GER** | **89** | **23%** | USA | 56 | 14% |

Table 3: The First 10 Modern Olympic Games (The host nation results are shown in red)

Since a controlling infrastructure was lacking, the host nations were given the task of picking sports that they thought fit the interests of their country and likely visitors. The number of events, equal to the number of gold medals, assuming no ties, increased from 44 in 1896 to 101 in 1912, prior to WW1. The number of events then jumped to 156 at Antwerp in 1920, just after WW1, and ended at 130 in 1936, prior to WW2, nearly three times the number of events in 1896.

Women were not allowed to compete in 1896, but began their Olympic journey in 1900. Swimming began for women in 1912 and athletics followed in 1928. In today’s Olympics, the number of male and female athletes has equalized. In both the Ancient and Modern Olympics, women successfully fought conscientiously and effectively for respect and inclusion.

## 4. DEMONSTRATION SPORTS: 1948-2000

After WW2, the IOC allowed host nations to select non-medal demonstration sports that were of special importance and interest to that country, like Finnish Baseball and Australian Rules Football. If a demonstration sport proved popular enough, it could then become a continuing medal-offering core sport. The information in Table 4 was taken from Wikipedia Olympic Demonstration Sports (2015, October 20). The right panel of Table 4 shows the 15 demonstration sports offered by hosts from 1948-1992. The six sports in red became continuing core sports, as seen in the left panel: handball, tennis, badminton, baseball, judo for women and taekwondo. Six sports were added directly to continuing status in the left panel: judo for men, volleyball, archery, table tennis, softball, and triathlon. Polo, handball, and judo for men were dropped in the middle panel, then judo for men and handball were both reinstated. Allowing demonstration sports was a very successful way of getting important new sports into the Olympics, due to exemplary action by the host nations.

The Olympic Games became too large to include demonstration sports after 1992, but the influence of demonstration sports remained through 2000, when a previous demonstration sport moved to continuing status. The size of the Olympic Games then led the IOC to seek a way to limit the number of sports federations, covered next.

| Year | City | IOC | | Organizing Committee |
|---|---|---|---|---|
| | | Sports Added | Sports Dropped | Demonstration Sports |
| **Demonstration Sports Could Become Continuing Sports** | | | | |
| 1948 | London | | Polo, Handball | Lacrosse, Swedish Gymnastics |
| 1952 | Helsinki | | | Finnish Baseball, **Handball** |
| 1956 | Melbourne | | | Australian Rules Football, **Baseball** |
| 1960 | Rome | | | |
| 1964 | Tokyo | Judo, Volleyball | | **Baseball**, Budo |
| 1968 | Mexico City | | Judo | Jai Alai, **Tennis** |
| 1972 | Munich | Archery**, Handball**, Judo | | **Badminton**, Water Skiing |
| 1976 | Montreal | | | |
| 1980 | Moscow | | | |
| 1984 | Los Angeles | | | **Baseball, Tennis** |
| 1988 | Seoul | Table Tennis, **Tennis** | | **Badminton, Baseball,** Bowling, **Judo(W), Taekwondo** |
| 1992 | Barcelona | **Badminton, Baseball, Judo (W)** | | Jai Alai, Roller Hockey, **Taekwondo** |
| 1996 | Atlanta | Softball, | | |
| 2000 | Sydney | **Taekwondo**, Triathlon | | |
| **Sports Added and Dropped by IOC Action** | | | | |
| 2004 | Athens | | | |
| 2008 | Beijing | | | |
| 2012 | London | | Baseball, Softball | |
| 2016 | Rio | Golf, Rugby 7s | | |

Table 4: Sports Added, Sports Dropped and Demonstration Sports 1948-2016

## 5. LIMITING SPORTS FEDERATIONS TO 28: 2004-2016

While the use of demonstration sports operated quite well, the IOC created a very cumbersome system in 2004 to limit the number of sports federations that organize Olympic sports to 28. Every four years the IOC had to either continue or discontinue each of the 28 sports federations that produce Olympic sports. The only way to add a new sport was to drop an existing federation and its organized sport and then add a new federation and its organized sport. The IOC decided to drop baseball and softball, effective in 2012 and then to add golf and rugby 7s effective in 2016.

In 2013, the IOC wanted to add a new sport, effective 2020. They voted to drop the modern pentathlon which was the only sport created specifically for the Modern Olympics, which was done by Baron de Coubertin, the creator of the Modern Olympics, His descendants convinced the IOC not to drop that sport. Hurriedly, the IOC voted to drop wrestling, a sport from the Ancient Olympics. Later, under huge pressure to fix that mistake, the IOC voted to reinstate wrestling. The IOC then abandoned the system for limiting sports federations.

## 6. OLYMPIC AGENDA 2020: 2020-2032

To replace the cumbersome system of trying to limit the number of sports federations, the IOC created Olympic Agenda 2020, which cleverly combines the best attributes of the two methods exemplified in Table 4, Olympic Agenda 2020 (2014). Instead of having the host offer non-medal demonstration sports, medal-earning sports can be added temporarily for sports whose federations are recognized by the IOC, with a possibility of becoming continuing core sports. The IOC recognizes about as many sports not currently in the Olympics as the number of sports that are currently in the Olympics, as shown in Table 5, based on information in Wikipedia List of International Sports Federations (2024).

Under Olympic Agenda 2020, the size of the Olympic Games is to be held constant in terms of an IOC-generated athlete total to be implemented by each host nation, rather than by a fixed number of sport federations, implemented by the IOC. The host’s requested use of Olympic Agenda 2020 must be approved by the IOC five years prior to that Olympic Games.

| Recognition | Sports Federations | Sports (Disciplines) | e-Sport | Mind Sports | Physical Sports | | |
|---|---|---|---|---|---|---|---|
| | | | | | Combat | Independent | Object |
| **IOC Summer** | 35 | 47 | 0 | 0 | 7 | 26 | 14 |
| **IOC Winter** | 7 | 15 | 0 | 0 | 0 | 13 | 2 |
| **Olympic Total** | 42 | 62 | 0 | 0 | 7 | 39 | 16 |
| **IOC Recognized Not in Olympics** | **37** | **66** | **0** | **2** | **6** | **29** | **29** |

Tabel 5: Sports in the Olympics Continuously and IOC Recognized Sports that are Eligible for Inclusion

Under Olympic Agenda 2020, the Tokyo 2020 Olympics (which took place in 2021 due to COVID) could add new IOC recognized sports whose combined athlete total is limited to 500, while separately limiting core sports athletes to 10,500. However, starting with Paris 2024, organizers and all later organizers must include any new sports athletes as part of the 10,500-athlete limit: that is, for every athlete involved in a new sport, one athlete would have to be dropped from a core sport. The tables for the Olympic Games to follow are taken from the IOC website and from Wikipedia's Games coverages.

TOKYO 2020

The left part of Table 6 shows how Tokyo 2020 organizers followed Olympic Agenda 2020 rules. Due to COVID, the Games took place in 2021. The core sport total was a bit over the 10,500-athlete limit while the new sport total was under 500. Total athlete count was just over 11,000. The Tokyo 2020 organizers selected three youth-oriented sports which proved to be very entertaining and popular; skateboarding, sport climbing and surfing. Also added were karate, baseball, and softball. We observed that the skateboarders were particularly entertaining to watch on TV, since they were so supportive of each other.

| **Tokyo 2020 (11090 Athletes) As Offered in 2021** | | | **Paris 2024 (10,500 Athletes)** | |
|---|---|---|---|---|
| **Core Sports** | **Athletes** | | **Core Sports** | **Athletes** |
| Total | 10616 | | Total | 10268 |
| | | | | |
| **New Sports** | **Athletes** | | **New Sports** | **Athletes** |
| Skateboarding | 80 | | Skateboarding | 88 |
| Sport Climbing | 40 | | Sport climbing | 68 |
| Surfing | 40 | | Surfing | 44 |
| Karate | 80 | | Breakdancing | 32 |
| Baseball and Softball | 234 | | | |
| Total | 474 | | Total | 232 |
| | | | | |
| **Grand Total** | **11090** | | **Grand Total** | **10500** |

Table 6: Tokyo 2020 and Paris 2024 Using Olympic Agenda 2020

Paris 2024

The Paris 2024 organizers began with the Tokyo 2020 athlete allocations in Table 6, but with the requirement of a 10.500 limit overall. Shown in the right part of Table 6, the great success of skateboarding, sport climbing and surfing led Paris 2024 to include all three. The fact that the youth-oriented and very entertaining sports of skateboarding, sport climbing and surfing would be offered again, led the IOC to move those three sports to core continuing status beginning with LA 2028. It is likely that the Paris 2024 organizers decided that the additional athletes needed to schedule karate, baseball and softball would have made it too difficult to drop so many athletes from core sports. Instead, breakdancing was added, a very entertaining and creative choice. The 10,500-athlete schedule of Table 6 was approved by IOC.

LA 2028

The beginning athlete budget in Table 7 was less well-defined for LA 2028 than was true for Paris 2024. The IOC had decided that animal cruelty in the modern pentathlon, the performance enhancing drug use in weightlifting, and concern with judging in boxing meant that all three sports would not be offered at LA 2028, giving LA 2028 organizers what seemed like 444 athlete positions to use. There was already room for the three new core sports of skateboarding, sport climbing and surfing. The LA 2028 Olympic Organizing Committee was approved to include baseball and softball, two popular sports in the USA; lacrosse which is gaining in popularity and was created by the Iroquois indigenous tribe, partially in the what is now the USA and partially in what is now Canada; flag football, a rapidly growing and safety-minded youth sport; and cricket and squash, two widely played international sports.

| **LA 2028 (10,500 Athletes)** | |
|---|---|
| **New Core Sports, But No Breakdancing** | **Athletes** |
| Total | 10468 |
| | |
| **Possible Deleted Sports** | **Athletes** |
| Modern Pentathlon (72, not deleted) | |
| Weightlifting (120, not deleted) | |
| Boxing | -252 |
| Possible Deleted Total | -252 |
| | |
| Cumulative Total | 10216 |
| | |
| **New Sports** | **Approximate Athletes** |
| Baseball and Softball | 234 |
| Lacrosse | 240 |
| Flag Football | 176 |
| Cricket | 240 |
| Squash | 80 |
| New Sport Total | 970 |
| | |
| **Possible Grand Total Before Reduction** | **11086 (10500+686)** |

Table 7: LA 2028 Assuming Values from Paris 2024 Using Olympic Agenda 2020

LA 2028 did not ask to offer breakdancing. As the new sport list was being approved, the IOC decided in favour of changes made by the federations organizing the modern pentathlon and weightlifting, so those sports were to be offered. Based on assumptions of the athlete count in Table 7, LA 2028 may have to delete 686 athletes from core sports to satisfy the 10,500-athlete limit.

BRISBANE 2032

By 2027 when the Brisbane 2032 program must be sent to the IOC for approval, the LA 2028 program of 10.500 athletes will have been finalized, which of course will have included skateboarding, sport climbing and surfing which must be offered. Since lacrosse, cricket and squash are popular in Australia, it is likely that those sports, already in that athlete budget, will likely be continued. Baseball and softball are popular enough that those two sports might be offered. Flag football is less likely to be offered., which would free up 176 positions (assuming 8 teams for men and 8 teams for women of 11 players each). If boxing is restored by the IOC, that would require 252 athletes, using the extra 176 and having to reduce core athletes by 76, Since netball is very popular in Australia, if it is included using 8 teams of 15 each, then 196 core athletes would have to be deleted. The Brisbane hosts have the right to request adding additional sports of interest.

## 7. CONCLUSIONS

By reviewing more than 1000 years of success in the planning and offering of the Ancient Olympics and a similar 130 years for the Modern Olympics, we can identify the factors creating that success. The hosts at Olympia began with one athletics running event of 600 feet in 776 BC. Over the next 200 years, they creatively added combat, chariot racing and equestrian racing sports which made the Games successful, relevant, and entertaining. That reputation led potential hosts at three other locations to create their own stadia and start their own Games from 582 BC to 573 BC: the Isthmian Games, the Nemean Games, and the Pythian Games at Delphi. The Games were staggered so that the Olympic Games and Phythian Games were held once every four years while the other two games were held twice every four years, making for six Games to be held over every four-year period. From that one running event emerged three more Games, due to clever and intuitive additions to the Games at Olympia. A very entertaining and popular sport, artistic performance competition. was added later.

At Olympia in Ancient Greece, women took part in their own Games, the Heraean Games, and were official winners of 8 events at the Olympic Games because they owned the winning horses in chariot and equestrian racing events. Women skilfully pursued their place in sport at Olympia and the other Games sites.

For the first 10 Games of the Modern Olympic era, beginning in 1896, the host nations were clever and forward thinking under the emerging International Olympic Committee. The number of events rose from 44 events (including the hugely successful and world-recognized marathon) tripling just before WW2. When the Games resumed, the IOC implemented a forward-looking scheme, allowing host cities to offer non-medal demonstration sports which might become continuing sports if successful. That lead to the movement of handball, baseball, tennis, badminton, judo for women and taekwondo to continuing status. Also directly added were judo for men, volleyball, archery, table tennis, softball, and triathlon, showing the value of clever hosts under flexible IOC rules.

The cumbersome and inflexible concept of freezing the number of sports federations was short-lived, but followed by Olympic Agenda 2020, which restored innovation to the hosts under reasonable limiting of the number of athletes. Three youth-oriented and popular sports moved to continuing status: skateboarding, sports climbing and surfing. Also, the following popular sports have been approved to be offered temporarily under this flexible and innovative policy: baseball, softball, karate, lacrosse, flag football, cricket, squash and even break dancing.

In the modern Olympics, women were not included in 1896, but began Olympic competition in 1900, followed by inclusion in swimming in 1912 and in athletics in 1928. Women now have equal athlete count compared to men. As in the Ancient Olympics, women have not settled for mediocrity but have skilfully and deservedly moved toward equality.

The key ingredients for Olympic success have been the clever and far-thinking expansion of relevant and entertaining sports by the hosts, and in the Modern Olympics, we have the effective and flexible IOC control.

## References


Miller, S.G. (2004). *Ancient Greek Athletics*, New Haven, CT: New Haven.

Olympic Agenda 2020 (2014), https://stillmed.olympic.org/Documents/Olympic_Agenda_2020/Olympic_Agenda_2020-20-20_Recommendations-ENG.pdf

Perseus Project (2024) Ancient Olympics Information, Retrieved from https://www.perseus.tufts.edu/Olympics/index.html

Romano, D.G. (1993). Athletics and Mathematics in Archaic Corinth: The Origins of the Greek Stadion. *Memoirs of the American Philosophical Society*, v. 206.

Stefani, R.T. (2014). Understanding the Velocity Ratio of Male and Female Olympic Champions in Running, Speed Skating, Rowing and Swimming, *Proceedings of the 12th Australasian Conference on Mathematics and Computers in Sport, Darwin Australia, 25-27 June, 2014.* pp 106-111.

Walleshinsky, D. and Loucky, L. (2012 and earlier editions). *The Complete Book of the Olympics*, Aurum Press Ltd, London, ISBN 978 1 84513 695 6.

Were Women Allowed at the Olympics? (2015, December). Retrieved from http://www.perseus.tufts.edu/Olympics/faq5..

Wikipedia Olympic Demonstration Sports (2015, October 20), Retrieved from https://en.wikipedia.org/wiki/Demonstration_sport..

Wikipedia Heraean Games. (2024). Retrieved from https://en.wikipedia.org/wiki/Heraean_Games

Wikipedia Isthmian Games. (2024). Retrieved from https://en.wikipedia.org/wiki/Isthmian_Games

Wikipedia List of Ancient Greek Victors. (2024). Retrieved from https://en.wikipedia.org/wiki/List_of_ancient_Olympic_victors

Wikipedia List of International Sports Federations (2024). Retrieved from https://en.wikipedia.org/wiki/List_of_international_sports_federations .

Wikipedia Pythian Games. (2024). Retrieved from https://en.wikipedia.org/wiki/Pythian_Games

# AUGMENTED DICTIONARY-BASED SENTIMENT ANALYSIS OF ENTITIES AND SENTENCES WITHIN THE SPORTS NEWS

Michael Kerr [a], Paul Bracewell [a,b,c,d]
[a] *Victoria University of Wellington, New Zealand*
[b] *Footmarx,* [c] *Endorsco*
[d] *Corresponding author: paul@endorsco.com*

## Abstract

Sentiment analysis, a key area within Natural Language Processing (NLP), has evolved significantly over the last two decades. Initially reliant on dictionary-based methods before the AI and machine learning era, the field has since shifted towards these advanced technologies, which now lead in terms of accuracy and efficiency. Despite the shift, traditional methods remain relevant due to their computational efficiency and satisfactory accuracy levels.

This paper presents a study analysing 42,300 news articles from six New Zealand outlets over the period June to November 2023. Utilizing the sentimentr package from R, the study benchmarks three sentiment dictionaries—SenticNet, SentiWordNet, and Jocker and Rinker's—against each other in measuring positive, negative, and neutral sentiments at the sentence level. The results demonstrate that Jocker and Rinker's dictionary outperforms the others with an overall accuracy of 59.4%, with neutral sentiments showing more precise evaluation metrics than positive or negative ones.

Moreover, the study employs named-entity recognition and coreference resolution to link entities to their respective sentence sentiments, revealing that sports players generally exhibit higher sentiment scores than other media figures. This observation underscores the nuanced performance of sentiment analysis tools across different contexts and entities.

In summary, our research highlights the continuing relevance of both traditional and machine learning-based sentiment analysis methods, providing insights into their application in media studies and the variances in sentiment attribution among different public figures.

**Keywords**: Machine Learning, Media Analysis

## 1. INTRODUCTION

News articles significantly influence public opinion by employing emotive phrases that engage readers. The power to shape social perceptions often resides subtly in the hands of journalists (Entman, 1989). Every adjective, adverb, and interjection carries some level of sentiment, which is then associated with any connected pronouns and proper nouns. This linkage forms the basis for sentiment analysis, a field that has gained prominence with the advancement of natural language processing (NLP).

Sentiment analysis, a highly relevant and extensively researched area, involves summarizing text based on its emotional content. It is closely tied to coreference resolution—the task of linking proper nouns to their respective pronouns within text. These tasks are crucial for accurately assessing the sentiment toward individuals or entities mentioned in articles. Modern challenges include differing sentiment expressions across media outlets, often influenced by practices like clickbait (Jung et al., 2022). For example, consider the sent ence: "Alice is a very good person, and Bob is not. He likes to party, but she likes to read." Computers must determine that "He" refers to Bob and "She" refers to Alice—a task straightforward for humans but complex for machines. Technical references, such as distinguishing between "King Charles III" and "King Charles II," require extensive background knowledge. Historically, the best models for coreference resolution were entity-based cross-task joint models (Durrett & Klein, 2014). Although large language models (LLMs) like GPT-4 have transformed NLP with innovations like transformers (Vaswani et al., 2017), they have not yet significantly impacted coreference resolution. Currently, the leading models are BERT (Devlin et al., 2019), which employs encoder-only models, and the latest seq2seq models by Bohnet et al. (2023).

State-of-the-art sentiment analysis now primarily utilizes machine learning and AI methods. However, older sentiment algorithms based on statistical and dictionary methods remain valuable. Though less accurate, they are more efficient (Wankhade et al., 2022) and capable of achieving sentence-level sentiment accuracy as high as 59.6%. The effectiveness of these methods depends significantly on the choice of dictionary used. These algorithms also allow for the association of entities with sentence sentiments, providing insights into how these entities are perceived in the media, though the accuracy of these sentiment attributions has not yet been systematically evaluated.

## 2. BACKGROUND

Sentiment analysis, even without coreference resolution, has a specific role where entity focus is not required. Samuels & McGonical (2020) applied traditional preprocessing steps, including the use of Term Frequency-Inverse Document Frequency (TF-IDF) to weight word frequencies (Ghag & Shah, 2014), followed by sentiment computation using dictionary-based methods from WordNet. While effective for general sentiment analysis, this approach falls short in accurately capturing sentiment related to specific entities mentioned within texts, a limitation highlighted by Atkinson & Escudero (2022).

In more complex scenarios, such as entity sentiment analysis, merely assigning sentiment based on word presence can lead to inaccuracies. This subfield, predominantly focused on human subjects, aims to assign sentiment to entities or their coreferences within a sentence. Luo & Mu (2022) critiqued traditional entity sentiment analysis methods for ignoring the contextual nuances of sentences, which could lead to inappropriate sentiment attributions to non-target subjects. To address these shortcomings, they introduced a Negative Sentiment Smoothing Model (NSSM), utilizing VADER (Hutto & Gilbert, 2014) to first identify negative paragraphs and the entities associated with them. NSSM refines how sentiment is attributed by smoothing out the impacts of negatively charged words associated with entities, ensuring that terms like 'suicide' do not automatically confer a negative sentiment to the associated entities.

