ORIGINAL ARTICLE

Embedded readiness indices and match outcome in youth football from U13 to U21: A reproducible re-evaluation and what the monitoring data can tell us

Rodrigo Villaseca-Vicuña . *School of Educational Sciences and Technology. Physical Education Pedagogy. Faculty of Education. Silva Henríquez Catholic University. Santiago, Chile.*

Pablo Galaz. *Department of Industrial Engineering. University of Chile. Santiago, Chile.*

Benjamín Villaseca-Vicuña. *Department of Industrial Engineering. University of Chile. Santiago, Chile.*

Pablo Merino-Muñoz. *Department of Physical Activity, Sports and Health Sciences. Faculty of Medical Sciences. University of Santiago de Chile (USACH). Santiago, Chile.*
Research Centre in Human Movement Sciences. Adventist University of Chile. Chile.

Jorge Pérez-Contreras. *School of Sports Science. Faculty of Health. Santo Tomás University. Chile.*
Department of Physical Education, Sport and Recreation. Faculty of Arts and Physical Education. Metropolitan University of Educational Sciences. Chile.

David Faúndez. *Department of Sport Science. Youth Football Academy. University of Chile Football Club. Santiago, Chile.*

Rodrigo Yáñez-Sepúlveda. *Faculty of Education and Humanities. School of Sport Sciences. Andrés Bello University. Viña del Mar, Chile.*
School of Medicine. Espíritu Santo University. Samborondón, Ecuador.

Santiago Zabaloy. *Faculty of Physical Activity and Sports. University of Flores. Buenos Aires, Argentina.*
Nucleus of High Performance in Sport (NAR). São Paulo, Brazil.
Graduate Program in Human Movement Sciences and Rehabilitation. Department of Human Movement Sciences. Federal University of São Paulo (UNIFESP). São Paulo, Brazil.

ABSTRACT

This study examined whether embedded readiness indices and match-day running were associated with match outcome across six youth football categories (U13–U21) in a professional Chilean academy. A retrospective longitudinal design included 873 eligible player–match records from 146 players across 140 matches; 141 official matches were recorded, but one contributed no eligible player–match records. The Fitness Index was defined as expected minus observed rating of perceived exertion on MD−3, and the Freshness Index as expected minus observed PlayerLoad on MD−1, with expected values estimated from GPS-derived external load using Random Forest models. Missing index values were multiply imputed. Logistic regression models with match-clustered standard errors and 95% confidence intervals (CI) examined associations with win versus non-win outcomes. None of the 49 model terms had a 95% CI excluding 1, with no evidence of differences between age categories. Adding readiness indices did not improve cross-validated predictive performance in the reported retrospective comparison. External load explained none of the out-of-player variance in MD−3 RPE (R^2 = −.04) but substantially explained MD−1 PlayerLoad (R^2 = .65). Moreover, only 23% of weekly countermovement-jump changes exceeded a provisional detectable-change threshold derived from within-test trial differences. Overall, there was no evidence that the embedded readiness indices were associated with match outcome, while their underlying monitoring indicators showed important limitations as weekly readiness signals.

Keywords: performance analysis, youth football, training load, readiness, rating of perceived exertion, PlayerLoad, countermovement jump, match outcome, reproducibility.

Cite this article as:
Villaseca-Vicuña, R., Galaz, P., Villaseca-Vicuña, B., Merino-Muñoz, P., Pérez-Contreras, J., Faúndez, D., Yáñez-Sepúlveda, R., & Zabaloy, S. (2027). Embedded readiness indices and match outcome in youth football from U13 to U21: A reproducible re-evaluation and what the monitoring data can tell us. *Journal of Human Sport and Exercise, 22*(1), 193-231. https://doi.org/10.55860/g4knrp32

Corresponding author. *School of Educational Sciences and Technology, Physical Education Pedagogy, Faculty of Education, Universidad Católica Silva Henríquez, Santiago 8280354, Chile..*
E-mail: rvillasecav@gmail.com

Submitted for publication September 01, 2026.
Accepted for publication November 02, 2026.
Published November 24, 2026.
Journal of Human Sport and Exercise. ISSN 1988-5202.

©Asociación Española de Análisis del Rendimiento Deportivo. Alicante. Spain.

doi: https://doi.org/10.55860/g4knrp32

INTRODUCTION

Physical performance in football has traditionally been monitored via GPS-derived external-load metrics, including total distance covered and distances accumulated at different speed thresholds, ranging from high-speed running to sprinting, as well as sprint counts (Villaseca-Vicuña, Barrera-Díaz, et al., 2026). Although operationally convenient, these variables have increasingly been questioned as key performance indicators (KPIs), as they are strongly influenced by tactical decisions, opponent behaviour, and match context rather than by a player's physical preparedness or the determinants of competitive success (Carling, 2013; Mendez-Villanueva et al., 2013; Oliva-Lozano et al., 2020; Oliva-Lozano et al., 2023). Recent applied work has explicitly argued against the use of running metrics as KPIs, emphasising their limited explanatory power with respect to match outcomes and football-specific actions (Mandorino et al., 2025).

In response to these limitations, applied sport science has progressively shifted towards a more holistic, readiness-oriented perspective, which contrasts with traditional, generalised approaches based on isolated indicators that fail to capture the multifactorial nature of football performance (Villaseca-Vicuña, Barrera-Díaz, et al., 2026). Readiness is broadly defined as the athlete's capacity to express high-quality football actions in a given competitive context (Mandorino et al., 2025). Within the "Sport Science 3.0" paradigm, this approach prioritises integrated and individualised monitoring frameworks, combining subjective (e.g., perceived wellness, fatigue, soreness, or mood) with objective metrics (e.g., neuromuscular performance indicators, accumulated external load, or acute performance fluctuations) to better capture the athlete's readiness (Gathercole et al., 2015; McLean et al., 2010; Settembre et al., 2024). More specifically, in youth football, such an integrated framework is particularly relevant, as training responses are highly heterogeneous and non-linear, and strongly influenced by growth- and maturation-related factors, as well as week-to-week variability in training and competition exposure (Buchheit & Mendez-Villanueva, 2013).

A methodological advancement aligned with this perspective is the embedded monitoring framework, which models expected physiological or mechanical responses based on external training-load metrics (e.g., total distance, high-speed running, accelerations, and decelerations) and interprets deviations as potential readiness signals, typically conceptualised through fitness- and freshness-related status (Mandorino et al., 2025). In its source, the framework was tested in 25 elite adult players of one club over 73 matches, with a Fitness Index derived from heart rate—an index previously validated against a submaximal running test (r = .70; Mandorino et al., 2024)—and a Freshness Index derived from PlayerLoad (Mandorino et al., 2025). These deviations are statistical residuals: their meaning depends on how well the expected response can be predicted and on how reliably the underlying indicators are measured, and their validity as readiness markers has rarely been examined in youth academies. In this sense, this framework explicitly incorporates match outcome as a contextual variable, enabling the examination of whether specific readiness–load configurations translate into observable competitive effects. This perspective is especially relevant within youth football development programmes, where immediate competitive performance must be balanced with long-term learning and athletic development goals, and where match-related demands may meaningfully shape players' long-term developmental trajectories (Buchheit et al., 2010; Ju et al., 2023; Settembre et al., 2024).

Despite the expanding scientific literature in football, longitudinal investigations in youth development settings remain scarce, as research efforts continue to be predominantly centred on professional footballers (Mandorino et al., 2025). Most existing research has examined isolated physical assessments or cross-sectional comparisons between age groups, with limited attention to the weekly evolution of readiness, internal load, and external load across competitive microcycles (Mendez-Villanueva et al., 2013).

 © 2027 ARD Asociación Española

Furthermore, multivariable approaches linking integrated readiness indicators to match outcome are scarce in academy settings. Although previous research, including Villaseca-Vicuña, Merino-Muñoz, et al. (2026), has investigated physical performance, contextual influences, and load-related responses in youth soccer, these studies were limited by cross-sectional methodologies, isolated domain analyses, and the lack of systematic week-by-week load monitoring. Similarly, the embedded monitoring framework proposed by Mandorino et al. (2025), despite integrating weekly load dynamics, was developed within professional populations, restricting its ecological transferability to youth developmental settings. Moreover, evidence from South American professional clubs—particularly within structured academy systems—remains markedly underrepresented in the literature, reinforcing the relevance of the present investigation.

An important gap in the literature is the limited investigation of readiness as an integrated, multidimensional construct across competitive microcycles and developmental stages in youth football, particularly with respect to its association with match outcomes. Therefore, the aim of the present study was to examine longitudinal associations between training load, embedded readiness indicators (fitness and freshness), and match outcome (win vs. non-win) across six youth development categories (U13–U21), analysed both collectively and according to developmental stage, within a professional Chilean football club. It was hypothesised that variations in training load and embedded readiness states across the competitive microcycle would be significantly associated with match outcome, with category-specific patterns reflecting maturational and developmental differences across the U13–U21 age range.

This revised version answers this question (Question 1) with an analysis that implements the methods as described; the original analysis contained coding errors whose effect is documented in Supplementary Table S1. Because an association between an index and match outcome can only be interpreted if the index measures what it is intended to measure, a second question was added (Question 2): what do the indicators that feed the indices—the RPE of a session, PlayerLoad and the countermovement jump—capture, and how much of their variation can be explained by external load?

METHODS

Study design

This study adopted a retrospective, observational and quantitative design, conducted over the 2024–2025 competitive season. The retrospective longitudinal approach enabled the examination of week-to-week variations in training load, readiness indicators, and competitive outcomes using historical monitoring data routinely collected within an elite football academy.

All data were obtained as part of the club's standard performance monitoring system for youth football players from an elite academy, with no interference in the technical–tactical planning or training structure implemented by the coaching staff. Accordingly, no experimental manipulation of training load or competitive exposure was performed, in line with best-practice recommendations for applied sport-science research in elite and developmental football contexts (Winter & Maughan, 2009).

The present analysis replaces that of the original submission, which did not implement several steps as described in the Methods (Supplementary Table S1). The revised analysis plan specifies the analysis described in the original Methods; it was written after the coding errors had been identified and after preliminary re-analyses of the same data with a different design, and it was recorded in a version-controlled repository before the outcome models reported here were fitted. Analyses added after the first results were seen are labelled post hoc throughout. The repository reproduces both the submitted and the revised analyses.

Participants

The sample consisted of male youth footballers from six developmental categories within a professional Chilean football club: under-13 (U13), under-14 (U14), under-15 (U15), under-16 (U16), under-18 (U18), and under-21 (U21). According to the club's physical-performance staff, players were monitored if they belonged to the competitive squad of their category and participated in at least 70% of training sessions during the study period. Contrary to what was stated in the original submission, complete and uninterrupted datasets were not available for all players and microcycles: GPS data were not recorded for every player in every training session, so the coverage of each indicator is reported (Figure 1) and missing values were handled by multiple imputation.

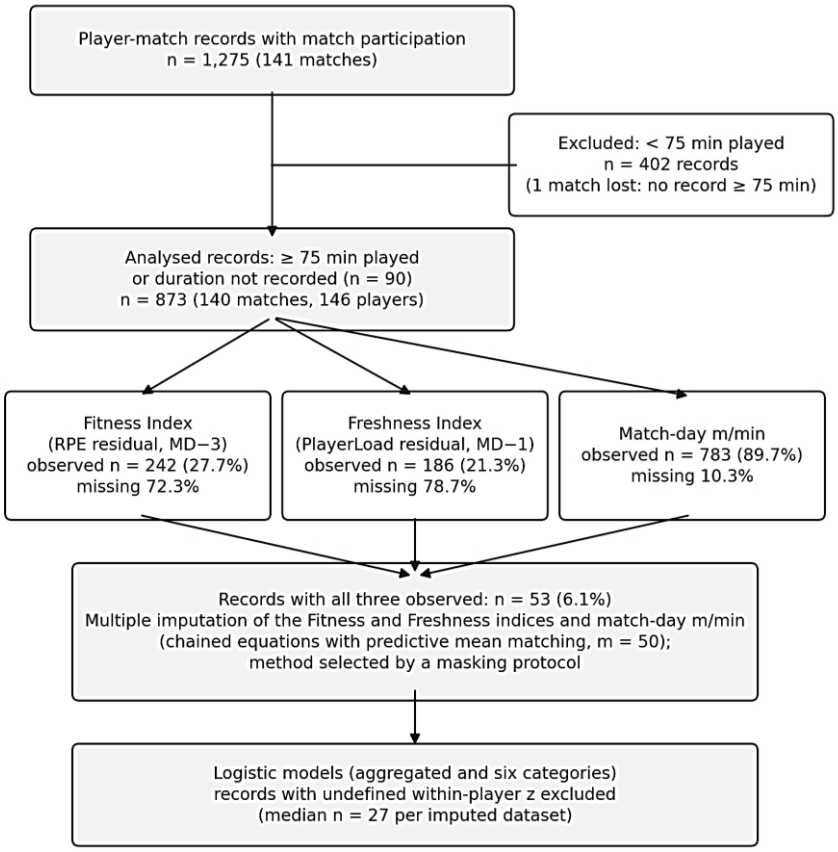


Figure 1. Flow of player–match records and coverage of each index. Exclusions for injury, return to play and attendance were applied by the staff before data export and are not shown. Percentages of observed values refer to the 873 analysed records. The imputation method was selected with the masking protocol fixed before the outcome models; the selection was repeated with the same rule after implementation errors were corrected and again after players were linked across seasons (Methods).

In this regard, players were excluded if they sustained injuries resulting in an absence of more than two consecutive weeks, were engaged in return-to-play protocols, or experienced prolonged periods of non-training or modified participation. These exclusion criteria were implemented to ensure that the analysed data reflected regular training and competition exposure, consistent with previous applied research in youth and professional football contexts (Rampinini et al., 2008). These criteria were applied by the staff before the data were exported, in a spreadsheet with no surviving record of the excluded players or the reasons. Because these exclusions may remove players with reduced readiness, selection bias cannot be ruled out (see Limitations). Analyses were conducted both at an aggregated level and stratified by developmental category to explore age-specific patterns across the youth development pathway.

 © 2027 ARD Asociación Española

Ethical considerations

Players aged 18 years or older signed a written informed consent. For players under 18, parents or legal guardians signed a written informed consent at the parents' meeting held at the start of each season, in which they were informed that the club's sport science department records data as part of routine training and evaluation practices and that these records may be used to guide training and for research reported as aggregated descriptive statistics. Players under 18 were informed at the start of the season that they would take part in regular evaluations by the sport science department, integrated into the club's routine training activities and involving no motor actions other than those of training itself.

No ethics committee review was sought, because the data were collected as part of the club's routine, non-invasive monitoring, without any procedure other than those of usual football training, and were analysed retrospectively after pseudonymisation. The Declaration of Helsinki (World Medical Association, 2013) was used as an ethical reference. Player names were replaced by pseudonymous codes when the data were loaded for analysis; the key linking codes to names is stored separately from the analysis files, and no analysis file or output contains names.

Procedures

Data were analysed from August 2024 to August 2025, a period that included a competitive break (no official matches between November 2024 and March 2025); official matches were played in 35 weeks. Player names were replaced by pseudonymous codes when the data were loaded. Because names were written in different formats in 2024 and 2025 (two words in 2024 and three in 2025), the same player received a different code in each season; records of the same player were therefore linked across seasons by name and category with an automatic, rule-based procedure that did not display names (83 links; the 477 codes in all data sources correspond to 394 players). Two codes may each include two different players with the same name and could not be separated (see Limitations). The 873 eligible player–match records belong to 146 players (Results). Training frequency differed by developmental category, with U13–U16 players completing four weekly sessions (Monday to Thursday), and U18–U21 players undertaking five sessions per week (Monday to Friday), reflecting the club's structured age-specific training model. Training sessions with GPS data lasted 55.3 ± 15.5 min (mean ± *SD* of 3,658 sessions; Table 1), and all teams competed in one official match per week.

Training sessions were coded relative to match day (MD-*k*, *k* days before the match). The Fitness Index uses the MD-3 session and the Freshness Index the MD-1 session; complete microcycles were not required, because the available data did not contain them for most players (Figure 1). In total, 141 official matches were played (84 wins, 9 draws, 48 losses).

To minimise confounding effects related to unequal match exposure and playing time, only matches in which players completed at least 75 minutes of effective playing time were included in the analysis, consistent with previous recommendations in football performance research (Mandorino et al., 2025). This criterion is now applied to both seasons; the original code did not apply it, and 52.6% of the 2025 records and 6.6% of the 2024 records analysed in the original submission had fewer than 75 min. Records from 2024 without a recorded match duration (n = 90) had been filtered by the staff before export and were retained. Data preparation for 2025 included two manual steps that were not recorded in the original code: player names in the RPE forms and in the CMJ sheets were reconciled by hand with the squad lists, and zero CMJ heights were set to missing. The resulting files are used as fixed, checksum-verified inputs.

