The Utility of Mixed Models in Sport Science: A Call for Further Adoption in Longitudinal Data Sets

Author

Newans, Tim, Bellinger, Phillip, Drovandi, Christopher, Buxton, Simon, Minahan, Clare

Published

2022

Journal Title

International Journal of Sports Physiology and Performance

Version

Accepted Manuscript (AM)

DOI

https://doi.org/10.1123/ijspp.2021-0496

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

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

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

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

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

Introduction

Highly-controlled, longitudinal, or cross-over designed research experiments are difficult to conduct in the high-performance sport environment. Researchers are regularly met with hesitation from both coaching staff and the athletes due to the potential disruption 'best-practice' research methodology poses to competition and their carefully configured training programs. Nonetheless, to advance the understanding of performance and adaptation in elite athletes, sports scientists, and researchers often interrogate routinely collected health[1–3], wellbeing[4–6], physical and physiological performance metrics, and locomotive movement data (i.e., movement patterns)[7–10] over multiple days, weeks, and years. Indeed, these observational studies have played a critical role in the understanding and development of sports performance and related disciplines over the last 20 years. However, these datasets are often characterized by multiple dependent observations (not only across matches and competitions, but within each athlete) and imbalanced data (e.g., athletes missing due to injuries, team selections, and/or rescheduling) that pose a significant challenge for this type of research. Nonetheless, the nuisances associated with dependent observations and imbalanced data are among the least acknowledged when conducting sports-science research.

Analysis of variance (ANOVA) with repeated measures has dominated as the most frequently utilized statistical method with which to analyse repeated measures datasets. However, ANOVA requires every participant to have a value in every observation and within a condition (e.g., position) every participant must have contributed an equal number of observations to avoid violating the relatively stringent assumptions[11]. Despite these assumptions, ANOVAs are still readily used within longitudinal sports science research[12–16]. A systematic review of contextual factors on match running in rugby league[17] reported that only seven of fifteen studies using the Global Navigational Satellite System (GNSS) to quantify movement patterns in athletes during match-play accounted for repeated measures within athletes (i.e., dependent observations). Of these seven, two studies eliminated data to run their ANOVA, while another study did not account for the athletes' natural variation in performance, leaving only four studies utilising their full dataset. Clearly, the characteristics of routinely collected health, wellbeing, and performance datasets in the high-performance sport environment can involve the manipulation of data, and depending on its severity, can undermine the confidence of analysis.

Obvious to the statistician and data analyst is that mixed models are likely a more appropriate statistical methodology for routinely collected health, wellbeing, and performance data such as the data in the present study. Indeed, within statistics literature, McElreath argues that mixed models deserve to be the default form of regression and that experiments that do not use a mixed model should justify not using this approach[18]. In his argument, he lists four reasons to use mixed models: i. To adjust estimates for repeat sampling, ii. To adjust estimates for imbalance in sampling, iii. To model variation among individuals or other groups, and iv. To avoid averaging, as some researchers pre-average data before running analyses[18]. Therefore, it is reasonable to suggest that mixed models are the most appropriate statistical methodology to analyse longitudinal datasets often acquired by sports scientists. This aligns with previous guidance by Hopkins and colleagues (2009) in encouraging sports science researchers to utilise mixed models rather than a repeated-measures ANOVA[19]. While mixed models have become more popular within sport-science research[4,7,8], the use of mixed models are often perceived as technical and requiring statistical expertise. Thus, the aim of this study was to make sports scientists aware of the utility of mixed models when analysing longitudinal datasets. To illustrate this, we described and compared the movement patterns of female rugby league athletes across three levels of competition. We hypothesised that match intensity, as determined by the mean speed, would be higher in Origin matches due to the higher quality players compared to NRLW matches and shorter duration compared to Test matches.

