Title:

High-intensity endurance capacity assessment as a tool for talent identification in elite youth female soccer

Preferred Running Head:

Talent identification in elite youth female soccer

Key Words:

Football, talent selection, fitness testing, logistic regression, retrospective analysis

Text Only Word Count:

4226

Abstract Word Count:

200

Number of Figures:

3

Number of Tables:

2

Disclosure Statement:

The authors report no conflict of interest

High-intensity endurance capacity assessment as a tool for talent identification in elite youth female soccer

Abstract

Talent identification and development programmes have received broad attention in the last decades, yet evidence regarding the predictive utility of physical performance in female soccer players is limited. Using a retrospective design, we appraised the predictive value of performance-related measures in a sample of 228 youth female soccer players previously involved in residential Elite Performance Camps (age range: 12.7 to 15.3 years). With 10-m sprinting, 30-m sprinting, counter-movement jump height, and Yo-Yo Intermittent Recovery Test Level 1 (IR1) distance as primary predictor variables, the Akaike Information Criterion (AIC) assessed the relative quality of four penalised logistic regression models for determining future competitive international squads U17-U20 level selection. The model including Yo-Yo IR1 was the best for predicting career outcome. Predicted probabilities of future selection to the international squad increased with higher Yo-Yo IR1 distances, from 4.5% (95% confidence interval, 0.8 to 8.2%) for a distance lower than 440 m to 64.7% (95% confidence interval, 47.3 to 82.1%) for a score of 2040 m. The present study highlights the predictive utility of high-intensity endurance capacity for informing career progression in elite youth female soccer and provides reference values for staff involved in the talent development of elite youth female soccer players.

Introduction

In recent years there has been an increased emphasis on the processes of talent identification and development (Johnston, Wattie, Schorer, & Baker, 2018). Talent identification refers to the recognition of individuals with potential to become elite, whereas talent development involves provision of an optimal environment to realise this potential (Reilly, Bangsbo, & Franks, 2000; A. M. Williams and Reilly, 2000). Effective talent identification and development not only increases the likelihood of team success but also generates high financial rewards via the transfer market (Mann, Dehghansai, & Baker, 2017). National teams do not have the option to purchase players via the transfer market; therefore, talent identification and development may be of greater importance to national governing bodies compared to domestic club teams. The exact composition of talent identification and development programmes will vary depending on the specific requirements of the sport. However, programmes are likely to consist of developing the technical/tactical, physiological, psychological and social skills required for success within a specific sport.

The early identification of individuals who will be successful at senior level is a complex and highly challenging process (Mann, et al., 2017). Traditional talent identification research has often focused on identifying characteristics which distinguish between elite and sub-elite youth performers (Breitbach, Tug, & Simon, 2014). Such a methodology assumes the most talented youth athletes will become the most talented senior athletes, i.e. that talent is static (Johnston, et al., 2018), although youth success is only a weak predictor of success at a senior level (Baker, Schorer, & Wattie, 2017; Kearney and Hayes, 2018). Identifying the characteristics worthy of investigation is complex, but a multi-dimensional approach including physical, psychological and sports-specific factors has been recommended to provide the most holistic methodology (Breitbach, et al., 2014). To further advance our understanding of potential factors contributing to senior success, it seems valuable to prospectively track players, or retrospectively trace long-term career progression (Till et al., 2015).