The intricacies of coreference resolution, essential for accurate sentiment analysis, involve aligning pronouns with their corresponding nouns within texts. De Clercq & Hoste (2020) found that coreference resolution could slightly improve accuracy in aspect-based sentiment analysis. The state-of-the-art in this area has been advanced by tools like FastCoreF and LingMess, with the latter showing superior performance despite computational limitations (Otmazgin et al., 2022). Training datasets like OntoNotes and WordNet, which provide extensive annotations and define a hierarchy of word senses, are crucial for training effective coreference resolution systems (Weischedel et al., 2010; Miller, 1995). These datasets emphasize the importance of high inter-annotator agreement to ensure reliability in training data.

The introduction of Large Language Models (LLMs) has marked a significant milestone in AI and machine learning, reshaping how sentiment analysis is performed. These models, leveraging architectures like transformers, have set new benchmarks in processing and understanding natural language (Vaswani et al., 2017). However, while LLMs excel in many NLP tasks, their application in sentiment analysis, particularly in nuanced entity-specific cases, is still evolving (Zhang et al., 2023; Luo & Gong, 2024).

## 3. DATA

DOT Loves Data maintains a comprehensive archive named 'The Pressroom,' which contains over 10 million articles from publicly accessible mainstream media platforms, dating back to 2005. This vast repository serves both as a resource for reporting current events and as a tool for analyzing trends in media reporting and public sentiment Bracewell (2022).

For this study, a subset of 42,300 articles was selected from notable New Zealand news outlets: Stuff, NZ Herald, RNZ, The Spinoff, Newsroom, and Newshub. The selected articles comprise various data points including article ID, headline, URL, outlet, article body, and multiple columns indicating the emotional and sentimental tones of the content, based on the methodologies explored by Bracewell (2022).

The selection of these outlets was strategically made to encompass a spectrum of political alignments and levels of editorial freedom, providing a balanced view across the political landscape. For instance, the NZ Herald is recognized for its center-right stance (NZ Library, 2023), while The Spinoff is known for its liberal and autonomous editorial approach (The Spinoff, 2024). This diverse selection aims to minimize bias and enhance understanding of how sentiment varies across different media perspectives.

| Outlet | N. Articles | Percentage |
|---|---|---|
| Newshub | 2,466 | 5.83% |
| Newsroom | 2,032 | 4.80% |
| NZ Herald | 11,731 | 27.7% |
| Radio NZ | 6,750 | 16.0% |
| The Spinoff | 608 | 1.44% |
| Stuff | 18,713 | 44.2% |

Table 1: Distribution of articles in the dataset by outlet.

To assess model performance, we randomly selected 1,000 sentences and manually annotated them as 'neutral,' 'positive,' or 'negative.' This sampling employed proportional stratified sampling techniques to maintain representative proportions from each news outlet. The breakdown of these annotations showed that 46.6% (458 sentences) were neutral, 24.6% (242 sentences) were positive, and 28.8% (283 sentences) were negative, reflecting a diverse range of sentiments across the sampled content.

## 4. METHODS

Exploratory Data Analysis (EDA) in Natural Language Processing (NLP) involves examining data to identify patterns, detect anomalies, and filter out irrelevant or inconsistent data. Much of the NLP data, derived from web scraping, usually does not contain many missing values. However, it does include a substantial amount of irrelevant content, like ads. For example, a significant portion of articles from the news outlet 'Stuff', especially those with fewer than 50 words, were found to contain such irrelevant content. Upon analysis, these articles were manually reviewed and removed to ensure the quality and relevance of the data used for further analysis.

The EDA process also helps understand how sentiment varies across different articles, reflecting the distinct editorial styles of various news outlets. For instance, the sentiment analysis revealed that the end of articles from 'Stuff' typically showed a drop in sentiment, primarily due to standard copyright statements that inherently carried negative sentiment. This kind of insight is crucial for adjusting sentiment analysis algorithms to account for structural idiosyncrasies in the data.

In terms of data pre-processing for NLP, it involves critical steps like lemmatization, which simplifies words to their base form, and tokenization, which breaks down text into manageable pieces. This simplification is vital for efficiently processing large volumes of text. Additionally, removing unnecessary punctuation and stop words (while retaining those crucial for maintaining context) helps in refining the dataset for more accurate analysis.

Part-of-speech tagging is another fundamental task in NLP pre-processing, where each word is tagged with its corresponding grammatical role, aiding in the structured analysis of texts. This tagging is essential for tasks like sentiment analysis, which rely on understanding the grammatical and contextual nuances of language.

The sentiment analysis itself is performed using sophisticated methods that include dictionary look-ups and contextual adjustments to accurately gauge the sentiment expressed in the text. These methods utilize both traditional and augmented dictionary approaches to provide a comprehensive analysis of sentiment across various texts.

Finally, selecting the appropriate dictionary for sentiment analysis is crucial as it significantly affects the outcomes. Various dictionaries, each with different strengths and weaknesses, are considered to ensure that the sentiment analysis is both accurate and reflective of the actual sentiment expressed in the texts.

## 5. RESULTS AND DISCUSSION

The sentiment analysis algorithm assigns continuous values to sentences, which necessitates categorizing these values into sentiment classes (positive, negative, neutral) based on the dictionary used. For instance, a neutral sentiment in Jocker and Rinker's dictionary ranges from [−0.05, 0.1), while in SenticNet it ranges from [−0.05, 0.2) and in SentiWordNet from [−0.1, 0.1). The performance of these dictionaries varies, as captured in Table 2, which provides a detailed breakdown of accuracy and F1 scores across different sentiment classes for each dictionary.

| Dictionary | Positive | F1 | Negative | F1 | Neutral | F1 | Total | F1 |
|---|---|---|---|---|---|---|---|---|
| J & R (lemmas) | 53.8% | 52.9% | 59.6% | 59.0% | 53.6% | 62.4% | 59.9% | 59.4% |
| J & R (tokens) | 44.7% | 50.5% | 44.1% | 44.6% | 37.6% | 41.6% | 50.8% | 53.1% |
| SentiWordNet | 45.8% | 25.2% | 33.3% | 52.6% | 55.0% | 46.5% | 44.1% | 55.4% |
| SenticNet | 37.9% | 42.8% | 25.2% | 37.9% | 42.8% | 44.7% | 37.9% | 50.5% |

Table 2: Performance Metrics for Different Dictionaries

The data shows that Jocker and Rinker's dictionary performs better with raw tokenized sentences than its lemmatized versions. This finding underscores the impact of text processing techniques on the effectiveness of sentiment analysis.

Entity sentiment analysis reveals significant variations in sentiment polarity among different individuals and groups mentioned in the media. Table 3 lists the top 15 entities with the highest and lowest sentiment scores using Jocker and Rinker's dictionary.

Table 3 presents the top 15 entities with the highest and lowest sentiment polarity using Jocker & Rinker's dictionary. Notable outliers include reporters such as Suzanne McFadden, Jonathan Milne, and Marc Daalder, whose coverage in their respective news outlets consistently conveys positive sentiment. Conversely, many of the entities with the lowest sentiment scores are convicted felons or victims of crimes, reflecting negative associations in the media coverage.

The entities with the highest sentiment scores are particularly interesting. They include New Zealand sports players like Ali Riley, [Jitka] Klimková, Hannah Wilkinson, Steve Hansen, and Shane van Gisbergen, suggesting that sports figures are often portrayed with a high degree of positive sentiment. Additionally, the late Tony Bennett, an iconic musical artist highly regarded in the industry who passed away in July 2023, also ranks among the top for positive sentiment. This indicates a strong positive bias in how sports stars and influential cultural figures are depicted in the media.

Table 3 lists the top entities with the highest and lowest sentiment polarity mentioned in news outlets over the past six months. Each entity listed has been mentioned at least 200 times, including public figures such as Dan Carter, Donald Trump, and Justin Bieber.

| Entity | Mean | Median | SD | 95% C.I. | N |
|---|---|---|---|---|---|
| Ramsay | -0.149 | -0.150 | 0.320 | (-0.181, -0.117) | 378 |
| Jaz | -0.125 | -0.139 | 0.297 | (-0.157, -0.092) | 317 |
| Lucy Letby | -0.113 | -0.144 | 0.318 | (-0.147, -0.078) | 324 |
| Matu Tangi Matua Reid | -0.112 | -0.135 | 0.272 | (-0.143, -0.08) | 289 |
| Jodie Shannon Hughes | -0.109 | -0.124 | 0.320 | (-0.146, -0.072) | 287 |
| Lauren Dickason | -0.108 | -0.129 | 0.313 | (-0.121, -0.095) | 2264 |
| Liané Letby | -0.108 | -0.129 | 0.326 | (-0.146, -0.07) | 284 |
| Erin Patterson | -0.106 | -0.143 | 0.320 | (-0.139, -0.072) | 357 |
| Jacob Mills Ramsay Reid | -0.105 | -0.125 | 0.261 | (-0.133, -0.077) | 333 |
| Lauren Anne Dickason | -0.104 | -0.127 | 0.307 | (-0.147, -0.062) | 202 |
| Roberto Jaz | -0.101 | -0.133 | 0.297 | (-0.122, -0.079) | 725 |
| Danny Jaz | -0.101 | -0.125 | 0.271 | (-0.125, -0.076) | 473 |
| Donald Trump | -0.098 | -0.120 | 0.294 | (-0.129, -0.068) | 360 |
| Grant Hansen | -0.097 | -0.146 | 0.306 | (-0.132, -0.062) | 292 |
| Dan Carter | -0.095 | -0.125 | 0.313 | (-0.129, -0.06) | 319 |
| Shane van Gisbergen | 0.158 | 0.167 | 0.281 | (0.122, 0.193) | 241 |
| Justin Bieber | 0.159 | 0.182 | 0.309 | (0.122, 0.196) | 267 |
| Matthew Scott Riley | 0.160 | 0.177 | 0.328 | (0.111, 0.209) | 175 |
| Tony Bennett | 0.169 | 0.178 | 0.269 | (0.132, 0.206) | 206 |
| Paul | 0.170 | 0.171 | 0.264 | (0.135, 0.206) | 215 |
| Hannah Wilkinson | 0.174 | 0.161 | 0.256 | (0.064, 0.285) | 23 |
| Klimková | 0.181 | 0.267 | 0.337 | (0.135, 0.227) | 207 |
| Ali Riley | 0.182 | 0.206 | 0.257 | (0.148, 0.217) | 213 |
| Marc Daalder | 0.183 | 0.183 | 0.322 | (0.144, 0.221) | 269 |
| Suzanne McFadden | 0.183 | 0.177 | 0.357 | (0.145, 0.221) | 339 |
| Jonathan Milne | 0.188 | 0.196 | 0.293 | (0.16, 0.215) | 439 |
| Matariki | 0.196 | 0.194 | 0.290 | (0.157, 0.235) | 216 |
| Justin Bieber | 0.213 | 0.222 | 0.252 | (0.183, 0.244) | 268 |
| Tony Bennett | 0.239 | 0.267 | 0.356 | (0.193, 0.284) | 236 |
| Ali Riley | 0.275 | 0.267 | 0.279 | (0.241, 0.309) | 354 |

Table 3: Top 15 most/least sentiment polarity scores for entities mentioned in news outlets from the past six months, given they have been mentioned at least 200 times. Also includes Dan Carter, Donald Trump and Justin Bieber.

The sentiment scores are derived from Jocker & Rinker's dictionary and reflect the mean, median, and standard deviation (SD) of sentiment values attributed to each entity. The count (N) reflects the number of times each entity was mentioned in the dataset.

Our sentiment analysis results indicate that Matariki has the highest sentiment score among all entities analyzed. Matariki, often regarded as the Māori New Year, symbolizes the reappearance of a star cluster that is visible for 11 months before it reappears in late June or early July, heralding the start of a new year. This event holds profound cultural significance in New Zealand. According to Bracewell (2022), an analysis of social media revealed distinct peaks in expressions of New Zealand self-identity coinciding with the Matariki celebration period in June and July. This underscores the deep cultural resonance and widespread recognition of Matariki as a cornerstone of New Zealand's national identity.

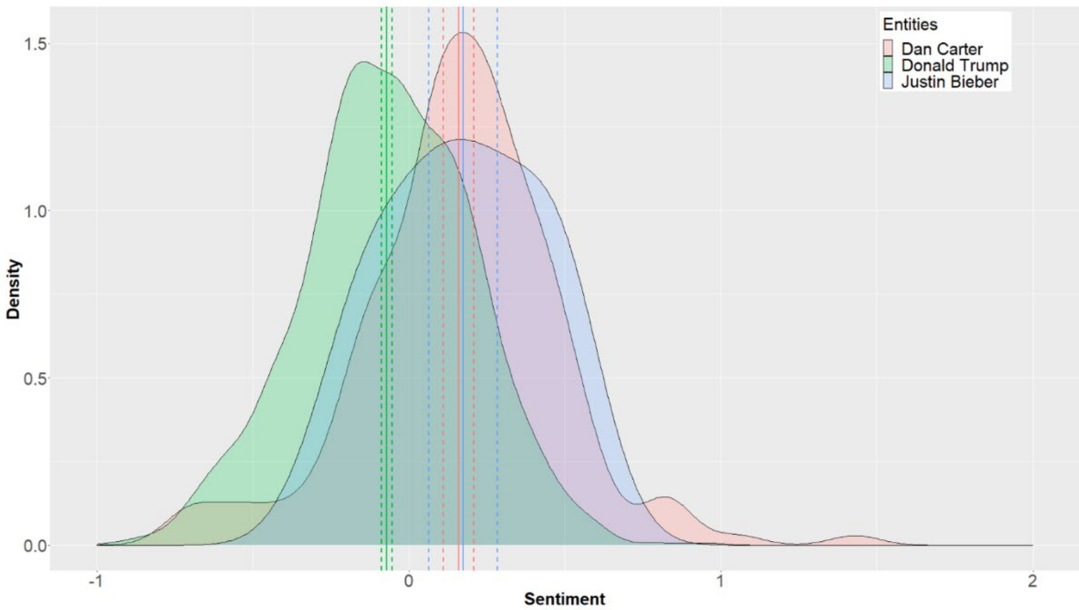



Figure 7: Example output of the sentiment analysis algorithm: Donald Trump vs Dan Carter vs Justin Bieber. Bold solid lines are the mean, and the small dotted lines are the 95% confidence intervals of the mean. The dictionary used is tokenized J & R.

In many instances, entities may display extremely high or low sentiment scores (values exceeding ±1), often because they are mentioned only once in the entire dataset, and that particular mention carries a significantly charged sentiment. This phenomenon affects both lesser-known entities and high-profile individuals alike. For example, Justin Bieber, despite his celebrity status, is mentioned relatively infrequently in the current media landscape—only 23 times, including references via pronouns, throughout our dataset. Consequently, his sentiment analysis might yield skewed results due to these few mentions. To address this variability and provide more reliable interpretations, we employ 95% confidence intervals where the standard deviation is not well-defined. We use the Student's t-distribution for calculating these intervals, ensuring a statistically robust approach to understanding sentiment trends across entities with limited mentions.

## 6. CONCLUSIONS

Here, we evaluated the effectiveness of various dictionaries used in sentiment analysis with a fixed algorithm. Our findings indicate that the dictionary by Jocker and Rinker, which integrates Jockers's 2015 dictionary and Hu & Liu's 2004 augmented dictionary, outperforms others, even without lemmatization. Other dictionaries assessed include SentiWordNet (Baccianella et al., 2010) and SenticNet (Cambria et al., 2016), as well as the lemmatized version of Jocker and Rinker's dictionary. Despite successful applications in different models, these dictionaries did not perform as well in our study.

Additionally, the study explores sentiment polarity among New Zealanders, identifying a trend where sports players generally receive higher sentiment ratings compared to others. Notably, no individuals received lower sentiment scores than those convicted of felonies. The entity with the highest sentiment was Matariki, correlating with its occurrence during the dataset's timeframe and its cultural significance in New Zealand.

The analysis also highlighted differences in sentiment expression across six New Zealand news outlets: Newshub, Newsroom, NZ Herald, Radio NZ, Stuff, and The Spinoff. Among these, The Spinoff exhibited sentences with higher sentiment polarity, whereas Newsroom showed the least.

This research acknowledges several limitations. Primarily, the model has not been trained on a dataset but rather demonstrates the efficacy of dictionaries within set parameters. To enhance model performance, future efforts should focus on extensive evaluation and training, including the use of a validation set to fine-tune hyperparameters. Our results indicate that most dictionaries struggle with classifying positive and negative sentiments accurately. Following the recommendations of Lazrig & Humpherys (2022), who advocate for excluding neutral sentiments to improve model accuracy, we might consider similar adjustments.

Another improvement area is the annotation process. Currently, annotations are conducted by a single individual, which could introduce bias. Typically, using multiple annotators helps mitigate personal bias and enhances the reliability of data (Rottger et al., 2022). If increasing the number of annotators is not feasible, employing a pre-existing annotated dataset could be an alternative, although such datasets are rare in news journalism.

Further analysis should focus on comparing sentiment across individual news outlets, expanding on initial observations regarding writing styles and potential clickbait tendencies. This could help identify which outlets tend to exhibit the most or least sentiment polarity.

Moving forward, integrating the sentiment analysis algorithm into a PyTorch framework for enhanced training and evaluation is a logical step. This approach will allow for the optimization of additional hyperparameters, such as adjusting the lexical scope to improve sentiment detection accuracy.

Moreover, revisiting the impact of coreference resolution on the accuracy of entity sentiment analysis could yield valuable insights. A comparative study to assess the effectiveness of models with and without coreference resolution could further refine our understanding of its benefits. By addressing these areas, we can significantly advance the accuracy and applicability of sentiment analysis tools in understanding and quantifying media sentiment.

## References


Al-Shabi, M. (2020). Evaluating the performance of the most important lexicons used to sentiment analysis and opinions mining. International Journal of Computer Science and Network Securiy, 20, 51–57.

Atkinson, J. & Escudero, A. (2022). Evolutionary natural-language coreference resolution for sentiment analysis. International Journal of Information Management Data Insights, 2(2), 100115.

Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, & D. Tapias (Eds.), Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10) Valletta, Malta: European Language Resources Association (ELRA).

Bohnet, B., Alberti, C., & Collins, M. (2023). Coreference resolution through a seq2seq transition-based system. Transactions of the Association for Computational Linguistics, 11, 212–226.

Bracewell, P. J. (2022). Quantifying the impact of olympics on the New Zealand national identity in mainstream media. In Proceedings of the 16th Australian Conference on Mathematics and Computers in Sports (pp. 55–60).: ANZIAM Mathsport.

Bracewell, P. J., Wells, J. D., & Craig, A. (2022). Detecting change in tone of tweets and the potential implications for monitoring mental health of athletes. In Proceedings of the 16th Australian Conference on Mathematics and Computers in Sports (pp. 48–54).: ANZIAM Mathsport.

Cambria, E., Poria, S., Bajpai, R., & Schuller, B. (2016). SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives. In Y. Matsumoto & R. Prasad (Eds.), Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers (pp. 2666–2677). Osaka, Japan: The COLING 2016 Organizing Committee.

De Clercq, O. & Hoste, V. (2020). It's absolutely divine! can fine-grained sentiment analysis benefit from coreference resolution? In M. Ogrodniczuk, V. Ng, Y. Grishina, & S. Pradhan (Eds.), Proceedings of the Third Workshop on Computational Models of Reference, Anaphora and Coreference (pp. 11–21). Barcelona, Spain (online): Association for Computational Linguistics.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In J. Burstein, C. Doran, & T. Solorio (Eds.), Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) (pp. 4171–4186). Minneapolis, Minnesota: Association for Computational Linguistics.

D'Oosterlinck, K., Bitew, S. K., Papineau, B., Potts, C., Demeester, T., & Develder, C. (2023). CAW-coref: Conjunction-aware word-level coreference resolution. In M. Ogrodniczuk, V. Ng, S. Pradhan, & M. Poesio (Eds.), Proceedings of The Sixth Workshop on Computational Models of Reference, Anaphora and Coreference (CRAC 2023) (pp. 8–14). Singapore: Association for Computational Linguistics.

Durrett, G. & Klein, D. (2014). A joint model for entity analysis: Coreference, typing, and linking. Transactions of the Association for Computational Linguistics, 2, 477–490.