Methods

Subjects

Over the 2018-2019 seasons, we were provided access to athlete positioning and timing data (i.e., movement patterns) transmitted by the GNSS for Australian female rugby league athletes including: club- (i.e., National Rugby League Women's; NRLW), state- (i.e., State of Origin; Origin), and international-level (i.e., Trans-Tasman Test; Test) competition. These data provided a unique opportunity to identify differences in movement patterns across three levels of rugby league competition previously observed in female athletes of other football codes[9,20]. Importantly, the rugby-league data examined in the present study included repeated match data for multiple athletes which varied in number and level of competition. That is, some athletes playing NRLW played more matches than others, and some also played Origin and Test matches. This created a data set of repeated and dependent measures as well as an imbalance in sampling among athletes and competition level. Therefore, careful consideration of the statistical approach was paramount to derive meaningful conclusions.

Figure 1 shows the sample dataset that contained 109 athletes in 12 NRLW matches during the 2018 and 2019 seasons, 49 athletes in the 2018 and 2019 Origin matches (2 matches), and 26 Australian 'Jillaroos' athletes in 2018 and 2019 Trans-Tasman Test matches (2 matches). Collectively, the present study included 115 unique athletes (age = 26.8 ± 5.4 yr; height = 1.68 ± 0.07 m; body mass = 76.7 ± 11.9 kg) and 472 match entries.

[INSERT FIGURE 1 ABOUT HERE]

Design

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

Methodology

The movement patterns of athletes competing in NRLW matches were collected using 10 Hz Optimeye S5 receivers (Catapult Sports, Victoria, Australia), while the Origin and Test movement pattern data were collected using 10 Hz GPSports EVO receivers (Catapult Sports). To measure match intensity, the mean speed (i.e., total distance divided by the time spent on the field) of each athlete was recorded. While it was not ideal that units by different manufacturers were used, only the total distance was required for this analysis which has been found to differ by 1.8% between these two devices[21]. As the mean speed of NRLW athletes has been found to decrease as function of the minutes played, a minimum threshold of 20 min of game time was applied to ensure the athlete spent an adequate time on the field. Additionally, this filter removed athletes playing for a short duration and therefore recording a high mean speed (i.e., metres/minute) if there were no stoppages in the short time they were on the field. Consequently, of the 472 match files, 380 were deemed eligible for analysis. The Griffith University Human Ethics Committee approved this study (GU Ref No: 2019/359).

Statistical Analysis

To compare the differences in analysis, both a repeated-measures ANOVA and a mixed model were fitted to the present study dataset. In both analyses, the effect of a player's position and level of competition on mean speed was assessed. For the ANOVA to be performed, we restructured the data to achieve a 'complete-case analysis'[22]. The mixed models were built using the *lme4*[23] package in R version 4.0.2[24], while the *emmeans*[25] and *sjPlot*[26] packages were used for pairwise comparisons and model diagnostics respectively. The *see*[27] and *performance*[28] packages were used to determine which model was the best performing model. The associated R script is attached as an appendix to this manuscript.
The dataset was arranged in 'long form' with each observation for each athlete on a new row. After loading in the *lme4* packages, the first model was built as seen in line 15 in Appendix 1. The following components were outlined:

1. Dependent Variable; the metric we were interested in explaining. In this case, mean speed which as expressed as a continuous variable.
2. Fixed Effects; the variables of interest that could explain variation in the dependent variables. Here, we are looking at 'level of competition' and 'playing position'. An interaction effect was also assessed, which would determine whether the difference in mean speed because of level of competition was uniform across all positions or whether the difference in mean speed because of level of competition is different for each position.
3. Random Effects; the variables that contain variation purely from the random sampling which could hide the influence of the fixed effects and would vary if the study were to be replicated[29]. Here, they are: i) 'athlete' as an individual athlete could have played up to 12 matches, ii) 'team' as the team that an athlete plays for (e.g., Broncos, Roosters, QLD Maroons, etc.) could have a confounding effect and is nested within the levels of competition, and iii) 'match' as there is variation match-to-match in the intensity of play, perhaps due to weather, for example. These can be included as random intercepts and/or as random slopes, in this example, all three variables were included as random intercepts. By including these random effects, we can account for the variability associated with these effects to reveal the underlying effect of 'level of competition' and 'position' on the dependent variable.
4. Distribution Shape; in this case, the data adequately met the assumptions for a normal (Gaussian) mixed model. This example used a dependent variable (mean speed) that met the assumptions of a normal (Gaussian) distribution which meant that a simpler linear mixed model could be used in this case. However, this may not always be the case and, therefore the dependent variable may need to be expressed as a factor or percent, or log-transformed to reduce non-normality of the data[19]. Alternatively, if there was frequency data (e.g., tackle count) or probability data, a Poisson distribution or binomial distribution respectively can be specified as the distribution family in the mixed model (using the *glmer* function rather than the simpler *lmer* function).