Currently there is a paucity of such data with a limited number of studies focusing on long-term career progression in a range of male team sports; rugby union (Fontana, Colosio, Da Lozzo, & Pogliaghi, 2017), rugby league (Till, et al., 2015), Australian football (Burgess, Naughton, & Hopkins, 2012) and soccer (Gonaus and Muller, 2012; le Gall, Carling, Williams, & Reilly, 2010). Each of these studies identified that a combination of anthropometric and physical performance characteristics discriminated between those athletes deemed successful and non-successful in their future careers. Collectively, these data suggest that fitness testing in youth male team sport athletes may provide useful information for predicting future career progression (Till, et al., 2015). However, to date, there is a lack of information available on female athletes and specifically female soccer players with no information currently available on predicting future career progression in females. International female match-play data demonstrates the high physical demands of the sport (Datson et al., 2017) and a substantial body of evidence has evaluated the physical capacity of female soccer players (Datson et al., 2014). Previous research has shown differences in physical performance characteristics based upon competitive playing standard (Mujika, Santisteban, Impellizzeri, & Castagna, 2009), player selection (Manson, Brughelli, & Harris, 2014) and age (Wright and Atkinson, 2017). However, the relative importance and influence that these characteristics have on future career progression has not been identified.

Therefore, using a retrospective design, our study aimed to ascertain the predictive value of relevant physical performance measures for determining future career progression in youth elite female soccer players.

Methods

Experimental Approach to the Problem

Anthropometric and field-based physical performance testing data were collected from youth elite female soccer players between 2011-2014, with testing sessions conducted as part of the Elite Performance Camps (EPC) programme for talented youth players. The English Football Association support a Girls' England Talent Pathway that aims to identify and develop youth players with potential. As part of the pathway, *talented* players aged 12-15 years attend residential EPCs for specialised training.

Data were retrospectively analysed and for the purposes of this study players were divided into two career progression levels for comparison: (1) selected for competitive international squads at U17-U20 level or (2) not selected for competitive international squads at U17-U20 level.

Prior to assessment, all players had previously completed each test on at least one previous occasion, which acted as their familiarisation. Physical performance tests were performed indoors and players wore shorts, t-shirt and football boots (except for the jumps when trainers were worn). Players performed a standardised generic warm-up prior to commencement of the physical assessments as well as specific warm-up routines prior to each performance test. To ensure consistency between testing occasions, National federation staff coached the warm-up activity and conducted all measurements.

All physical performance tests were completed at approximately the same time of day to reduce any circadian rhythm effect (Reilly and Brooks, 1986). Tests were completed in a single session and in the same order (anthropometry, jumps, linear speed and Yo-Yo Intermittent Recovery Test Level 1 [Yo-Yo IR1]) on each test occasion. The test order was designed in an attempt to minimise the influence of previous tests on subsequent performance. Players refrained from strenuous exercise in the 24 hours before fitness testing session and consumed their normal pre-training diet. To encourage maximal effort, players received consistent verbal encouragement throughout the physical performance tests. Usual appropriate ethics committee clearance was not required as data was collected as a condition of employment (Winter and Maughan, 2009) and all players had previously consented for their data to be used for research purposes. Nevertheless, all data were anonymized prior to analysis to ensure player confidentiality.

Participants

Data were collected from 284 youth elite female soccer players (612 separate observations; with a median of two testing occasions per player (range = 1-6). However, for analysis purposes, a complete dataset was required per player and therefore the effective sample size was reduced to 228 (13.9 ± 0.6 years). Where players were tested on multiple occasions, the *best* score for each performance test was included in the analysis.

All participants were part of the England Football Association's Talent pathway and as such they participated in a minimum of two football sessions per week and one match. In addition, players would complete up to two strength and conditioning practices per week and have access to specialist support.

Procedures

Anthropometric and Physical Performance Measures

Player height (m), sitting height and body mass (kg) were measured using a stadiometer (Seca 217, Germany) and calibrated digital scales (Seca 876, Germany), respectively. Skinfolds (mm) were taken as an estimate of adiposity and measured at eight sites: biceps, triceps, subscapular, iliac crest, supraspinale, abdominals, front thigh and medial calf using skinfold calipers (Harpenden, UK). An International Society for the Advancement of Kinanthropometry (ISAK) accredited anthropometrist performed all measurements, with ISAK guidelines followed (Jones et al., 2006). Height, sitting height and body mass were used to calculate maturity offset for each player on each testing occasion using the Mirwald (2002) equation.