Entman, R. M. (1989). How the media affect what people think: An information processing approach. The Journal of Politics, 51(2), 347–370.
Ghag, K. & Shah, K. (2014). SentiTFIDF – Sentiment Classification using Relative Term Frequency Inverse Document Frequency. International Journal of Advanced Computer Science and Applications, 5(2).
Gong, X., Ying, W., Zhong, S., & Gong, S. (2022). Text sentiment analysis based on transformer and augmentation. Frontiers in Psychology, 13.
Honnibal, M., Montani, I., Van Landeghem, S., & Boyd, A. (2020). spaCy: Industrial-strength Natural Language Processing in Python.
Hu, M. & Liu, B. (2004). Mining and summarizing customer reviews. In Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '04 (pp. 168–177). New York, NY, USA: Association for Computing Machinery.
Hutto, C. & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. Proceedings of the International AAAI Conference on Web and Social Media, 8(1), 216–225.
Jockers, M. L. (2015). Syuzhet: Extract sentiment and plot arcs from text. https://github.com/mjockers/syuzhet.
Jung, A.-K., Stieglitz, S., Kissmer, T., Mirbabaie, M., & Kroll, T. (2022). Click me. . . ! the influence of clickbait on user engagement in social media and the role of digital nudging. PLOS ONE, 17(6), 1–22.
Lazrig, I. & Humpherys, S. (2022). Using machine learning sentiment analysis to evaluate learning impact. Information Systems Education Journal, 20, 13–21.
Ligthart, A., Catal, C., & Tekinerdogan, B. (2021). Systematic reviews in sentiment analysis: a tertiary study. Artificial Intelligence Review, 54, 4997–5053.
Luo, M. & Mu, X. (2022). Entity sentiment analysis in the news: A case study based on negative sentiment smoothing model (nssm). International Journal of Information Management Data Insights, 2(1), 100060.
Luo, W. & Gong, D. (2024). Pre-trained large language models for financial sentiment analysis. Miller, G. A. (1995). Wordnet: A lexical database for english. Commun. ACM, 38(11), 39–41.
Musto, C., Semeraro, G., & Polignano, M. (2014). A comparison of lexicon-based approaches for sentiment analysis of microblog. CEUR Workshop Proceedings, 1314, 59–68.
Ng, V. (2017). Machine learning for entity coreference resolution: A retrospective look at two decades of research. Proceedings of the AAAI Conference on Artificial Intelligence, 31(1).
NZ Library (2023). Background: The New Zealander. Papers Past, . https://paperspast.natlib.govt.nz/newspapers/new-zealand-herald.
Otmazgin, S., Cattan, A., & Goldberg, Y. (2022). F-coref: Fast, accurate and easy to use coreference resolution. In W. Buntine & M. Liakata (Eds.), Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing: System Demonstrations (pp. 48–56). Taipei, Taiwan: Association for Computational Linguistics.
Passonneau, R. J., Salleb-Aoussi, A., Bhardwaj, V., & Ide, N. (2010). Word sense annotation of polysemous words by multiple annotators. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, & D. Tapias (Eds.), Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10) Valletta, Malta: European Language Resources Association (ELRA).
Rinker, T. W. (2021). sentimentr: Calculate text polarity sentiment. version 2.9.1. 23
Rodrı́guez-Ibáñez, M., Casáñez-Ventura, A., Castejón-Mateos, F., & Cuenca-Jiménez, P.-M. (2023). A review on sentiment analysis from social media platforms. Expert Systems with Applications, 223, 119862.
Rottger, P., Vidgen, B., Hovy, D., & Pierrehumbert, J. (2022). Two contrasting data annotation paradigms for subjective NLP tasks. In M. Carpuat, M.-C. de Marneffe, & I. V. Meza Ruiz (Eds.), Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 175–190). Seattle, United States: Association for Computational Linguistics.
Samuels, A. & Mcgonical, J. (2020). News sentiment analysis.
Schweinberger, M. (2021). Sentiment analysis in r. https://slcladal.github.io/sentiment.html.
Shridhar, K., Monath, N., Thirukovalluru, R., Stolfo, A., Zaheer, M., McCallum, A., & Sachan, M. (2023). Longtonotes: OntoNotes with longer coreference chains. In A. Vlachos & I. Augenstein (Eds.), Findings of the Association for Computational Linguistics: EACL 2023 (pp. 1428–1442). Dubrovnik, Croatia: Association for Computational Linguistics.
Silva, M. D. (2023). Preprocessing Steps for Natural Language Processing (NLP): A Beginner's Guide.
Tabinda Kokab, S., Asghar, S., & Naz, S. (2022). Transformer-based deep learning models for the sentiment analysis of social media data. Array, 14, 100157.
The Spinoff (2024). About The Spinoff. https://thespinoff.co.nz/about-us.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need.
Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods. Artificial Intelligence Review, 55, 5731–5780.
Weischedel, R., Marcus, M., Palmer, M., Hovy, E., Belvin, R., Pradhan, S., & Ramshaw, L. (2010). Ontonotes: A large training corpus for enhanced processing. In Handbook of Natural Language Processing and Machine Translation. New York, NY: Springer.
Williams, J. (2013). Puaka and matariki: The māori new year. The Journal of the Polynesian Society, 122(1), 7–19.

Zhang, W., Deng, Y., Liu, B., Pan, S. J., & Bing, L. (2023). Sentiment analysis in the era of large language models: A reality check.

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., Liu, P., Nie, J.-Y., & Wen, J.-R. (2023). A survey of large language models.

# USING EVENT DATA TO CHARACTERISE POSSESSION CHAINS IN ELITE SOCCER MATCH-PLAY

Jack Lee[a,b,c], Matthew Varley[a], Andrew Govus[a], David Carey[a]

[a] *La Trobe University, Melbourne, Victoria, Australia*
[b] *Melbourne City Football Club, Melbourne, Victoria, Australia*
[c] *Corresponding author: jack.lee@latrobe.edu.au*

**Submitted for consideration as:**

- ☒ Oral presentation
- ☐ Poster

**Abstract**

Teams in soccer implement strategies to secure a positive match outcome. As a result, player and team performance has been quantified using aggregated match statistics. Extensive research has been conducted on aggregated attacking metrics, where increased goals scored, shots, passes, time in ball possession and more in-depth statistics such as expected goals generally increase the likelihood of a positive match outcome. Alternatively, analysis has also been conducted on aggregated defensive metrics such as duels, tackles, and interceptions to quantify performance. However, these aggregated metrics can be difficult to apply directly to coaching practice and typically show large match-to-match variation due to differing opposition strategies and styles of play. To gain more in-depth and coachable insights, it is essential to understand the sequence of actions and events of a team once they gain possession to when they lose possession known as a possession chain. By analysing the outcomes, duration, number of passes and meterage gained of each unique possession chain, an understanding of the strategies of each team can be concluded. Play-by-play transactional event data from the 2020/21, 2021/22 and 2022/23 English Premier League, German Bundesliga, Spanish La Liga, Italian Serie A and French Ligue 1 seasons was analysed. The results and methods provide a framework to analyse both attacking and defensive performance at possession chain level to gain a deeper understanding into differing styles of play and their relationship with further in-match success or match outcome.

*Keywords*: Soccer, performance analysis, possession, possession chain, success

**Acknowledgements**

We wish to thank City Football Group; and StatsBomb for providing the appropriate data.

**References**

Castellano, J., Casamichana, D., & Lago, C. (2012). The use of match statistics that discriminate between successful and unsuccessful soccer teams. *Journal of human kinetics*, *31*, 139.

Forcher, L., Altmann, S., Forcher, L., Jekauc, D., & Kempe, M. (2022). The use of player tracking data to analyze defensive play in professional soccer-A scoping review. *International journal of sports science & coaching*, *17*(6), 1567-1592.

Forcher, L., Beckmann, T., Wohak, O., Romeike, C., Graf, F., & Altmann, S. (2023). Prediction of defensive success in elite soccer using machine learning - Tactical analysis of defensive play using tracking data and explainable AI. *Sci Med Footb*, 1-16. https://doi.org/10.1080/24733938.2023.2239766.

Goes, F. R., Brink, M. S., Elferink-Gemser, M. T., Kempe, M., & Lemmink, K. A. (2021). The tactics of successful attacks in professional association football: large-scale spatiotemporal analysis of dynamic subgroups using position tracking data. *Journal of sports sciences*, *39*(5), 523-532.

Winter, C., & Pfeiffer, M. (2016). Tactical metrics that discriminate winning, drawing and losing teams in UEFA Euro 2012®. *Journal of sports sciences*, *34*(6), 486-492.

Wright, C., Atkins, S., Polman, R., Jones, B., & Sargeson, L. (2011). Factors associated with goals and goal scoring opportunities in professional soccer. *International Journal of Performance Analysis in Sport*, *11*(3), 438-449.

# Unveiling the Mental Game: Leveraging YODA Psychometry in Football Performance Analysis

Sadanand Venkataraman [a,c], Sundharakumar K B[a], Bharathi Malakreddy A[b], Santhi Natarajan[a]
[a] *Shiv Nadar University Chennai, India*
[b] *Research & Consultancy Cell, BMS Institute of Technology & Management, Bengaluru*
[c] *Corresponding author:* saadanandv@gmail.com

## Abstract

In the realm of professional soccer, the confluence of mental agility and physical skill dictates the success of players on the field. This study introduces a groundbreaking approach, "Your Offence and Defence Analysis" (YODA), to delve into the often-neglected psychological aspect of soccer players. YODA, a psychometric tool, immerses players in simulated match scenarios, meticulously mapping their psychological responses to a range of traits. These traits, encompassing cognitive patterns and emotional resilience, are then analyzed through advanced AI and data analytics methodologies.

Our research employed YODA to examine soccer players from varying competitive levels. The pilot cohort of YODA includes over 120 players, ranging from amateur players to seasoned veterans of the sport. This comprehensive analysis illuminates the cognitive landscapes of players, offering a nuanced understanding of their impact on in-game performance. YODA translates intricate psychological data into practical, performance-enhancing insights for coaches and analysts alike. These insights were used to train an Indian College soccer team over a season. A significant improvement was seen throughout the course of the season amongst players as well as the overall team. Post YODA, the team had their best season in over 20 years, being crowned the State Champions at the end of the season.

This exploration, the first of its kind, has profound implications for soccer coaching and team management. It advocates for a more personalized coaching approach, attuned to the unique psychological makeup of each player, thereby revolutionizing traditional training methodologies. Our findings contribute to a richer, more holistic understanding of player performance, extending beyond physical metrics to include the pivotal role of mental attributes in soccer.

**Keywords**: Football, Psychometry, Performance Analysis, Cognitive Patterns

## 1. INTRODUCTION

Football, often referred to as "the beautiful game," transcends mere athletic competition, embodying a rich tapestry of cultural, social, and economic narratives. It has evolved from simple kickabouts to a global phenomenon, showcasing a complex blend of strategy, skill, and physical prowess. Football Analytics has become an integral part of this evolution, providing insights that refine tactics, boost player performance, and inform strategic decisions. However, traditional analytics mainly focus on quantifiable aspects like physical stamina, tactical execution, and technical skills, frequently neglecting the psychological dimensions that significantly affect in-game performance.

The importance of psychological factors in football cannot be overstated. Mental agility, decision-making under pressure, emotional stability, and resilience are pivotal elements [1] that determine the success of players and teams. These aspects, though less visible than physical attributes, often dictate the outcome of matches by influencing players' ability to cope with stress, maintain concentration, and execute strategies effectively. Research in sports psychology suggests that the mental state of a player is as crucial as physical fitness [2] in achieving peak performance, emphasizing the need for a holistic approach to player analysis [1], [2] and development.

Enter "Your Offence and Defence Analysis" (YODA), a groundbreaking psychometric tool crafted to bridge this gap in sports analytics. YODA is designed to delve into the mental corridors of football players, assessing a spectrum of psychological traits through simulated match scenarios. This innovative approach enables a comprehensive mapping of players' cognitive patterns, emotional resilience, and decision-making capabilities, offering a nuanced understanding of their psychological impact on performance.

This paper's primary objective is to examine the effectiveness of YODA in boosting football performance by analyzing players' psychological foundations. We conducted a methodical assessment of a diverse cohort from various competitive levels to uncover the correlation between psychological traits and on-field performance.

This paper unfolds the journey of YODA's inception, its methodological framework, the empirical analysis of the initial cohort, and case studies that highlight its application. Subsequent sections will detail the results, discussing the distribution and implications of psychological traits and personality patterns within the cohort. Two Players' analysis through YODA is also showcased in this paper. The discussion section explores the ramifications of these findings for coaching strategies and player development, culminating in a conclusion that reflects on the future trajectory of YODA in the realm of sports analytics and psychological assessment in football.

## 2. YODA METHODOLOGY

FRAMEWORK OF YODA

The YODA methodology encapsulates the psychological intricacies of football players. It merges traditional sports science with advanced psychometric analysis [5], [6], [7]. This multidisciplinary approach integrates cognitive, behavioural, and emotional assessments [7] to provide a holistic view of a player's mental framework (Figure 1). We base the framework on simulated real-game scenarios to draw out authentic psychological reactions. This structure aids in deeply understanding players' mental attributes, such as decision-making, pressure management, and emotional control.

YODA TRAITS & SUBTRAITS

YODA evaluates a comprehensive range of traits and subtraits, each category is meticulously crafted to align with the demands and situations encountered in football. We map a player's traits based on their responses to each scenario. These traits and their detailed descriptions appear in Table 1.

THE YODA ASSESSMENT

The assessment through YODA is a meticulously crafted process. It revolves around four pillars, each designed to shed light on specific traits and their corresponding sub-traits:

- Ethics and In-Game Decisions,
- Team Sense,
- Emotional State, and
- Lifestyle

The tool immerses the candidate in 48 varied situations, each meticulously crafted to gauge their reactions. These scenarios, reflective of real-game situations, offer invaluable insights into a player's mental orientation and potential on-field reactions.

YODA integrates the esteemed Big Five personality traits as listed in Table 2—Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism—into its assessment framework. Each personality trait helps analyze specific psychometric characteristics relevant to football. For example, a player's level of 'Openness' may affect their adaptability to different playing styles, while 'Conscientiousness' might reflect their commitment to training and strategy.

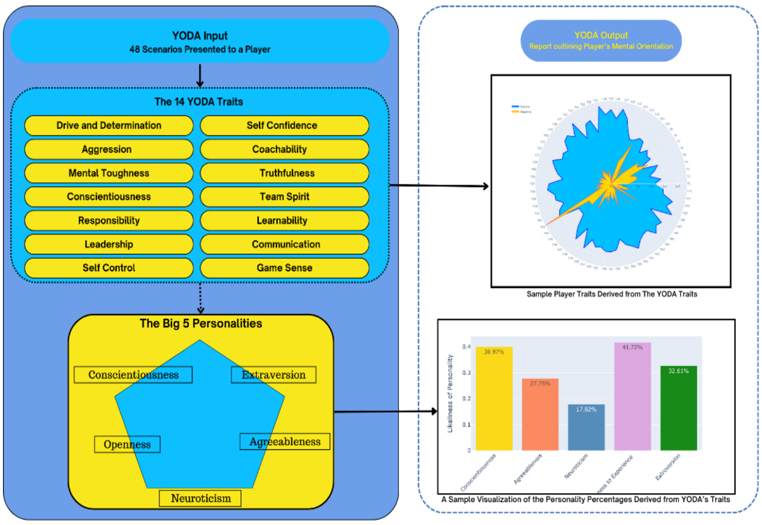



Figure 1: The YODA Framework

Table 1: The YODA Traits and their Descriptions

| Trait Number | Trait | Description |
|---|---|---|
| 1 | Drive and Determination | The relentless motivation to meet objectives and overcome obstacles, characterized by persistent effort and a strong desire to succeed in the face of challenges. |
| 2 | Aggression | A controlled form of behavior aimed at asserting dominance or achieving a goal, without intent to harm, essential for competitive scenarios in football. |
| 3 | Mental Toughness | The capacity to consistently perform at a high level, regardless of pressures or competitive circumstances, showing resilience and focus. |
| 4 | Conscientiousness | A trait depicting one's commitment to discipline, planning, and goal-oriented behavior, ensuring reliability and precision in performance. |
| 5 | Responsibility | The readiness to initiate actions and be accountable for them, driving personal and team success through proactive leadership. |
| 6 | Leadership | The ability to guide and inspire teammates towards common objectives, fostering unity and collective effort for team success. |
| 7 | Self-Control | The regulation of one’s emotions and behaviors, maintaining focus and composure in various game situations. |
| 8 | Self Confidence | The belief in one's skills and judgment, crucial for making decisive actions and maintaining a positive mindset in competition. |
| 9 | Coachability | The openness to receive and act on feedback, demonstrating a commitment to personal improvement and adaptability. |
| 10 | Truthfulness | The capacity to be honest and express genuine feelings and experiences, enhancing team trust and communication. |
| 11 | Team Spirit | A cooperative and enthusiastic commitment to working with others towards shared goals, embodying unity and camaraderie. |
| 12 | Learnability | The ability to quickly absorb and apply new knowledge and skills, essential for adapting to evolving game strategies and roles. |
| 13 | Communication | The effective exchange of information, both verbally and non-verbally, crucial for strategic execution and team synergy. |
| 14 | Game Sense | The understanding of game dynamics and the ability to make smart decisions, reflecting a player’s strategic acumen and situational awareness. |

## 3. INITIAL COHORT

INSTRUMENT

The YODA assessment utilized a detailed questionnaire to gauge the football-relevant psychological traits and subtraits of players. Starting with 45 scenarios, the tool expanded to 48 after consultations with an AIFF-associated coach. These scenarios, mirroring real-game conditions, prompted players to respond naturally, reflecting their instinctive reactions and decisions. Developed as a PEBL Battery, the instrument offered ease in administration and precise response capture, with answers rated on a Likert scale from "Strongly Agree" to "Strongly Disagree" to quantify each trait.

CANDIDATES

The initial cohort comprised 68 football players, representing diverse competition levels and experiences [8], including amateur, youth, and semi-professional leagues. Participants came from institutions like BMS Institute of Technology & Management, Bengaluru, and TKM College of Engineering, Kollam, among others, providing a rich insight spectrum into the players' psychological landscape. We collected demographic data, such as age, playing position, and experience, to assess its impact on the psychological traits measured by YODA.

REGISTRATION & CONSENT

We facilitated study engagement through dialogues with team captains, coaches, and administrators, clearly stating the research goals and inviting player participation. Players then signed consent forms before their YODA sessions, ensuring ethical compliance and informed involvement.

PROCEDURE

Participants supplied initial metadata like name, age, experience, and preferred position before their YODA session. This was followed by a comprehensive briefing on the session's structure and goals, preparing the

Table 2: The Big 5 Personality Traits

| Personality Trait | Description |
|---|---|
| Openness | Open individuals are curious about both inner and outer worlds, and their lives are experientially richer. They are willing to entertain novel ideas and unconventional values, and they experience both positive and negative emotions more keenly than closed individuals. Openness to Experience includes active imagination, aesthetic sensitivity, attentiveness to feelings, a preference for variety, intellectual curiosity and independence of judgment. |
| Agreeableness | An agreeable person is fundamentally altruistic, sympathetic to others and eager to help them, and in return believes that others will be equally helpful. The disagreeable/antagonistic person is egocentric, sceptical of others' intentions, and competitive rather than cooperative. |
| Conscientiousness | The conscientious person is purposeful, strong-willed, and determined. Conscientiousness is manifested in achievement orientation (hardworking and persistent), dependability (responsible and careful) and orderliness (planful and organised). |
| Neuroticism | Neuroticism is a dimension of normal personality indicating the general tendency to experience negative effects such as fear, sadness, embarrassment, anger, guilt and disgust. High scorers may be at risk of some kinds of psychiatric problems. A high Neuroticism score indicates that a person is prone to having irrational ideas, being less able to control impulses, and coping poorly with stress. A low Neuroticism score is indicative of emotional stability. These people are usually calm, even-tempered, relaxed and able to face stressful situations without becoming upset. |
| Extraversion | Extraverts are energetic and optimistic. Introverts are reserved rather than unfriendly, independent rather than followers, and even-paced rather than sluggish. Extraversion is characterised by positive feelings and experiences and is therefore seen as a positive effect. Extraversion includes traits such as sociability, assertiveness, activity and talkativeness. |

participants thoroughly. They then tackled the 48 scenario-based questions in the YODA questionnaire, with sessions recorded both in person and virtually, upon obtaining explicit consent, allowing for an extensive data analysis.

## INITIAL COHORT ANALYSIS

The initial cohort analysis offered a robust data set for exploring the psychological profiles and traits of the 68 football players involved (Figure 2 and Figure 3).

**Demographic Overview**

- **Age Distribution and Trends**: The age range of the cohort was broad, from *11 to 46 years*, with an average age of approximately *19.9 years*, primarily clustering players in their mid-twenties. This age diversity combined youthful vigor with experienced maturity, reflecting the dynamic nature of football.
- **Positional Breakdown**: Analysis showed a predominance of offensive roles, with *strikers representing 28%* of the cohort, underscoring the attacking prowess. *Midfielders accounted for 26%*, crucial for game tempo control and linking defense to offense, while the inclusion of defenders and goalkeepers ensured a balanced team dynamic.
- **Team Affiliation**: The cohort had significant representations from *BMS Institute of Technology & Management*, *Bengaluru*, and *TKM College of Engineering, Kollam*, along with youth players from *Crescent FC* and students from *Christ (Deemed to be) University*, *Mount Carmel College*, highlighting a diverse footballing background.

**Response Analysis Over Traits**

- **Positive Response**: The average positive response rate was *0.68*, indicating a generally optimistic or agreeable attitude among players toward the scenarios presented in the YODA assessment.
- **Neutral Response**: Averaging *0.66*, neutral responses suggested a cautious or balanced approach in decision-making processes among the players.
- **Negative Response**: The average rate of *0.38* for negative responses pointed to areas where players might experience psychological stress or disagreement, highlighting opportunities for targeted mental conditioning and support.