Variables and Instruments

External load

External load was quantified using WIMU Pro™ GPS devices operating at a sampling frequency of 10 Hz, which have demonstrated validity and reliability for monitoring locomotor demands in football (Gómez-Carmona et al., 2019). The following variables were extracted for each training session: total distance covered (TD, m), metres per minute (MM), high-speed running distance (HSR, m; > 20 $km \cdot h^{-1}$ in both seasons), accelerations (ACC, > 3 $m \cdot s^{-2}$), decelerations (DEC, < −3 $m \cdot s^{-2}$), maximum velocity (Vmax, $km \cdot h^{-1}$), and PlayerLoad (PL, arbitrary units). These metrics collectively characterise the volume and intensity of external locomotor load, as well as changes in movement dynamics imposed during training.

Internal load

Internal load was assessed using ratings of perceived exertion (RPE), collected approximately 30 minutes after each training session using the CR-10 scale. Session internal load was calculated as the product of RPE and session duration (minutes), following the method originally proposed by Foster (1998) and subsequently validated in football populations (Impellizzeri et al., 2004). The Fitness Index, however, uses the RPE of the MD-3 session itself (Eq. 1.a), not session-RPE, as in the original analysis.

Countermovement jump

Countermovement-jump (CMJ) height was recorded on MD−4 and MD−1 as part of routine monitoring; two trials were performed and the best one was recorded (in the 2024 data, in which both trials were available, the recorded height equalled the best trial in at least 99.7% of the 360 tests of each day). The relative change from MD-4 to MD-1 was computed for each player and week; changes larger than 30% in absolute value were treated as missing. The CMJ was used for the convergent validity of the indices and for a post hoc alternative Freshness Index.

Freshness Index (FR)

The Freshness Index was designed to capture the player's relative neuromuscular freshness on MD-1. To derive this indicator, PlayerLoad recorded on MD-1 was modelled as a function of external load variables from the same session (TD, MM, HSR, ACC, DEC, Vmax; Eq. 2.a). This modelling approach enabled estimation of the residual mechanical stress immediately MD-1, adjusted for the specific load profile of the session. Operationally, the Freshness Index is the expected minus the observed MD-1 PlayerLoad (Eq. 2.b): positive values indicate less accelerometer load than expected for the locomotor load recorded by GPS in the same session, which the source framework interprets as favourable. It is therefore a residual of a statistical model, not a direct measure of neuromuscular freshness; what it captures is examined in the Results (Question 2). Final values were standardised using individual *z*-scores (within each player, across all of the player's records from both seasons; Eq. 2.c) to facilitate both intra- and inter-category comparisons (Mandorino et al., 2025).

Fitness Index (FI)

The Fitness Index was conceptualised as an indicator of accumulated physiological load during the central phase of the competitive microcycle, reflecting the player's residual fatigue or preparedness prior to the tapering period. In accordance with the embedded monitoring framework proposed by Mandorino et al. (2025), the RPE of the MD-3 session (three days before match day), typically corresponding to one of the highest loading sessions of the week, was modelled using external-load variables from the same session (session duration, MM, HSR, ACC, DEC, Vmax, PL; Eq. 1.a). Operationally, the Fitness Index is the expected minus the observed MD-3 RPE (Eq. 1.b): positive values indicate lower perceived exertion than expected for the external load of the session (favourable), and negative values higher perceived exertion than expected,

 © 2027 ARD Asociación Española

which may be read as unexpected internal strain rather than as a lack of fitness. Although the concept refers to accumulated load, the index uses a single session (MD-3), as in the original analysis. Both indices follow the sign convention of the source framework (expected − observed, so that positive values are favourable; Mandorino et al., 2025), which the original code also used for the Fitness Index; the original Equation 1.b and text had stated observed − expected and interpreted the index in that direction, so the equations have been corrected. Reversing the sign would only invert the odds ratios ($OR \rightarrow 1/OR$) and would not change any inference. As with the Freshness Index, Fitness Index values were standardised using individual *z*-scores (Eq. 1.c).

Readiness framework

Readiness was conceptualised as a multidimensional construct reflecting the player's capacity to tolerate accumulated internal and external training loads and to express optimal performance during competition. In line with contemporary applied monitoring frameworks, readiness was operationalised through the interaction of two embedded indicators: (i) a Fitness Index, representing the RPE response to the MD-3 session relative to its external load, corresponding to the central loading phase of the competitive microcycle; and (ii) a Freshness Index, representing the PlayerLoad response to the MD-1 session relative to its locomotor load (Mandorino et al., 2025). In addition to the embedded readiness indicators (fitness and freshness), readiness was also assessed from a subjective perspective, acknowledging that the athlete's perceived recovery and psycho-physiological status represent a complementary dimension of functional readiness. Subjective readiness was monitored using a wellness questionnaire adapted from McLean et al. (2010), comprising five items assessing perceived fatigue, muscle soreness, sleep quality, stress, and mood. Total scores ranged from 5 to 25 points, with lower scores indicating poorer perceived recovery and readiness. Wellness data were available only for the 2024 season and did not enter the outcome models; they were used to examine the convergent validity of the indices.

Match outcome

Competitive performance was defined based on official match outcome and dichotomized as win or non-win (i.e., draw or loss). This classification facilitated examination of the association between weekly readiness–load profiles and team competitive success, consistent with contemporary analytical approaches in applied football research (Mandorino et al., 2025; Settembre et al., 2024). Because match outcome is a team-level variable shared by all players of a match, the study examines associations with team results, not with individual player performance.

Statistical analysis

Derivation of embedded monitoring indicators

In line with the embedded monitoring framework described by Mandorino et al. (2025), Fitness Index and Freshness Index values were derived using machine-learning models trained to predict expected physiological or mechanical responses from external-load inputs of the same session. Random Forest algorithms were employed for this purpose, and readiness indicators were computed as the difference between expected and observed responses (expected − observed, so that positive values are favourable). All derived indices were subsequently normalised using *z*-scores. Figure 2 shows this chain conceptually, together with match-day running, which is concurrent with the result rather than a pre-match predictor (see Discussion). A Random Forest (Breiman, 2001) averages many regression trees, each grown on a random subsample of the data, and can capture non-linear relationships without specifying them in advance. The hyperparameters of the original analysis were kept (100 trees, other settings at the software defaults, random seed 42). One model was trained for each response with all observed sessions of the corresponding day: 525 MD-3 sessions with RPE (105 players) and 656 MD-1 sessions with PlayerLoad (115 players). A model

that predicts the same observations it was trained on overfits them (in the original analysis, residuals were computed in sample). Expected values were therefore obtained out of player: players (linked across seasons) were split into five groups, the sessions of each group were predicted by a model trained on the other groups, and this was repeated with 20 different random splits whose predictions were averaged. Model performance was summarised by the out-of-player R^2, root mean square error (RMSE) and calibration slope (the slope of observed on predicted values; 1 indicates good calibration).

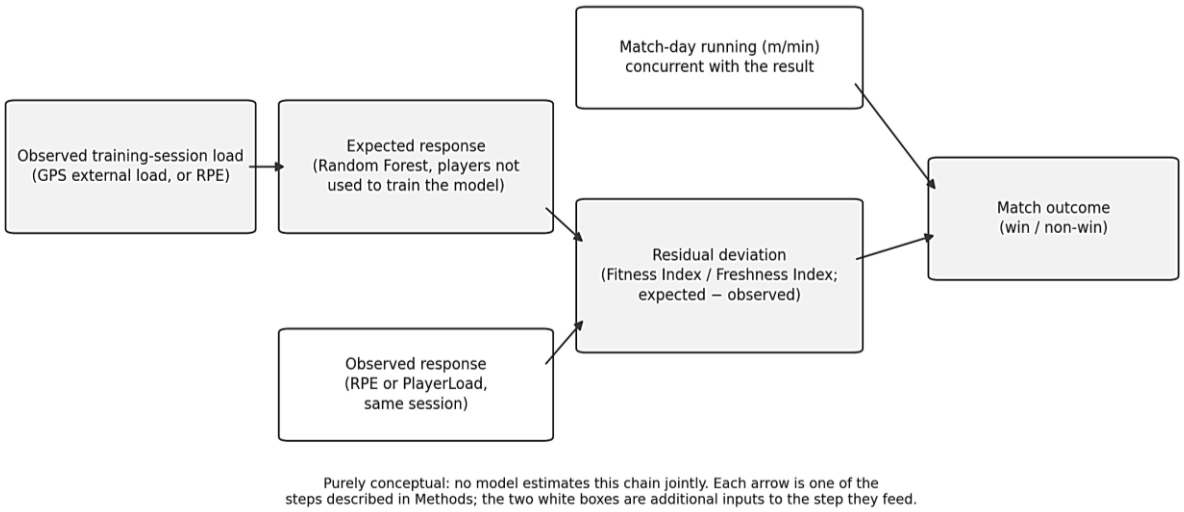


Figure 2. Conceptual framework relating observed training load, the expected-response model, the residual (Fitness Index / Freshness Index), match-day running and match outcome (reviewer request, comment R2.3). This diagram is schematic and illustrates the steps described in Methods; no model estimates the whole chain jointly.

In addition to main effects, second- and third-order interaction terms were constructed (Fitness × Freshness, Freshness × MD Load, Fitness × MD Load, and Fitness × Freshness × MD Load) to capture combined and conditional effects between readiness components and competitive demands. The equations are:

(1.a) Expected RPE(MD-3) = *f*(session duration, MM, HSR, ACC, DEC, Vmax, PL of the MD-3 session)
(1.b) FI = expected RPE(MD-3) − RPE(MD-3)
(1.c) $z_{FI} = (FI - \mu_{FI,i}) / \sigma_{FI,i}$

(2.a) Expected PL(MD-1) = *f*(TD, MM, HSR, ACC, DEC, Vmax of the MD-1 session)
(2.b) FR = expected PL(MD-1) − PL(MD-1)
(2.c) $z_{FR} = (FR - \mu_{FR,i}) / \sigma_{FR,i}$
(3) $z_{MD} = (MM_{MD} - \mu_{MD,i}) / \sigma_{MD,i}$

(4) logit $P(\text{win}) = \beta_0 + \beta_1 \cdot z_{FI} + \beta_2 \cdot z_{FR} + \beta_3 \cdot z_{MD} + \beta_4 \cdot z_{FI} \times z_{FR} + \beta_5 \cdot z_{FI} \times z_{MD} + \beta_6 \cdot z_{FR} \times z_{MD} + \beta_7 \cdot z_{FI} \times z_{FR} \times z_{MD}$

Here, *f* is the function learned by the Random Forest, MM_{MD} is the match-day metres per minute, and μ and σ denote the mean and *SD* of player *i*'s values across all of the player's records from both seasons; the subscript *i* identifies the player. Equation 4 corrects an erratum in the original submission, which listed total distance and a repeated Fitness × Freshness term; the fitted model has always been the one in Equation 4. A *z*-score is undefined for a player with a single record or without variation; such records were excluded from the outcome models.

 © 2027 ARD Asociación Española

Missing data

The Fitness Index, the Freshness Index and match-day running were missing in 72.3%, 78.7% and 10.3% of the 873 eligible records, respectively, and all three were observed in 53 records (6.1%; Figure 1); missing values were therefore not sporadic. Multiple imputation creates several plausible versions of the missing values, analyses each completed dataset and combines the results so that the uncertainty due to missing data is reflected in the confidence intervals (Rubin, 1987; van Buuren, 2018). The imputation method was chosen with a masking protocol, fixed before any outcome model was fitted, that does not use the outcome model: 20% of the observed values of each variable were hidden at random (20 repetitions), each candidate method imputed them, and methods were compared by the normalised root mean square error between hidden and imputed values and by the ratio of imputed to true variance (a ratio between 0.8 and 1.2 was taken to preserve variability). Candidates were the category × season mean (the method of the original analysis), the player's own mean in other records, k-nearest neighbours, and multiple imputation by chained equations with predictive mean matching or with Bayesian regression. The protocol was first applied with players identified by season: it selected chained equations with predictive mean matching, and a first complete run of the analysis was made with it. An independent verification of the code then found implementation errors in the candidate methods (the Bayesian regression was fitted to unstandardised variables, category entered the auxiliary variables as linear age, and there was an error in the fallback of the player-mean method); after they were corrected, the selection was repeated with the same rule and, because no method then preserved the variance of all three variables, the fallback rule chose chained equations with Bayesian regression. After players were linked across seasons, the selection was repeated once more with the same protocol: chained equations with predictive mean matching was the only method that preserved the variance of the three variables (variance ratios 1.07, 0.85, 0.98 for the Fitness Index, the Freshness Index and match-day running) and also had the lowest mean normalised RMSE (0.808, vs 0.816 for chained equations with Bayesian regression), so it was selected by the main rule (Supplementary Table S3). The analysis with chained equations with Bayesian regression is reported as a post hoc sensitivity analysis and led to the same conclusions. Auxiliary variables were category, season, playing position, minutes played, match venue, the other imputed variables and the player's mean in other records; match outcome was also included, as recommended for multiple imputation (van Buuren, 2018). In total, 50 imputed datasets were analysed and the estimates were combined with Rubin's rules. Within-player *z*-scores were computed in each imputed dataset (median 27 records excluded per dataset because the *z*-score was undefined).

Outcome model

The association between readiness indicators, training load, and match outcome was evaluated using binary logistic regression, with match outcome coded as win (1) or non-win (0), including an intercept and the seven terms of Equation 4. Because all players of a team share the outcome of a match, standard errors were clustered by match (cluster-robust or "sandwich" standard errors; Cameron & Miller, 2015), so that the match, not the player record, is treated as the independent unit. Odds ratios (*OR*) are reported per within-player *SD* with 95% confidence intervals (CI) and exact *p* values; 90% CIs, used in the original submission following magnitude-based recommendations (Batterham & Hopkins, 2006; Hopkins et al., 2009), are provided in Supplementary Table S11. Model fit is reported as the log-likelihood, McFadden's pseudo-R^2 and a likelihood-ratio test against the intercept-only model (medians across imputations). Because the likelihood-ratio test assumes independent records, a joint Wald test of the seven terms with match-clustered standard errors, combined across imputations with the *D*1 statistic (Li et al., 1991), was added post hoc.

To account for the longitudinal and hierarchical structure of the data, supplementary analyses were conducted using logistic generalised linear mixed models (GLMM) with random intercepts for player and developmental category, fitted to the aggregated data by variational Bayes, an approximate Bayesian

method; the linear mixed models mentioned in the original submission are not appropriate for a binary outcome and had not been fitted. A random intercept for match is not identifiable, because the outcome is constant within a match; the dependence within matches is handled by the clustered standard errors of the main model.

Stability was examined with a bootstrap that resamples whole matches (2,000 replicates, distributed across the imputed datasets), reporting percentile 95% CIs and the proportion of replicates with the same sign as the estimate, and with a reduced model containing only the three main effects. Events per parameter (EPV) were counted in matches: the number of matches with the less frequent outcome divided by the number of estimated parameters (Peduzzi et al., 1996). To interpret the interactions, predicted probabilities of winning from the aggregated model were computed across the 5th–95th percentiles of the observed (not imputed) values of each index, with the other variables at their 10th, 50th or 90th percentile (Figure 3).

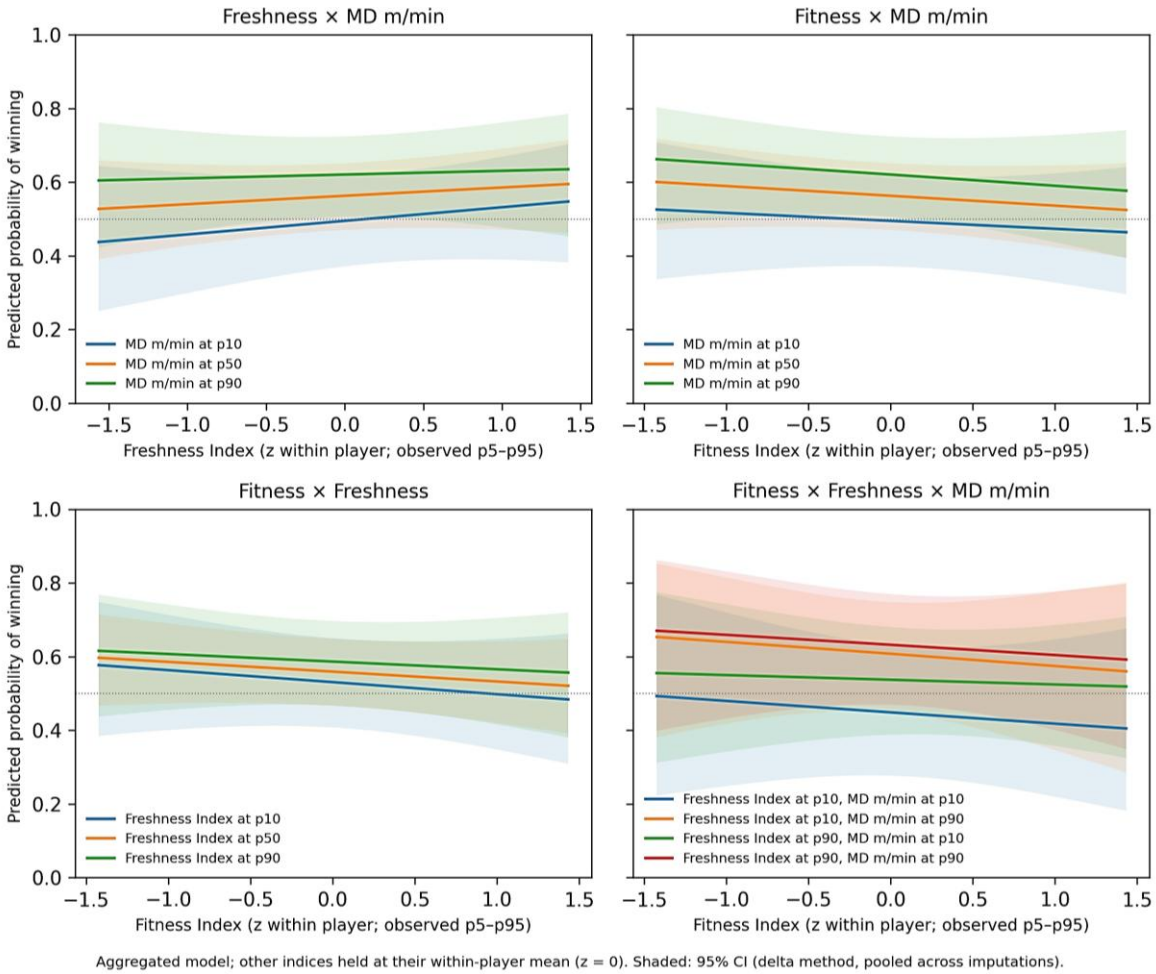


Figure 3. Predicted probability of winning (95% CI, shaded) from the aggregated original model, across the 5th–95th percentiles of the observed values of the varied index, with the other variables at the indicated percentiles (p10, p50, p90) or at their within-player mean (z = 0). Estimates pooled across 50 imputed datasets; CIs by the delta method with match-clustered standard errors. The dotted line marks a probability of 0.5.