Estimations of player's lower limb muscular power were assessed via a countermovement jump (CMJ) on a jump mat (KMS Innervations, Australia). The jump mat was placed on a firm, concrete surface at the edge of the indoor third-generation turf pitch. Following generic and jump-specific warm-up activity, players were permitted an additional practice jump on the mat before performing three recorded trials. Players were instructed to step on to the mat and place their feet in the middle of the mat (a comfortable distance apart) and with their hands on their hips. Starting from an upright position, players were instructed to jump as high as possible while keeping their hands on their hips and legs straight when in the air and refraining from bringing their legs into a pike position or flicking their heels. The highest jump height recorded to the nearest 0.1 cm was retained for analysis. Linear speed times were measured using electronic timing gates (Brower TC Timing System, USA) over distances of 0-30 m. A 50 m steel tape measure (Stanley, UK) was used to measure the 30 m distance and markers were placed at 0, 10 m and 30 m; in addition, a marker was placed 1 m behind the zero line. Tripods were placed directly over each marker at a height of 87 cm above ground level and a timing gate (transmitter) was fitted to each tripod. Opposite each tripod, at a distance of 2 m, another tripod and timing gate (receiver) was positioned. Following a speed-specific warm-up activity, players were permitted an additional practice sprint through the course before performing three recorded trials. Players commenced each sprint with their preferred foot on a line 1 m behind the first timing gate. Each sprint was separated by a 3-min recovery period. The fastest time at each distance to the nearest 0.001 s was retained for analysis   Player's high-intensity endurance capacity was assessed via Yo-Yo IR1 (Krustrup et al., 2003). The reliability of each of the anthropometric and physical performance measures have previously been established in a similar sample to the present study (Datson, 2016).

Statistical Analysis

Data are presented as mean ± standard deviation (SD) for continuous variables, and frequency or percentages for categorical variables. To derive consistent estimates for the predicted probabilities of future selection (Grant, 2014), four penalized logistic regression models included 10-m sprinting (s), 30-m sprinting (s), counter-movement jump height (cm), and Yo-Yo IR1 distance (m) as distinct primary predictor variables controlling for differences in maturity offset and adjusting for chronological age and anthropometric characteristics (Coveney, 2008; Firth, 1993). To provide reference value that might inform staff members involved in talent identification and development processes, predicted probabilities were derived for the 1^{st}, 2.5^{th}, 25^{th}, $50^{th,}$ $75^{th,}$ 97.5^{th} and 99^{th} percentiles of each performance measure (Williams, 2012). To examine accuracy of the estimated models, we appraised the internal calibration of derived probabilities using a novel method based on a calibration belt approach (Nattino, Lemeshow, Phillips, Finazzi, & Bertolini, 2017). By definition, internal calibration refers to the degree of agreement between the estimated probabilities and observed outcome rates in the sample in which the model was developed (Austin and Steyerberg, 2014). As an alternative to commonly used tests and graphic tools (Steyerberg et al., 2010), the confidence band around the curve (i.e., the calibration belt) is a measure of uncertainty in the estimate of the curve and enables a formal internal calibration appraisal (Nattino, Finazzi, & Bertolini, 2014b). A model correctly predicts the frequency of events if the calibration belt contains the bisector of the axes (Nattino, et al., 2017).

The Akaike Information Criterion (AIC) assessed relative quality of each model in the set of candidate models. The Akaike difference (ΔAIC) from the estimated best model (i.e., the model with the lowest AIC value; ΔAIC = 0) was evaluated according to the following scale: 0-2, essentially equivalent; 2-7, plausible alternative; 7-14, weak support; > 14, no empirical support (Burnham, Anderson, & Huyvaert, 2011). Predicted probabilities are presented as point estimates with the related disposition (95% confidence interval, CI) and model internal validation was illustrated for the best/essentially equivalent models. Analyses were performed using R (version 3.6.0, R Foundation for Statistical Computing, Vienna, Austria) and Stata (StataMP v14.0; StataCorp LP, College Station, TX).