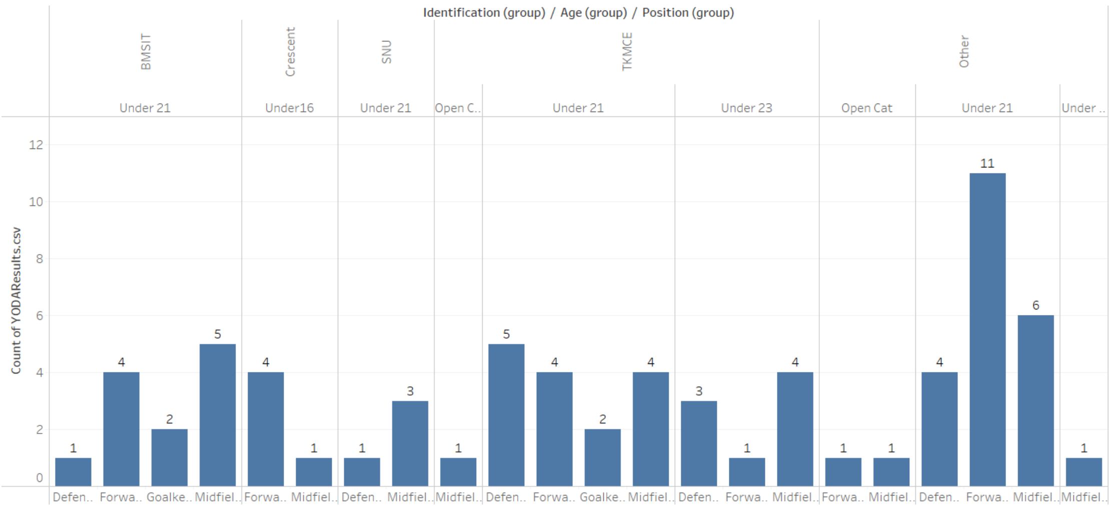



Figure 2: Demographic Overview of Initial Cohort

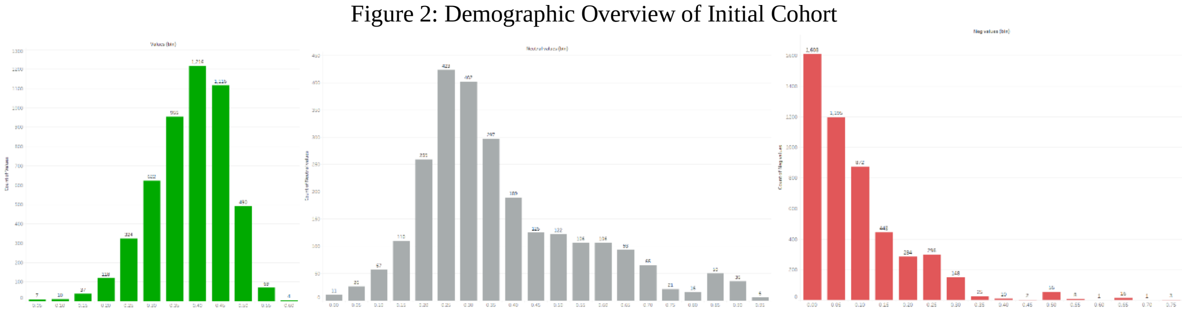


Figure 3: (a) Positive Response Distribution(b) Neutral Response Distribution (c) Negative Response Distribution

**Personality Trait Analysis**

- **Overview of Personality Traits**: The cohort displayed moderate levels of *Conscientiousness and Openness to Experience*, with respective average values of *0.41 and 0.40*, suggesting a disciplined, diligent, and tactically adaptable player base.
- **Correlation with Responses**: A strong correlation was observed between certain personality traits and response patterns. For instance, players with higher Conscientiousness typically had more positive responses, indicating a potential linkage between disciplined behavior, diligent work ethics, and positive psychological orientation.

## 4. CASE STUDY

BACKGROUND AND PROFILE

The cohort analysis, focusing on forward players, provides a nuanced understanding of the psychological profiles pivotal for offensive roles in football. These players, hailing from BMS Institute of Technology & Management, Bengaluru [referred to as BMSIT&M] and TKM College of Engineering, Kollam[referred to as TKMCE], showcase diverse mental attributes essential for the attacking phases of the game. Table 3 contains the metadata collected from this subgroup of the cohort.

The YODA assessment unveiled a multifaceted psychological landscape for these forwards [9], highlighting an array of strengths and potential areas for growth tailored to their pivotal roles in scoring and offensive play aligning with the detailed player profiles explored by Demediuk et al. [9] Players like YODA30 and YODA35 exemplify the cohort's drive and analytical prowess, showcasing a clear-headed understanding of the game, complemented by a robust work ethic and a strategic approach to football.

The Trait and Personality results for each player along with the overall distribution of both for this subgroup is depicted in Figure 4 and Figure 5 respectively, illustrating the individual YODA outcomes for a representative player.

YODA ASSESSMENT RESULTS

During the YODA assessment, all candidates were subjected to 48 scenarios, testing a range of psychological traits. The forwards displayed a varied range of psychological traits, with "Learnability" and "Coachability" emerging as dominant traits. This suggests a strong inclination towards continuous improvement and a receptive

Table 3: Metadata of Forwards from Cohort

| YODAID | Age | Preferred Position | Team | Test Date |
|---|---|---|---|---|
| YODA030 | 19 | Forward | BMSIT&M | October 9, 2021 |
| YODA035 | 18 | Forward | BMSIT&M | October 9, 2021 |
| YODA037 | 18 | Forward | BMSIT&M | October 13, 2021 |
| YODA049 | 22 | Forward | TKMCE | August 25, 2021 |
| YODA055 | 21 | Forward | TKMCE | August 28, 2021 |
| YODA057 | 20 | Forward | TKMCE | August 30, 2021 |
| YODA058 | 20 | Forward | TKMCE | August 30, 2021 |
| YODA064 | 21 | Forward | TKMCE | October 09, 2021 |

attitude to coaching feedback, essential for adapting and excelling in dynamic game situations. “Drive and Determination” and “Responsibility” also scored highly, indicating a relentless pursuit of goals and a proactive approach to on-field responsibilities. Notably, “Aggression” registered lower, hinting at a more calculated and less confrontational style of play among the forwards.

Variability analysis revealed that while some traits like Learnability and Team Spirit (T11) showed consistency, others like Self-Control (T7) and Communication (T13) varied more significantly across the cohort, indicating areas for targeted developmental interventions.

Correlation analysis highlighted the interplay between various psychological aspects, such as the negative correlation between Neutral and Positive responses, underscoring the complexity of mental and emotional dynamics in football performance.

IMPACT ON PERFORMANCE

These psychological traits contribute significantly to the performance of forwards, affecting their decision-making, adaptability, and overall effectiveness in attacking roles. For example:

- High **Learnability** and **Coachability** facilitate rapid tactical adaptation and strategic execution.
- The balanced approach to **Aggression** suggests a strategic method of play, prioritizing intelligent positioning and timing over sheer physical confrontations.

DEVELOPMENTAL INSIGHTS

For optimal development, training programs should focus on enhancing traits like Mental Toughness (T3) and Game Sense (T14). Tailored coaching strategies, emphasizing scenario-based learning and cognitive training, can further refine their decision-making and situational awareness.

FUTURE IMPLICATIONS

These findings underscore the importance of psychological profiling in developing forward players. By leveraging insights from YODA, coaches can craft personalized development plans that address individual psychological needs, enhance team dynamics, and optimize on-field performance.

This comprehensive analysis of forward players in the cohort through YODA psychometry provides valuable insights into the psychological underpinnings that influence performance in football. It establishes a foundation for integrating psychological assessment with technical and tactical training, paving the way for a holistic approach to player development and team strategy formulation.

## 5. RESULTS

INITIAL COHORT TRAIT DISTRIBUTION AND ANALYSIS

The analysis of the initial cohort, comprising 68 athletes, both male and female, across various competition levels, utilized the YODA Assessment to examine 72 psychological traits. The traits were evaluated through responses to 48 tailored scenarios, yielding a rich dataset for analysis.

**Trait Distribution**

The distribution of traits showcased a wide range, reflecting the diverse psychological profiles within the cohort. For instance, the mean scores for traits such as Drive and Determination and Mental Toughness indicated a moderate level of these attributes across the cohort, with standard deviations suggesting variability in how these traits manifest among individuals.

**Analysis Highlights**

- The cohort displayed a balanced distribution in traits like Leadership and Self-Control, indicating a good mix of strategic thinking and emotional regulation among the players.
- Certain traits, such as Coachability and Team Spirit, scored higher on average, suggesting a prevalent disposition towards teamwork and coachability in the cohort.

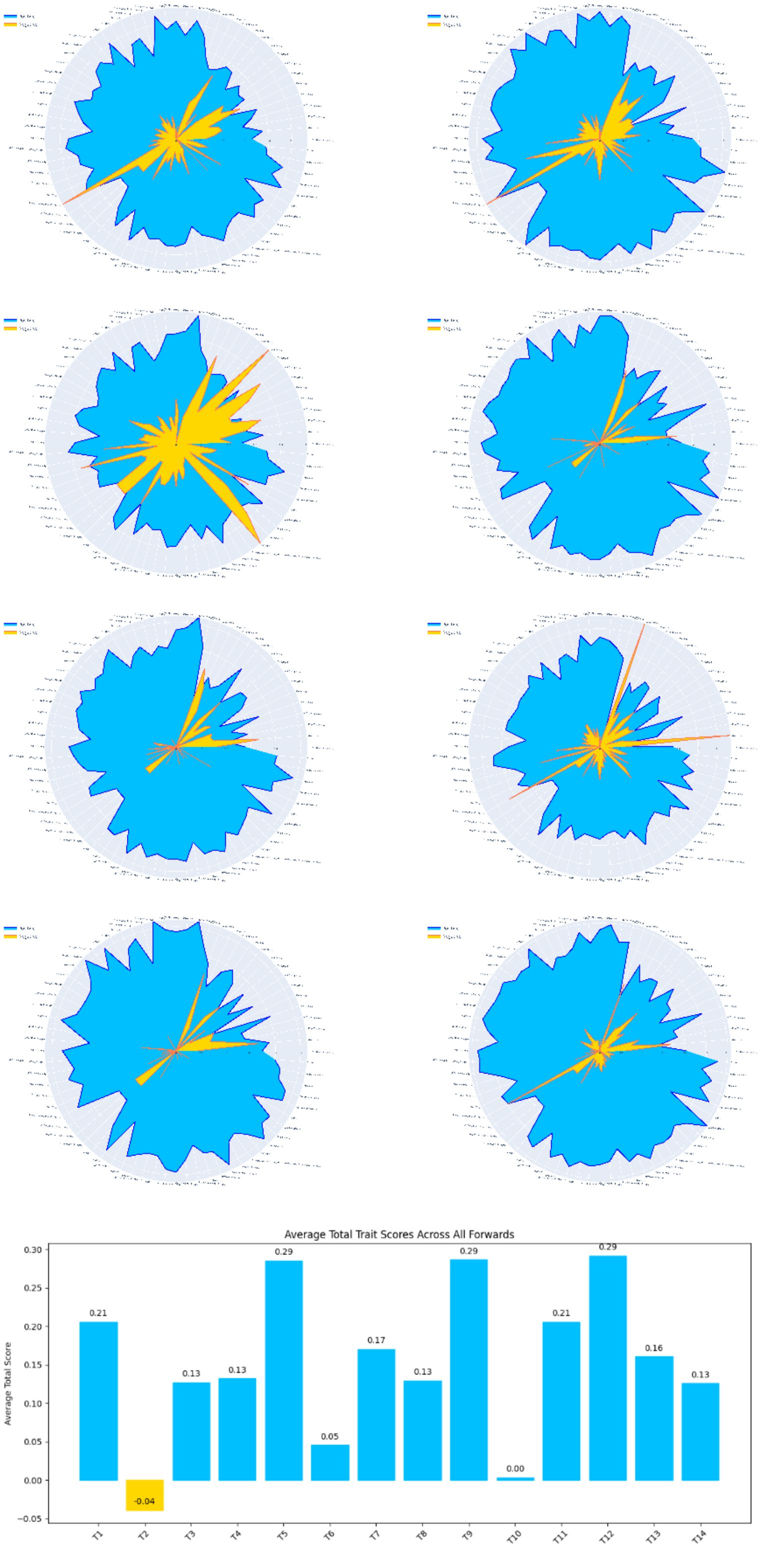



Figure 4: YODA Trait Analysis of Forwards in Cohort

INITIAL COHORT PERSONALITY DISTRIBUTION AND ANALYSIS

The personality analysis, grounded in the Big Five model, correlated the assessed traits with broader personality dimensions, offering a nuanced view of the cohort's psychological makeup.

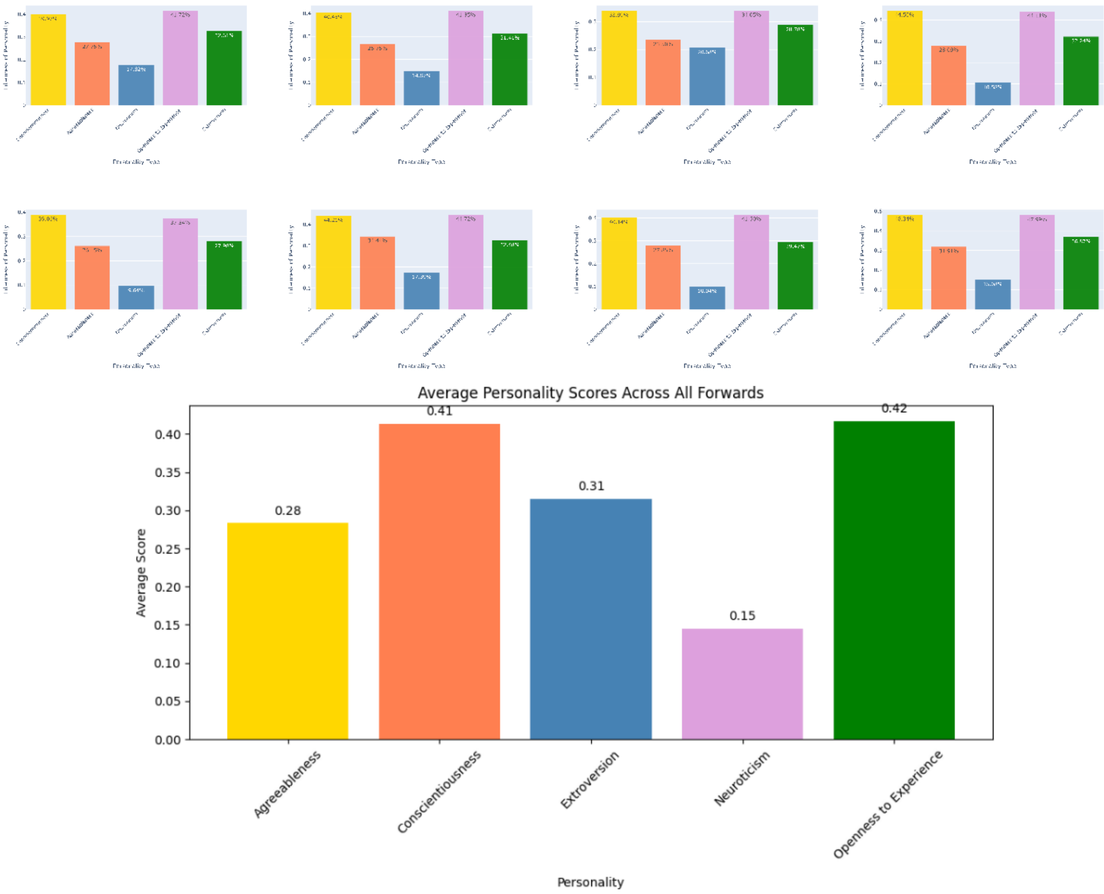



Figure 5: YODA Personality Analysis of Forwards in Cohort

**Personality Correlations**

- High scores in Conscientiousness and Openness to Experience were common, reflecting a cohort that is both disciplined and open to new experiences.
- Extraversion and Agreeableness had varied scores, indicating a diverse range of social dynamics and cooperative behavior among the players.

**Personality Profile Insights**

- A notable correlation was found between the traits related to tactical awareness and decision-making (like Game Sense and Anticipation) and the Big Five dimension of Openness, suggesting that more open individuals may possess better adaptability on the field.
- The trait of Mental Toughness showed a moderate correlation with Conscientiousness, highlighting that players with higher discipline and orderliness tend to be mentally tougher.

## 6. DISCUSSION

ANALYSIS OF YODA'S IMPACT ON PLAYER DEVELOPMENT

YODA's comprehensive evaluation of psychological traits enables the creation of targeted training programs, enhancing players' mental resilience and performance, especially under pressure. By understanding individual psychological profiles, coaches can develop personalized development plans, optimizing each player's mental and emotional growth alongside their physical and technical skills. This approach ensures that players are equipped to handle the demands of competitive play, fostering a mindset conducive to peak performance and long-term career success.

YODA'S ROLE IN ENHANCING TACTICAL AND STRATEGIC DECISIONS

YODA's insights into players' psychological traits are pivotal in informing tactical decisions, allowing for strategic player positioning and game planning that capitalize on psychological strengths. This understanding enhances team dynamics and improves on-field communication, essential for executing game strategies effectively.

Furthermore, YODA's ability to predict player responses in high-pressure scenarios [5] aids in injury prevention, as mental readiness can mitigate risk factors for physical injuries. Additionally, incorporating psychological training into the rehabilitation process addresses the mental and emotional challenges associated with recovery, facilitating a holistic approach to player welfare.

### IMPLICATIONS FOR COACHING AND PLAYER DEVELOPMENT

The integration of psychological assessments like YODA into coaching methodologies [10], [11], [12], [13] revolutionizes player development. Coaches gain deeper insights [3], [4], [14] into the mental makeup of their players, enhancing player-coach relationships and fostering a team environment that prioritizes mental health and psychological growth. This shift towards a more psychologically informed coaching practice promotes a balanced development model, where mental fortitude is given equal importance as physical prowess, leading to more resilient and adaptable athletes.

### INTEGRATION OF PSYCHOLOGICAL ASSESSMENT IN SPORTS ANALYTICS

The incorporation of psychological assessments into sports analytics represents a significant advancement in player scouting and selection. By bridging the gap between physical and psychological data, a more holistic view of player potential is achieved, enhancing the accuracy and depth of player evaluations. As the industry evolves, future trends in sports analytics are expected to increasingly incorporate psychological assessments, recognizing the essential role of mental attributes in player performance and long-term success. This integration signifies a move towards more comprehensive and multidimensional analytics in sports, ensuring that both physical and psychological aspects are considered in player development and performance optimization.

## 7. CONCLUSIONS AND FUTURE WORK

YODA's exploration into the psychological profiling of football players has yielded significant insights [5], [6], [7]. It has established a clear correlation between psychological traits and on-field performance, underlining the importance of mental attributes in football. The initial cohort analysis revealed diverse psychological profiles, with traits such as Drive and Determination, Mental Toughness, and Game Sense being prevalent. The case study of YODA035 further exemplified how individualized psychological assessments can illuminate pathways for player development, highlighting the tool's precision in identifying areas for growth and enhancement.

### LIMITATIONS

Future studies should aim to expand the research scope by including a broader and more diverse cohort, encompassing professional players to gauge the YODA tool's efficacy across different competitive tiers. Longitudinal research is crucial to understand the sustained impact of psychological training on player performance over time. Integrating the YODA assessment with physiological and performance data can offer a more comprehensive view of a player’s capabilities, enhancing the predictive accuracy of their potential success and longevity in the sport.

### IMPLICATIONS FOR FOOTBALL ANALYTICS

The incorporation of psychological assessments like YODA in football analytics heralds a new era in player development and scouting. By offering a nuanced understanding of a player's psychological framework, YODA can significantly enhance tactical and strategic decision-making, promote mental resilience, and optimize team dynamics. As football continues to evolve, integrating psychological metrics with traditional physical and technical analysis will become imperative, ensuring a holistic approach to player development and team performance optimization. This study underscores the potential of psychological assessments in enriching football analytics, advocating for a more integrated and comprehensive approach to understanding and nurturing football talent.

### FUTURE RESEARCH DIRECTIONS

Future studies should aim to expand the research scope by including a broader and more diverse cohort, encompassing professional players to gauge the YODA tool's efficacy across different competitive tiers. Longitudinal research is crucial to understand the sustained impact of psychological training on player performance over time. Integrating the YODA assessment with physiological and performance data can offer a more comprehensive view of a player’s capabilities, enhancing the predictive accuracy of their potential success and longevity in the sport.

**Acknowledgements**
We extend our sincere gratitude to the Sports Department at BMS Institute of Technology, Bengaluru and TKM College of Engineering, Kollam, for their invaluable support and for providing us the opportunity in facilitating the YODA assessment with their football teams.