Analysis by developmental category

In addition to the aggregated model including all age groups, separate models were estimated for each developmental category (U13–U21). This stratified approach was intended to explore category-specific

 © 2027 ARD Asociación Española

readiness–performance patterns across the youth pathway, consistent with applied research in elite football populations (Mandorino et al., 2025). Because separate models do not test whether associations differ, differences between categories were tested within one model with category × index interactions (category × Fitness Index, × Freshness Index and × match-day running; 15 terms), using a joint Wald test with match-clustered standard errors combined across imputations.

Sensitivity analyses

Planned sensitivity analyses were: (i) the model adjusted for match venue (home vs away), the only contextual variable available for all matches; (ii) the category-mean imputation of the original analysis, as a reference; (iii) the main-effects model and the match bootstrap described above; and (iv) the GLMM. The following were added post hoc, after the first results had been seen: (v) the same analysis with multiple imputation by chained equations with Bayesian regression, the other multiple-imputation method of the protocol; (vi) all candidate imputation methods and complete cases only; (vii) a Freshness Index computed from the CMJ (observed minus expected change in CMJ height from MD-4 to MD-1, so that positive values are again favourable, with external load accumulated from MD-4 to MD-2 as input), because it is closer to the neuromuscular construct; (viii) adjustment for playing position and for the opponent club, the latter both as fixed effects with match-clustered standard errors and as a random intercept in the GLMM; (ix) grouping the categories into the three developmental stages used in the original conclusions (U13–U14, U15–U16 and U18–U21), into the two training structures (U13–U16 vs U18–U21) and into U13–U15 vs U16–U21; (x) multilevel multiple imputation with a random intercept per player; and (xi) the analysis with players identified by season, the unit used before players were linked across seasons. All results are reported, whatever their direction.

Comparison with simple indicators

To assess whether the indices add predictive value, models were compared by cross-validation grouped by match (10 folds, repeated 20 times): a base model with venue only; + match-day running; + the two indices; + the raw indicators from which they are derived (MD-3 RPE and MD-1 PlayerLoad); and the original seven-term model plus venue. Performance was measured by log-loss and Brier score (lower is better), the area under the ROC curve (AUC; .5 indicates no discrimination) and the calibration slope. The imputation for this comparison did not use the outcome.

Convergent validity and signal diagnostics

Convergent validity (planned) was examined with within-player correlations of the indices with the 2024 wellness items of the same day and with the weekly CMJ change, estimated by repeated-measures correlation (Bakdash & Marusich, 2017), which removes stable differences between players and quantifies how two variables co-vary within the same player. Signal diagnostics were added post hoc; they are descriptive and do not use match outcome. To understand why the expected-response models did or did not learn, we estimated (a) the share of the variance of RPE and PlayerLoad that lies between players, with linear mixed models fitted to all training sessions (random intercepts for player, week and session; fixed effects for category, season and session day); (b) within-player correlations of MD-3 RPE with the PlayerLoad, duration and distance of the same session, and of MD-1 PlayerLoad with distance and duration, as a reference for a relationship that the model does learn; and (c) for the CMJ, the within-player *SD* of the weekly change, the within-player correlation between the MD-4 and MD-1 jumps of the same week, the correlation between the changes of consecutive weeks, the correlation of the change with the load accumulated from MD-4 to MD-2, the technical error between the two trials of a test (2024, when both trials were recorded), the minimal detectable change (MDC95 = 1.96 × √2 × technical error) and the share of weekly changes exceeding it (Hopkins, 2000).

Software and reproducibility

Analyses were conducted in Python 3.10.0 (pandas 2.2.2, scikit-learn 1.5.1, statsmodels 0.14.4) with a scripted, deterministic and version-controlled pipeline that includes automated tests; each results file records the version of the code that produced it.

RESULTS

Data coverage and sample

Of 1,275 player–match records with match-day GPS data (141 matches), 873 from 146 players met the ≥75-min criterion, covering 140 matches (Figure 1). The Fitness Index could be computed for 242 records (27.7%), the Freshness Index for 186 (21.3%) and match-day running was available for 783 (89.7%); all three were observed in 53 records (6.1%). Coverage varied between categories (Fitness Index 11.3–40.7%; Freshness Index 8.6–37.9%), and no record had all three observed in U15 or U13 (Table 1). The aggregated model had 7.0 events per parameter; the category models, with 20–25 matches each, had 0.8–1.4, far below the commonly recommended minimum of 10 (Peduzzi et al., 1996).

Table 1. Sample, results and coverage of each index by category (panel A).

Category	Matches (W/D/L)	Players*	Records*	Fitness Index observed	Freshness Index observed	Matchday m·min^{-1} observed	All three observed	Records analysed†	EPV‡
U21	24 (13/2/9)	41	177	36 (20.3%)	67 (37.9%)	177 (100.0%)	18 (10.2%)	171	1.4
U18	24 (16/2/6)	36	166	51 (30.7%)	62 (37.3%)	156 (94.0%)	29 (17.5%)	162	1.0
U16	23 (11/4/8)	33	139	36 (25.9%)	12 (8.6%)	139 (100.0%)	4 (2.9%)	134	1.4
U15	20 (14/1/5)	30	124	14 (11.3%)	15 (12.1%)	124 (100.0%)	0 (0.0%)	121	0.8
U14	25 (14/0/11)	36	132	50 (37.9%)	16 (12.1%)	94 (71.2%)	2 (1.5%)	128	1.4
U13	24 (16/0/8)	32	135	55 (40.7%)	14 (10.4%)	93 (68.9%)	0 (0.0%)	130	1.0
All	140 (84/9/47)	146	873	242 (27.7%)	186 (21.3%)	783 (89.7%)	53 (6.1%)	846	7.0

Note. * Players (linked across seasons) and player–match records with ≥ 75 min played (or, in 2024, without a recorded duration); one of the 141 official matches had no eligible record. Observed: *n* (%) of records in which the index, or match-day running, could be computed from observed data. † Median number of records per imputed dataset entering the original model, after excluding records with undefined within-player *z*-scores. ‡ Events per parameter of the original model (eight parameters), counted in matches. W/D/L, wins/draws/losses.

Table 1 (continued). Training and match loads by category (panel B).

Category	Training sessions with GPS	Training session duration (min)	Match minutes played	RPE, MD−3 session (CR-10)	PlayerLoad, MD−1 session (a.u.)	Match-day m·min^{-1}
U21	789	56.7 ± 16.2	94.5 ± 8.9	6.0 ± 1.6	45.0 ± 14.2	102.1 ± 13.9
U18	653	58.5 ± 16.3	94.9 ± 8.1	6.4 ± 1.7	50.0 ± 12.6	100.6 ± 13.6
U16	408	59.3 ± 15.8	94.0 ± 6.4	5.4 ± 1.8	50.5 ± 10.9	103.0 ± 14.9
U15	454	56.7 ± 14.6	93.5 ± 6.4	4.5 ± 2.0	59.0 ± 12.5	102.5 ± 11.8
U14	692	51.6 ± 14.5	85.9 ± 5.2	6.0 ± 1.3	59.1 ± 15.3	99.3 ± 15.5
U13	662	51.2 ± 13.1	84.1 ± 3.6	4.7 ± 1.3	55.4 ± 11.5	96.5 ± 13.8
All	3,658	55.3 ± 15.5	92.1 ± 8.1	5.6 ± 1.7	50.1 ± 14.0	101.0 ± 14.0

Note. Mean ± *SD*. Training sessions: all training sessions with GPS data of each category (both seasons). Match minutes: eligible records with a recorded duration. RPE, PlayerLoad and match-day m·min^{-1}: observed values in the eligible player–match records (*n* as in panel A).

Question 1: association with match outcome

The original model

In the aggregated model (846 records per imputed dataset, 140 matches, 84 won), none of the seven terms had a 95% CI excluding 1 (Table 2). The main-effect estimates were: Fitness Index, *OR* = 0.90, 95% CI [0.69, 1.16], *p* = .413; Freshness Index, *OR* = 1.10, 95% CI [0.84, 1.43], *p* = .479; and match-day running, *OR* = 1.24, 95% CI [0.95, 1.61], *p* = .110. The same was true for the 42 terms of the six category models, whose intervals were wide (upper limits between 1.42 and 3.71). With 90% CIs, 1 of 49 terms excluded 1 (Supplementary Table S11). The aggregated model explained little of the outcome (McFadden pseudo-R^2 = .016; likelihood-ratio *p* = .012). The likelihood-ratio test, which assumes independent records, was below .05 in the aggregated model (*p* = .012) and the U18 model (*p* = .013); the post hoc joint Wald tests with match-

 © 2027 ARD Asociación Española

clustered standard errors gave $p \geq .87$ in all seven models. In the match bootstrap, no percentile interval excluded 1, and in the aggregated model the proportion of replicates with the same sign as the estimate ranged from .52 to .95. The main-effects model gave 0 of 21 terms with a 95% CI excluding 1 (Supplementary Table S4).

Table 2. Original model (Equation 4): odds ratios per within-player SD, aggregated and by category, pooled across 50 imputed datasets.

Term	All categories	U21	U18	U16	U15	U14	U13
Fitness Index (FI)	0.90	0.94	0.85	0.77	0.83	0.90	0.77
95% CI	0.69–1.16	0.55–1.61	0.51–1.42	0.41–1.42	0.40–1.74	0.47–1.72	0.39–1.49
Freshness Index (FR)	1.10	1.16	1.21	1.01	1.09	0.97	1.11
95% CI	0.84–1.43	0.66–2.02	0.68–2.16	0.52–1.95	0.52–2.29	0.49–1.91	0.56–2.20
Match-day m·min^{-1} (MD)	1.24	1.37	1.83	1.25	1.44	1.01	0.89
95% CI	0.95–1.61	0.75–2.53	0.93–3.63	0.61–2.57	0.56–3.71	0.51–2.00	0.46–1.70
FI × FR	1.02	1.01	1.04	0.94	1.02	0.99	0.93
95% CI	0.80–1.30	0.60–1.72	0.53–2.05	0.48–1.85	0.50–2.12	0.44–2.23	0.44–1.96
FI × MD	0.98	0.89	0.96	1.07	0.98	1.00	1.00
95% CI	0.77–1.26	0.52–1.51	0.55–1.69	0.54–2.15	0.41–2.36	0.49–2.07	0.46–2.20
FR × MD	0.96	1.08	0.95	0.94	0.88	1.00	0.93
95% CI	0.76–1.20	0.62–1.87	0.58–1.54	0.45–1.99	0.37–2.10	0.48–2.07	0.44–1.97
FI × FR × MD	0.99	1.17	0.88	1.03	1.05	1.00	1.05
95% CI	0.79–1.25	0.57–2.38	0.53–1.43	0.47–2.26	0.41–2.68	0.35–2.87	0.46–2.40
Records analysed	846	171	162	134	121	128	130
Matches (won / not won)	140 (84 /56)	24 (13 / 11)	24 (16 / 8)	23 (11 / 12)	20 (14 / 6)	25 (14 / 11)	24 (16 / 8)
Events per parameter	7.0	1.4	1.0	1.4	0.8	1.4	1.0

Note. 95% CIs from standard errors clustered by match. Bold would indicate an interval excluding 1; none did. FI, Fitness Index; FR, Freshness Index; MD, match-day m·min^{-1}. Records analysed: median per imputed dataset. 90% CIs are in Supplementary Table S11 and bootstrap intervals in Supplementary Table S4.

Across the observed range of the indices (5th–95th percentiles), predicted probabilities of winning from the aggregated model varied between .40 and .67, with 95% CIs between .18 and .86 (Figure 3). There was no evidence that any combination of the indices and match-day running changed the probability of winning, although changes of moderate size cannot be excluded.

Differences between categories and developmental stages

The joint test of the 15 category × index interactions gave no evidence that associations differed between categories ($p > .999$). Post hoc, grouping the categories into the three developmental stages of the original conclusions (U13–U14, U15–U16, U18–U21) gave 0 of 21 terms with a 95% CI excluding 1 and no evidence of differences between stages ($p = .87$); the same held for the two training structures (U13–U16 vs U18–U21: 0 of 14 terms; $p = .63$) and for U13–U15 vs U16–U21 (0 of 14; $p = .78$) (Supplementary Table S7). With 21 terms, about one interval excluding 1 would be expected by chance alone.

Sensitivity analyses

In the planned venue-adjusted model, 0 of 49 index and match-load terms had a 95% CI excluding 1 (aggregated Fitness Index, $OR = 0.87$, 95% CI [0.67, 1.14]). The estimate for playing at home was compatible with a home advantage but imprecise, with a lower limit at about 1 ($OR = 2.11$, 95% CI [0.999, 4.46], $p = .050$). The GLMM (50 of 50 fits converged) gave 1 of 7 terms with a 95% credible interval excluding 1 (match-day running, $OR = 1.28$, 95% credible interval [1.07, 1.52]; Supplementary Table S5); this model does not account for the clustering of players within matches and its variational approximation underestimates the variance, so its intervals are narrower than those of the main model, in which the interval for match-day running included 1 ($OR = 1.24$, 95% CI [0.95, 1.61]). Post hoc, the conclusions were unchanged with chained equations with Bayesian regression (0 of 49 terms; category test $p > .999$), with players identified by season instead of linked across seasons (208 player–season identifiers; 0 of 49 terms; category test $p > .999$; Supplementary Table S12), with multilevel imputation (0 of 70 terms, including the main-effects models;

category test $p > .999$), with the CMJ-based Freshness Index (0 of 49 terms; category test $p > .999$; Supplementary Table S8), after adjusting for venue and playing position (0 of 49 terms; category models with EPV ≤ 0.8) and with opponent fixed effects and match-clustered standard errors (0 of 7 terms; 97 of 140 matches, after excluding the 10 opponent clubs that were always or never beaten; Supplementary Table S6). With players identified by season, the conclusions for the indices were the same, although the venue interval and the match-day running intervals of the mixed models changed enough to cross 1 (see Supplementary Table S12 and Limitations). In the GLMM with a random intercept for the opponent club (24 clubs), 1 of 7 terms had a credible interval excluding 1, and the interval for match-day running excluded 1 (*OR* = 1.29, 95% credible interval [1.02, 1.64]); this model does not cluster by match and variational Bayes underestimates the variance, so its intervals are too narrow, and the comparable inference is the fixed-effects model (match-day running, *OR* = 1.27, 95% CI [0.92, 1.76], $p = .140$). With single imputation methods, which understate uncertainty, isolated category-specific terms excluded 1 and changed with the method (category mean, 3 of 49; category mean with the original convention, 4 of 49; k-nearest neighbours, 3 of 49; player mean, 1 of 42, with 6 of 7 models estimable), and none of them was in the aggregated model. With complete cases (53 records with the three variables observed; 34 with defined *z*-scores) only 2 of 7 models were estimable; 1 term excluded 1 (U18 FI × MD, estimated from 23 records in 11 matches), too few matches for a reliable clustered standard error (Supplementary Table S3).

Comparison with simple indicators. Neither the indices nor the raw indicators improved cross-validated log-loss or Brier score over venue alone, and the AUC stayed close to .5 in all models (Supplementary Table S9): log-loss 0.688 with venue, 0.690 adding match-day running, 0.691 adding the two indices and 0.692 adding the raw MD-3 RPE and MD-1 PlayerLoad instead; the AUC was .49, .53, .54 and .54, respectively, and .54 for the original seven-term model with venue (log-loss 0.695).