Results

From the original sample size, 228 players with valid performance and maturity measures at the time of assessment over the examined observation period were eligible. Of these players, 50 players were selected for future competitive international squads at U17-U20 level and 178 players not selected. The range for chronological age, body weight, height, and sum-of-skinfolds was 12.7 to 15.3 years, 33.4 to 85.6 kg, 141.5 to 188 cm, 39.5 to 166.9 mm, respectively.

Summary characteristics for each of the examined variables are illustrated graphically in dot-and-violin plots, with the bulk of data values describing the centre of the distribution (Figure 1). For the selected players, the mean 10-m sprinting, 30-m sprinting, CMJ height, Yo-Yo IR1 distance was 1.805 (± 0.121), 4.623 (± 0.197), 29.79 (± 3.45), and 1393 (± 365), respectively. For the unselected players, the mean 10-m sprinting, 30-m sprinting, CMJ height, Yo-Yo IR1 distance was 1.841 (± 0.103), 4.724 (± 0.232), 28.64 (3.81), and 1077 (353), respectively. The point estimate and likely range of compatible values for the mean difference in the measure of interest between selected versus unselected players in the international squad at U17-U20 level was –0.036 s (95%CI, –0.070 s to –0.002 s) for 10-m sprinting, –0.101 s (–0.172 s to –0.030 s) for 30-m sprinting, 0.44 cm (–0.67 cm to 1.55 cm) for CMJ height, 189 m (93 m to 285 m) for Yo-Yo IR1 distance.

Figure 1 near here

Table 1 near here

Comparison of separate logistic regression models with penalized maximum likelihood on information theory grounds revealed that the model including Yo-Yo IR1 distance as primary predictor was the best of the four candidates for determining probabilities of international squad selection in later career stages (Table 1). Additionally, sensitivity analyses revealed a trivial main effect for biological maturity offest (P = 0.664) and for Yo-Yo IR1 distance × biological maturity offest interaction term (P = 0.673) in the model. The probabilities for a player of future international squads U17-U20 level selection increased with higher Yo-Yo IR1 distances, from 4.5% (95% confidence interval, 0.8 to 8.2%) for a distance lower than 440 m to 64.7% (95% confidence interval, 47.3 to 82.1%) for a score of 2040 m (Figure 2). Table 2 illustrates the probabilities of future selection by Yo-Yo IR1 distance percentile. With the dataset randomly split into developmental and validation subsets of 166 and 66 players, the 95% calibration belt encompassed the bisector over the whole range of the predicted probabilities suggesting acceptable model internal calibration (Figure 3). The penalized logistic regression models including other performance-related variables were empirically unsupported (Table 1).

Table 2 near here

Figure 2 near here

Figure 3 near here

Discussion

For the first time, we ascertained the predictive value of physical performance measures to determine future career progression in a sample of elite youth female soccer players. Our results show players with higher Yo-Yo IR1 scores are more likely to be selected for the competitive international squad at U17-U20 level independent of playing position. The present study highlights the predictive value of high-intensity endurance capacity for informing career progression in elite youth female soccer.

The present data showed that 22% of EPC players progressed into competitive international squads at U17-U20 level. This low to moderate success rate is similar to that observed in male soccer (Gonaus and Muller, 2012) and across multiple Olympic sports (Vaeyens, Gullich, Warr, & Philippaerts, 2009). Such a relatively low conversion rate reflects that female soccer in England adopts a pyramid model for talent development and therefore, due to squad sizes, it would never be possible for all players to progress from EPCs to competitive international squads. Indeed, 86% of England's bronze medal winning U20 2018 World Cup squad were part of the EPC programme and were analysed in this dataset as ~14 year old players. Nevertheless, further analysis of players selected into competitive international squads at U17-U20 level whom did not progress through the EPC would be worthy of future research to highlight alternative development pathways (Till et al., 2015).