**References**

[1] S. Lazarević, J. Lukić, and V. Mirkovic, "ROLE OF FOOTBALL SCOUTS IN PLAYER TRANSFORMATION PROCESS: FROM TALENTED TO ELITE ATHLETE," SPORT - Science and Practice, vol. 10, pp. 65-79, 2020.
[2] S.A. Fathima, V.P. Sumathi, and S. Sumanth, "Data analytics in football sport to identify gaps for the improvement of quality opportunities throughout world-wide teams," International Journal of Recent Technology and Engineering (IJRTE), vol. 7, pp. 364-368, 2018.
[3] B. Cwiklinski et al., "Who Will Score? A Machine Learning Approach to Supporting Football Team Building and Transfers," Entropy, 2021.
[4] K. Tuyls et al., "Game Plan: What AI can do for Football, and What Football can do for AI," Journal of Artificial Intelligence Research, vol. 71, 2021.
[5] S. Venkataraman, K. Sundharakumar, A. B. Malakreddy, and S. Natarajan, "Decoding the psyche: Engineering psychological profiles in football through yoda analysis," in ISEA 2024 – The Engineering of Sport 15, 2024.
[6] S. Venkataraman, K. B. Sundharakumar, A. B. Malakreddy, and S. Natarajan, "Yuva-sq: A cognitive scouting model for the beautiful game," in International Conference on Innovative Trends in Information Technology 2024, 2024.
[7] B. Malakreddy A, S. Venkataraman, M. S. Khan, Nidhim S. Padmanabhuni, and S. Natarajan, "Optimizing Semantic Segmentation for Enhanced Football Analytics: A Pixel-Level Approach," in International Conference on Machine Learning and Data Engineering (ICMLDE), 2023.
[8] J. Babu and N. Nimkar, "Study on Talent Identification of Soccer for Young Indian Male Players," Annals of Tropical Medicine & Public Health, vol. 23, 2020.
[9] S. Demediuk et al., "Performance Index: A New Way To Compare Players," 2021.
[10] Sherman-Ash and K. Goldner, "Sharpstack: Cholesky Correlations for Building Better Lineups," 2021.
[11] Mazzelli and R. Nason, "Player Benchmarking and Outcomes: A Behavioral Science Approach," 2021.
[12] D. Finigan, B.M. Mills, and D.F. Stone, "Pulling Starters," 2021.
[13] L. Bransen and J. Van Haaren, "Player Chemistry: Striving for a Perfectly Balanced Soccer Team," in Proceedings of the 2021 MIT Sloan Sports Analytics Conference, Boston, Massachusetts, United States, 2021.
[14] S. Tolan et al., "Measuring the occupational impact of AI tasks cognitive abilities and AI benchmarks," Journal of Artificial Intelligence Research, vol. 71, pp. 191-236, 2021.

# METHODS TO ENCOURAGE ELITE MALE SINGLES PLAYERS TO COMPETE IN DOUBLES AND/OR MIXED DOUBLES IN GRAND SLAM TOURNAMENTS

Tristan Barnett [ae], Graham Pollard [b], Anthony Bedford [c], Sam Robertson [d]
[a] *Macquarie University*
[b] *Faculty of Science and Technology, University of Canberra*
[c] *University of the Sunshine Coast*
[d] *profsamrobertson.com*
[e] *Corresponding author: tristan.barnett@students.mq.edu.au*

## Abstract

This study investigates the use of alternate game scoring structures in tennis to reduce the length of men's singles grand slam matches. In doing so, it is envisaged that some/many elite singles players would be more likely to compete in Grand Slam doubles as well, and some possibly even in mixed doubles events as well (or alternatively). This would enhance the quality of the overall tournament and provide enhanced opportunities for players who are at an elite level in singles, doubles and mixed doubles. The mathematics outlined in this paper suggests that the following tournament modifications might reasonably be considered at the Grand Slam tournaments:

- Extend the length of Grand Slam tournaments from 14 days to 15 or 16 days.
- Replace the deuce and No-ad games with the 50-40 game (server must win 4 points and receiver only has to win 3 points) in all Grand Slam tennis matches.
- Adopt a final advantage set in men's and women's singles Grand Slam events.

**Keywords**: Grand Slams, scoring systems, tennis

## 1. INTRODUCTION

There are arguably equal skills in all three forms of tennis (singles, doubles and mixed doubles), albeit each requires somewhat different skillsets. Grand Slams are the pinnacle tournaments in tennis and have all three forms of the game. Elite male singles players tend not to play in men's doubles and mixed doubles Grand Slam events due to the (possible) need for physical recovery from singles matches and the risk of injury given that men's singles Grand Slams are the best-of-5 sets matches. In contrast women's singles are the best-of-3 sets matches. This reluctance to play doubles is evident in the dominance of Djokovic, Nadal and Federer in winning a total of 66 Grand Slam singles events, but competing in only a handful of Grand Slam doubles events and this was early in their playing careers. In non-Grand Slam doubles events Federer won the Tennis Masters Series in Miami 2003, CA Tennis Trophy in Vienna 2003, Gerry Weber Open in Halle 2005 and the Olympics in Beijing in 2008. Djokovic won the AEGON Championships in London 2010 and the ATP Cup in Brisbane 2020. Nadal won the International Championships in Croatia 2003, Tata Open in Chennai 2004, Qatar Open in Doha 2005, Master Series Monte-Carlo in Monte-Carlo in 2008, Qatar ExxonMobil Open in Doha 2009, BNP Paribas Open in Indian Wells in 2010, Qatar ExxonMobil Open in Doha 2011, BNP Paribas Open in Indian Wells in 2012, Qatar ExxonMobil Open in Doha 2015, Olympics in Rio de Janeiro 2016 and the China Open in Beijing 2016. Of note, a lot of these tournament wins were early in these players' careers. In contrast, Serena Williams won 23 singles Grand Slams, 14 doubles Grand Slams and 2 mixed doubles Grand Slams. Hence, it is reasonable to believe that if the number of days played in Grand Slam tournaments was extended and the singles matches were expected to be of shorter duration, then the elite men's singles players such as Djokovic, Nadal and Federer might have competed in more Grand Slam doubles tournaments and potentially won some doubles Grand Slams. This would probably enhance spectator interest, sponsorship, prize money and benefit the tournament. How many days is reasonable to extend Grand Slam tournaments? If the length was extended by say 1 week, then you would probably get several elite singles players competing in doubles. However, there could be scheduling issues with future tournaments and TV coverage rights given that the schedule is already very tight. Thus, there is a sliding scale of extending the length from 1 day to 1 week. All four Grand Slams typically start on a Monday and go for 14 days with the last event being the men's singles final on a Sunday. It could be feasible to start Grand Slams on the prior Sunday or Saturday and increase the length of a Grand Slam tournament by 1-2 days, and hence reduce the length of the tournament leading up to the Grand Slam by 1-2 days. This was evident in the recent 2024 Australian Open where the 1st round was competed over 3 days starting on the prior Sunday. (This was also done previously in the French Open). It would appear to be self-evident that the main reason elite men singles players are not competing in doubles is the fact that men's singles play best-of-5 sets matches (which can be very physically demanding), compared to women who play best-of-3 sets matches. Is it then possible to slightly modify the scoring structure of men's singles matches such that the length of the match is reduced but the 5 sets match structure is maintained, and 4 levels of hierarchy (points, games, sets, match), whilst keeping the probabilities of the better player winning for the new structure and the current structure quite comparable?

This article demonstrates that this is possible by replacing the current deuce game with either the 50-40 game (the server must win 4 points to win the game whilst the receiver only needs to win 3 points to win) or the No-ad (once the score reaches 40-40 the winner of the next point wins the game). This 50-40 game was first studied by Pollard and Noble (2004). Note that the No-ad game is currently used in doubles on the main tour (excluding Grand Slams) and Grand Slam mixed doubles.

## 2. METHODS

Brown et. al. (2022) calculated characteristics for the 50-40 and No-ad games and compared them with those of the current deuce game. These characteristics were:

1) Probability of the better player winning a match
2) Mean number of points played
3) Standard deviation of the total number of points played
4) Coefficient of skewness of the total number of points played
5) 95%, 99% and 99.5% points of the cumulative distribution of the total number of points played.

Their results indicated that the mean, standard deviation, coefficient of skewness and 95%, 99% and 99.5% points of this cumulative distribution are reduced for the 50-40 and No-ad games when compared to the deuce game, whilst the probabilities of winning the match are reasonably comparable for all three scoring systems.

For this paper, the following characteristics are obtained and are presented in Tables 1 to 3.

a) Probability of the better player winning a match - 5 sets
b) Probability of the better player winning a match – 3 sets
c) Mean amount of time played for best of 5 sets (hours)
d) Standard deviation of the amount of time played for best of 5 sets (hours)
e) Coefficient of skewness of the amount of time played for best of 5 sets (hours)
f) Percentage of best of 5 set matches going beyond 3 hours
g) Percentage of best of 5 set matches going beyond 4 hours
h) Percentage of best of 5 set matches going beyond 5 hours.

The probabilities of reaching a game score of 6-6, 12-12, 18-18, …in an advantage set were evaluated for a set between two extremely strong servers. (Table 4)

The time taken to play a point was estimated at 8 seconds and the time between points (which also incorporates the break time at change of ends) was estimated at 35 seconds. These estimates are based on historical data (Barnett 2012). These estimates were used to ascertain the difference in the time taken for a match under the alternative game structures.

## 3. RESULTS

The results for 3 scenarios are given in Tables 1 to 3. Table 2 gives the results for a typical match between two men with average serving success ($p_A$=0.66 and $p_B$=0.62), where player A is the better player. Table 1 gives the corresponding results for the case when both players have a lower success rate on service ($p_A$=0.62 and $p_B$=0.58, where player A is again the better player), and Table 3 gives the corresponding results for the case where both players have a higher success rate on service.

| Game | Prob B5 sets | Prob B3 sets | Mean B5 sets | SD | Skew | 3 hours | 4 hours | 5 hours |
|---|---|---|---|---|---|---|---|---|
| Deuce | 0.741 | 0.697 | 3.11 | 0.74 | 0.15 | 54.6% | 12.9% | 0.3% |
| No-ad | 0.721 | 0,681 | 2.74 | 0.63 | 0.11 | 35.1% | 2.3% | 0.0% |
| 50-40 | 0.715 | 0.675 | 2.34 | 0.55 | 0.13 | 12.9% | 0.0% | 0.0% |

Table 1 Match characteristics when $p_A$=0.62 and $p_B$=0.58

| Game | Prob B5 sets | Prob B3 sets | Mean B5 sets | SD | Skew | 3 hours | 4 hours | 5 hours |
|---|---|---|---|---|---|---|---|---|
| Deuce | 0.734 | 0.691 | 3.12 | 0.73 | 0.14 | 55.0% | 12.9% | 0.3% |
| No-ad | 0.719 | 0.679 | 2.77 | 0.64 | 0.11 | 36.7% | 2.7% | 0.0% |
| 50-40 | 0.718 | 0.678 | 2.37 | 0.55 | 0.13 | 14.2% | 0.0% | 0.0% |

Table 2 Match characteristics when $p_A$=0.66 and $p_B$=0.62

| Game | Prob B5 sets | Prob B3 sets | Mean B5 sets | SD | Skew | 3 hours | 4 hours | 5 hours |
|---|---|---|---|---|---|---|---|---|
| Deuce | 0.725 | 0.684 | 3.14 | 0.73 | 0.12 | 56.4% | 13.5% | 0.3% |
| No-ad | 0.717 | 0.677 | 2.82 | 0.64 | 0.11 | 39.5% | 3.5% | 0.0% |
| 50-40 | 0.721 | 0.680 | 2.41 | 0.56 | 0.13 | 16.1% | 0.1% | 0.0% |

Table 3 Match characteristics when $p_A$=0.70 and $p_B$=0.66

Whilst there are many similarities in these tables, the percentage of 'long' matches is reduced very considerably by using the No-ad and 50-40 games (see columns 7 for example).

Table 4 shows the probability that an advantage set reaches certain scores (6-6, 12-12, 18-18, …) for parameters relevant to the case of extremely strong servers (and/or weak receivers). The parameters used are the observed statistics in the final advantage set from the famous John Isner and Nicholas Mahut match at Wimbledon in 2010 ($p_A$=0.82 for Isner and $p_B$=0.77 for Maht). For these parameters, when using the 50-40 game, extremely long matches are most unlikely to occur.

| Score line | Deuce | No-ad | 50-40 |
|---|---|---|---|
| 6-6 | 74.1% | 64.6% | 34.7% |
| 12-12 | 54.3% | 40.7% | 9.5% |
| 18-18 | 39.8% | 25.6% | 2.6% |
| 24-24 | 29.1% | 16.1% | 0.7% |
| 30-30 | 21.3% | 10.2% | 0.2% |
| 36-36 | 15.6% | 6.4% | 0.1% |
| 42-42 | 11.5% | 4.0% | 0.0% |
| 48-48 | 8.4% | 2.5% | 0.0% |
| 54-54 | 6.1% | 1.6% | 0.0% |
| 60-60 | 4.5% | 1.0% | 0.0% |
| 66-66 | 3.3% | 0.6% | 0.0% |
| 68-68 | 3.0% | 0.5% | 0.0% |

Table 4: Chances of reaching a game score line in an advantage set when $p_A$=0.82 and $p_B$=0.77

Table 5 gives the corresponding results when using 50-40 games and a 5th set advantage, whilst the first 4 sets are tiebreak sets. 4-hour matches are most unlikely to occur, thereby opening the possibility for using an advantage fifth set.

| | Prob | Mean | SD | Skew | 3 hours | 4 hours | 5 hours |
|---|---|---|---|---|---|---|---|
| $p_A$=0.62 and $p_B$=0.58 | 0.715 | 2.35 | 0.55 | 0.18 | 13.2% | 0.1% | 0.0% |
| $p_A$=0.66 and $p_B$=0.62 | 0.719 | 2.38 | 0.56 | 0.20 | 14.6% | 0.1% | 0.0% |
| $p_A$=0.70 and $p_B$=0.66 | 0.722 | 2.41 | 0.57 | 0.23 | 16.6% | 0.2% | 0.0% |

Table 5: Characteristics for a best of 5 sets match with 50-40 games and a 5th advantage set

## 4. DISCUSSION

It can be seen from Tables 1, 2 and 3 that the probability that the better player wins the match is only "marginally" reduced by using the No-ad or 50-40 game instead of the deuce game. More specifically, it can be shown that the probability of the better player winning a match in a best of 5 sets using a 50-40 game is always greater than the probability of the better player winning a match in a best of 3 sets using a deuce game. For example, from table 1 the probability of the better player winning a best of 5 sets match using a 50-40 game is 0.715 and the probability of the better player winning a best of 3 sets match using a deuce game is 0.697 (as used in women's singles matches).

As expected, both the mean and standard deviation of the total time played in the match are reduced relative to the deuce game when using either the No-ad or 50-40 game. For example, the No-ad and 50-40 games reduce the mean duration by 21 mins and 45 mins respectively when $p_A$=0.66 and $p_B$=0.62. As expected, the probabilities that the best of 5 sets match goes beyond 3, 4 and 5 hours are significantly reduced for the No-ad and 50-40 games. In particular, the chance that a typical men's tennis match lasts more than 4 hours using a deuce game is about 13% at present (see Table 2), whereas it would occur less than 0.01% of the time using 50-40 games. Thus the 50-40 game not only reduces the mean and the standard deviation of the total time played from both deuce and No-ad games, but allows for the possibility of playing an advantage final set and keeping with tradition, which is not 'possible' using deuce of No-games games. Also, for both best of 3 and best of 5 sets matches, the probability of the stronger player winning the match when using No-ad or 50-40 games is comparable to its value when using deuce games. Based on the above observations the 50-40 and No-ad games could be considered "preferable" scoring systems to the deuce game and the 50-40 game could be considered a "preferable" scoring system to the No-ad game.

There are examples of continuous quality improvement in tennis. Firstly, the development of the tiebreak game is one example. One type of tiebreak was invented by Jimmy Van Alen in 1965 to reduce the length of matches, and in 1970, the US Open became the first of the Grand Slam tournaments to use the tiebreak set. Jimmy Van Alen talked about the tiebreak for a decade before the best of 9 points was used in an invitation professional tournament at the then US Hall of Fame at Newport in 1965.

Before the No-ad game was used in Grand Slam mixed doubles tournaments, it was trialed at 'lower-level' tournaments and in particular doubles on the non-Grand Slam main tour. This change in scoring systems was brought about in 2006 along with a first-to-10 tiebreak game as the deciding final set. When the No-ad game was found to be a useful and practical game structure for these tournaments, it was adopted for mixed doubles at the Grand Slam events. Given that the 50-40 game has (mathematical) scoring system properties that make it quite attractive compared with both advantage games and No-ad games, it follows that, *in the interest of continuous improvement in the game of tennis*, *it would seem appropriate that the 50-40 game be trialed in some tournaments, initially in men's and women's doubles.* It is noted that *such a scoring system would also have greater measures of importances and excitement* than the present games scoring formats (Pollard (2017)).

*There is another scoring system modification that could well be trialed simultaneously. It is the first to 9 points tiebreak game (lead by 2 points), denoted by TB9.* It has been shown mathematically to be fairer than the TB7 system (and the TB10 system) presently used (Pollard (2005), Pollard and Noble (2002, 2003, and 2004b)). This slightly longer tiebreak game (at the end of each set) would 'go well with' the shorter 50-40 games during the set as it would reduce the small difference noted above in the probability that the better player wins, at least within a set when reaching a tiebreak at 6 games all. There would appear to be no reason why these two changes could not be trialed simultaneously.

Another example of continuous quality control in tennis, although arguably quite slow in its adoption, was the use of a tiebreak final set in all Grand Slam matches. Organizers and governing bodies reacted from 'long' matches occurring as a result of the advantage final set. For example, Ivanisevic defeating Krajicek 15-13 at Wimbledon in 1998, Philippoussis defeating Schalken 20–18 at Wimbledon in 2000, Roddick defeating El Aynaoui 21–19 at the 2003 Australian Open, and Isner defeating Mahut 70-68 at Wimbledon in 2010. Several alternative tiebreak fifth sets were used before the present 10-point tiebreak was adopted. Interestingly, this tiebreak game can be somewhat unfair in doubles (just think of it as approximately the best of 18 points, and 18 is not divisible by 4, the number of players on the court) It can also be unfair in singles in the presence of an end effect (e.g. when wind/sun operate differently at each end of the court). The 9-point tiebreak game (best of 16 points) does not have these two shortcomings, or these shortcomings are substantially reduced. The reader is referred to papers by Pollard (2005), Pollard and Noble (2002, 2003, 2004b), and Pollard, Noble and Pollard (2022).

Table 4 uses the probabilities of winning points on serve, as occurred in the Inser vs Mahut match, to represent the most 'extreme' serving probabilities that can occur in men's tennis. It can be shown that there was a 3.0% chance of the match reaching 68-68 in such an advantage set. If a No-ad game was played instead, an estimate of the match reaching 68-68 is 0.5%. However, if a 50-40 game was played then an estimate of the match reaching 68-68 is 0.0%. In fact, it is highly unlikely the match would even go beyond 30-30 (0.2%) with a 50-40 game. *Therefore, by adopting the 50-40 game, not only is the possibility of very long match durations significantly reduced but the playing of an advantage set might be considered feasible, in keeping with tradition.* As can be seen in the table, this is not possible with the current deuce or No-ad game. Thus, *using the 50-40 game in all matches in men's and women's doubles (rather than the No-ad game) would be a useful trial.*

Table 5 gives characteristics for a 50-40 game with a $5^{th}$ advantage set, the first 4 sets being tiebreak sets. As expected, the probabilities of the stronger player winning are increased in comparison to playing all tiebreak sets. The probability of a match going beyond 5 hours is virtually zero, and going beyond 4 hours only 0.1%-0.2%. Thus, four tiebreak sets and a $5^{th}$ advantage set all using 50-40 games could be trialed as a replacement to using deuce games in 5 tiebreak sets.

Finally, table 6 gives a comparison of winning a match from 6-games all in the fifth set using the 7 point tiebreak game with playing a final advantage set with 50-40 games. The stronger player has between a 1.8%-2.3% increase in winning a match by playing 50-40 games in a final advantage set in comparison to playing the tiebreak game.

| | **deuce game tiebreak set** | **50-40 game advantage set** |
|---|---|---|
| $p_A$=0.62 and $p_B$=0.58 | 56.5% | 58.3% |
| $p_A$=0.66 and $p_B$=0.62 | 56.6% | 58.6% |
| $p_A$=0.70 and $p_B$=0.66 | 56.9% | 59.2% |

Table 6: Comparison of two scoring systems in winning a match from 6-games all in the final set

## 5. CONCLUSIONS

By analyzing various characteristics of scoring systems, there is mathematical evidence to suggest that the current deuce game could be replaced by either the No-ad or 50-40 game to reduce the length of men's singles Grand Slam matches. By adopting the 50-40 game, the mean length of a match is reduced by on average 45 minutes and playing an *advantage* $5^{th}$ set becomes feasible in keeping with tradition (and this is not the case with the deuce or No-ad games). For these reasons it is recommended that by replacing the deuce game with the 50-40 game in men's singles, would encourage elite men's singles players to also compete in doubles Grand Slams (and possibly even mixed doubles), which should enhance spectator interest, sponsorship, prize money and improve the quality of the doubles events. Further, by also adopting the 50-40 game in men's doubles and mixed doubles Grand Slams would reduce the length of doubles and mixed doubles matches, which would further increase the likelihood of elite men's singles players competing in doubles and mixed doubles Grand Slams. Likewise adopting a 50-40 game in women's singles and doubles Grand Slams would also increase the likelihood of elite women's singles players competing in doubles and mixed doubles Grand Slams. For consistency it could be possible to adopt the 5040 game in every tennis match Grand Slam and non-Grand Slam, which would help players adapt to the unbalanced structure of the 5040 game, and furthermore encourage elite singles players to compete in non-Grand Slam doubles. With any system it could be practical to trial the 5040 game at 'lower' tournaments such as non-Grand Slam doubles given the current No-ad game structure.