Once the full model (i.e., all fixed and random effects included, see line 15 in Appendix 1) was established, the full model was compared to models without the fixed and random effects to assess the different model fits (see Line 23 in Appendix 1). When comparing models using the Akaike Information Criterion (AIC), it was determined that the full model was the preferred model as it displayed the lowest AIC. AIC was chosen as it includes a model complexity term (twice the number of model parameters) to penalise models containing variables that did not contribute to the model. The R^2 and Root-Mean-Square Error (RMSE) were also reported for the reader but were not used for model selection. As the models were nested, a likelihood ratio test could also have been used.

To meet the assumptions of a mixed model, a histogram of the residuals was generated to assess the normality of the residuals, Q-Q plots for each random effect were generated to assess the normality of the random effects, and the model's residuals were plotted against its fitted value to assess homoscedasticity (see Line 25 in Appendix 1). None of these assumptions were violated so, therefore, it was deemed appropriate to proceed with the analysis.

Results

In order to meet the assumptions of a repeated-measures ANOVA, we restructured the data to achieve a 'complete-case analysis'[22], which eliminated any athlete that did not feature in all three levels of competition. This initial filter eliminated 48.7% of the available data. Of the remaining data, each athlete varied from one to seven observations within each level of competition which continued to threaten the validity of the analysis as multiple observations of one athlete could bias the results. Secondly, it would not be suitable to answer the hypothesis as eliminating any athlete that did not play in the Test matches would result in the average NRLW mean speed being compiled of only athletes that had played in both Origin and Test matches, rather than the collective NRLW cohort. Alternatively, if we used the mean for each athlete in each level of competition, this reduced the number of observations to 157 (i.e., eliminated 58.7% of the available data). However, when averaging each athlete's values, it would not account for the fact that there is less uncertainty in the mean speed of an athlete that had seven observations compared to an athlete that only has one observation. Therefore, it was determined that it was not appropriate to run an ANOVA across this dataset. Due to the use of ANOVA being deemed inappropriate to adequately analyse these data, only the results of the mixed model are presented here.

Model Comparisons

Seven mixed models were assessed: the full model (i.e., containing all fixed and random effects), three models with each model removing one random effect, and three models removing the interaction effect, the competition level, and position respectively. The full model displayed the lowest AIC, the equal-highest R^2, and the lowest RMSE and, therefore, was chosen as the preferred model. The results are displayed in Table 1. We can write the selected model in short form as:

$$\begin{aligned}\mathrm{MeanSpeed}_{pmti} &= \beta_0 + \beta_1 \times \mathrm{Competition}_{pmti} + \beta_2 \times \mathrm{Position}_{pmti} \\ &\quad + \beta_3 \times \mathrm{Competition}_{pmti} \times \mathrm{Position}_{pmti} + \gamma_p^{\mathrm{player}} + \gamma_m^{\mathrm{match}} + \gamma_t^{\mathrm{team}} + \epsilon_{pmti},\end{aligned}$$

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

[INSERT TABLE 1 ABOUT HERE]

Random Effects

When considering the model comparison, the full model with player, team, and match random effects was preferred over the reduced models that removed each of the random effects, as evidenced by the full model having the lowest AIC value. Similarly, when assessing the variance explained by each of the random effects, it was determined that all three random effects should remain in the model. The estimated random effects variances are displayed in Table 2, together with the estimated residual variance. The conditional means for each level of random effect are displayed in Figure 2. The player conditional means differed between -6.9 to 5.9 $m \cdot min^{-1}$ from the marginal mean, the team conditional means differed between -2.7 to 2.1 $m \cdot min^{-1}$, and the match conditional means differed between -8.7 to 5.8 $m \cdot min^{-1}$ from the marginal mean.