From a real-world perspective, our findings are not surprising as previous studies revealed difference in Yo-Yo IR1 score to distinguish between competitive level in females, with elite players out-performing their sub-elite counterparts (Mujika, et al., 2009). Furthermore, high-intensity endurance capacity represents an important aspect of soccer performance in elite female players, with an increased Yo-Yo IR1 test score largely associated with a higher match running performance (Krustrup, Mohr, Ellingsgaard, & Bangsbo, 2005). Translated into a soccer-specific context, a greater

Yo-Yo IR1 performance may allow players to out-run their opponents and if coupled with a sufficient tactical understanding, may allow a player to have a greater influence on the match (Young and Pryor, 2007).

The fact that the penalized logistic regression model with Yo-Yo IR1 as primary predictor variable emerged as the best in the candidate pool highlighted the greater relative importance of this aspect relevant to female soccer performance than sprinting and jumping qualities. The limited predictive value of linear sprint performance to determine future youth international career outcome supports previous research which observed no differences in 15-m sprint performance between elite and sub-elite female soccer players (Mujika, et al., 2009). Nonetheless, our results are in contrast with previous research in male youth soccer players where superior jumping and sprinting performance characteristics were observed in successful versus unsuccessful career progression in Austrian and French players (Gonaus and Muller, 2012; le Gall, et al., 2010). An explanation for these gender differences might be related to the greater talent pool in male players, thus potentially placing increased emphasis on physiological and performance measures to help discriminate between talented male players. Indeed, in the study by Gonaus and Muller (2012) there were a similar number of players per year compared to the present study. However, the players were attending one of the twelve National youth academies in Austria and hence the total number of players in the National programme was likely to be ~12 times greater than the female EPC programme evaluated in the current study. However, we also point out that the study by Gonaus and Muller (2012) adopted modelling approaches different from our logistic regression analyses, which, therefore, may limit the extent of any comparison with our study outcomes. Furthermore, it should also be considered that talent development programmes generally start at a younger age for males compared to females with structured academy programmes starting for boys from the age of 9 years (Goto, Morris and Nevill, 2015).

A further novel aspect of our study was that we provide reference values to help inform and guide decisions of staff members involved in a talent identification process in youth female soccer. For example, to illustrate the practical value of our data, consider a new youth female soccer player aged 13.5 years, who has been selected for an elite camp, and registered a Yo-Yo IR1 score of 1890 m. According to our data, this value would occur in fewer than 3 players in 100 and indicates that, at approximately the 97.5th percentile, this new player has an average predicted probability of future international career ranging from 40.4% to 71.5% (Table 2). From a real-world perspective, given the multifactorial nature of soccer performance (Impellizzeri and Marcora, 2009), our study results suggest that high-intensity endurance capacity assessment can serve as a valuable complementary tool for talent identification in youth female soccer players.

In general, an underlying purpose of gathering physical performance data is to provide coaches and practitioners with information which may guide talent identification and development programmes. Within this particular context and facing similar challenges to those of the clinician with diagnosis and prognosis (Steyerberg, et al., 2010), a coach may be interested in to know how this may translate to meaningful real-world impact either in the short (identification) or long (development) term. Therefore, a critical appraisal of decision-analytic measures as indices of model internal calibration is fundamental to ascertain the validity and accuracy of the estimated probabilities

(Steyerberg and Harrell, 2016). Unlike the current practices for alternative regression modelling strategies illustrated in the sports science literature (Carey, Ong, Morris, Crow, & Crossley, 2016; Jaspers et al., 2018; Woods, Raynor, Bruce, McDonald, & Robertson, 2016; Woods, Veale, Fransen, Robertson, & Collier, 2018), we adopted a novel approach which outperforms the commonly used yet stringent graphical approaches for model internal calibration of the predicted probabilities emerging from our model (Nattino, Finazzi, & Bertolini, 2014a). If a predictive model is not carefully evaluated for nor fails to show acceptable internal calibration, any probability prediction lacks empirical support and real-world practical value for coaches and staff members involved in talent identification and development processes (Austin and Steyerberg, 2014; Nattino, et al., 2017; Steyerberg and Harrell, 2016; Steyerberg, et al., 2010).