**Acknowledgements**

We wish to thank Geoff Pollard from Tennis Australia for insightful comments towards this paper.

**References**

Barnett T. and Clarke S.R. (2005). Combining player statistics to predict outcomes of tennis matches. *IMA Journal of Management Mathematics*. 16 (2), 113-120.

Barnett T. and Cross R. (2007). How the tennis scoreboard affects player performance. *Medicine and Science in Tennis*. 12(2), 6-8.

Barnett T. (2012). Analyzing tennis scoring systems: from the origins to today. *Journal of Medicine and Science in Tennis*. 17(2), 68-77.

Brown A., Barnett T., Pollard G.H., Pollard G.N. and Ejov V. (2022). No-Ad game scoring within a best of 5 sets tennis matches. *Proceedings of the 16th Australian Conference on Mathematics and Computers in Sport*.

Pollard G.H. (2005). On solving an aspect of unfairness in the tennis doubles scoring system, *Chance*, 18(4), 17-19.

Pollard G.H. (2017). Measuring excitement in Sport. *Journal of Sports Analytics*, 3, 37-43.

Pollard G.H. and Noble K.J. (2002). A solution to the unfairness of the tiebreak game when used in tennis doubles. *Proceedings of the 6th Australian Conference on Mathematics and Computers in Sport*. University of Technology, Sydney, edited by Cohen, G. and Langtry, T., 221-226.

Pollard G. H. and Noble K.J. (2003). A solution to the unfairness of tiebreak tennis doubles. *Tennis Science and Technology 2*, International Tennis Federation, edited by S. Miller, 325-332.

Pollard GH (2005). On solving an aspect of unfairness in the tennis doubles scoring system, *Chance*, 18(4), 17-19.

Pollard GH (2017). Measuring excitement in Sport. *Journal of Sports Analytics*, 3, 37-43.

Pollard GH and Noble KJ (2002). A solution to the unfairness of the tiebreak game when used in tennis doubles. *Proceedings of the 6th Australian Conference on Mathematics and Computers in Sport*. University of Technology, Sydney, edited by Cohen, G. and Langtry, T., 221-226.

Pollard GH and Noble KJ (2003). A solution to the unfairness of tiebreak tennis doubles. *Tennis Science and Technology 2*, International Tennis Federation, edited by S. Miller, 325-332.

Pollard GH and Noble KJ (2004a). The benefits of a new game scoring system in tennis, the 50-40 game. *Proceedings of the 7th Australasian Conference on Mathematics and Computers in Sport*. Massey University, New Zealand, edited by R Hugh Morton and S Ganesalingam, 262-265.

Pollard GH and Noble KJ (2004b). Some attractive properties of the 16-point tiebreak game in tennis. *Proceedings of the 7th Australasian Conference on Mathematics and Computers in Sport*. Massey University, New Zealand, edited by R Hugh Morton and S Ganesalingam, 257-261.

Pollard GH, Noble KJ and Pollard GN (2022). New Scoring Systems to reduce unfairness in men's doubles. *Journal of Sports Analytics*, 8, 291-298.

# PERFORMANCE ANALYSIS IN REVERSE - CREATING THE WORLD'S FIRST GLOBAL (VIRTUAL) FOOTBALL LEAGUE

Karl Jackson [arc], Joshua Robinson [b]

[a] *One Future Football - Melbourne, Australia*
[b] *Champion Data - Melbourne, Australia*
[c] *Corresponding author: karl@onefuturefootball.com*

**Submitted for consideration as (select one):**
- Oral presentation

**Abstract**

Performance Analysis generally involves the process of observing an athletic performance, systematically measuring the observations, then performing notational analysis to make assumptions about skill sets and tactical intent. With the creation of One Future Football our process was in reverse - creating skill profiles and tactical intents for virtual athletes and teams, and using stochastic processes to create a performance. One Future Football is the world's first global virtual football league. Players with names, nationalities, back-stories, skill profiles and personality profiles were created and assigned to 12 inaugural teams. These teams then compete in simulated football matches, grouped into Rounds and Seasons in an effort to be crowned champions of One Future Football. Fans follow the league, teams, and players on social media, compete in fantasy football competitions and influence the outcome of matches by contributing to the development of their favourite players via a virtual training mechanic. Careful consideration went into the creation of the players and the randomised Match Engine to ensure that outcomes were in line with stadium football leagues and players. Artificial Intelligence and process automation allows for the creation of vast amounts of content for the purpose of performance analysis, social media publication and internal operations.

*Keywords*: Simulation, Process Automation, Performance Analysis

# MATH _ SPORT:

## The Preposition Proposition

Steven Stern
Professor of Data Science
Centre for Data Analytics
Bond University
Gold Coast, AUSTRALIA

## MATH **OF** SPORT

- Measure numerical aspects of:
  - performance/outcome of individual games/players
  - overall competitions/careers.
- Widely viewed as "beginning" with SABRmetrics Era.
  - Not really true. Math/Numbers always part of sport and its "mystique":
    E.g., Do you recognise:
    - 99.94 = Bradman's career test cricket batting average
    - 9.58 = Bolt's 100m world record
    - 1,360 = Lockett's career AFL goals
    - 56 = DiMaggio's MLB game hitting streak
    - 18 = Nicklaus' career total golf major championships
    - 2,857 = Gretzky's career NHL points total
    - 3,667 = Maravich's NCAA men's basketball career total (surpassed by C Clark this year)

## MATH **FOR** SPORT

- Use measure OF sport to innovate and improve:
  - Gambling/Fantasy Leagues
  - Playing Strategies (both in-game & competition choice)
  - Team Composition (both one-game and list management)
  - Player Wellness & Preparation
  - Rating & Ranking (in-tournament, in-year & over time)
- "Moneyball" & Analytics Driven In-Game Coaching/Broadcasting
  - Careful, can oversimplify (e.g., NFL: go for 2, go for 4th down)
  - Some broadcast analytics are poor/simplistic (e.g., Cricket Run Projector, WinProb Projectors)
- Can have unintended consequences:
  QUESTION: Is Stephen Curry better for team than Shaquille O'Neal/Wilt Chamberlain?
  Legit debate now, but without Possession Efficiency metrics changing gameplay?

## MATH **IN** SPORT

- Part of the "Rules"
- Competition/Structural Rules
  - Secondary ranking criteria
  - Ball-tracking ("Shrödinger's cricket ball")
  - Offside VAR in soccer
- In-Game Rules
  - Very few examples
    - LBW projected path
    - DLS
    - Ice Skating/Gymnastics/Diving scoring systems (i.e., leave out high/low)

## MATH **IN** SPORT

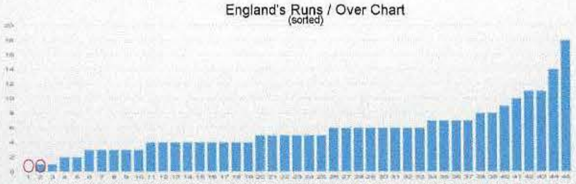
- As progress OF → FOR → IN, crucial to do "full context" diagnostics
- Unintended Consequences REALLY IMPORTANT
  - 1992 CWC: Most productive overs method - ENG 253 (45), SA 231/6 (42.5), needed 22 (2.1), lost 2

England's Runs / Over Chart

# MATH IN SPORT

- As progress OF → FOR → IN, crucial to do "full context" diagnostics
- Unintended Consequences REALLY IMPORTANT
  - 1992 CWC: Most productive overs method - ENG 253 (45), SA 231/6 (42.5), needed 22 (2.1), lost 2

England's Runs / Over Chart (sorted)

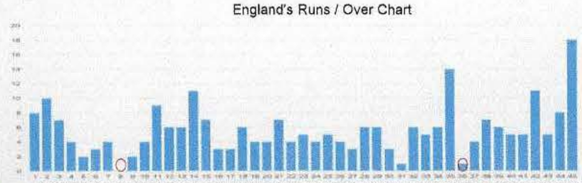


IDEA: If X overs lost in chase, deduct lowest scoring X overs of 1st innings from target [New target 21 (0.1)]

# MATH IN SPORT

- As progress OF → FOR → IN, crucial to do "full context" diagnostics
- Unintended Consequences REALLY IMPORTANT
  - 1992 CWC: Most productive overs method - ENG 253 (45), SA 231/6 (42.5), needed 22 (2.1), lost 2

England's Runs / Over Chart

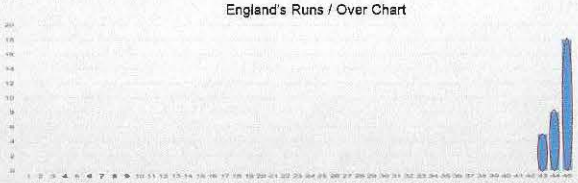


IDEA: If X overs lost in chase, deduct lowest scoring X overs of 1st innings from target [New target 21 (0.1)]
MODIFICATION: Use X lowest scoring among overs after interruption over.

# MATH IN SPORT

- As progress OF → FOR → IN, crucial to do "full context" diagnostics
- Unintended Consequences REALLY IMPORTANT
  - 1992 CWC: Most productive overs method - ENG 253 (45), SA 231/6 (42.5), needed 22 (2.1), lost 2

England's Runs / Over Chart



IDEA: If X overs lost in chase, deduct lowest scoring X overs of 1st innings from target [New target 21 (0.1)]
MODIFICATION: Use X lowest scoring among overs after interruption over. [New target - Avg: 3 (0.1)]

# MATH IN SPORT

- As progress OF → FOR → IN, crucial to do "full context" diagnostics
- Unintended Consequences REALLY IMPORTANT
  - 1992 Cricket World Cup: If "modified" most productive overs method used?
    - Might not have been "outcry" and no D/L?
    - But, still not a sensible method (e.g., same target whether SA were 231/6 or 231/9 or 231/2)
    - So, likely would have eventually led to "catastrophe" (which is what most sporting bodies wait for before making changes, especially for math-based rules!)
    - If not careful, we just end up playing "Error Whack-a-Mole"

## MATH IN SPORT

- As progress OF → FOR → IN, crucial to do "full context" diagnostics
- Need set of principles to guide assessment of model/technique
  - Rain Rule Principles:
    1. More runs scored must not disadvantage batting team
    2. More wickets must not disadvantage the bowling team
    3. Only overall innings scores determine the match outcome (i.e., runs before an interruption count equally to runs afterwards)
    4. (a) A chase at par when it is interrupted should remain at par upon resumption; and,
       (b) Given initial chase, victory requirements from *par* are unique
       (i.e., "expected" score for a range of overs is the same whether they are played or lost)

## MATH IN SPORT

- As progress OF → FOR → IN, crucial to do "full context" diagnostics
- Need set of principles to guide assessment of model/technique
  - Rain Rule Principles:
    3. Only overall innings scores determine the match outcome (i.e., runs before interruption count equally to runs afterwards)

One consequence: Probability-based methods are ruled out, as they violate (3)

IDEA: Team 1 scores $S$, chase interrupted, Team 2 target revised to maintain Pr(Win)

(i) Team 2 at par, $P$, at interruption; new target $T_P$ [set so Pr(Win) = 50%]; chase innings ends on $T_P$ off last ball: Team 2 WIN

(ii) Above par, $P$+$x$, at interruption; new target $T_{P+x} > T_P$ [shortening favours team ahead, so larger target maintains Pr(Win)]; chase innings ends on $T_P$: Team 2 LOSE

So, outcome of probability-based method changes depending on when runs scored. [i.e., in (i) chase score was $P + (T_P - P) = T_P$ for win and in (ii) chase score was $P + x + (T_P - P - x) = T_P$ for loss]

## MATH IN SPORT

- Let's consider another "math in the game" example:
- A Quick Review of Ranking Systems:
  - Key Distinction: Purely Win/Loss Based vs Team Statistics Based
    - Win/Loss better (Team stats say who is better "on paper")
    - But, should use strength of outcome (big win worth more)
      - SIDE EFFECT: No disincentive for high-ranked vs low-ranked team match-up, as low-ranked can improve even with a narrow loss.

## MATH IN SPORT

- A Quick Review of Ranking Systems:
  - Kendix Cricket Ranking System
    - ELO-based, but no margin of victory used (as directed by ICC)

IDEA: Each team has current rating, $R_{team}$ = average match results over timeframe. Individual match results between $A$ and $B$ determined as: (assume $R_A > R_B$)

| Case: | Team | Ranking Points Awarded for: | | |
|---|---|---|---|---|
| | | Win | Loss | Tie |
| $R_A - R_B \le 40$ | A | $R_B + 50$ | $R_B - 50$ | $R_B$ |
| | B | $R_A + 50$ | $R_A - 50$ | $R_A$ |
| $R_A - R_B > 40$ | A | $R_A + 10$ | $R_A - 90$ | $R_A - 40$ |
| | B | $R_B + 90$ | $R_B - 10$ | $R_B + 40$ |

## MATH IN SPORT

- A Quick Review of Ranking Systems:
  - *Proposed Stern* Cricket Ranking System
    - ELO-based, with "margin of victory" used (more on MoV later)

IDEA: Each team has current rating, $R_{team}$ = average match results over timeframe. Individual match results between $A$ and $B$ determined as: (assume $R_A > R_B$)

| Case: | Team | Ranking Points Awarded for: Win | Loss | Tie |
|---|---|---|---|---|
| $R_A - R_B \le 40$ | A | $R_B + 20 + \text{MoV}$ | $R_B - 20 + \text{MoV}$ | $R_B$ |
| | B | $R_A + 20 + \text{MoV}$ | $R_A - 20 + \text{MoV}$ | $R_A$ |
| $R_A - R_B > 40$ | A | $R_A - 20 + \text{MoV}$ | $R_A - 60 - \text{MoV}$ | $R_A - 40$ |
| | B | $R_B + 60 + \text{MoV}$ | $R_B + 20 + \text{MoV}$ | $R_B + 40$ |

## MATH IN SPORT

- A Quick Review of Ranking Systems:
  - Comparison of Cricket Ranking System (capped MoV version shown)


## MATH IN SPORT

- Let's consider another "math in the game" example:
  - Secondary ranking criteria for round robin tournament phases.
  - I'll focus again on cricket: current standard is

$$\text{Net Run Rate} = \frac{\text{Total}^{\dagger}\text{ Runs Scored}}{\text{Total}^{\dagger}\text{ Overs Batted}} - \frac{\text{Total}^{\dagger}\text{ Runs Conceded}}{\text{Total}^{\dagger}\text{ Overs Bowled}}$$

†[For interrupted matches, 1st innings runs = Target − 1 & 1st innings overs = 2nd innings max available overs. Also, bowled out = used all available overs]

- Benefits? Easy to calculate. "Intuitive", at least many claim. The "incumbent".
- Drawbacks? Doesn't account for wickets! Matches contribute unequally.

## MATH IN SPORT

- EXAMPLE: Consider a ODI WC Qualifier League with results:

| TEAM | ROUND 1 | 2 | 3 | 4 | 5 | W-L | NRR |
|---|---|---|---|---|---|---|---|
| A | 244 v 98 | 164 v 180 | 198/8 (47.4) v 195 | 246 v 146 | 195/3 (26) v 192 | 4 - 1 | +1.347 |
| B | 176/4 (38) v 174 | 180 v 164 | 115/2 (22.5) v 114 (47) | 267 v 74 | - | 4 - 0 | +1.919 |
| C | 353 v 68 | 273 v 254 | 253 v 187 | 146 v 245 | - | 3 - 1 | +1.355 |

Major Win | Minor Win | Minor Loss | Major Loss

- Teams B & C play final match. Team B scores 162; Team C reach 150/5 (32).
- If Team C win ⟹ three-way tie (only 2 move on to next round).
- Let's examine some scenarios:

# MATH IN SPORT

- EXAMPLE: Consider a ODI WC Qualifier League with results:

| TEAM | ROUND 1 | ROUND 2 | ROUND 3 | ROUND 4 | ROUND 5 | W-L | NRR |
|---|---|---|---|---|---|---|---|
| A | 244 v 98 | 164 v 180 | 198/8 (47.4) v 195 | 245 v 146 | 195/3 (26) v 192 | 4 - 1 | +1.347 |
| B | 176/4 (38) v 174 | 180 v 164 | 115/2 (22.5) v 114 (47) | 267 v 74 | 162 v 163/5 (34) | 4 - 1 | *+1.286* |
| C | 353 v 68 | 273 v 254 | 253 v 187 | 146 v 245 | 163/5 (34) v 162 | 4 - 1 | +1.413 |

Major Win | Minor Win | Minor Loss | Major Loss

- Teams B & C play final match. Team B scores 162; Team C reach 150/5 (32).
- If Team C win $\Longrightarrow$ three-way tie (only 2 move on to next round).
- Let's examine some scenarios: Team C only need 2 more overs

# MATH IN SPORT

- EXAMPLE: Consider a ODI WC Qualifier League with results:

| TEAM | ROUND 1 | ROUND 2 | ROUND 3 | ROUND 4 | ROUND 5 | W-L | NRR |
|---|---|---|---|---|---|---|---|
| A | 244 v 98 | 164 v 180 | 198/8 (47.4) v 195 | 245 v 146 | 195/3 (26) v 192 | 4 - 1 | +1.347 |
| B | 176/4 (38) v 174 | 180 v 164 | 115/2 (22.5) v 114 (47) | 267 v 74 | 162 v 163/5 (36) | 4 - 1 | *+1.312* |
| C | 353 v 68 | 273 v 254 | 253 v 187 | 146 v 245 | 163/5 (36) v 162 | 4 - 1 | +1.370 |

Major Win | Minor Win | Minor Loss | Major Loss

- Teams B & C play final match. Team B scores 162; Team C reach 150/5 (32).
- If Team C win $\Longrightarrow$ three-way tie (only 2 move on to next round).
- Let's examine some scenarios: Team C need 4 more overs

# MATH IN SPORT

- EXAMPLE: Consider a ODI WC Qualifier League with results:

| TEAM | ROUND 1 | ROUND 2 | ROUND 3 | ROUND 4 | ROUND 5 | W-L | NRR |
|---|---|---|---|---|---|---|---|
| A | 244 v 98 | 164 v 180 | 198/8 (47.4) v 195 | 245 v 146 | 195/3 (26) v 192 | 4 - 1 | +1.347 |
| B | 176/4 (38) v 174 | 180 v 164 | 115/2 (22.5) v 114 (47) | 267 v 74 | 162 v 163/5 (38) | 4 - 1 | +1.337 |
| C | 353 v 68 | 273 v 254 | 253 v 187 | 146 v 245 | 163/5 (38) v 162 | 4 - 1 | *+1.328* |

Major Win | Minor Win | Minor Loss | Major Loss

- Teams B & C play final match. Team B scores 162; Team C reach 150/5 (32).
- If Team C win $\Longrightarrow$ three-way tie (only 2 move on to next round).
- Let's examine some scenarios: Team C need 6 more overs *(no way for A to miss out!)*

# MATH IN SPORT

- EXAMPLE: Consider a ODI WC Qualifier League with results:

| TEAM | ROUND 1 | ROUND 2 | ROUND 3 | ROUND 4 | ROUND 5 | W-L | NRR |
|---|---|---|---|---|---|---|---|
| A | 244 v 98 | 164 v 180 | 198/8 (47.4) v 195 | 245 v 146 | 195/3 (26) v 192 | 4 - 1 | +1.347 |
| B | 176/4 (38) v 174 | 180 v 164 | 115/2 (22.5) v 114 (47) | 267 v 74 | - | 4 - 0 | +1.919 |
| C | 353 v 68 | 273 v 254 | 253 v 187 | 146 v 245 | - | 3 - 1 | +1.355 |

Major Win | Minor Win | Minor Loss | Major Loss

- Teams B & C play final match. Team B scores 162; Team C reach 150/5 (32).
- Unless ...
- Match interrupted, 4 overs lost: New Target = 157, Team C 159/5 (33.3)
  [NOTE: In DLS matches, NRR gives team batting first Target – 1 in max available overs]

# MATH IN SPORT

- EXAMPLE: Consider a ODI WC Qualifier League with results:

| TEAM | ROUND | | | | | W-L | NRR |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | | |
| A | 244 v 98 | 164 v 180 | 198/8 (47.4) v 195 | 245 v 146 | 195/3 (26) v 192 | 4 - 1 | *+1.347* |
| B | 176/4 (38) v 174 | 180 v 164 | 115/2 (22.5) v 114 (47) | 267 v 74 | 156 (46) v 159/5 (33.3) | 4 - 1 | **+1.351** |
| C | 353 v 68 | 273 v 254 | 253 v 187 | 146 v 245 | 159/5 (33.3) v 156 (46) | 4 - 1 | +1.371 |

Major Win | Minor Win | Minor Loss | Major Loss

  - Teams B & C play final match. Team B scores 162; Team C reach 150/5 (32).
  - Unless ...
  - Match interrupted, 4 overs lost: New Target = 157, Team C 159/5 (33.3)
    [RECALL: In DLS matches, NRR gives team batting first Target – 1 in max available overs]

# MATH IN SPORT

- EXAMPLE: Consider a ODI WC Qualifier League with results:

| TEAM | ROUND | | | | | W-L | NRR |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | | |
| A | 244 v 98 | 164 v 180 | 198/8 (47.4) v 195 | 245 v 146 | 195/3 (26) v 192 | 4 - 1 | *+1.347* |
| B | 176/4 (38) v 174 | 180 v 164 | 115/2 (22.5) v 114 (47) | 267 v 74 | 156 (46) v 159/5 (33.3) | 4 - 1 | **+1.351** |
| C | 353 v 68 | 273 v 254 | 253 v 187 | 146 v 245 | 159/5 (33.3) v 156 (46) | 4 - 1 | **+1.371** |

Major Win | Minor Win | Minor Loss | Major Loss

  - So, why is NRR still used? Wasn't this seen as a "catastrophe"?
  - Teams A, B & C are Singapore, Nepal and USA (needed England, South Africa and India!)
    [ASIDE: Interruption was a political protest pitch invasion by Nepalese crowd. Any conspiracy theorists out there?]