Question 2: what the indicators capture

Validity of the indices

Within players, there was no evidence that the indices were correlated with the wellness items of the same day: Fitness Index correlations ranged from $r = -.14$ to .06 (154 observations from 41 players), and Freshness Index correlations ranged from $r = .01$ to .15 (66 observations from 23 players). There was also no evidence of correlations with weekly CMJ change (Fitness Index, $r = -.12$, 95% CI [−.30, .07]; Freshness Index, $r = -.01$, 95% CI [−.28, .26]). The largest absolute correlation was .15 and 0 of 14 correlations had $p < .05$, but with 66–154 observations the 95% CIs ranged from −.31 to .43, so moderate correlations cannot be excluded (Table 3; Supplementary Table S10).

Signal diagnostics (post hoc)

External load did not predict MD−3 RPE out of player ($R^2 = -.04$; calibration slope 0.33; Table 3). In a mixed model of all training sessions, about 57% of the variance of RPE lay between players after accounting for category, season and session day, and within players RPE was weakly to moderately correlated with the load of the same session ($r = .33$ with PlayerLoad, .17 with duration and .34 with distance). A model evaluated on players it has not seen cannot learn each player's habitual level of RPE, so the Fitness Index probably reflects the observed RPE (with its sign reversed) more than a deviation from a load-based expectation. In contrast, external load predicted MD-1 PlayerLoad well ($R^2 = .65$; calibration slope 0.98), with about 12% of its variance between players; within players, PlayerLoad was strongly correlated with distance ($r = .83$), so the Freshness Index captures the part of accelerometer load not explained by the GPS-derived locomotor variables.

 © 2027 ARD Asociación Española

The weekly CMJ change (2,490 player-weeks) had *M* = 2.2% and *SD* = 6.9% (within-player *SD* = 6.6%). The MD−4 and MD−1 jumps of the same week were moderately correlated within players (*r* = .52). Changes in consecutive weeks showed no positive correlation (*r* = −.04, 95% CI [−.09, .01]), and the change was only weakly related to the load accumulated from MD−4 to MD−2 (*r* = .13 with PlayerLoad). The technical error between the two trials of a test was 1.04 cm (3.2%), so the minimal detectable change was 2.89 cm; weekly changes had *M* = 0.65 cm and *SD* = 2.31 cm, and only 23% of them exceeded the minimal detectable change. In a mixed model of the jump heights (post hoc), the estimated day effect (MD−1 vs MD−4) was 0.67 cm, 95% CI [0.50, 0.85], a small systematic increase. Accordingly, in the post hoc CMJ-based Freshness Index, external load did not predict the CMJ change out of player (R^2 = −.05).

Individual trajectories (post hoc)

Because the Fitness and Freshness indices were observed in only 27.7% and 21.3% of records, individual trajectories were built from observed values only, excluding players with fewer than 4 observed records or without within-player variability among them (27 players for the Fitness Index and 18 for the Freshness Index; Supplementary Figure S1). Extreme deviations (within-player |*z*| > 1.5) were similarly frequent before wins and non-wins (Fitness Index: 5/78 (6.4%) vs 2/69 (2.9%); Freshness Index: 2/45 (4.4%) vs 3/43 (7.0%); Supplementary Table S13), with samples too small for a formal comparison.

Table 3. What the indicators feeding the indices capture: expected-response models, signal diagnostics and convergent validity.

Indicator and quantity	Estimate (95% CI)	Observations (players)
RPE of the MD−3 session		
Out-of-player R^2 of the expected-response model (calibration slope 0.33)	−.04	525 sessions (105)
Share of variance between players (post hoc mixed model, all training sessions)	.57	2,473 sessions (187)
Within-player *r* with PlayerLoad of the same session	.33 [.25, .42]	507 (87)
Within-player *r* with duration of the same session	.17 [.07, .26]	507 (87)
Within-player *r* with total distance of the same session	.34 [.25, .42]	507 (87)
PlayerLoad at MD−1 (response of the Freshness Index)		
Out-of-player R^2 of the expected-response model (calibration slope 0.98)	.65	656 sessions (115)
Share of variance between players (post hoc mixed model, all training sessions)	.12	3,658 sessions (240)
Within-player *r* with total distance of the same session	.83 [.81, .86]	634 (93)
Within-player *r* with duration of the same session	.59 [.53, .64]	634 (93)
CMJ height change, MD−4 → MD−1 (post hoc alternative Freshness response)		
Out-of-player R^2 of the expected-response model (calibration slope 0.23)	−.05	723 player-weeks (116)
Mean weekly change, % (*SD*; within-player *SD*)	2.2 (6.9; 6.6)	2,490 player-weeks
Technical error between the two trials of a test, cm (*CV*, %)	1.04 (3.2)	721 tests (2024)
Minimal detectable change (MDC_{95}), cm; weekly changes exceeding it	2.89; 23%	2,490 player-weeks
Within-player *r* between MD−4 and MD−1 jumps of the same week	.52 [.49, .55]	2,485 (184)
Within-player *r* between changes of consecutive weeks	−.04 [−.09, .01]	1,501 (160)
Within-player *r* with PlayerLoad accumulated MD−4 to MD−2	.13 [.05, .21]	706 (99)
Within-player *r* with distance accumulated MD−4 to MD−2	.13 [.05, .20]	706 (99)
Convergent validity of the indices (planned)		
Fitness Index vs 2024 wellness items, same day (range of *r*)	−.14 to .06	154 (41)
Freshness Index vs 2024 wellness items, same day (range of *r*)	.01 to .15	66 (23)
Fitness Index vs weekly CMJ change	−.12 [−.30, .07]	153 (45)
Freshness Index vs weekly CMJ change	−.01 [−.28, .26]	80 (29)

Note. *r*, within-player (repeated-measures) correlation. Share of variance between players: from a linear mixed model with random intercepts for player, week and session and fixed effects for category, season and session day, fitted to all training sessions with the variable. $MDC_{95} = 1.96 \times \sqrt{2} \times$ technical error. R^2 values are out of player; negative values indicate predictions worse than the mean. Signal diagnostics and the CMJ-based model are post hoc and descriptive (they do not use match outcome); convergent validity was planned. Wellness: five items (sleep, fatigue, stress, mood and muscle soreness) and the total score, available only in 2024.

DISCUSSION

This re-evaluation, which corrects the coding errors of the original analysis, found no evidence that the Fitness Index, the Freshness Index, match-day running or their interactions were associated with match outcome, either across categories or within any category, in models that account for the clustering of players

within matches, and no evidence that associations differed between categories or developmental stages. The category-specific findings of the original submission did not survive the correction of the Freshness Index (none of the five remained after this step; Supplementary Table S1). The second part of the study, based largely on post hoc diagnostics, may help explain the null result: with the data available, RPE varied to a large extent between players and could not be predicted from external load for players not used to train the model, PlayerLoad was largely explained by distance, and most weekly jump changes did not exceed the minimal detectable change (only 23% did); as measured, these variables therefore may have carried little week-to-week information beyond stable differences between players and the GPS-derived load.

Comparison with the source of the framework

The embedded framework was proposed and tested in 25 elite adult players of one club over 73 matches (Mandorino et al., 2025). That study derived the Fitness Index from heart rate, whose residual had been validated against a submaximal running test (r = .70; Mandorino et al., 2024), whereas the original submission and this re-evaluation derived it from RPE, for which no comparable validation is available and which, in our data, varied to a large extent between players. The source study modelled match outcome with a mixed multinomial model with a random intercept for player, without accounting for the clustering of players within matches, and reported three-way interactions and median-split combinations without correction for multiple comparisons. Our null result is therefore best read as an informative contrast—a different population (youth players), a different internal-load measure (RPE instead of heart rate) and a more conservative inference (match-clustered standard errors and all terms reported)—rather than as a refutation of the framework, which remains to be tested prospectively with validated measures.

Absence of evidence is not evidence of absence

A non-significant result does not show that an association is absent (Altman & Bland, 1995); the informative quantity is the range of effects compatible with the data. In the aggregated model, the 95% CIs of the three main effects lay between 0.69 and 1.61 per within-player *SD*: associations larger than these limits are not compatible with the data, given the imputation model (72% and 79% of the Fitness and Freshness values were imputed), whereas smaller associations, including moderate ones, cannot be excluded. In the comparison with simple indicators, the residual indices and the raw indicators from which they derive performed almost identically (log-loss 0.691 vs 0.692; Brier score .249 vs .249; AUC .54 vs .54; calibration slope 0.37 vs 0.36), and neither improved log-loss or Brier score over venue alone, so in this dataset the machine-learning step added complexity without improving discrimination or calibration. Expressed as probabilities, predicted probabilities of winning across the observed range of the indices varied between .40 and .67, with CIs spanning .18 to .86. Within categories, the intervals were much wider (upper limits up to 3.71), so the present data can neither confirm nor rule out moderate or large category-specific associations; they show only that the original category-specific findings were not supported once the analysis was corrected. Likewise, an AUC close to .5 indicates that the indices did not discriminate wins from non-wins in these data, not that readiness is irrelevant to performance.

Possible reasons why the framework did not work with these data

Three features of the data may explain much of the null result. The first is coverage. Training-session data were missing for most records (72% for the Fitness Index and 79% for the Freshness Index), so most index values used in the models were imputed, and no imputation method, including the post hoc multilevel method, recovered hidden Freshness values much better than their mean (normalised RMSE 0.92–1.03; Supplementary Table S3). Multiple imputation carries this uncertainty into wide CIs, whereas single imputation, as in the original analysis, understates it: single-imputation methods produced isolated category-specific terms excluding 1 that changed with the method.

 © 2027 ARD Asociación Española

The second is the expected-response model. The embedded framework assumes that external load predicts the internal or mechanical response, so that the residual isolates the part attributable to the player's state. For RPE this assumption failed for new players: external load explained none of the out-of-player variance, and about 57% of the variance of RPE across training sessions lay between players after accounting for category, season and session day (post hoc). In these youth players, RPE therefore behaved to a large extent as an individual scale—some players habitually reported higher values than others—and within players it was only moderately related to session load. A model that predicts RPE for players it has not seen cannot learn these individual levels. For PlayerLoad the assumption held, but because PlayerLoad depended strongly on distance, its residual reflects the part of accelerometer load not explained by GPS locomotor variables, which may include differences in movement type, device placement or measurement error and need not be related to neuromuscular freshness. Consistent with this, there was no evidence that either index correlated with wellness or with the CMJ change, although the intervals do not exclude moderate correlations. Out-of-sample validation of the expected-response model is therefore a prerequisite for interpreting any residual index (Impellizzeri et al., 2019).

The third is the reliability of the neuromuscular measure (post hoc). The technical error between the two trials of a CMJ test was 1.04 cm (3.2%), which gives a minimal detectable change of 2.89 cm (Hopkins, 2000). The average weekly change from MD−4 to MD−1 was M = 0.65 cm, SD = 2.31 cm, and only 23% of weekly changes exceeded the minimal detectable change. Within players the change varied by about 6.6% from week to week, and changes of consecutive weeks showed no positive correlation. Most weekly changes therefore did not exceed the minimal detectable change, although a mixed model of the jump heights estimated a small systematic increase from MD−4 to MD−1 (day effect = 0.67 cm, 95% CI [0.50, 0.85]), which may reflect the lower load before the match. This may explain why a CMJ-based Freshness Index (post hoc) could not be predicted from load and did not change the conclusions.

What would be needed

Testing the embedded framework in academies would require: (i) complete GPS and RPE collection for all players in the sessions used by the indices (here, MD-3 and MD-1); (ii) expected-response models that account for each player's own baseline—for example, player-specific intercepts or within-player standardisation of the response before modelling—validated out of sample before residuals are interpreted; (iii) neuromuscular measures whose error is small relative to the expected weekly change—for example, more trials or more frequent tests, with the technical error estimated in the same setting—so that changes can be distinguished from measurement noise (Hopkins, 2000); (iv) more matches per category, or a prospective multi-season design, given that the category models had 0.8–1.4 events per parameter; and (v) outcomes closer to the player than the team result, such as availability, coach ratings or individual match contribution. A prospective, pre-registered evaluation of discrimination and calibration, with explicit decision thresholds, would be needed before any index could inform training decisions; even then, an association would not show that modifying the load associated with an index increases the probability of winning.

A possible club-level procedure for an extreme residual, positive or negative, would be: (i) check the record, to rule out errors of measurement, load or player identification; (ii) review the player, discussing the signal with him and contrasting it with his usual evolution, wellness, soreness, recent load and training behaviour; and (iii) decide jointly, so that, if the full assessment suggests difficulties with recovery or tolerance, the staff could adjust volume, repetitions or rest, recording the decision and the subsequent response. The index alone would not determine the adjustment. This procedure could be evaluated in two stages: first, a prospective period observing without intervening on the index, with complete records, to evaluate whether the deviations identify relevant changes in recovery, training tolerance or availability, with thresholds defined

and validated before they are used for decisions; second, if that stage is favourable, a pilot comparing usual monitoring with usual monitoring plus an index-supported review protocol, randomised if feasible, with feasibility, availability and training tolerance as primary outcomes and match outcome as a secondary, exploratory outcome. This is a proposal for validation, not a procedure demonstrated in the source framework (Mandorino et al., 2025).

Developmental explanations as hypotheses

The original submission attributed category-specific patterns to maturation-related differences (Cumming et al., 2017; Lloyd & Oliver, 2012; Lloyd et al., 2014; Towlson & Cumming, 2022; Tumkur Anil Kumar et al., 2021). Biological maturation was not measured, chronological category is a poor proxy for maturity status, particularly around U14–U16 (Lloyd et al., 2014), and the corrected analysis found no differences between categories or developmental stages. These explanations are therefore presented only as hypotheses for future studies that include maturity offset or growth velocity (Mirwald et al., 2002): in U13–U14, high inter-player variability would be expected from differences in growth, maturation and training history; in U15–U16, trajectories would be expected to diverge, with some players still undergoing marked growth changes and others at a more advanced stage, so that individual growth status may explain part of the response better than chronological category; and in U18–U21, the relative influence of accelerated growth would be expected to be smaller for most players, with training history, competitive exposure and recovery becoming more relevant, while individual differences persist. Chronological categories cannot establish each player's maturity status (Bergeron et al., 2015).

Match-day running and match context

Match-day running is concurrent with the result: it is influenced by score evolution, tactics, the opponent and playing position, so its coefficient describes an association during the match, not a pre-match predictor. There was no evidence of an association in the aggregated model with match-clustered standard errors (*OR* = 1.24, 95% CI [0.95, 1.61], *p* = .110), although the interval was compatible with modest positive associations. In the mixed models, the credible interval of match-day running excluded 1 in the planned GLMM (*OR* = 1.28, 95% credible interval [1.07, 1.52]) and excluded 1 in the post hoc GLMM with an opponent random intercept (*OR* = 1.29, 95% credible interval [1.02, 1.64]); these models do not cluster by match and their variational approximation underestimates the variance, so their intervals are too narrow. The comparable model with opponent fixed effects and match-clustered standard errors gave *OR* = 1.27, 95% CI [0.92, 1.76]. In the venue-adjusted model, the estimate for playing at home (*OR* = 2.11, 95% CI [0.999, 4.46], *p* = .050) was compatible with the home advantage described in the literature (Pollard, 1986) but imprecise, with a lower limit at about 1, so it should not be read as a finding of this study. The only external-load variable in the outcome models was match-day running. Across all 19 aggregated estimates with multiple imputation (primary and sensitivity analyses, including the mixed models and the CMJ-based and alternative-imputation versions), its odds ratio ranged from 1.15 to 1.30, always above 1, and its interval included 1 whenever standard errors were clustered by match (it excluded 1 only in some mixed models). These data therefore neither support nor contradict the use of running-based indicators as performance metrics.

Limitations

Several limitations should be acknowledged. First, the study comes from one club over two partial seasons, with 140 matches and 20–25 matches per category. Second, coverage of training GPS data was low and multiple imputation cannot fully compensate: imputed values carry the assumptions of the imputation model, which was additive (it did not include the interaction terms, which tends to shrink them toward the null), and no method recovered hidden Freshness values much better than their mean; the post hoc analyses with other imputation methods led to the same conclusions. Third, the exclusions applied by the staff before data

 © 2027 ARD Asociación Española

export were not documented, so selection bias cannot be ruled out. Fourth, biological maturation was not measured, and direct physiological or biochemical markers such as creatine kinase, blood lactate concentration, or hormonal responses were not available. Fifth, contextual information was partial: opponent quality was recorded only for part of the 2025 matches and on a mixed scale (league positions and cup rounds), goals were recorded only in 2024, and possession and tactical variables were not available. Sixth, wellness was available only for 2024, and the reliability of the CMJ protocol also limits the interpretation of the CMJ change. Seventh, the outcome is a team-level result, which depends on the interaction of the team, the opponent and the match context, and does not show that an individual player was better prepared. Eighth, players were linked across seasons by an automatic, rule-based procedure on names and categories that was not confirmed by the club, and two codes that may each include two players with the same name could not be separated. The analysis with players identified by season gave the same conclusions for the indices, but outside the indices the change was enough for intervals to cross 1: the venue interval excluded 1 with players identified by season (*OR* = 2.14, 95% CI [1.01, 4.51]) but included 1 with linked players (*OR* = 2.11, 95% CI [0.999, 4.46]). The match-day running intervals of the two mixed models excluded 1 only with linked players (GLMM: *OR* = 1.17, 95% credible interval [0.99, 1.39], vs *OR* = 1.28, 95% credible interval [1.07, 1.52]; GLMM with opponent club: *OR* = 1.26, 95% credible interval [1.00, 1.58], vs *OR* = 1.29, 95% credible interval [1.02, 1.64]) (Supplementary Table S12). Finally, the observational design precludes causal inference.