Nonetheless, our study is not without limitations. The predicted probabilities of future youth international career outcome were estimated using one-time-only (best score) retrospective performance data gathered in the previous years. Arguably, future research based on repeated high-intensity endurance data could potentially advance further the understanding of what longitudinal increment in Yo-Yo IR1 should be targeted to increase the probability of future international career in female soccer. Such an investigation may be possible only following a model external validation, with any potential study involving an adequate sample of player and a consistent number of multiple assessments over subsequent years. However, we maintain that, due to the nature of this and other talent development programmes, it is unlikely to be possible to include repeated measures over a number of years since players are regularly deselected from the development programme. Furthermore, given the multifactorial nature of soccer performance (Impellizzeri and Marcora, 2009), measures of technical ability were not examined in this study (Impellizzeri et al., 2008).

Conclusions

Our findings substantiate novel evidence regarding the utility of physical performance variables to determine future international career in elite youth female soccer players. This study highlights the value of high-intensity endurance capacity as an important aspect relevant to elite youth female soccer performance and illustrates predicted probabilities for Yo-Yo IR1 centiles that can inform talent identification and development processes.

References

Austin, P. C., & Steyerberg, E. W. (2014). Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. *Statistics in Medicine, 33*(3), pp. 517-535. doi:10.1002/sim.5941

Baker, J., Schorer, J., & Wattie, N. (2017). Compromising Talent: Issues in Identifying and Selecting Talent in Sport. *Quest, 70*(1), pp. 48-63. doi:10.1080/24733938.2018.1427883

Breitbach, S., Tug, S., & Simon, P. (2014). Conventional and genetic talent identification in sports: will recent developments trace talent? *Sports Medicine, 44*(11), pp. 1489-1503. doi:10.1007/s40279-014-0221-7

Burgess, D., Naughton, G., & Hopkins, W. (2012). Draft-camp predictors of subsequent career success in the Australian Football League. *J Sci Med Sport, 15*(6), pp. 561-567. doi:10.1016/j.jsams.2012.01.006

Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behavioral Ecology and Sociobiology, 65*(1), pp. 23-35. doi:10.1007/s00265-010-1029-6 Retrieved from <Go to ISI>://WOS:000285786000003

Carey, D. L., Ong, K. L., Morris, M. E., Crow, J., & Crossley, K. M. (2016). Predicting ratings of perceived exertion in Australian football players: methods for live estimation. *Int J Comput Sci Sport, 15*(2), pp. 64-77. doi:10.1515/ijcss-2016-0005

Coveney, J. (2008). firthlogit: Stata module to calculate bias reduction in logistic regression. Statistical Software Components S456948, Department of Economics, Boston College. http://econpapers.repec.org/software/bocbocode/s456948.htm. Accessed June 23, 2019.

Datson, N., Drust, B., Weston, M., Jarman, I. H., Lisboa, P. J., & Gregson, W. (2017). Match physical performance of elite female soccer players during international competition. *J Strength Cond Res, 31*(9), pp. 2379-2387. doi:10.1519/JSC.0000000000001575

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika, 80*(1), pp. 27-38. doi:10.2307/2336755

Fontana, F. Y., Colosio, A. L., Da Lozzo, G., & Pogliaghi, S. (2017). Player's success prediction in rugby union: From youth performance to senior level placing. *Journal of Science and Medicine in Sport*, *20*(4), pp. 409-414. doi:10.1016/j.jsams.2016.08.017