# MATH IN SPORT

- Need Principles
  - Individual matches should contribute according to the margin of victory
    - E.G., 5th Singapore match score was 195/3 (26) v 192, which is a very big margin of Victory, but it contributes to NNR exactly the same as if the match score had been 195/9 (26) v 192, a much narrower win.
  - Matches with equivalent outcomes should contribute equally
    - E.G., 1st USA match score was 353 v 68, which is an NRR = $\frac{353}{50} - \frac{68}{50} = 5.7$.
      Suppose instead it was 282 (40) v 54 (40). Same individual match NRR, but overall NRR changes from +1.371 to +1.183
      (i.e., wins in longer match, more beneficial. Thus, wins batting first generally better)
  - Contribution shouldn't depend on match structure
    - E.G., Interrupted matches contribute differently (i.e., fewer overs, implicit accounting for wickets via DLS)

# MATH IN SPORT

- (More Recent) EXAMPLE:
  - SUPER SIX stage of most recent ODI CWC Qualifiers:

| TEAM | MATCH | | | | | W-L | NRR |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | | |
| NED | 315 v 31[illegible]/4 (40.5) | 374† v 374 | 192 v 213 | 320 (44) v 246 (44) | 278/6 (42.5) v 277 | 3 - 2 | +0.160 |
| SCO | 320 v 244 | 163 v 245 | 185/3 (43.3) v 181 | 234 v 203 | 277 v 278/6 (42.5) | 3 – 2 | +0.102 |

† NED won their 2nd match in a Super Over

  - Both one big loss, NED benefits batting first (opponent rate in fewer overs)
  - Equivalent rate batting 2nd: 390.6 (50) v 315 (50) $\Longrightarrow$ NED NRR = +0.095
    - So, SCO would have gone through!

# MATH IN SPORT

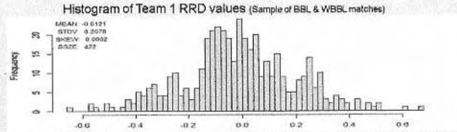
- In 2008 (at this conference), I introduced alternative based on
  - Relative Resource Difference

    RRD = Proportion of winner's available resources they did not need

    $= |S_2 - P_2| / \max\{S_2, P_2\}$

    (where $S_2$ = Final score of chasing team & $P_2$ = DLS par score of chasing team at end)
  - Margin is RRD for winner, –RRD for loser.
    - RRD $\in [0,1]$, so each match equal importance
    - RRD accounts for wickets as well as runs (as uses DLS par score)
    - RRD equivalent for batting first or second

Histogram of Team 1 RRD values (Sample of BBL & WBBL matches)




# MATH IN SPORT

- (More Recent) EXAMPLE:
  - SUPER SIX stage of most recent ODI CWC Qualifiers:

| TEAM | MATCH | | | | | W-L | NRR |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | | |
| NED | 315 v 319/4 (40.5) | 374† v 374 | 192 v 213 | 320 (44) v 246 (44) | 278/6 (42.5) v 277 | 3 - 2 | +0.160 |
| SCO | 320 v 244 | 163 v 245 | 185/3 (43.3) v 181 | 234 v 203 | 277 v 278/6 (42.5) | 3 – 2 | +0.102 |

† NED won their 2nd match in a Super Over

- So, replace NRR with sRRD = sum of RRD values

# MATH IN SPORT

- (More Recent) EXAMPLE:
  - SUPER SIX stage of most recent ODI CWC Qualifiers:

| TEAM | MATCH | | | | | W-L | sRRD |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | | |
| NED | 315 v 319/4 (40.5) | 374† v 374 | 192 v 213 | 320 (44) v 246 (44) | 278/6 (42.5) v 277 | 3 - 2 | +0.053 |
| SCO | 320 v 244 | 163 v 245 | 185/3 (43.3) v 181 | 234 v 203 | 277 v 278/6 (42.5) | 3 – 2 | +0.086 |

† NED won their 2nd match in a Super Over

- SCO would have gone through, as it should have been!
  - Not high profile enough, I guess.

*Math FOR Sport* widely accepted and utilised

*Math IN Sport* maybe a little ways to go!

# MATH _ SPORT:

## The Preposition Proposition

***THANKS!***

***& SOME MORE FAMOUS SPORTS NUMBERS***

***QUESTIONS?***

Steven Stern
Professor of Data Science
Centre for Data Analytics
Bond University
Gold Coast, AUSTRALIA

# USE OF TECHNOLOGY BY VOLUNTEER JUNIOR CRICKET COACHES IN AUSTRALIA

Anthony Bedford [a,d] , Erica Mealy [a], Peter Embleton [a]
[a] *University of the Sunshine Coast*
[d] *Corresponding author:* *abedford@usc.edu.au*

**Abstract**
The majority of coaches in Australian amateur, youth and children's sports are volunteers. Whilst a significant contribution is made, and needed for the sport, we consider what this cohort's understanding is of technology. Specifically, this study considers how this cohort uses technology in cricket via a survey instrument.
The survey was broken into four principal areas: demographics, use of apps and video for coaching, confidence in identifying a dangerous bowling action and facilities available at their training venues. While the use of apps and video is widespread, there is still a usage gap. We explore those gaps through our results and highlight variation in both app and video use. Significant opportunities regarding the development of training and services provided by state and national organisations have been identified. There were also opportunities identified in the use of video as both an immediate and long-term feedback mechanism. We find there is the opportunity for leadership in this space to develop best practices for the use of the recorded footage, its subsequent analysis, and the way it is stored.

*Keywords*: Cricket, coaching, app, video

## 1. INTRODUCTION
Most coaches within junior cricket (primarily under 11 – under 17) in Australia are volunteers: they are typically parents of players in the team who may have had some cricket experience in their past. These people are typically time-poor and are also very commonly the winter sports coach (soccer, rugby etc). Studies looking at this demographic are very scarce, possibly due to the difficulty of engaging with people who already give up so much of their time in volunteer activities.
This study is primarily focused on the use of technology-based tools (primarily apps and video) that are used for coaching. As part of this focus, the participants were asked to self-identify their level of comfort with troubleshooting technology, how they may choose to learn new skills and if they felt they had any speciality in coaching different skills. These extra dimensions were added to provide a clear context to the answers and help identify any lessons learned from the use of these technologies and why they stopped, if they had stopped.
Additionally, coaches were asked to self-assess their ability to identify a dangerous bowling action, the response to this result was particularly interesting given the amount of disagreement in the medical and professional areas about what this constitutes.
The research questions to be answered by this survey are focused on the coaches' previous experience with technology in their coaching, with an app being the most likely tool they will have used. There will be interesting outcomes in terms of the lived experience of amateur coaches with video technologies when compared with the overwhelmingly positive results found in the literature review.

## 2. METHODS
Prior to starting the ethics process, the state cricket association, Queensland Cricket, was approached for support in distribution. The help from Queensland Cricket in getting the survey out was invaluable and the team thank and acknowledge them for their help.
After obtaining ethical approval, information was gathered via an anonymous survey. The survey was designed to be time efficient and accessible on mobile devices. The participants are perceived as being very time poor and not overly motivated to participate in surveys. The desire to be respectful of the coach's time led to most of the questions framed as multiple choice with a final option to provide "other" information if they desired.
The survey was conducted in 2022, during the changeover of the traditional Winter and Summer sports (in Australia this is August and September). This time frame was chosen to pick up both the coaches that continue with cricket over winter and to pick up the ones that only coach over summer, before they are potentially too busy to respond. While the analysis of the data was de-identified, there was an opportunity for the participants

to leave a contact email address if they were interested in receiving more information from the project or trying out new video technologies in their coaching.

The online survey platform Qualtrics (Qualtrics, Provo, UT) was used to present the survey and gather the data. With the help of Queensland Cricket, coaches from all over Queensland were sent the invitation to participate. The qualifying criteria was “coached a junior cricket team in the past 2 years” to try and make the inclusion as broad as possible while still getting up to date information. There were 49 responses to the survey, of which 42 contained responses that were analysed.

## 3. RESULTS

SURVEY FINDINGS

Basic demographic and experience levels are shown below:

| What is your age range? | Total |
|---|---|
| 18-25 | 2 |
| 25-35 | 2 |
| 35-45 | 15 |
| 45-55 | 21 |
| 55+ | 2 |
| **Total** | **42** |

| How many years have you been coaching? | Total |
|---|---|
| 1-2 years | 7 |
| 2-3 years | 5 |
| 4-5 years | 10 |
| 5 + years | 20 |
| **Total** | **42** |

**Highest qualification**

| Highest qualification | Total |
|---|---|
| I don't have any formal qualifications | 3 |
| Level 1 Community Coach | 22 |
| Level 1.5 Community coach | 4 |
| Level 2 Representative Coach | 11 |
| Other | 2 |
| **Total** | **42** |

Table 1: (a) Ages (b) Coaching Years and (c) Qualification

Cricket Australia runs a certification scheme for junior cricket coaches with many associations that specifies all coaches should be at least “Level 1 Community Coach” certified. To help achieve this, within Queensland, there are several free clinics every year, run by Queensland Cricket, to assist coaches certified to this level. These certifications are known interchangeably by either a level number or a description of the accreditation. For the survey, these two nomenclatures were combined to form a single descriptor that would be meaningful regardless of the local name of the certification:

| Survey option | Cricket Australia description |
|---|---|
| Level 1 Community Coach | Community Coach |
| Level 1.5 Community Coach | Advanced Coach |
| Level 2 Representative Coach | Representative Coach |

Table 2: Levels of coach

The 1.5 Community Coach (Advanced Coach) is a new certification, and currently, few individuals hold this certification. While most coaches in junior cricket are volunteers, it was expected that paid coaches were likely to use technology in different ways. To explore this, the survey recruitment criteria did not exclude paid coaches, however, their data was analysed separately. Only three of the forty-two were paid coaches.

DANGEROUS BOWLING

A side note to this work, was the self-rated ability of coaches to determine a dangerous bowling action. Respondents were asked their level of confidence in detecting an unsafe bowling action. This is shown in Figure 1, with the years coaching mapped against this criterion.

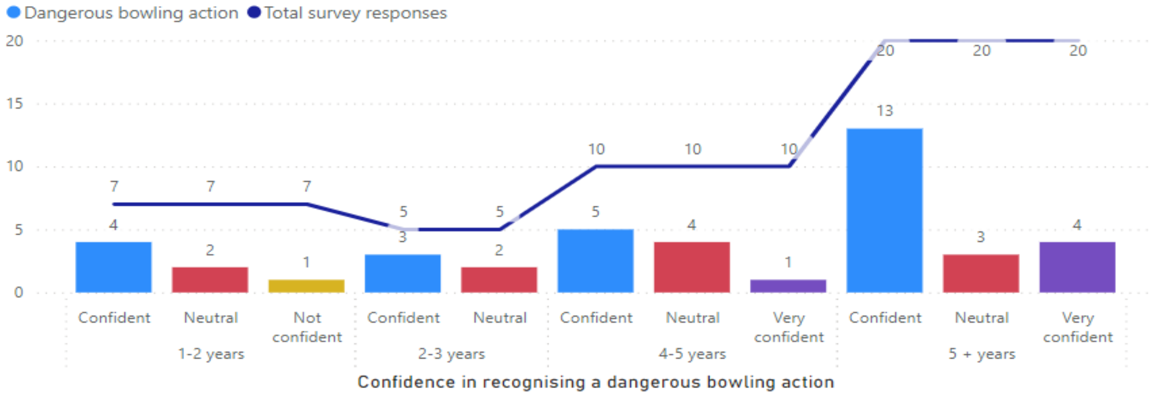



Figure 1: Bowling Action by Years Coaching

UTILITY OF APPS

Establishing a baseline in our work is important for understanding app usage by coaching level and age group. Figure 2(a) & (b) illustrate these dimensions.

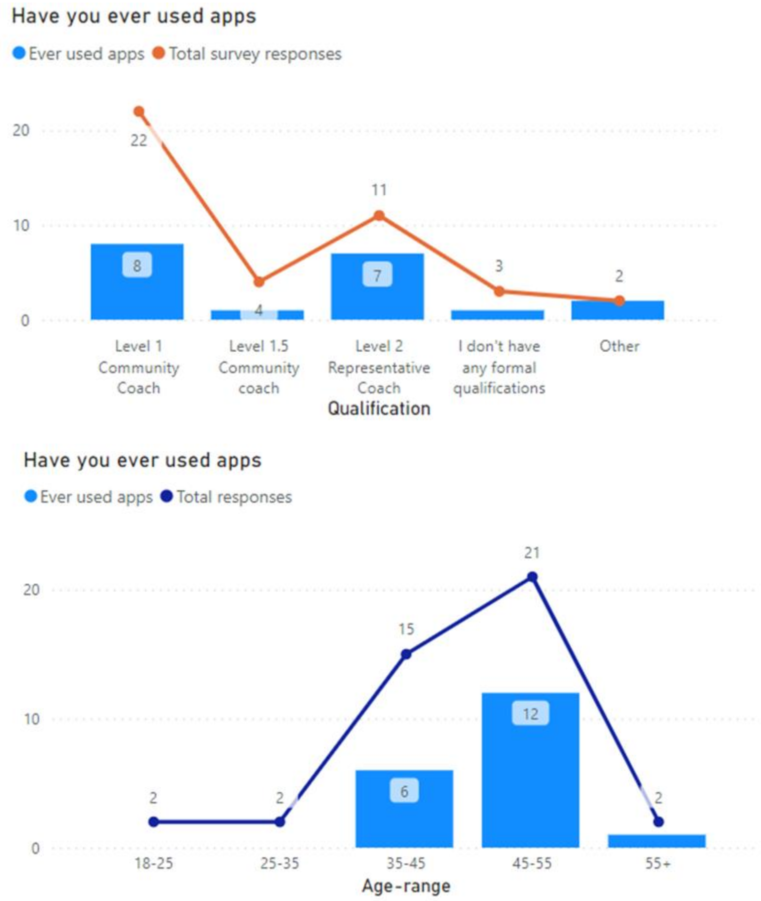



Figure 2: (a) App use by Level of Coaching (b) App use by Age

Whilst coaches described a variety of apps, Table 3 shows both the types of apps used and how many coaches utilised them. Most respondents reported limited use: twelve used only one app, two used two apps and one used three.

| App name | Count | Type |
|---|---|---|
| CA Coach App | 9 | Coaching skills (no video analysis) |
| Coaches eye | 4 | Video analysis |
| NX Cricket HD | 1 | Scoring / match reporting |
| Cricket Coach plus | 1 | Video analysis |
| Kinovea | 1 | Video analysis |
| Fulltrack AI | 1 | Video analysis |
| Frogbox | 1 | Scoring / match reporting |
| Unspecified | 6 | Did not describe the app |

Table 3: Apps used by coaches, and their type

Figure 3 isolates the utility of the app by each Coaching preference. As seen, the perceived effectiveness of app use was rated as “Neutral”, “Somewhat effective” or “Very effective”, irrespective of style.

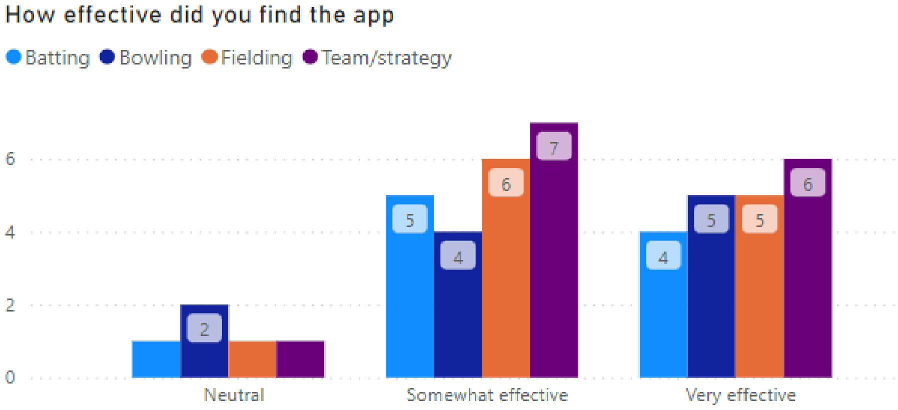



Figure 3: Effectiveness by Coaching preference

An interest was the ongoing use of apps, with results in Figure 4 showing the findings of use by Level of coaching.


Figure 4: Use of apps by Level

The way the app was used was of interest, with one respondent stating *Not enough time in training.* Five respondents said *Hard to use while coaching*, whilst another two stated *Didn’t find it effective with the group.* Thus, we look at the years coaching by the responses, as shown in Figure 5.

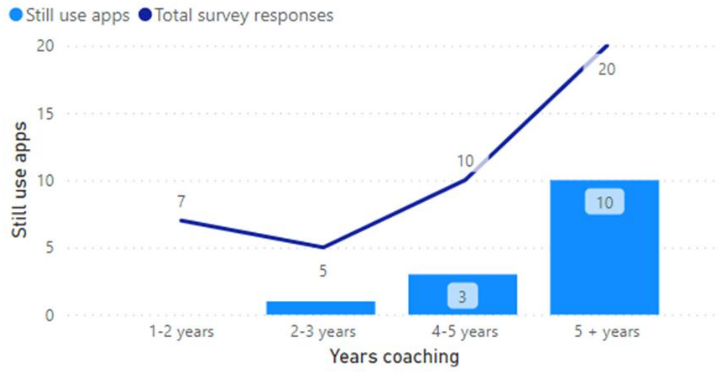



Figure 5: Use of apps by years coaching

Now, we explore the utility of recorded footage – the contention of further work which will pose estimation to remove all visual noise from a video of a player. Figure 6 shows the relationship between the level of coaching and the use of the video, as well as age. We have included effectiveness in the age plot.

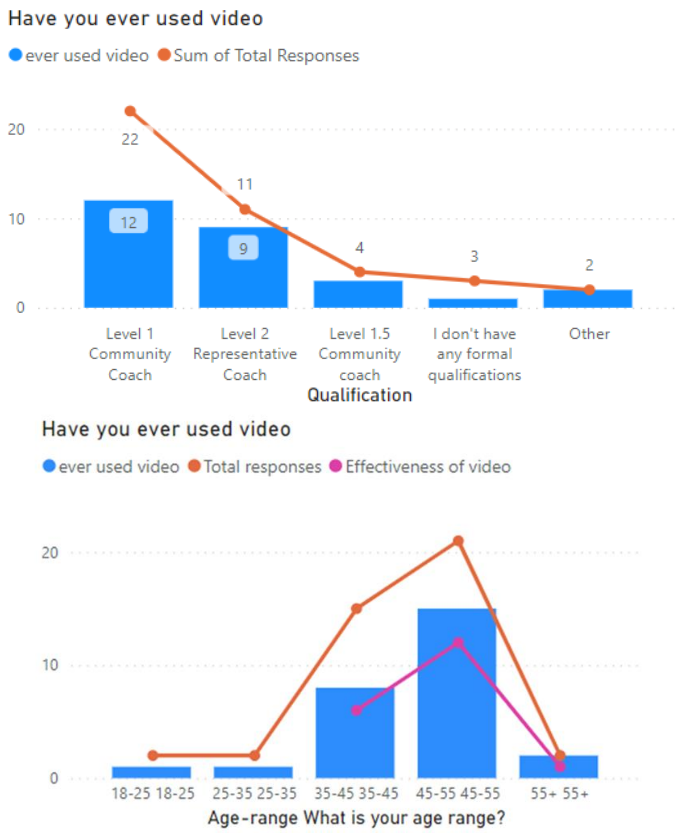



Figure 6: (a) Use of recorded footage by coaching level (b) Use of recorded footage by age

## 4. DISCUSSION

The basic demographic findings show a predominance of coaches with over four years' experience – 30 of the 42 respondents fell into this category. Most of the respondents held a Level 1 Community Coach certification or higher, and coaches of all levels reported feeling confident/very confident in detection of dangerous bowling action (Figure 1). When considering the use of apps, only 36.4% of Level 1 coaches reported using apps, while coaches with higher certifications did so in greater proportions (Figure 2(a)). CA Coach App was reported as the most used app (Table 3). Figure 3 outlines app effectiveness by coaching preference, with Team/Strategy coaches reporting the highest levels of effectiveness. Figure 4 highlights that app use and longevity vary among certification level, where ongoing use of apps remains limited, especially in Level 1 and 1.5 Community Coaches.