Practical applications

Supported by the present data: (i) in this academy, the Fitness and Freshness indices, as implemented, should not be used to inform match-related decisions; (ii) before computing embedded indices, practitioners should check the coverage of the required sessions—here, the indices could be computed for only 27.7% and 21.3% of player–match records; (iii) expected-response models should be validated on players not used for training before their residuals are interpreted, and for RPE, which varied mostly between players, within-player baselines are needed; and (iv) most weekly CMJ changes based on the best of two trials did not exceed the minimal detectable change (only 23% exceeded 2.89 cm), so they are not sufficient on their own to track weekly changes in the athlete's state. Principles from the literature, not tested in this study: monitoring should consider individual responses and maturation status, and load progression should be gradual and individualised (Lloyd & Oliver, 2012); these are offered as hypotheses for decision support, not as recommendations derived from our data. An unexpected deviation could be a reason to talk with the player and review his situation with the staff, contrasting the signal with wellness, soreness, recent exposure and on-field observations, and individualising decisions by the player's own trajectory rather than applying the same response to an entire category; these applications, too, require future evaluation.

CONCLUSIONS

After the coding errors of the original analysis were corrected, there was no evidence that embedded readiness indices derived from routine monitoring (the Fitness and Freshness indices) or match-day running were associated with match outcome in youth football from U13 to U21, either overall or by category, in models that account for the clustering of players within matches, and no evidence of differences between developmental stages. For the three main effects, the aggregated 95% CIs excluded odds ratios above 1.61 or below 0.69 per within-player *SD* but not moderate associations, and these limits depend on the imputation model, because 72% and 79% of the Fitness and Freshness values were imputed. As collected in this academy (post hoc diagnostics), RPE varied to a large extent between players and could not be predicted from external load for players not used to train the model, PlayerLoad was largely explained by distance, and most weekly jump changes did not exceed the minimal detectable change (only 23% did). Embedded

readiness frameworks in youth football require complete monitoring data, player-specific baselines, reliable neuromuscular measures and out-of-sample validation before they can inform training or competition decisions.

AUTHOR CONTRIBUTIONS

All authors meet the criteria for authorship in accordance with established ethical guidelines. Contributions are specified according to the CRediT (Contributor Roles Taxonomy) as follows:

Conceptualisation: Rodrigo Villaseca-Vicuña. Methodology: Rodrigo Villaseca-Vicuña, Pablo Galaz. Formal analysis: Rodrigo Villaseca-Vicuña, Pablo Galaz. Investigation: Rodrigo Villaseca-Vicuña, Benjamín Villaseca-Vicuña, Pablo Merino-Muñoz, Jorge Perez-Contreras, David Faundez, Rodrigo Yáñez-Sepúlveda. Data curation: Rodrigo Villaseca-Vicuña, David Faundez. Writing – original draft: Rodrigo Villaseca-Vicuña. Writing – review & editing: Rodrigo Villaseca-Vicuña, Pablo Galaz, Benjamín Villaseca-Vicuña, Pablo Merino-Muñoz, Jorge Perez-Contreras, David Faundez, Rodrigo Yáñez-Sepúlveda, Santiago Zabaloy. Supervision: Rodrigo Villaseca-Vicuña, Santiago Zabaloy.

All authors have critically reviewed and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CONFLICT OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this manuscript.

ETHICAL APPROVAL AND CONSENT TO PARTICIPATE

No ethics committee review was sought, because the data were collected as part of the club's routine, non-invasive monitoring, without any procedure other than those of usual football training, and were analysed retrospectively after pseudonymisation. The Declaration of Helsinki (World Medical Association, 2013) was used as an ethical reference. Players aged 18 years or older signed a written informed consent; for players under 18, parents or legal guardians signed a written informed consent at the parents' meeting held at the start of each season, which covered the use of the monitoring data for research reported as aggregated statistics. Players under 18 were informed that they would take part in regular evaluations integrated into the club's routine training activities.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analysed during the current study are not publicly available due to the confidentiality policies of the professional football club involved and because they involve minors but are available from the corresponding author on reasonable request. The complete analysis code, with pinned

 © 2027 ARD Asociación Española

software environments that reproduce both the submitted and the revised analyses, is available at https://doi.org/10.5281/zenodo.23607423

AI USE DISCLOSURE

In accordance with current publishing ethics and transparency recommendations, artificial intelligence (AI) tools were used solely to assist with translation and language editing, with the aim of improving clarity and readability. No AI tools were used in the generation of scientific content, including the study design, data collection, analysis, interpretation of results, or the formulation of conclusions. The authors retain full responsibility for the content of the manuscript and confirm its originality, integrity, and accuracy.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to the professional football academy involved in this study for providing access to the data and supporting the development of this research within a high-performance environment. Special thanks are extended to the technical staff and the sport science department across all youth development categories for their collaboration, commitment, and continuous support throughout the monitoring and data collection processes. Their contribution was instrumental in enabling the implementation of an integrated monitoring system within a real-world applied setting, ensuring high ecological validity and facilitating the translation of scientific knowledge into practical decision-making in youth football development.

REFERENCES

Altman, D. G., & Bland, J. M. (1995). Absence of evidence is not evidence of absence. *BMJ*, *311*(7003), 485. https://doi.org/10.1136/bmj.311.7003.485

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated measures correlation. *Frontiers in Psychology*, *8*, Article 456. https://doi.org/10.3389/fpsyg.2017.00456

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

Bergeron, M. F., Mountjoy, M., Armstrong, N., Chia, M., Côté, J., Emery, C. A., Faigenbaum, A., Hall, G., Jr., Kriemler, S., Léglise, M., Malina, R. M., Pensgaard, A. M., Sánchez, A., Soligard, T., Sundgot-Borgen, J., van Mechelen, W., Weissensteiner, J. R., & Engebretsen, L. (2015). International Olympic Committee consensus statement on youth athletic development. *British Journal of Sports Medicine*, *49*(13), 843–851. https://doi.org/10.1136/bjsports-2015-094962

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Buchheit, M., & Mendez-Villanueva, A. (2013). Reliability and stability of anthropometric and performance measures in highly-trained young soccer players: Effect of age and maturation. *Journal of Sports Sciences*, *31*(12), 1332–1343. https://doi.org/10.1080/02640414.2013.781662

Buchheit, M., Mendez-Villanueva, A., Simpson, B. M., & Bourdon, P. C. (2010). Match running performance and fitness in youth soccer. *International Journal of Sports Medicine*, *31*(11), 818–825. https://doi.org/10.1055/s-0030-1262838

Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, *50*(2), 317–372. https://doi.org/10.3368/jhr.50.2.317

Carling, C. (2013). Interpreting physical performance in professional soccer match-play: Should we be more pragmatic in our approach? *Sports Medicine*, *43*(8), 655–663. https://doi.org/10.1007/s40279-013-0055-8

Cumming, S. P., Lloyd, R. S., Oliver, J. L., Eisenmann, J. C., & Malina, R. M. (2017). Bio-banding in sport: Applications to competition, talent identification, and strength and conditioning of youth athletes. *Strength and Conditioning Journal*, *39*(2), 34–47. https://doi.org/10.1519/SSC.0000000000000281

Foster, C. (1998). Monitoring training in athletes with reference to overtraining syndrome. *Medicine & Science in Sports & Exercise*, *30*(7), 1164–1168. https://doi.org/10.1097/00005768-199807000-00023

Gathercole, R., Sporer, B., Stellingwerff, T., & Sleivert, G. (2015). Alternative countermovement-jump analysis to quantify acute neuromuscular fatigue. *International Journal of Sports Physiology and Performance*, *10*(1), 84–92. https://doi.org/10.1123/ijspp.2013-0413

Gómez-Carmona, C. D., Bastida-Castillo, A., García-Rubio, J., Ibáñez, S. J., & Pino-Ortega, J. (2019). Static and dynamic reliability of WIMU PRO™ accelerometers according to anatomical placement. *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology*, *233*(2), 238–248. https://doi.org/10.1177/1754337118816922

Hopkins, W. G. (2000). Measures of reliability in sports medicine and science. *Sports Medicine*, *30*(1), 1–15. https://doi.org/10.2165/00007256-200030010-00001

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

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

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

Ju, W., Doran, D., Hawkins, R., Evans, M., Laws, A., & Bradley, P. (2023). Contextualised high-intensity running profiles of elite football players with reference to general and specialised tactical roles. *Biology of Sport*, *40*(1), 291–301. https://doi.org/10.5114/biolsport.2023.116003

Li, K. H., Raghunathan, T. E., & Rubin, D. B. (1991). Large-sample significance levels from multiply imputed data using moment-based statistics and an F reference distribution. *Journal of the American Statistical Association*, *86*(416), 1065–1073. https://doi.org/10.2307/2290525

Lloyd, R. S., & Oliver, J. L. (2012). The youth physical development model: A new approach to long-term athletic development. *Strength and Conditioning Journal*, *34*(3), 61–72. https://doi.org/10.1519/SSC.0b013e31825760ea

Lloyd, R. S., Oliver, J. L., Faigenbaum, A. D., Myer, G. D., & De Ste Croix, M. B. A. (2014). Chronological age vs. biological maturation. *Journal of Strength and Conditioning Research*, *28*(5), 1454–1464. https://doi.org/10.1519/JSC.0000000000000391

Mandorino, M., Clubb, J., & Lacome, M. (2024). Predicting soccer players' fitness status through a machine-learning approach. *International Journal of Sports Physiology and Performance*, *19*(5), 443–453. https://doi.org/10.1123/ijspp.2023-0444

Mandorino, M., Lacome, M., Verheijen, R., & Buchheit, M. (2025). *Time to drop running as a KPI in elite football: Football fitness and freshness as match-day preconditions* (Report No. 254, Version 1). Sport Performance & Science Reports. https://sportperfsci.com/time-to-drop-running-as-a-kpi-in-elite-football-football-fitness-and-freshness-as-match-day-preconditions/

McLean, B. D., Coutts, A. J., Kelly, V., McGuigan, M. R., & Cormack, S. J. (2010). Neuromuscular, endocrine, and perceptual fatigue responses during different length between-match microcycles in professional

 © 2027 ARD Asociación Española

rugby league players. *International Journal of Sports Physiology and Performance*, *5*(3), 367–383. https://doi.org/10.1123/ijspp.5.3.367

Mendez-Villanueva, A., Buchheit, M., Simpson, B., & Bourdon, P. C. (2013). Match play intensity distribution in youth soccer. *International Journal of Sports Medicine*, *34*(2), 101–110. https://doi.org/10.1055/s-0032-1306323

Mirwald, R. L., Baxter-Jones, A. D. G., Bailey, D. A., & Beunen, G. P. (2002). An assessment of maturity from anthropometric measurements. *Medicine & Science in Sports & Exercise*, *34*(4), 689–694. https://doi.org/10.1249/00005768-200204000-00020

Oliva-Lozano, J. M., Martínez-Puertas, H., Fortes, V., López-Del Campo, R., Resta, R., & Muyor, J. M. (2023). Is there any relationship between match running, technical-tactical performance, and team success in professional soccer? A longitudinal study in the first and second divisions of LaLiga. *Biology of Sport*, *40*(2), 587–594. https://doi.org/10.5114/biolsport.2023.118021

Oliva-Lozano, J. M., Rojas-Valverde, D., Gómez-Carmona, C. D., Fortes, V., & Pino-Ortega, J. (2020). Worst case scenario match analysis and contextual variables in professional soccer players: A longitudinal study. *Biology of Sport*, *37*(4), 429–436. https://doi.org/10.5114/biolsport.2020.97067

Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., & Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, *49*(12), 1373–1379. https://doi.org/10.1016/S0895-4356(96)00236-3

Pollard, R. (1986). Home advantage in soccer: A retrospective analysis. *Journal of Sports Sciences*, *4*(3), 237–248. https://doi.org/10.1080/02640418608732122

Rampinini, E., Impellizzeri, F. M., Castagna, C., Azzalin, A., Ferrari Bravo, D., & Wisløff, U. (2008). Effect of match-related fatigue on short-passing ability in young soccer players. *Medicine & Science in Sports & Exercise*, *40*(5), 934–942. https://doi.org/10.1249/MSS.0b013e3181666eb8

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Wiley. https://doi.org/10.1002/9780470316696

Settembre, M., Buchheit, M., Hader, K., Hamill, R., Tarascon, A., Verheijen, R., & McHugh, D. (2024). Factors associated with match outcomes in elite European football - insights from machine learning models. *Journal of Sports Analytics*, *10*(1), 1–16. https://doi.org/10.3233/JSA-240745

Towlson, C., & Cumming, S. P. (2022). Bio-banding in soccer: Past, present, and future. *Annals of Human Biology*, *49*(7–8), 269–273. https://doi.org/10.1080/03014460.2022.2129091

Tumkur Anil Kumar, N., Oliver, J. L., Lloyd, R. S., Pedley, J. S., & Radnor, J. M. (2021). The influence of growth, maturation and resistance training on muscle-tendon and neuromuscular adaptations: A narrative review. *Sports*, *9*(5), Article 59. https://doi.org/10.3390/sports9050059

van Buuren, S. (2018). *Flexible imputation of missing data* (2nd ed.). CRC Press. https://doi.org/10.1201/9780429492259

Villaseca-Vicuña, R., Barrera-Díaz, J., Otero-Saborido, F., Zabaloy, S., Palorames, E., & Sarmento, H. (2026). Key performance indicators in elite female football: Associations between sprints, duels, and match results. *International Journal of Sports Physiology and Performance*, *21*(3), 448–455. https://doi.org/10.1123/ijspp.2025-0408

Villaseca-Vicuña, R., Merino-Muñoz, P., Cortes-Rocco, G., Escobar, N., Lara, M. M., Sepúlveda, R. Y., Barrera-Díaz, J., & Pérez-Contreras, J. (2026). Differences in physical performance according to contextual variables in U21 football players. *Physiologia*, *6*(1), Article 8. https://doi.org/10.3390/physiologia6010008

Winter, E. M., & Maughan, R. J. (2009). Requirements for ethics approvals. *Journal of Sports Sciences*, *27*(10), 985. https://doi.org/10.1080/02640410903178344

World Medical Association. (2013). World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA*, *310*(20), 2191–2194. https://doi.org/10.1001/jama.2013.281053



This work is licensed under a Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0 DEED).

 © 2027 ARD Asociación Española

SUPPLEMENTARY MATERIAL

All tables are generated by code from the analysis outputs. Unless stated otherwise, models are the original model (Equation 4 of the manuscript) fitted to 873 eligible player–match records with multiple imputation (chained equations with predictive mean matching, $m = 50$), standard errors clustered by match and 95% confidence intervals (CI); odds ratios (*OR*) are per within-player *SD*. Whether an interval excludes 1 can be read from its lower (*LL*) and upper (*UL*) limits; limits that round to 1.00 are shown with three decimals. Analyses added after the first results had been seen are labelled post hoc. Players are linked across seasons (Methods). FI, Fitness Index; FR, Freshness Index; MD, match-day m·min^{-1}; n/e, not estimable; CrI, credible interval; EPV, events per parameter.

Effect of correcting the errors of the original analysis

Each step adds one correction to the previous one, so that the effect of each error can be seen. The seven terms are those reported as statistically significant in the original submission. S0 re-runs the submitted code with its original software environment. S1–S4 keep the submitted data (category-mean imputation); S5 and S6 use the rebuilt data without imputation (records with all variables observed), still with in-sample residuals and undefined within-player *z*-scores set to 0, as in the original code. S7 combines four changes in one step—players linked across seasons, out-of-player residuals, records with undefined within-player *z*-scores excluded instead of set to 0, and multiple imputation instead of complete cases—and is the revised primary analysis (Table 2 of the manuscript; rows are the median per imputed dataset). The intermediate step without imputation (S6 with out-of-player residuals and undefined *z*-scores excluded) is the complete-case analysis of Supplementary Table S14. All steps are expressed in the sign convention of the revised analysis (expected − observed for both indices, positive = favourable), which is that of the source framework and of the submitted code. The audit included a step S2 that recomputed the Fitness Index as observed − expected, as the original Equation 1.b was written; in the final convention that step leaves every estimate unchanged and is not shown. Where an earlier step computed an index as observed − expected (the Fitness Index in S3–S6 and the Freshness Index residual in S3–S6), the odds ratios of terms with an odd number of index factors were inverted (*OR* → 1/*OR*, CI → 1/upper–1/lower), which does not change whether the interval excludes 1. S0–S6 identify players by season; S7 links them across seasons.

Table S1. Cumulative effect of the corrections on the seven associations reported as significant in the original submission.