[INSERT TABLE 2 ABOUT HERE]

[INSERT FIGURE 2 ABOUT HERE]

Fixed Effects

Table 3 displays the additive interaction effects to calculate each permutation within the dataset. The intercept (i.e., 74.5 $m \cdot min^{-1}$) represents the mean speed for a Test adjustable player on average. From here, the coefficients can be added or subtracted depending on what level of competition and what position was being played. For example, for the mean speed of an Origin forward athlete, you would start with the intercept of 75.0 $m \cdot min^{-1}$. From here, you will add 8.7 $m \cdot min^{-1}$ for an Origin-level athlete, then add another 1.9 $m \cdot min^{-1}$ for a forward, and then subtract 2.2 $m \cdot min^{-1}$ for the interaction effect of an Origin-level forward, resulting in an estimated mean of 83.4 $m \cdot min^{-1}$.

[INSERT TABLE 3 ABOUT HERE]

Figure 3 presents the estimated marginal means for mean speed recorded during NRLW, Origin, and Test matches. Origin matches recorded, on average the highest mean speed, followed by NRLW matches, and then Test matches. When considering position, in both NRLW and Origin matches, adjustables recorded the highest mean speed; however, forwards recorded the highest mean speed in Test matches. Meanwhile, backs recorded the lowest mean speed in NRLW matches, with interchange recording the lowest mean speed in both Origin and Test matches. As the changes in mean speed across competitions were not uniform across all competition, this confirmed the presence of an interaction effect.

[INSERT FIGURE 3 ABOUT HERE]

Discussion

The present study displayed why mixed models should be the more heavily-adopted when analysing sport science field-based datasets with repeated measures, like the rugby league women's dataset used in this study. Mixed models were able to account for multiple observations of the same individuals in the dataset, players changing positions between matches, and account for inter-player, inter-match, and inter-team dependencies to extract the true effects of position and level of competition on mean speed and the appropriate quantification of the uncertainty of these estimates. One strength of the mixed model in analysing data was its ability to account for the differing number of observations of athletes. In the dataset, athletes ranged between one and nine matches played with 24 athletes playing only one match, while five athletes played nine matches each. This is not ideal when applying repeated-measures ANOVA as either some athletes or some time points (i.e., games) would need to be excluded from the analysis. While removing an athlete if they do not have an observation in every time point or removing a time point if many athletes are missing is statistically appropriate if the missing data points are proved to be missing completely at random (i.e., no bias in missing data)[30], this method can cause unnecessary deletion of large amounts of data. If there are many participants and many time points, it can sometimes eliminate so much data that no analysis can be performed on the remaining dataset; for example, 48.7% of the data collected for the present study would have been omitted when filtering out players that did not participate in Origin or Test matches. However, mixed models can attribute a level of uncertainty to each athlete dependent on their sample size and, therefore, the model can more accurately quantify the true mean speed of a player with nine observations compared to a single observation of another athlete. In doing so, mixed models create flexibility in the datasets that can be utilised that more closely align to datasets seen in sports due to injuries, squad selection, and access to athletes on a given day. The mixed model provided an analysis that could retain all available data points and did not require the elimination of data points to retain a 'complete case analysis' dataset[22], nor did it require data imputation to estimate missing data[30].

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

When considering the findings of this study, the reduced overall mean speed recorded for the Test (80 min), compared with the Origin (60 min) and NRLW (60 min) matches in the present study could be explained by the longer Test-match duration. We have previously reported that the mean speed when travelling >12 km$\cdot$h^{-1} of athletes recorded during international matches declines by ~40% within the first half of the match[8]. We also previously demonstrated that there were no significant differences in the relative distances covered in any of the speed zones, as well as the overall mean speed when comparing the first and second half of NRLW matches[7]. These results contrast with those seen in other codes, with significantly increased total distance and high-speed running in international football compared to domestic football[9]. On closer examination of mean speed in the present study, it was evident that the position contributed to the model with forwards recording the highest intensity and the interchange athletes delivering the lowest intensity. This could be due to a disparity in playing ability between the starting forwards and the interchange replacements. We previously established that the mean speed of interchange athletes significantly reduced as their playing duration increased[7], which could explain why at the international level, when interchange athletes are required to play increased minutes due to the longer format, they cannot sustain the same intensity as the starting forwards.