Gonaus, C., & Muller, E. (2012). Using physiological data to predict future career progression in 14- to 17-year-old Austrian soccer academy players. *J Sports Sci, 30*(15), pp. 1673-1682. doi:10.1080/02640414.2012.713980

Goto, H., Morris, J. and Nevill, M.E. (2015). Match analysis of U9 and U10 Englist Premier League academy soccer players using a global positioning system: relevance for talent identification and development. Journal of Strength and Conditioning Research, 29(4), pp. 954-963. doi: 10.1519/JSC.0b013e3182a0d751

Grant, R. L. (2014). Converting an odds ratio to a range of plausible relative risks for better communication of research findings. *BMJ, 348*, p f7450. doi:10.1136/bmj.f7450

Impellizzeri, F. M., & Marcora, S. M. (2009). Test validation in sport physiology: lessons learned from clinimetrics. *International Journal of Sports Physiology and Performance, 4*(2), pp. 269-277.

Impellizzeri, F. M., Rampinini, E., Maffiuletti, N. A., Castagna, C., Bizzini, M., & Wisloff, U. (2008). Effects of aerobic training on the exercise-induced decline in short-passing ability in junior soccer players. *Appl Physiol Nutr Metab, 33*(6), pp. 1192-1198. doi:10.1139/H08-111

Jaspers, A., De Beeck, T. O., Brink, M. S., Frencken, W. G. P., Staes, F., Davis, J. J., & Helsen, W. F. (2018). Relationships Between the External and Internal Training Load in Professional Soccer: What Can We Learn From Machine Learning? *Int J Sports Physiol Perform, 13*(5), pp. 625-630. doi:10.1123/ijspp.2017-0299

Johnston, K., Wattie, N., Schorer, J., & Baker, J. (2018). Talent Identification in Sport: A Systematic Review. *Sports Medicine, 48*(1), pp. 97-109. doi:10.1007/s40279-017-0803-2

Kearney, P. E., & Hayes, P. R. (2018). Excelling at youth level in competitive track and field athletics is not a prerequisite for later success. *J Sports Sci, 36*(21), pp. 2502-2509. doi:10.1080/02640414.2018.1465724

Krustrup, P., Mohr, M., Ellingsgaard, H., & Bangsbo, J. (2005). Physical demands during an elite female soccer game: importance of training status. *Medicine and Science in Sports and Exercise, 37*(7), pp. 1242-1248.

le Gall, F., Carling, C., Williams, M., & Reilly, T. (2010). Anthropometric and fitness characteristics of international, professional and amateur male graduate soccer players from an elite youth academy. *J Sci Med Sport, 13*(1), pp. 90-95. doi:10.1016/j.jsams.2008.07.004

Mann, D. L., Dehghansai, N., & Baker, J. (2017). Searching for the elusive gift: advances in talent identification in sport. *Curr Opin Psychol, 16*, pp. 128-133. doi:10.1016/j.copsyc.2017.04.016

Manson, S. A., Brughelli, M., & Harris, N. K. (2014). Physiological characteristics of international female soccer players. *J Strength Cond Res, 28*(2), pp. 308-318. doi:10.1519/JSC.0b013e31829b56b1

Mujika, I., Santisteban, J., Impellizzeri, F. M., & Castagna, C. (2009). Fitness determinants of success in men's and women's football. *J Sports Sci, 27*(2), pp. 107-114. doi:10.1080/02640410802428071

Nattino, G., Finazzi, S., & Bertolini, G. (2014a). Comments on 'Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers' by Peter C. Austin and Ewout W. Steyerberg. *Statistics in Medicine, 33*(15), pp. 2696-2698. doi:10.1002/sim.6126

Nattino, G., Finazzi, S., & Bertolini, G. (2014b). A new calibration test and a reappraisal of the calibration belt for the assessment of prediction models based on dichotomous outcomes. *Statistics in Medicine*, *33*(14), pp. 2390-2407. doi:10.1002/sim.6100