Step	CI level (%)	Records	Matches	Model	Term	*OR*	CI *LL*	CI *UL*	CI excludes 1	Terms excluding 1 at this step (of 7)
Published	90	1,569	141	All	FI	0.88	0.80	0.97	Yes	7
Published	90	1,569	141	All	MD	1.09	1.00	1.20	Yes	7
Published	90	1,569	141	U21	FI × FR × MD	0.70	0.51	0.96	Yes	7
Published	90	1,569	141	U18	FR × MD	1.47	1.13	1.92	Yes	7
Published	90	1,569	141	U14	FR × MD	0.71	0.52	0.96	Yes	7
Published	90	1,569	141	U14	FI × FR × MD	1.48	1.04	2.12	Yes	7
Published	90	1,569	141	U13	FR × MD	0.59	0.43	0.81	Yes	7
S0	90	1,569	141	All	FI	0.88	0.80	0.97	Yes	6
S0	90	1,569	141	All	MD	1.09	0.998	1.20	No	6
S0	90	1,569	141	U21	FI × FR × MD	0.70	0.51	0.96	Yes	6
S0	90	1,569	141	U18	FR × MD	1.47	1.13	1.92	Yes	6
S0	90	1,569	141	U14	FR × MD	0.71	0.52	0.96	Yes	6
S0	90	1,569	141	U14	FI × FR × MD	1.48	1.04	2.12	Yes	6
S0	90	1,569	141	U13	FR × MD	0.59	0.43	0.81	Yes	6
S1	90	1,569	141	All	FI	0.87	0.79	0.96	Yes	7
S1	90	1,569	141	All	MD	1.10	1.003	1.21	Yes	7
S1	90	1,569	141	U21	FI × FR × MD	0.71	0.52	0.97	Yes	7
S1	90	1,569	141	U18	FR × MD	1.60	1.20	2.14	Yes	7
S1	90	1,569	141	U14	FR × MD	0.72	0.53	0.98	Yes	7
S1	90	1,569	141	U14	FI × FR × MD	1.43	1.001	2.04	Yes	7
S1	90	1,569	141	U13	FR × MD	0.65	0.46	0.91	Yes	7
S3	90	1,569	141	All	FI	0.88	0.80	0.97	Yes	2
S3	90	1,569	141	All	MD	1.11	1.01	1.22	Yes	2
S3	90	1,569	141	U21	FI × FR × MD	1.02	0.76	1.37	No	2
S3	90	1,569	141	U18	FR × MD	1.00	0.78	1.28	No	2
S3	90	1,569	141	U14	FR × MD	1.24	0.91	1.68	No	2
S3	90	1,569	141	U14	FI × FR × MD	1.12	0.76	1.65	No	2
S3	90	1,569	141	U13	FR × MD	0.72	0.51	1.02	No	2
S4	95	1,569	141	All	FI	0.88	0.77	1.01	No	0
S4	95	1,569	141	All	MD	1.11	0.89	1.38	No	0
S4	95	1,569	141	U21	FI × FR × MD	1.02	0.80	1.32	No	0
S4	95	1,569	141	U18	FR × MD	1.00	0.84	1.19	No	0
S4	95	1,569	141	U14	FR × MD	1.24	0.90	1.70	No	0
S4	95	1,569	141	U14	FI × FR × MD	1.12	0.80	1.56	No	0
S4	95	1,569	141	U13	FR × MD	0.72	0.40	1.29	No	0
S5	95	54	24	All	FI	2.51	0.74	8.51	No	0
S5	95	54	24	All	MD	1.18	0.43	3.25	No	0

Step	CI level (%)	Records	Matches	Model	Term	*OR*	CI *LL*	CI *UL*	CI excludes 1	Terms excluding 1 at this step (of 7)
S5	95	54	24	U21	FI × FR × MD	unstable	—	—	—	0
S5	95	54	24	U18	FR × MD	unstable	—	—	—	0
S5	95	54	24	U14	FR × MD	n/e	—	—	—	0
S5	95	54	24	U14	FI × FR × MD	n/e	—	—	—	0
S5	95	54	24	U13	FR × MD	n/e	—	—	—	0
S6	95	53	23	All	FI	1.93	0.69	5.43	No	0
S6	95	53	23	All	MD	0.77	0.22	2.67	No	0
S6	95	53	23	U21	FI × FR × MD	no conv.	—	—	—	0
S6	95	53	23	U18	FR × MD	unstable	—	—	—	0
S6	95	53	23	U14	FR × MD	n/e	—	—	—	0
S6	95	53	23	U14	FI × FR × MD	n/e	—	—	—	0
S6	95	53	23	U13	FR × MD	n/e	—	—	—	0
S7	95	846	140	All	FI	0.90	0.69	1.16	No	0
S7	95	846	140	All	MD	1.24	0.95	1.61	No	0
S7	95	846	140	U21	FI × FR × MD	1.17	0.57	2.38	No	0
S7	95	846	140	U18	FR × MD	0.95	0.58	1.54	No	0
S7	95	846	140	U14	FR × MD	1.00	0.48	2.07	No	0
S7	95	846	140	U14	FI × FR × MD	1.00	0.35	2.87	No	0
S7	95	846	140	U13	FR × MD	0.93	0.44	1.97	No	0

Note. Steps: Published, values reported in the submitted manuscript; S0, Original analysis: submitted code re-run; S1, + intercept (A5); S3, + Freshness Index computed as a PlayerLoad residual (A1); S4, + match-clustered SE (A6); S5, + rebuilt data, observed values only (A3); S6, + ≥ 75-min criterion in 2025 (A7); S7, revised specification (players linked across seasons, out-of-player residuals, undefined z excluded, multiple imputation by predictive mean matching m = 50, match-clustered SE). n/e = not estimable (too few rows for the parameters); unstable = fitted with fewer than one event per parameter; no conv. = did not converge. 'Published values' reproduces the submitted manuscript, and its number of terms excluding 1 is as reported there; S0 reproduced these estimates (the aggregated match-day running term had a lower 90% limit just below 1). S7 values are read from the primary analysis and coincide with Table 2. A1, Freshness Index computed as expected PlayerLoad minus the percentage change in CMJ height; A2, Fitness Index computed as expected − observed (the source convention, kept in the revision) while Eq. 1.b and the text stated observed − expected, so the text interpreted it in the opposite direction; A3, category-mean imputation of most records; A5, logistic models without intercept; A6, standard errors ignoring the clustering of records within matches; A7, ≥ 75-min criterion not applied in 2025.

Expected-response models (random forest)

Table S2. Out-of-player performance of the expected-response models.

Response	Session	Inputs	Sessions	Players	R^2	RMSE	Calibration slope	R^2 per split *M*	R^2 per split *SD*
RPE	MD−3	Session duration, MM, HSR, ACC, DEC, Vmax, PL	525	105	−.04	1.79	0.33	−.07	.03
PlayerLoad	MD−1	TD, MM, HSR, ACC, DEC, Vmax	656	115	.65	11.57	0.98	.64	.01
CMJ change	MD−4 MD−1	Sums of TD, HSR, ACC, DEC and PL; mean MM; maximum Vmax	723	116	−.05	0.068	0.23	−.08	.01

Note. RPE (MD−3) is the response of the Fitness Index, PlayerLoad (MD−1) that of the Freshness Index, and the CMJ change that of the CMJ-based Freshness Index (post hoc). Random Forest with the hyperparameters of the original analysis (100 trees, other settings at the scikit-learn defaults, seed 42). Predictions out of player: 5-fold cross-validation grouped by player, repeated with 20 random splits whose predictions were averaged; R^2 per split is the R^2 of each single split (M, mean; SD, standard deviation). Inputs are from the same session; for the CMJ change, they are accumulated from MD−4 to MD−2, and sessions are player-weeks. RMSE in CR-10 units (RPE), arbitrary units (PlayerLoad) and proportion (CMJ change). Negative R^2 indicates predictions worse than the mean.

Imputation: Selection protocol and sensitivity to the method

Selection protocol (fixed before any outcome model was fitted): 20% of the observed values of each variable were hidden at random in 20 repetitions (m = 10 for multiple methods); the normalised RMSE (RMSE / SD of the observed values; 1 is the error of predicting the mean) and the ratio of imputed to true variance were averaged over repetitions. A method preserves the variance if the ratio is between 0.8 and 1.2 for the three variables. With players identified by season, the first application of the protocol selected chained equations with predictive mean matching, and a first complete run was made with it; after implementation errors in the candidate methods were corrected (unstandardised variables in the Bayesian regression, category entered as linear age among the auxiliary variables, and an error in the fallback of the player-mean method), the selection was repeated with the same rule and, because no method preserved the variance, the fallback rule chose chained equations with Bayesian regression. After players were linked across seasons, the selection was repeated once more: chained equations with predictive mean matching was the only method that preserved the variance of the three variables and had the lowest mean normalised RMSE, so it was chosen by the main rule. The table shows this final selection; chained equations with Bayesian regression is reported as a post hoc sensitivity analysis. The multilevel method was added post hoc and evaluated with the same protocol. Sensitivity of the results to the imputation method is shown in Supplementary Tables S14 and S15.

Table S3. Imputation methods evaluated on deliberately hidden observed values.

Method	Variable	Normalised RMSE	*SD*	Variance ratio	Mean normalised RMSE	Preserves variance
Category × season mean	Fitness Index	0.915	0.071	0.27	0.936	No
Category × season mean	Freshness Index	0.919	0.451	0.27	0.936	No
Category × season mean	Match-day m·min^{-1}	0.975	0.115	0.05	0.936	No

 © 2027 ARD Asociación Española

Method	Variable	Normalised RMSE	*SD*	Variance ratio	Mean normalised RMSE	Preserves variance
Player's own mean	Fitness Index	0.836	0.059	0.78	0.830	No
Player's own mean	Freshness Index	0.999	0.445	0.62	0.830	No
Player's own mean	Match-day m·min⁻¹	0.655	0.115	0.69	0.830	No
k-nearest neighbours (k = 10)	Fitness Index	0.896	0.072	0.20	0.820	No
k-nearest neighbours (k = 10)	Freshness Index	0.949	0.442	0.19	0.820	No
k-nearest neighbours (k = 10)	Match-day m·min⁻¹	0.616	0.050	0.63	0.820	No
Chained equations, predictive mean matching (primary)	Fitness Index	0.848	0.066	1.07	0.808	Yes
Chained equations, predictive mean matching (primary)	Freshness Index	0.978	0.424	0.85	0.808	Yes
Chained equations, predictive mean matching (primary)	Match-day m·min⁻¹	0.598	0.055	0.98	0.808	Yes
Chained equations, Bayesian regression (post hoc sensitivity)	Fitness Index	0.851	0.075	1.09	0.816	No
Chained equations, Bayesian regression (post hoc sensitivity)	Freshness Index	0.994	0.406	1.53	0.816	No
Chained equations, Bayesian regression (post hoc sensitivity)	Match-day m·min⁻¹	0.602	0.054	0.99	0.816	No
Multilevel chained equations (post hoc)	Fitness Index	0.851	—	1.20	0.825	No
Multilevel chained equations (post hoc)	Freshness Index	1.025	—	1.84	0.825	No
Multilevel chained equations (post hoc)	Match-day m·min⁻¹	0.599	—	1.00	0.825	No

Note. SD: standard deviation of the normalised RMSE across repetitions (multiple methods). Mean normalised RMSE: mean over the three variables. Preserves variance: variance ratio between 0.8 and 1.2 for the three variables.

Stability: Match bootstrap, model fit and main-effects model

Table S4. Original model with match-bootstrap intervals (2,000 replicates distributed across the imputed datasets).

Model	Term	*OR*	95% CI *LL*	95% CI *UL*	*p*	Bootstrap 95% CI *LL*	Bootstrap 95% CI *UL*	Same sign	Valid replicates	Total replicates
All	FI	0.90	0.69	1.16	.413	0.70	1.15	.80	2,000	2,000
All	FR	1.10	0.84	1.43	.479	0.83	1.41	.77	2,000	2,000
All	MD	1.24	0.95	1.61	.110	0.96	1.61	.95	2,000	2,000
All	FI × FR	1.02	0.80	1.30	.876	0.81	1.31	.53	2,000	2,000
All	FI × MD	0.98	0.77	1.26	.891	0.77	1.28	.58	2,000	2,000
All	FR × MD	0.96	0.76	1.20	.706	0.75	1.22	.65	2,000	2,000
All	FI × FR × MD	0.99	0.79	1.25	.935	0.78	1.28	.52	2,000	2,000
U21	FI	0.94	0.55	1.61	.818	0.50	1.64	.60	2,000	2,000
U21	FR	1.16	0.66	2.02	.602	0.59	2.17	.68	2,000	2,000
U21	MD	1.37	0.75	2.53	.308	0.69	3.02	.82	2,000	2,000
U21	FI × FR	1.01	0.60	1.72	.958	0.54	1.88	.50	2,000	2,000
U21	FI × MD	0.89	0.52	1.51	.653	0.43	1.62	.66	2,000	2,000
U21	FR × MD	1.08	0.62	1.87	.782	0.53	2.14	.63	2,000	2,000
U21	FI × FR × MD	1.17	0.57	2.38	.664	0.53	3.13	.67	2,000	2,000
U18	FI	0.85	0.51	1.42	.530	0.49	1.59	.69	2,000	2,000
U18	FR	1.21	0.68	2.16	.513	0.65	2.45	.73	2,000	2,000
U18	MD	1.83	0.93	3.63	.082	0.87	4.59	.95	2,000	2,000
U18	FI × FR	1.04	0.53	2.05	.911	0.46	2.27	.56	2,000	2,000
U18	FI × MD	0.96	0.55	1.69	.894	0.46	1.87	.58	2,000	2,000
U18	FR × MD	0.95	0.58	1.54	.827	0.51	2.02	.53	2,000	2,000
U18	FI × FR × MD	0.88	0.53	1.43	.594	0.37	1.84	.68	2,000	2,000
U16	FI	0.77	0.41	1.42	.396	0.32	1.45	.81	2,000	2,000
U16	FR	1.01	0.52	1.95	.976	0.46	2.20	.50	2,000	2,000
U16	MD	1.25	0.61	2.57	.542	0.58	3.39	.73	2,000	2,000
U16	FI × FR	0.94	0.48	1.85	.856	0.39	2.44	.56	2,000	2,000
U16	FI × MD	1.07	0.54	2.15	.839	0.49	2.99	.59	2,000	2,000
U16	FR × MD	0.94	0.45	1.99	.879	0.38	2.62	.53	2,000	2,000
U16	FI × FR × MD	1.03	0.47	2.26	.931	0.37	3.25	.52	2,000	2,000
U15	FI	0.83	0.40	1.74	.628	0.36	2.23	.71	1,995	2,000
U15	FR	1.09	0.52	2.29	.812	0.40	2.77	.56	1,995	2,000
U15	MD	1.44	0.56	3.71	.449	0.47	5.63	.78	1,995	2,000
U15	FI × FR	1.02	0.50	2.12	.947	0.39	2.84	.52	1,995	2,000
U15	FI × MD	0.98	0.41	2.36	.970	0.32	3.02	.50	1,995	2,000
U15	FR × MD	0.88	0.37	2.10	.766	0.20	2.22	.64	1,995	2,000
U15	FI × FR × MD	1.05	0.41	2.68	.920	0.29	4.13	.54	1,995	2,000

Model	Term	*OR*	95% CI *LL*	95% CI *UL*	*p*	Bootstrap 95% CI *LL*	Bootstrap 95% CI *UL*	Same sign	Valid replicates	Total replicates
U14	FI	0.90	0.47	1.72	.741	0.41	1.82	.62	2,000	2,000
U14	FR	0.97	0.49	1.91	.934	0.45	2.14	.56	2,000	2,000
U14	MD	1.01	0.51	2.00	.970	0.47	2.59	.55	2,000	2,000
U14	FI × FR	0.99	0.44	2.23	.981	0.33	2.52	.50	2,000	2,000
U14	FI × MD	1.00	0.49	2.07	.989	0.39	2.26	.51	2,000	2,000
U14	FR × MD	1.00	0.48	2.07	.992	0.34	2.43	.50	2,000	2,000
U14	FI × FR × MD	1.00	0.35	2.87	.999	0.22	4.13	.53	2,000	2,000
U13	FI	0.77	0.39	1.49	.429	0.31	1.63	.77	2,000	2,000
U13	FR	1.11	0.56	2.20	.760	0.54	2.61	.61	2,000	2,000
U13	MD	0.89	0.46	1.70	.714	0.36	1.92	.63	2,000	2,000
U13	FI × FR	0.93	0.44	1.96	.839	0.35	2.40	.59	2,000	2,000
U13	FI × MD	1.00	0.46	2.20	.998	0.34	2.58	.52	2,000	2,000
U13	FR × MD	0.93	0.44	1.97	.846	0.32	2.25	.56	2,000	2,000
U13	FI × FR × MD	1.05	0.46	2.40	.912	0.32	3.40	.55	2,000	2,000

Note. Same sign: proportion of bootstrap replicates with the same sign as the pooled estimate. Valid replicates exclude fits that did not converge or had extreme coefficients. Model fit and the main-effects model are shown in Supplementary Tables S16 and S17.

Generalised linear mixed models

Table S5. Logistic GLMM (aggregated data; variational bayes): posterior mean OR and approximate 95% credible interval.