Conclusion

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

Practical Applications

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

References

1 Ferris D, Gabbett T, McLellan C, Minahan C. Basal markers of inflammation, muscle damage, and performance during five weeks of pre-season training in elite youth Rugby League players. *J Athl. Enhance* 2018; 7(2).

2 Bartlett J, Hatfield M, Parker B, Roberts L, Minahan C, Morton J, Thornton H. DXA-derived estimates of energy balance and its relationship with changes in body composition across a season in team sport athletes. *Eur J Sport Sci* 2020; 20(7):859–867.

3 Halson S. Monitoring training load to understand fatigue in athletes. *Sports Med* 2014; 44(2):139–147.

4 Bellinger P, Ferguson C, Newans T, Minahan C. No influence of prematch subjective wellness ratings on external load during elite Australian Football match play. *Int J Sports Physiol Perform* 2020; 15(6):801–807.

5 Gallo T, Cormack S, Gabbett T, Lorenzen C. Self-reported wellness profiles of professional Australian football players during the competition phase of the season. *J Strength Cond Res* 2017; 31(2):495–502.

6 Govus A, Coutts A, Duffield R, Murray A, Fullagar H. Relationship between pretraining subjective wellness measures, player load, and rating-of-perceived-exertion training load in American college football. *Int J Sports Physiol Perform* 2018; 13(1):95–101.

7 Newans T, Bellinger P, Buxton S, Quinn K, Minahan C. Movement patterns and match statistics in the National Rugby League Women's (NRLW) Premiership. *Front Sports Act Living* 2021; 3.

8 Quinn K, Newans T, Buxton S, Thomson T, Tyler R, Minahan C. Movement patterns of players in the Australian Women's Rugby League team during international competition. *J Sci Med Sport* 2020; 23(3):315–319.
9 Griffin J, Newans T, Horan S, Keogh J, Dodd K, Andreatta M, Minahan C. Acceleration and high-speed running profiles of women's international and domestic football matches. *Front Sports Act Living* 2021; 3:71.
10 Thornton H, Armstrong C, Rigby A, Minahan C, Johnston R, Duthie G. Preparing for an Australian Football League Women's league season. *Front Sports Act Living* 2020; 2:216.
11 Kenny D, Judd C. Consequences of violating the independence assumption in analysis of variance. *Psychol Bull* 1986; 99(3):422–431.
12 Haugnes P, Torvik P-Ø, Ettema G, Kocbach J, & Sandbakk Ø. The effect of maximal speed ability, pacing strategy, and technique on the finish sprint of a sprint cross-country skiing competition. *Int J Sports Physiol Perform* 2019; 14(6):788–795.
13 Rago V, Krustrup P, Martín-Acero R, Rebelo A, Mohr M. Training load and submaximal heart rate testing throughout a competitive period in a top-level male football team. *J Sports Sci* 2020; 38(11–12):1408–1415.
14 Tierney P, Blake C, Delahunt E. Physical characteristics of different professional rugby union competition levels. *J Sci Med Sport* 2021; 24(12):1267–1271.
15 Hannon MP, Coleman NM, Parker LJ, McKeown J, Unnithan VB, Close GL, ... & Morton JP. Seasonal training and match load and micro-cycle periodization in male Premier League academy soccer players. *J Sports Sci* 2021; 39(16):1838–1849.
16 McCaskie CJ, Young WB, Fahrner BB, Sim M. Association between preseason training and performance in elite Australian football. *Int J Sports Physiol Perform* 2019; 14(1):68–75.
17 Dalton-Barron N, Whitehead S, Roe G, Cummins C, Beggs C, Jones B. Time to embrace the complexity when analysing GPS data? A systematic review of contextual factors on match running in rugby league. *J Sports Sci* 2020:1–20.
18 McElreath R. *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman and Hall/CRC, 2018.
19 Hopkins WG, Marshall SW, Batterham AM, et al. Progressive Statistics for Studies in Sports Medicine and Exercise Science. *Med Sci Sports Exerc* 2009; 41(1):3–12.
20 Andersson H, Randers M, Heiner-Møller A, Krustrup P, Mohr M. Elite female soccer players perform more high-intensity running when playing in international games compared with domestic league games. *J Strength Cond Res* 2010; 24(4):912–919.
21 Thornton H, Nelson A, Delaney J, Serpiello F, Duthie G. Interunit reliability and effect of data-processing methods of global positioning systems. *Int J Sports Physiol Perform* 2019; 14(4):432–438.
22 Nakai M, Ke W. Review of the methods for handling missing data in longitudinal data analysis. *Int J Math Anal* 2011:13.
23 Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Softw* 2015; 67(1):1–48.
24 R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, R Foundation for Statistical Computing, 2019.
25 Lenth R. *emmeans: Estimated Marginal Means, aka Least-Squares Means*. 2020.
26 Lüdecke D. *sjPlot: Data Visualization for Statistics in Social Science*. 2020.
27 Lüdecke D, Patil I, Ben-Shachar MS, Wiernik B, Waggoner P, Makowski D. see: An R Package for Visualizing Statistical Models. *J Open Source Softw* 2021; 6(64):3393.
28 Lüdecke D, Ben-Shachar MS, Patil I, Waggoner P, Makowski D. performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *J Open Source Softw* 2021; 6(60):3139. Doi: 10.21105/joss.03139.