Nattino, G., Lemeshow, S., Phillips, G. S., Finazzi, S., & Bertolini, G. (2017). Assessing the calibration of dichotomous outcome models with the calibration belt. *Stata Journal, 17*(4), pp. 1003-1014. doi:10.1177/1536867X1801700414

Reilly, T., Bangsbo, J., & Franks, A. (2000). Anthropometric and physiological predispositions for elite soccer. *J Sports Sci, 18*(9), pp. 669-683. doi:10.1080/02640410050120050

Reilly, T., & Brooks, G. A. (1986). Exercise and the circadian variation in body temperature measures. *Int J Sports Med, 7*(6), pp. 358-362. doi:10.1055/s-2008-1025792

Steyerberg, E. W., & Harrell, F. E., Jr. (2016). Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol, 69*, pp. 245-247. doi:10.1016/j.jclinepi.2015.04.005

Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., . . . Kattan, M. W. (2010). Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology, 21*(1), pp. 128-138. doi:10.1097/EDE.0b013e3181c30fb2

Till, K., Cobley, S., O'Hara, J., Morley, D., Chapman, C., & Cooke, C. (2015). Retrospective analysis of anthropometric and fitness characteristics associated with long-term career progression in Rugby League. *J Sci Med Sport, 18*(3), pp. 310-314. doi:10.1016/j.jsams.2014.05.003

Vaeyens, R., Gullich, A., Warr, C. R., & Philippaerts, R. (2009). Talent identification and promotion programmes of Olympic athletes. *J Sports Sci, 27*(13), pp. 1367-1380. doi:10.1080/02640410903110974

Williams, A. M., & Reilly, T. (2000). Talent identification and development in soccer. *J Sports Sci, 18*(9), pp. 657-667. doi:10.1080/02640410050120041

Williams, R. (2012). Using the margins command to estimate and interpret adjusted predictions and marginal effects. *Stata Journal, 17*(1), pp. 308–331. doi:10.1177/1536867X1201200209

Winter, E. M., & Maughan, R. J. (2009). Requirements for ethics approvals. *J Sports Sci, 27*(10), p 985. doi:10.1080/02640410903178344

Woods, C. T., Raynor, A. J., Bruce, L., McDonald, Z., & Robertson, S. (2016). The application of a multi-dimensional assessment approach to talent identification in Australian football. *J Sports Sci, 34*(14), pp. 1340-1345. doi:10.1080/02640414.2016.1142668

Woods, C. T., Veale, J., Fransen, J., Robertson, S., & Collier, N. F. (2018). Classification of playing position in elite junior Australian football using technical skill indicators. *J Sports Sci, 36*(1), pp. 97-103. doi:10.1080/02640414.2017.1282621

Wright, M. D., & Atkinson, G. (2017). Changes in sprint-related outcomes during a period of systematic training in a girls' soccer academy. *J Strength Cond Res*doi:10.1519/JSC.0000000000002055

Young, W. B., & Pryor, L. (2007). Relationship between pre-season anthropometric and fitness measures and indicators of playing performance in elite junior Australian Rules football. *J Sci Med Sport, 10*(2), pp. 110-118. doi:10.1016/j.jsams.2006.06.003

Figure legends

Figure 1. Dot-and-violin plots for physical performance variables in elite youth female soccer players. Green represents individuals selected for competitive international squads at U17-U20 level and red represents individuals not selected for competitive international squads at U17-U20 level.

Figure 2. Predicted probabilities of selection by Yo-Yo IR1 distance.

Figure 3. Calibration belt (95% confidence level) plot and calibration statistic for the relationship between the model's fit probabilities and the observed proportions of the response.

Table legends

Table 1. Relative quality of the four candidate models

Table 2. Predicted probability of selection by Yo-Yo IR1 distance centile