Specification	Parameter	Estimate	95% CrI *LL*	95% CrI *UL*	Converged fits	Total fits
GLMM: player + category (planned)	Fitness Index (FI), OR	0.87	0.67	1.13	50	50
GLMM: player + category (planned)	Freshness Index (FR), OR	1.11	0.86	1.44	50	50
GLMM: player + category (planned)	Match-day m·min⁻¹ (MD), OR	1.28	1.07	1.52	50	50
GLMM: player + category (planned)	FI × FR, OR	1.02	0.80	1.30	50	50
GLMM: player + category (planned)	FI × MD, OR	0.98	0.76	1.26	50	50
GLMM: player + category (planned)	FR × MD, OR	0.96	0.76	1.21	50	50
GLMM: player + category (planned)	FI × FR × MD, OR	0.99	0.78	1.26	50	50
GLMM: player + category (planned)	SD of random intercept: player	0.17	—	—	50	50
GLMM: player + category (planned)	SD of random intercept: category	0.40	—	—	50	50
+ opponent club (post hoc)	Fitness Index (FI), OR	0.88	0.62	1.24	50	50
+ opponent club (post hoc)	Freshness Index (FR), OR	1.13	0.81	1.56	50	50
+ opponent club (post hoc)	Match-day m·min⁻¹ (MD), OR	1.29	1.02	1.64	50	50
+ opponent club (post hoc)	FI × FR, OR	1.02	0.73	1.43	50	50
+ opponent club (post hoc)	FI × MD, OR	0.97	0.72	1.32	50	50
+ opponent club (post hoc)	FR × MD, OR	1.00	0.74	1.36	50	50
+ opponent club (post hoc)	FI × FR × MD, OR	0.97	0.70	1.34	50	50
+ opponent club (post hoc)	SD of random intercept: player	0.22	—	—	50	50
+ opponent club (post hoc)	SD of random intercept: category	0.82	—	—	50	50
+ opponent club (post hoc)	SD of random intercept: opponent (club or club × category)	3.22	—	—	50	50
+ club + club × category (post hoc)	Fitness Index (FI), OR	0.76	0.44	1.30	50	50
+ club + club × category (post hoc)	Freshness Index (FR), OR	1.12	0.70	1.81	50	50
+ club + club × category (post hoc)	Match-day m·min⁻¹ (MD), OR	1.30	0.91	1.87	50	50
+ club + club × category (post hoc)	FI × FR, OR	1.08	0.60	1.92	50	50
+ club + club × category (post hoc)	FI × MD, OR	0.89	0.51	1.56	50	50
+ club + club × category (post hoc)	FR × MD, OR	1.01	0.65	1.57	50	50
+ club + club × category (post hoc)	FI × FR × MD, OR	0.93	0.56	1.54	50	50
+ club + club × category (post hoc)	SD of random intercept: player	0.59	—	—	50	50
+ club + club × category (post hoc)	SD of random intercept: category	0.24	—	—	50	50
+ club + club × category (post hoc)	SD of random intercept: opponent (club or club × category)	5.77	—	—	50	50
+ club + club × category (post hoc)	SD of random intercept: club × category (nested)	4.44	—	—	50	50
Opponent × category (post hoc; not identifiable)	Fitness Index (FI), OR	0.74	0.42	1.31	50	50
Opponent × category (post hoc; not identifiable)	Freshness Index (FR), OR	1.12	0.68	1.85	50	50
Opponent × category (post hoc; not identifiable)	Match-day m·min⁻¹ (MD), OR	1.28	0.88	1.87	50	50
Opponent × category (post hoc; not identifiable)	FI × FR, OR	1.09	0.58	2.03	50	50

 © 2027 ARD Asociación Española

Specification	Parameter	Estimate	95% CrI *LL*	95% CrI *UL*	Converged fits	Total fits
Opponent × category (post hoc; not identifiable)	FI × MD, OR	0.87	0.48	1.61	50	50
Opponent × category (post hoc; not identifiable)	FR × MD, OR	1.02	0.64	1.61	50	50
Opponent × category (post hoc; not identifiable)	FI × FR × MD, OR	0.92	0.54	1.59	50	50
Opponent × category (post hoc; not identifiable)	SD of random intercept: player	0.71	—	—	50	50
Opponent × category (post hoc; not identifiable)	SD of random intercept: category	0.24	—	—	50	50
Opponent × category (post hoc; not identifiable)	SD of random intercept: opponent (club or club × category)	8.65	—	—	50	50

Note. Estimates are posterior mean ORs for the fixed effects and SDs for the random intercepts. The GLMM cannot include a random intercept for match, because the outcome is constant within a match; it therefore does not account for the clustering of records within matches, and variational Bayes underestimates posterior variance, so its intervals are too narrow. The comparable inference for the opponent adjustment is the fixed-effects model with match-clustered standard errors (Supplementary Table S18). Opponent club: 24 clubs (median 4.0 matches per club). Opponent × category: 92 levels, 60 with a single match, so this random intercept is practically a match intercept; that specification is not interpretable and is shown only for completeness.

Adjustment for venue, playing position and opponent

Table S6. Original model adjusted for match venue (planned) and for venue and playing position (post hoc).

Adjustment	Model	Term	*OR*	95% CI *LL*	95% CI *UL*	Records	Matches	Won	Not won	EPV
Venue (planned)	All	Fitness Index (FI)	0.87	0.67	1.14	846	140	84	56	6.2
Venue (planned)	All	Freshness Index (FR)	1.07	0.82	1.40	846	140	84	56	6.2
Venue (planned)	All	Match-day m·min^{-1} (MD)	1.18	0.91	1.53	846	140	84	56	6.2
Venue (planned)	All	FI × FR	1.02	0.80	1.30	846	140	84	56	6.2
Venue (planned)	All	FI × MD	0.97	0.76	1.24	846	140	84	56	6.2
Venue (planned)	All	FR × MD	0.96	0.76	1.21	846	140	84	56	6.2
Venue (planned)	All	FI × FR × MD	1.00	0.79	1.26	846	140	84	56	6.2
Venue (planned)	All	Home venue	2.11	0.999	4.46	846	140	84	56	6.2
Venue (planned)	U21	Fitness Index (FI)	0.86	0.48	1.53	171	24	13	11	1.2
Venue (planned)	U21	Freshness Index (FR)	1.15	0.64	2.07	171	24	13	11	1.2
Venue (planned)	U21	Match-day m·min^{-1} (MD)	1.19	0.71	2.01	171	24	13	11	1.2
Venue (planned)	U21	FI × FR	0.99	0.56	1.77	171	24	13	11	1.2
Venue (planned)	U21	FI × MD	0.89	0.51	1.56	171	24	13	11	1.2
Venue (planned)	U21	FR × MD	1.12	0.62	2.03	171	24	13	11	1.2
Venue (planned)	U21	FI × FR × MD	1.15	0.55	2.42	171	24	13	11	1.2
Venue (planned)	U21	Home venue	3.93	0.62	24.82	171	24	13	11	1.2
Venue (planned)	U18	Fitness Index (FI)	0.83	0.48	1.43	162	24	16	8	0.9
Venue (planned)	U18	Freshness Index (FR)	1.19	0.67	2.13	162	24	16	8	0.9
Venue (planned)	U18	Match-day m·min^{-1} (MD)	1.72	0.90	3.28	162	24	16	8	0.9
Venue (planned)	U18	FI × FR	1.04	0.53	2.06	162	24	16	8	0.9
Venue (planned)	U18	FI × MD	0.94	0.55	1.63	162	24	16	8	0.9
Venue (planned)	U18	FR × MD	0.95	0.59	1.53	162	24	16	8	0.9
Venue (planned)	U18	FI × FR × MD	0.89	0.55	1.45	162	24	16	8	0.9
Venue (planned)	U18	Home venue	1.71	0.25	11.64	162	24	16	8	0.9
Venue (planned)	U16	Fitness Index (FI)	0.76	0.41	1.42	134	23	11	12	1.2
Venue (planned)	U16	Freshness Index (FR)	1.01	0.53	1.93	134	23	11	12	1.2
Venue (planned)	U16	Match-day m·min^{-1} (MD)	1.25	0.61	2.56	134	23	11	12	1.2
Venue (planned)	U16	FI × FR	0.94	0.48	1.84	134	23	11	12	1.2
Venue (planned)	U16	FI × MD	1.07	0.54	2.14	134	23	11	12	1.2
Venue (planned)	U16	FR × MD	0.94	0.45	1.99	134	23	11	12	1.2
Venue (planned)	U16	FI × FR × MD	1.04	0.47	2.27	134	23	11	12	1.2
Venue (planned)	U16	Home venue	1.00	0.15	6.47	134	23	11	12	1.2
Venue (planned)	U15	Fitness Index (FI)	0.88	0.42	1.84	121	20	14	6	0.7
Venue (planned)	U15	Freshness Index (FR)	1.12	0.53	2.38	121	20	14	6	0.7
Venue (planned)	U15	Match-day m·min^{-1} (MD)	1.59	0.65	3.88	121	20	14	6	0.7
Venue (planned)	U15	FI × FR	1.02	0.49	2.10	121	20	14	6	0.7
Venue (planned)	U15	FI × MD	0.99	0.40	2.41	121	20	14	6	0.7
Venue (planned)	U15	FR × MD	0.89	0.37	2.16	121	20	14	6	0.7
Venue (planned)	U15	FI × FR × MD	1.07	0.42	2.72	121	20	14	6	0.7
Venue (planned)	U15	Home venue	0.43	0.05	3.72	121	20	14	6	0.7
Venue (planned)	U14	Fitness Index (FI)	0.88	0.45	1.71	128	25	14	11	1.2
Venue (planned)	U14	Freshness Index (FR)	0.91	0.46	1.80	128	25	14	11	1.2

Adjustment	Model	Term	*OR*	95% CI *LL*	95% CI *UL*	Records	Matches	Won	Not won	EPV
Venue (planned)	U14	Match-day m·min⁻¹ (MD)	0.96	0.48	1.90	128	25	14	11	1.2
Venue (planned)	U14	FI × FR	1.00	0.44	2.29	128	25	14	11	1.2
Venue (planned)	U14	FI × MD	0.99	0.47	2.09	128	25	14	11	1.2
Venue (planned)	U14	FR × MD	0.97	0.45	2.09	128	25	14	11	1.2
Venue (planned)	U14	FI × FR × MD	1.00	0.31	3.16	128	25	14	11	1.2
Venue (planned)	U14	Home venue	3.31	0.47	23.18	128	25	14	11	1.2
Venue (planned)	U13	Fitness Index (FI)	0.67	0.32	1.40	130	24	16	8	0.9
Venue (planned)	U13	Freshness Index (FR)	1.07	0.53	2.13	130	24	16	8	0.9
Venue (planned)	U13	Match-day m·min⁻¹ (MD)	0.81	0.40	1.66	130	24	16	8	0.9
Venue (planned)	U13	FI × FR	0.93	0.40	2.18	130	24	16	8	0.9
Venue (planned)	U13	FI × MD	0.96	0.42	2.19	130	24	16	8	0.9
Venue (planned)	U13	FR × MD	0.92	0.41	2.07	130	24	16	8	0.9
Venue (planned)	U13	FI × FR × MD	1.08	0.44	2.65	130	24	16	8	0.9
Venue (planned)	U13	Home venue	8.41	0.81	87.23	130	24	16	8	0.9
Venue and playing position (post hoc)	All	Fitness Index (FI)	0.87	0.66	1.14	846	140	84	56	4.0
Venue and playing position (post hoc)	All	Freshness Index (FR)	1.07	0.82	1.40	846	140	84	56	4.0
Venue and playing position (post hoc)	All	Match-day m·min⁻¹ (MD)	1.18	0.91	1.53	846	140	84	56	4.0
Venue and playing position (post hoc)	All	FI × FR	1.02	0.79	1.31	846	140	84	56	4.0
Venue and playing position (post hoc)	All	FI × MD	0.97	0.75	1.24	846	140	84	56	4.0
Venue and playing position (post hoc)	All	FR × MD	0.96	0.76	1.21	846	140	84	56	4.0
Venue and playing position (post hoc)	All	FI × FR × MD	1.00	0.78	1.26	846	140	84	56	4.0
Venue and playing position (post hoc)	All	Home venue	2.10	0.99	4.45	846	140	84	56	4.0
Venue and playing position (post hoc)	U21	Fitness Index (FI)	0.85	0.47	1.54	171	24	13	11	0.8
Venue and playing position (post hoc)	U21	Freshness Index (FR)	1.15	0.63	2.08	171	24	13	11	0.8
Venue and playing position (post hoc)	U21	Match-day m·min⁻¹ (MD)	1.20	0.71	2.03	171	24	13	11	0.8
Venue and playing position (post hoc)	U21	FI × FR	0.99	0.55	1.80	171	24	13	11	0.8
Venue and playing position (post hoc)	U21	FI × MD	0.88	0.49	1.58	171	24	13	11	0.8
Venue and playing position (post hoc)	U21	FR × MD	1.13	0.61	2.09	171	24	13	11	0.8
Venue and playing position (post hoc)	U21	FI × FR × MD	1.15	0.53	2.49	171	24	13	11	0.8
Venue and playing position (post hoc)	U21	Home venue	4.00	0.62	25.91	171	24	13	11	0.8
Venue and playing position (post hoc)	U18	Fitness Index (FI)	0.83	0.47	1.46	162	24	16	8	0.6
Venue and playing position (post hoc)	U18	Freshness Index (FR)	1.21	0.67	2.20	162	24	16	8	0.6
Venue and playing position (post hoc)	U18	Match-day m·min⁻¹ (MD)	1.72	0.87	3.40	162	24	16	8	0.6
Venue and playing position (post hoc)	U18	FI × FR	1.04	0.51	2.13	162	24	16	8	0.6
Venue and playing position (post hoc)	U18	FI × MD	0.96	0.54	1.69	162	24	16	8	0.6
Venue and playing position (post hoc)	U18	FR × MD	0.97	0.59	1.59	162	24	16	8	0.6
Venue and playing position (post hoc)	U18	FI × FR × MD	0.86	0.51	1.45	162	24	16	8	0.6
Venue and playing position (post hoc)	U18	Home venue	1.77	0.25	12.43	162	24	16	8	0.6
Venue and playing position (post hoc)	U16	Fitness Index (FI)	0.76	0.39	1.45	134	23	11	12	0.8
Venue and playing position (post hoc)	U16	Freshness Index (FR)	1.02	0.52	1.99	134	23	11	12	0.8
Venue and playing position (post hoc)	U16	Match-day m·min⁻¹ (MD)	1.26	0.59	2.71	134	23	11	12	0.8
Venue and playing position (post hoc)	U16	FI × FR	0.93	0.45	1.95	134	23	11	12	0.8
Venue and playing position (post hoc)	U16	FI × MD	1.08	0.52	2.27	134	23	11	12	0.8
Venue and playing position (post hoc)	U16	FR × MD	0.94	0.43	2.08	134	23	11	12	0.8
Venue and playing position (post hoc)	U16	FI × FR × MD	1.04	0.45	2.37	134	23	11	12	0.8
Venue and playing position (post hoc)	U16	Home venue	0.98	0.14	6.67	134	23	11	12	0.8
Venue and playing position (post hoc)	U15	Fitness Index (FI)	0.87	0.40	1.89	121	20	14	6	0.4
Venue and playing position (post hoc)	U15	Freshness Index (FR)	1.13	0.51	2.53	121	20	14	6	0.4
Venue and playing position (post hoc)	U15	Match-day m·min⁻¹ (MD)	1.60	0.63	4.05	121	20	14	6	0.4
Venue and playing position (post hoc)	U15	FI × FR	1.01	0.47	2.17	121	20	14	6	0.4
Venue and playing position (post hoc)	U15	FI × MD	0.98	0.38	2.53	121	20	14	6	0.4
Venue and playing position (post hoc)	U15	FR × MD	0.86	0.33	2.27	121	20	14	6	0.4
Venue and playing position (post hoc)	U15	FI × FR × MD	1.07	0.39	2.91	121	20	14	6	0.4
Venue and playing position (post hoc)	U15	Home venue	0.41	0.05	3.69	121	20	14	6	0.4
Venue and playing position (post hoc)	U14	Fitness Index (FI)	0.86	0.43	1.72	128	25	14	11	0.8
Venue and playing position (post hoc)	U14	Freshness Index (FR)	0.92	0.45	1.85	128	25	14	11	0.8
Venue and playing position (post hoc)	U14	Match-day m·min⁻¹ (MD)	0.97	0.47	1.97	128	25	14	11	0.8

© 2027 ARD Asociación Española

Adjustment	Model	Term	*OR*	95% CI *LL*	95% CI *UL*	Records	Matches	Won	Not won	EPV
Venue and playing position (post hoc)	U14	FI × FR	1.00	0.41	2.42	128	25	14	11	0.8
Venue and playing position (post hoc)	U14	FI × MD	1.01	0.45	2.28	128	25	14	11	0.8
Venue and playing position (post hoc)	U14	FR × MD	0.97	0.43	2.19	128	25	14	11	0.8
Venue and playing position (post hoc)	U14	FI × FR × MD	1.00	0.29	3.38	128	25	14	11	0.8
Venue and playing position (post hoc)	U14	Home venue	3.48	0.47	25.51	128	25	14	11	0.8
Venue and playing position (post hoc)	U13	Fitness Index (FI)	0.67	0.30	1.51	130	24	16	8	0.6
Venue and playing position (post hoc)	U13	Freshness Index (FR)	1.08	0.52	2.26	130	24	16	8	0.6
Venue and playing position (post hoc)	U13	Match-day m·min^{-1} (MD)	0.78	0.37	1.65	130	24	16	8	0.6
Venue and playing position (post hoc)	U13	FI × FR	0.93	0.36	2.37	130	24	16	8	0.6
Venue and playing position (post hoc)	U13	FI × MD	0.92	0.38	2.18	130	24	16	8	0.6
Venue and playing position (post hoc)	U13	FR × MD	0.96	0.40	2.30	130	24	16	8	0.6
Venue and playing position (post hoc)	U13	FI × FR × MD	1.08	0.40	2.91	130	24	16	8	0.6
Venue and playing position (post hoc)	U13	Home venue	10.49	0.96	114.38	130	24	16	8	0.6

Note. Reference position: midfielder. Position coefficients are not shown. Category models have up to 14 parameters for 20–25 matches (events per parameter ≤ 0.8), so their clustered standard errors are unreliable. Records: median across imputed datasets; Won and Not won: matches; EPV, events per parameter. The adjustment for the opponent club is shown in Supplementary Table S18.