*Use of Video in coaching*

As a secondary outcome of the survey, the utility of recorded footage was also of interest. While there is a level of overlap with apps, the prevalence of smart phones with built in cameras can, in many cases, provide highly capable coaching tools using built in software. Cummiskey (2011) notes that "As equipment prices drop, mobile phone are ubiquitous and new processing technologies become mainstream, the use of video and phone apps are becoming more mainstream'.

Wilson (2008) finds "Sometimes, athletes who make small changes to technique feel that they have made large adjustments, yet the coach continues to press for larger changes. Viewing a video of their performance with the coach allows the athlete to see exactly how much change has been made".

While the paper focused on professional coaches in specialist spaces, it shows why volunteer coaches may use recorded footage in their coaching strategies. Overall, there was an increase in reported use of recorded footage compared to other use of apps. However, given the ubiquity of mobile devices with high quality in-built cameras, to have over 35% of volunteer coaches report having never used a video recording from their mobile phone (or other) was surprising (Figures 5 and 6). In contrast, 2 out of the 3 paid coaches reported using video for their coaching.

Of the 76% respondents who reported recorded footage as an effective coaching tool, only 69% of coaches reported utilising it, highlighting a gap between value and practicality. Further, only 41% of those who use recorded footage reported that it is reviewed by "Others", leaving another gap in the how footage is utilised. Additionally, given that most respondents reported coaching skills in "Team/Strategy" and "Fielding" rather than the more technical areas, there may be value from external support for skills such as Batting and Bowling.

## 5. CONCLUSION~~S~~

While the data is not fine-grained or a large enough sample to support definitive conclusions, the relatively small number of new app users may suggest a decline in app uptake. Given that most respondents rated apps as the "Somewhat effective" or higher, there is potential gap here that can be addressed.

**Acknowledgements**

Queensland Cricket for their support of this research through the distribution of the survey. Ethics approved by The University of the Sunshine Coast, S221720.

.

**References**

Cummiskey, M. (2011). There's an app for that: Smartphone use in health and physical education. Journal of Physical Education, Recreation & Dance, 82(8), 24–30.

Wilson, B. D. (2008). Developments in video technology for coaching. *Sports Technology, 1*(1), 34–40.

# STAKING AND BETTING BEHAVIOURS OF PROFESSIONAL PUNTERS

Anthony Bedford [b,c], Erica Mealy [b], Alicia Velcich[b], Tristan Barnett[a]
[a] *Macquarie University*
[b] *University of the Sunshine Coast*
[c] *Corresponding author: abedford@usc.edu.au*

**Abstract**
Gambling and wagering are risk-taking behaviours in which a punter seeks reward above that which is staked. But when does the size of the potential loss influence their risk-taking behaviour, and how can it be mitigated? In this work, we build on previous research examining the role of data in betting decisions (Bedford & Barnett, 2023). We expand through analysis of survey responses to ascertain behavioural changes between professional and amateur punters under differing conditions.

This two-phase study began with respondents undertaking the Investment Risk Tolerance Assessment to establish their baseline risk tendencies. In the second phase, respondents answered a questionnaire covering three key factors: betting and trading decisions, the influence of external factors, and the psychological underpinnings of betting behaviour. These responses were analysed to understand how self-perceived betting behaviours change under different risk conditions, such as wagering with one's own money, their employer's, or of an unrelated third party.

Our analysis revealed behavioural trends and differences between professional and amateur punters. Drawing on an industry case study, traders employed a model in which the betting decision process is intentionally abstracted from direct outcomes. This structure, likened to the simulated detachment in Orson Scott Card's, Ender's Game, creates psychological distance between the bet and the emotional weight of the loss. Notably, using a per-unit approach rather than focusing on absolute stake size highlighted significant shifts in trades and bets when the potential cost of loss was considered.

We conclude by proposing operational structures designed to mitigate the psychological effects of losses in betting decisions. These include the introduction of feedback mechanisms aimed at supporting more rational decision making among traders, analyst and other key stakeholders involved in wagering and trading environments.

*Keywords*: Betting, wagering, staking, sport, horseracing, sports betting

## Introduction

Gambling and wagering are forms of risk-taking behaviour, where the prospect of gain outweighs the risk of loss. These behaviours are influenced by psychological, biological, and social factors, with outcomes shaped not only by chance, but also by an individual's perceptions and decision-making (Griffiths, 1991).

Gambling holds substantial economic and psychological implications. The Australian Government Productivity Commission (2010) reported that gambling industries generate considerable income, contributing to national revenue and employment. However, these benefits are offset by significant psychological costs, including addiction, financial stress, and mental health issues.

A critical component in understanding gambling behaviour is risk perception—how individuals evaluate potential losses and gains. These evaluations are shaped by emotional states, cognitive biases, and personal experiences, all of which influence decision-making under uncertainty (Clark, 2010). A core element of this is loss aversion, a well-established concept in behavioural economics, which describes the tendency for individuals to assign greater weight to losses than to equivalent gains (Kahneman & Tversky, 1979). This can lead to irrational behaviour, particularly when punters attempt to recover from losses.

Decision-making in gambling is also influenced by the emotional attachment to betting outcomes. Orson Scott Card's Ender's Game offers a useful metaphor: decisions made in a simulated environment, are psychologically detached from the reality of their consequences. Similarly, in professional gambling environments, decisions are often made within structured frameworks—using abstracted models or unit-based systems that create separation between the bet itself and the emotional weight of winning or losing. This detachment may reduce the impact of risk perception and loss aversion and support more strategic decision-making in staking environments.

*Submitted to MATHSPORT 2024*

This study compares two distinct groups operating within different psychological and structural conditions: professional traders, who predominately make data-driven decisions within structured environments, and amateur punters, who gamble recreationally and are potentially more emotionally driven. It is hypothesised that these groups differ not only in their approach to risk, but also in how they respond to model or expert recommendations, financial stakes, and decision making.

While prior research has explored the psychological factors of risk perception and loss aversion in gambling contexts, there is limited work directly comparing professionals and amateur punters under varying risk conditions. Few studies have examined how factors such as stake ownership, model use, and abstraction (e.g., units vs. dollars) influence behaviour across these distinct groups. This study aims to explore behavioural differences between professional and amateur punters, focusing on how risk conditions, stake abstraction, and modelling influence decision-making under differing risk conditions.

**Methods**

This study utilised a two-phase design. First, participants completed the Investment Risk Tolerance Assessment (IRTA) to establish their baseline financial risk tolerance; second, they completed a questionnaire on gambling related decision making.

Participants were recruited through purposive sampling to ensure individuals had exposure in either professional or recreational betting. A total of nineteen participants were recruited for the study and categorised as either professional traders – individuals who engage in structured, model-based betting environments, or as amateur punters – individuals who gamble recreationally with less formal structure.

To assess participants tolerance for financial risk, the IRTA was utilised. The IRTA is derived from the Grable and Lytton (G&L) risk tolerance scale, which has been used in over 160,000 cases and demonstrates high internal consistency ($\alpha \approx 0.77$) and construct validity, providing a robust method to distinguish individual differences in financial risk (Kuzniak et al., 2015). The IRTA scale categorises respondents into one of five levels of risk tolerance based on their accumulative score (Table 1).

Participants then completed a questionnaire covering betting decisions, the influence of external factors, and psychological aspects of gambling under varying risk conditions. The survey included questions related to model recommendations, behaviours towards different staking and perceived influences on betting behaviours. Responses from the survey were analysed using chi-square tests to assess group level differences.

Table 1: IRTA Scale Categories of Financial Risk Tolerance

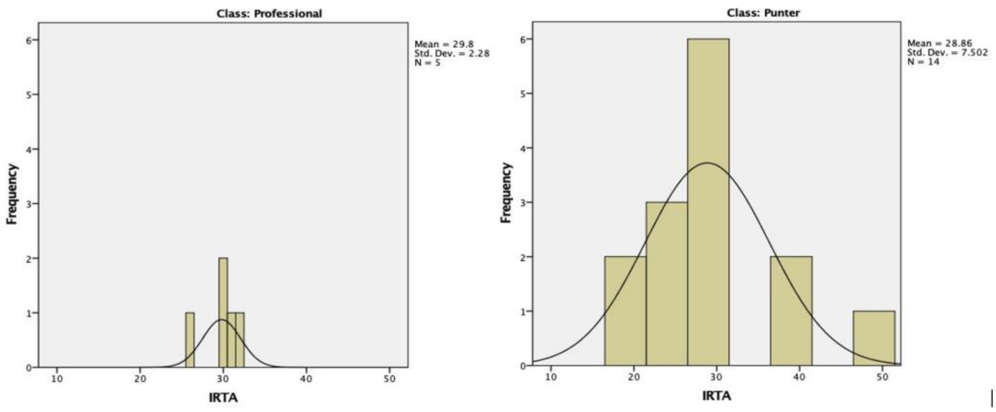
| Score | Level |
|---|---|
| 33 – 47 | High tolerance for risk |
| 29 – 32 | Above-average tolerance for risk |
| 23 – 28 | Average/moderate tolerance for risk |
| 19 – 22 | Below average tolerance for risk |
| 0 – 18 | Low tolerance for risk |

Table 1 shows the IRTA scale categories to measure an individual's financial risk tolerance

**Results**

The distribution of IRTA scores for punters and professionals illustrates differences in financial risk tolerance across the two groups.

Figure 1: Distribution of IRTA Scores for Professionals and Punters




The mean IRTA score was slightly higher for professionals ($\overline{x}$ = 29.8, s = 2.28, n = 5) compared to punters ($\overline{x}$ = 28.86, s = 7.50, n = 14). Both groups indicate an above average tolerance risk.

Survey question results have been compiled into a table below to highlight key chi-squared measures, including group means and significant values.

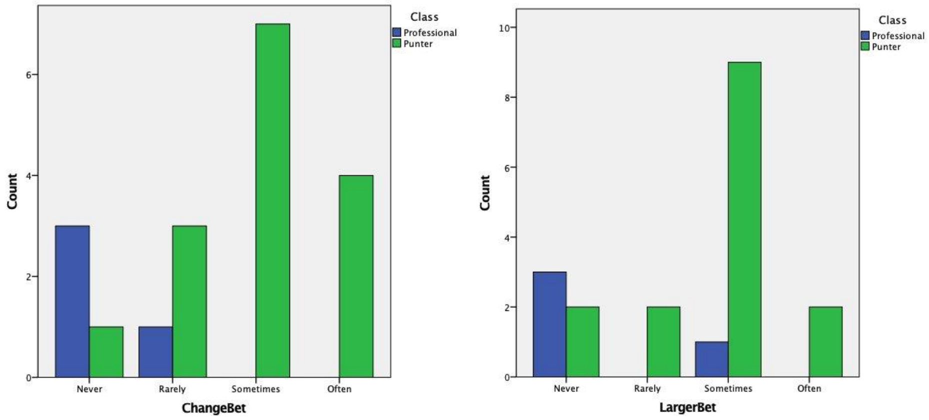
Table 2: Survey Responses Summary

| **Survey Question** | $\chi^2$ | ***p*-value** | $\overline{x}_{\mathbf{prof}}$ | $\overline{x}_{\mathbf{punt}}$ | **Significant** |
|---|---|---|---|---|---|
| Bet/traded differently with own vs others' money | 4.388 | 0.356 | 2.80 | 3.60 | No |
| Made bet/trade decision different to model due recommended to stake size | 15.132 | 0.002 | - | - | Yes |
| Felt pressured to trade/bet more due to external factors | 5.205 | 0.157 | 1.80 | 2.57 | No |
| Importance of bet/trade size in decision making | 4.285 | 0.371 | 3.00 | 4.07 | No |
| Made a bet/trade larger than model suggestion | 10.233 | 0.017 | - | - | Yes |
| Made a bet/trade smaller than model suggestion | 0.889 | 0.828 | 2.40 | 2.71 | No |
| Belief that model's market resembles actual market | 11.909 | 0.018 | - | - | Yes |
| Changed bet/trade due to market conditions, contrary to model suggestion | 5.002 | 0.172 | 3.20 | 2.86 | No |
| Changed bet/trade due to news/events, contrary to model suggestion | 0.835 | 0.841 | 2.60 | 3.00 | No |
| Changed bet/trade due to client demands, contrary to model suggestion | 1.859 | 0.602 | 2.20 | 2.57 | No |
| Changed bet/trade due to internal factors, contrary to model suggestion | 0.827 | 0.661 | 2.20 | 1.86 | No |
| Psychological difficulty on bet/trade for underdogs model vs favourites model | 5.763 | 0.218 | 2.60 | 3.07 | No |
| Psychological ease on bet/trade using units vs dollars | 2.239 | 0.692 | 3.40 | 2.71 | No |
| Most appealing bet/trade model (Bet and dollars / bet and units / bet only) | 13.842 | 0.001 | - | - | Yes |

Summary of chi-squared, group means and *p*-values for survey responses for professional and amateur punters. Statistically significant results ($p < 0.05$) are indicated.

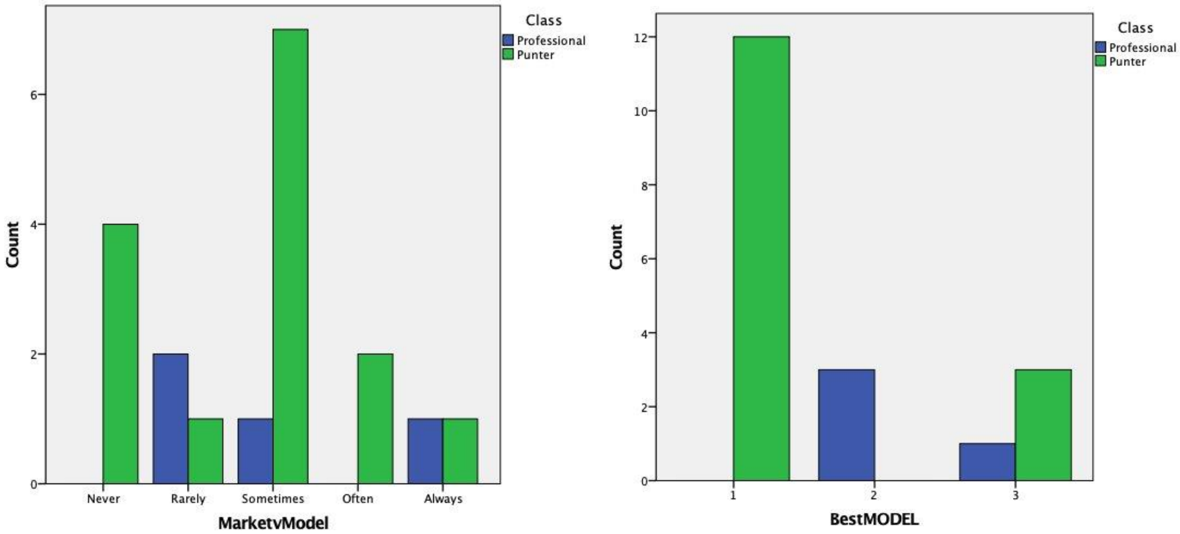
The following figures present survey responses that yielded statistically significant results. These grouped bar charts illustrate differences between professional and amateur punters in relation to key decision making and psychological factors in betting and trading settings.

Figure 2: Survey Responses to Model Recommendations
(a) “Made bet/trade decision different to model due recommended to stake
(b) “Made a bet/trade larger than model suggestion”
size”




(a) Professionals reported “rarely” or “never” deviating from model recommendations, whereas punters demonstrated more flexibility on stake size ($\chi^2$= 15.132, *p* = 0.002).
(b) Professionals reported either “never” or “sometimes” increasing bet/trade size beyond model recommendations, whereas punters reported more varied responses ($\chi^2$= 10.233, *p* = 0.017).

Figure 3: Survey Responses to Model Accuracy and Preference
(a): “Belief that model’s market resembles actual market”
(b): “Most appealing bet/trade model”




(a) Professionals were relatively evenly distributed across “rarely”, “sometimes” and “always” believing in a model's market prediction. In contrast, punters most frequently reported “sometimes” or “never” believing in a model's market prediction ($\chi^2$= 11.909, *p* = 0.018).
(b) Model preferences are indicated by 1 = "Bet and Dollars", 2 = "Bet and Units” and 3 = "Bet Only”. Punters preferred the “Bet and Dollars” model, whereas professionals were more varied, with a greater proportion preferring “Bet and Units” ($\chi^2$= 13.842, *p* = 0.001).

**Discussion**

In team sports, athletes are trained through simulation and strategy in preparation for real competition. In professional betting environments, the approach is similar—modelling and simulating outcomes before betting decisions are finalised. The difference lies in the consequences, particularly how loss is experienced. Or does it?

An operational betting setup observed during this study found analysts and traders worked in close physical proximity, sharing a workspace where trades were made, monitored, and adjusted, with immediate feedback. This lack of physical and psychological separation (between those creating models and those executing decisions) led to visible tensions. Bets were occasionally changed, creating an atmosphere of distrust and frustration, as decisions became reactive rather than strategic.

Over time, this model was revised: analysts were relocated to a separate office space, and a trade manager was introduced to oversee the link between strategy and execution. Importantly, professionals transitioned to using a unit-based system, abstracting financial decisions away from raw dollar values. This shift created a buffer between emotional reactions and rational decision-making, allowing traders to operate with a clearer focus and enhanced strategy. This case study aligns with the quantitative findings of this research. Professionals in our survey showed a strong preference for unit-based betting models and rarely deviated from model-driven decisions due to stake size perceptions. Punters, by contrast, preferred dollar-based models and frequently adjusted decisions when financial amounts were perceived to be too high or too low. This reflects a deeper entanglement with emotional aspects of money.

The separation of roles—between "risk makers" (analysts) and "risk takers" (traders)—appears to be crucial. In this sense, the betting environment mirrors the conceptual world of Ender's Game, where simulations feel detached from reality, but outcomes are deeply consequential. Professionals preferred to operate within abstracted systems which removed the immediacy of loss and gain, focusing instead on strategic consistency. In this setup, winning is paramount not for ego, but because failure compromises the survival of the business.

This notion—separating the mode of winning from the cost of winning—is central to what we describe as the Ender's Game effect. By disentangling dollar values from decision-making and fostering psychological distance through structure, professional environments can reduce the effects of loss aversion and promote stable, rational risk-taking.

**Conclusion**

This study revealed clear behavioural differences between professional traders and amateur punters in high-risk betting contexts. Professionals demonstrated greater consistency in following model recommendations and showed a strong preference for unit-based staking, allowing them to remain more psychologically detached from monetary values. In contrast, punters were more emotionally reactive, often adjusting their decisions based on the size of the stake and demonstrating less trust in model recommendations overall.

Observational findings further supported this divide. When analysts and traders were separated and bets reframed into units, team members reported reduced emotional interference and improved trust. While not measured directly, these insights suggest that role separation and abstraction may support more effective decision-making.

The findings of this research highlight the value of psychological and structural tools in shaping betting behaviours. Whether among professional traders or recreational punters, rational risk-taking begins not with better models, but with better psychological distance. Like in Ender's Game, the most effective decisions are made when the weight of emotional investment is lifted from the decision – not because the consequences of the decision are insignificant, but because the clarity of strategy depends on it.

**Acknowledgements**

We wish to acknowledge Tristan Barnett's assistance on previous work in this area, and Alicia Velcich for compiling results.

**References**

Bedford, A, & Barnett, T (2023). Stake Size and Wagering in a Professional Betting Environment–when data affects decision making, 18th International Conference on Gambling and Risk Taking.

Kuzniak, S, Rabbani, A, Heo, W, Ruiz-Menjivar, J & Grable, J.E (2015). The Grable and Lytton risk-tolerance scale: A 15-year retrospective. Financial Services Review. 24. pp 177 – 192

Australian Government Releases Productivity Commission Report Into Gambling 2010, , Washington, D.C.

Griffiths, M. (1991). Psychobiology of the Near-Miss in Fruit Machine Gambling. The Journal of Psychology, 125(3), 347–357. https://doi.org/10.1080/00223980.1991.10543298

Clark, L. (2010). Decision-making during gambling: An integration of cognitive and psychobiological approaches. *Philosophical Transactions of the Royal Society B: Biological Sciences, 365*(1538), 319–330. https://doi.org/10.1098/rstb.2009.0147

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263–291. https://doi.org/10.2307/1914185