29 Yarkoni T. The generalizability crisis. *Behav Brain Sci* 2019:1–37.
30 Borg D, Nguyen R, Tierney N. Missing Data: Current practice in football research and recommendations for improvement. *Sci Med Footb* 2021.

Figure 1. Flowchart of women's rugby league athletes playing at various competition levels indicating the levels of dependency within the dataset. N.B. NRLW = National Rugby League Women's; Origin = State of Origin; Test = Trans-Tasman Test.

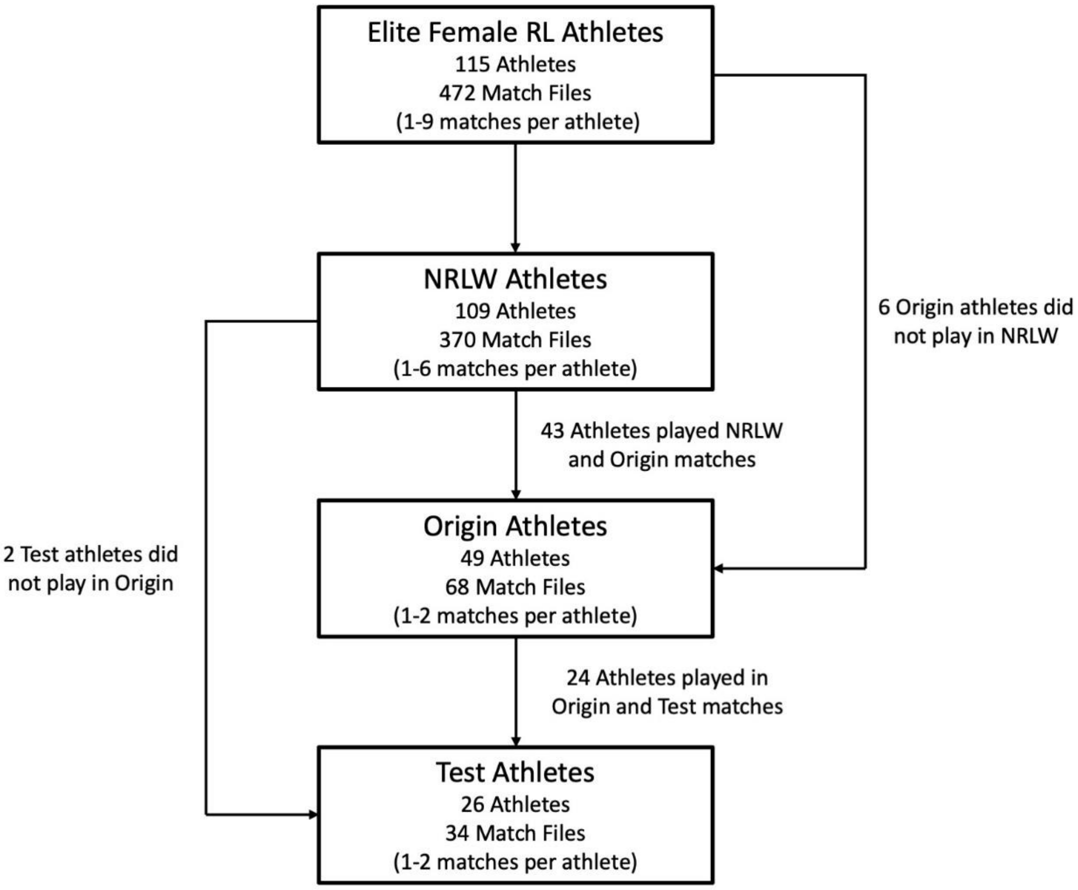

Figure 2. Estimated conditional means for each random effect level of player, team, and match.

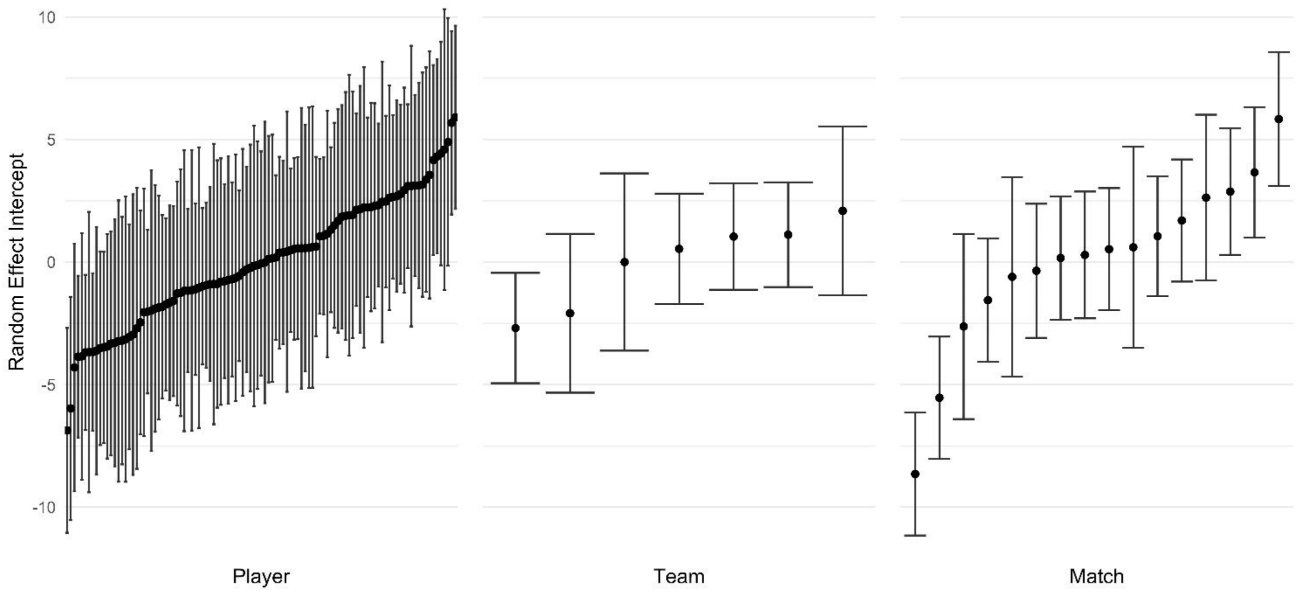

Figure 3. Mean speed of women's rugby league athletes by competition level and position. N.B. NRLW = National Rugby League Women's; Origin = State of Origin; Test = Trans-Tasman Test.