Developmental Stages and Groupings of Categories (Post Hoc)

Table S7. Original model by developmental stage (used in the original conclusions), by training structure (four vs. five sessions per week) and for U13–U15 vs. U16–U21

Grouping	Group	Term	*OR*	95% CI *LL*	95% CI *UL*	Records	Matches	Won	Not won	EPV
Developmental stage	U13–U14	Fitness Index (FI)	0.86	0.56	1.34	258	49	30	19	2.4
Developmental stage	U13–U14	Freshness Index (FR)	1.02	0.66	1.59	258	49	30	19	2.4
Developmental stage	U13–U14	Match-day m·min^{-1} (MD)	0.94	0.60	1.46	258	49	30	19	2.4
Developmental stage	U13–U14	FI × FR	0.99	0.61	1.62	258	49	30	19	2.4
Developmental stage	U13–U14	FI × MD	1.00	0.62	1.61	258	49	30	19	2.4
Developmental stage	U13–U14	FR × MD	0.97	0.63	1.51	258	49	30	19	2.4
Developmental stage	U13–U14	FI × FR × MD	1.02	0.61	1.70	258	49	30	19	2.4
Developmental stage	U15–U16	Fitness Index (FI)	0.90	0.60	1.35	255	43	25	18	2.3
Developmental stage	U15–U16	Freshness Index (FR)	1.02	0.66	1.59	255	43	25	18	2.3
Developmental stage	U15–U16	Match-day m·min^{-1} (MD)	1.25	0.76	2.04	255	43	25	18	2.3
Developmental stage	U15–U16	FI × FR	1.01	0.67	1.53	255	43	25	18	2.3
Developmental stage	U15–U16	FI × MD	1.05	0.67	1.63	255	43	25	18	2.3
Developmental stage	U15–U16	FR × MD	0.90	0.56	1.46	255	43	25	18	2.3
Developmental stage	U15–U16	FI × FR × MD	1.01	0.64	1.59	255	43	25	18	2.3
Developmental stage	U18–U21	Fitness Index (FI)	0.90	0.61	1.32	333	48	29	19	2.4
Developmental stage	U18–U21	Freshness Index (FR)	1.21	0.83	1.77	333	48	29	19	2.4
Developmental stage	U18–U21	Match-day m·min^{-1} (MD)	1.53	0.98	2.39	333	48	29	19	2.4
Developmental stage	U18–U21	FI × FR	1.05	0.69	1.58	333	48	29	19	2.4
Developmental stage	U18–U21	FI × MD	0.91	0.63	1.32	333	48	29	19	2.4
Developmental stage	U18–U21	FR × MD	0.99	0.70	1.41	333	48	29	19	2.4
Developmental stage	U18–U21	FI × FR × MD	0.96	0.68	1.36	333	48	29	19	2.4
Training structure	U13–U16	Fitness Index (FI)	0.88	0.65	1.20	513	92	55	37	4.6
Training structure	U13–U16	Freshness Index (FR)	1.03	0.75	1.40	513	92	55	37	4.6
Training structure	U13–U16	Match-day m·min^{-1} (MD)	1.08	0.78	1.49	513	92	55	37	4.6
Training structure	U13–U16	FI × FR	1.00	0.74	1.34	513	92	55	37	4.6
Training structure	U13–U16	FI × MD	1.03	0.74	1.42	513	92	55	37	4.6
Training structure	U13–U16	FR × MD	0.93	0.68	1.27	513	92	55	37	4.6
Training structure	U13–U16	FI × FR × MD	1.01	0.73	1.41	513	92	55	37	4.6
Training structure	U18–U21	Fitness Index (FI)	0.90	0.61	1.32	333	48	29	19	2.4
Training structure	U18–U21	Freshness Index (FR)	1.21	0.83	1.77	333	48	29	19	2.4
Training structure	U18–U21	Match-day m·min^{-1} (MD)	1.53	0.98	2.39	333	48	29	19	2.4
Training structure	U18–U21	FI × FR	1.05	0.69	1.58	333	48	29	19	2.4
Training structure	U18–U21	FI × MD	0.91	0.63	1.32	333	48	29	19	2.4
Training structure	U18–U21	FR × MD	0.99	0.70	1.41	333	48	29	19	2.4
Training structure	U18–U21	FI × FR × MD	0.96	0.68	1.36	333	48	29	19	2.4
U13–U15 vs. U16–U21	U13–U15	Fitness Index (FI)	0.89	0.62	1.28	379	69	44	25	3.1
U13–U15 vs. U16–U21	U13–U15	Freshness Index (FR)	1.03	0.71	1.49	379	69	44	25	3.1

Grouping	Group	Term	*OR*	95% CI *LL*	95% CI *UL*	Records	Matches	Won	Not won	EPV
U13–U15 vs. U16–U21	U13–U15	Match-day m·min⁻¹ (MD)	1.07	0.72	1.58	379	69	44	25	3.1
U13–U15 vs. U16–U21	U13–U15	FI × FR	1.00	0.70	1.44	379	69	44	25	3.1
U13–U15 vs. U16–U21	U13–U15	FI × MD	1.01	0.68	1.51	379	69	44	25	3.1
U13–U15 vs. U16–U21	U13–U15	FR × MD	0.92	0.64	1.33	379	69	44	25	3.1
U13–U15 vs. U16–U21	U13–U15	FI × FR × MD	1.03	0.67	1.58	379	69	44	25	3.1
U13–U15 vs. U16–U21	U16–U21	Fitness Index (FI)	0.89	0.65	1.21	467	71	40	31	3.9
U13–U15 vs. U16–U21	U16–U21	Freshness Index (FR)	1.16	0.82	1.63	467	71	40	31	3.9
U13–U15 vs. U16–U21	U16–U21	Match-day m·min⁻¹ (MD)	1.40	0.98	2.01	467	71	40	31	3.9
U13–U15 vs. U16–U21	U16–U21	FI × FR	1.04	0.75	1.45	467	71	40	31	3.9
U13–U15 vs. U16–U21	U16–U21	FI × MD	0.98	0.72	1.33	467	71	40	31	3.9
U13–U15 vs. U16–U21	U16–U21	FR × MD	0.97	0.72	1.33	467	71	40	31	3.9
U13–U15 vs. U16–U21	U16–U21	FI × FR × MD	0.97	0.73	1.29	467	71	40	31	3.9

Note. Developmental stages: joint test of the 6 stage × index interactions, p = .87; with 21 terms, about one interval excluding 1 would be expected by chance. Training structure: joint test of group × index interactions, p = .63; main-effects model, 0 of 6 terms excluding 1. U13–U15 vs. U16–U21: joint test of group × index interactions, p = .78; main-effects model, 0 of 6 terms excluding 1. Records: median across imputed datasets; Won and Not won: matches; EPV, events per parameter. The main-effects model by developmental stage is shown in Supplementary Table S19.

Freshness INDEX COMPUTED FROM THE COUNTERMOVEMENT JUMP (Post Hoc)

The Freshness Index was recomputed as the observed minus the expected change in CMJ height from MD−4 to MD−1, with the external load accumulated from MD−4 to MD−2 as input (Supplementary Table S2). It could be computed for 325 of 873 records (37.2%; 139 records with the three variables observed); imputation used the same method as the primary analysis. Convergent validity with the CMJ is not reported for this index because it would be circular.

Table S8. Original model with the CMJ-based freshness index.

Model	Term	*OR*	95% CI *LL*	95% CI *UL*	Records	Matches	Won	Not won	EPV
All	Fitness Index (FI)	0.88	0.67	1.17	846	140	84	56	7.0
All	Freshness Index (FR)	0.97	0.76	1.23	846	140	84	56	7.0
All	Match-day m·min⁻¹ (MD)	1.21	0.93	1.57	846	140	84	56	7.0
All	FI × FR	0.99	0.79	1.24	846	140	84	56	7.0
All	FI × MD	0.99	0.79	1.23	846	140	84	56	7.0
All	FR × MD	0.85	0.68	1.07	846	140	84	56	7.0
All	FI × FR × MD	0.99	0.77	1.28	846	140	84	56	7.0
U21	Fitness Index (FI)	0.88	0.50	1.57	171	24	13	11	1.4
U21	Freshness Index (FR)	0.99	0.61	1.61	171	24	13	11	1.4
U21	Match-day m·min⁻¹ (MD)	1.38	0.73	2.61	171	24	13	11	1.4
U21	FI × FR	0.93	0.51	1.70	171	24	13	11	1.4
U21	FI × MD	0.95	0.54	1.67	171	24	13	11	1.4
U21	FR × MD	0.65	0.33	1.27	171	24	13	11	1.4
U21	FI × FR × MD	1.11	0.51	2.40	171	24	13	11	1.4
U18	Fitness Index (FI)	0.82	0.47	1.41	162	24	16	8	1.0
U18	Freshness Index (FR)	0.83	0.49	1.40	162	24	16	8	1.0
U18	Match-day m·min⁻¹ (MD)	1.74	0.91	3.32	162	24	16	8	1.0
U18	FI × FR	0.97	0.54	1.74	162	24	16	8	1.0
U18	FI × MD	0.87	0.50	1.51	162	24	16	8	1.0
U18	FR × MD	0.85	0.50	1.43	162	24	16	8	1.0
U18	FI × FR × MD	0.91	0.44	1.88	162	24	16	8	1.0
U16	Fitness Index (FI)	0.72	0.37	1.38	134	23	11	12	1.4
U16	Freshness Index (FR)	1.16	0.60	2.24	134	23	11	12	1.4
U16	Match-day m·min⁻¹ (MD)	1.22	0.61	2.43	134	23	11	12	1.4
U16	FI × FR	0.98	0.52	1.83	134	23	11	12	1.4
U16	FI × MD	1.02	0.49	2.12	134	23	11	12	1.4
U16	FR × MD	1.00	0.51	1.97	134	23	11	12	1.4
U16	FI × FR × MD	1.15	0.48	2.73	134	23	11	12	1.4
U15	Fitness Index (FI)	0.84	0.40	1.75	121	20	14	6	0.8
U15	Freshness Index (FR)	0.92	0.45	1.87	121	20	14	6	0.8
U15	Match-day m·min⁻¹ (MD)	1.48	0.58	3.81	121	20	14	6	0.8
U15	FI × FR	1.10	0.53	2.28	121	20	14	6	0.8
U15	FI × MD	0.97	0.46	2.07	121	20	14	6	0.8
U15	FR × MD	0.78	0.32	1.90	121	20	14	6	0.8
U15	FI × FR × MD	0.86	0.35	2.10	121	20	14	6	0.8

 © 2027 ARD Asociación Española

Model	Term	*OR*	95% CI *LL*	95% CI *UL*	Records	Matches	Won	Not won	EPV
U14	Fitness Index (FI)	0.90	0.43	1.87	128	25	14	11	1.4
U14	Freshness Index (FR)	0.94	0.49	1.79	128	25	14	11	1.4
U14	Match-day m·min^{-1} (MD)	0.97	0.48	1.93	128	25	14	11	1.4
U14	FI × FR	0.93	0.51	1.70	128	25	14	11	1.4
U14	FI × MD	1.10	0.55	2.17	128	25	14	11	1.4
U14	FR × MD	0.76	0.41	1.42	128	25	14	11	1.4
U14	FI × FR × MD	1.00	0.47	2.13	128	25	14	11	1.4
U13	Fitness Index (FI)	0.80	0.44	1.48	130	24	16	8	1.0
U13	Freshness Index (FR)	1.02	0.54	1.92	130	24	16	8	1.0
U13	Match-day m·min^{-1} (MD)	0.89	0.47	1.69	130	24	16	8	1.0
U13	FI × FR	1.03	0.47	2.22	130	24	16	8	1.0
U13	FI × MD	1.02	0.50	2.09	130	24	16	8	1.0
U13	FR × MD	1.07	0.51	2.23	130	24	16	8	1.0
U13	FI × FR × MD	0.91	0.34	2.39	130	24	16	8	1.0

Note. Joint test of the 15 category × index interactions: p > .999. Post hoc joint Wald tests of the seven terms: p ≥ .71 in all models. Records: median across imputed datasets; Won and Not won: matches; EPV, events per parameter.

Comparison with simple indicators

Table S9. Cross-validated performance (10-fold cross-validation grouped by match, repeated 20 times; averaged across 50 imputed datasets).

Freshness Index	Model	Log-loss	Brier score	AUC	Calibration slope
PlayerLoad (PL)	Venue only	0.688	.247	.49	0.45
PlayerLoad (PL)	+ match-day running	0.690	.248	.53	0.41
PlayerLoad (PL)	+ match-day running + indices (FI, FR)	0.691	.249	.54	0.37
PlayerLoad (PL)	+ match-day running + raw indicators (MD−3 RPE; MD−1 PL or CMJ change)	0.692	.249	.54	0.36
PlayerLoad (PL)	Original seven-term model + venue	0.695	.250	.54	0.31
CMJ	Venue only	0.688	.247	.49	0.45
CMJ	+ match-day running	0.690	.248	.53	0.40
CMJ	+ match-day running + indices (FI, FR)	0.692	.249	.54	0.35
CMJ	+ match-day running + raw indicators (MD−3 RPE; MD−1 PL or CMJ change)	0.693	.249	.54	0.34
CMJ	Original seven-term model + venue	0.693	.249	.54	0.35

Note. PlayerLoad (PL): primary analysis (Freshness Index from PlayerLoad; raw indicators MD−3 RPE and MD−1 PlayerLoad). CMJ: post hoc analysis with the CMJ-based Freshness Index (raw indicators MD−3 RPE and CMJ change). Lower log-loss and Brier score and higher AUC are better; AUC .5 indicates no discrimination; calibration slope 1 indicates good calibration. Imputation did not use the outcome. Records: 836 (140 matches) in the PlayerLoad analysis and 834 (140 matches) in the CMJ analysis.

Convergent validity of the indices

Table S10. Within-player (repeated-measures) correlations of the indices with wellness and with the weekly CMJ change.

Index	Measure	Session	Observations	Players	*r*	95% CI *LL*	95% CI *UL*	*p*
Fitness Index	2024 wellness: sleep	MD−3	154	41	−.02	−.21	.16	.810
Fitness Index	2024 wellness: fatigue	MD−3	154	41	.01	−.17	.19	.906
Fitness Index	2024 wellness: stress	MD−3	154	41	−.14	−.31	.05	.143
Fitness Index	2024 wellness: mood	MD−3	154	41	.06	−.12	.25	.497
Fitness Index	2024 wellness: muscle soreness	MD−3	153	41	.02	−.17	.20	.849
Fitness Index	2024 wellness: total wellness	MD−3	154	41	.00	−.19	.18	.963
Freshness Index	2024 wellness: sleep	MD−1	66	23	.01	−.29	.31	.944
Freshness Index	2024 wellness: fatigue	MD−1	66	23	.14	−.16	.42	.364
Freshness Index	2024 wellness: stress	MD−1	66	23	.08	−.22	.37	.608
Freshness Index	2024 wellness: mood	MD−1	66	23	.04	−.26	.34	.784
Freshness Index	2024 wellness: muscle soreness	MD−1	66	23	.12	−.18	.40	.440
Freshness Index	2024 wellness: total wellness	MD−1	66	23	.15	−.15	.43	.327
Fitness Index	Weekly CMJ change	MD−4 → MD−1	153	45	−.12	−.30	.07	.200
Freshness Index	Weekly CMJ change	MD−4 → MD−1	80	29	−.01	−.28	.26	.948

Note. Wellness was recorded only in 2024. Session-level residuals were correlated with the wellness items of the same day; weekly indices with the CMJ change of the same week. The direction of the wellness scale is described in the Methods.

Original model with 90% confidence intervals

Table S11. Original model: OR per within-player SD With 90% CIs (as used in the original submission).

Model	Term	*OR*	90% CI *LL*	90% CI *UL*	Records	Matches	Won	Not won	EPV
All	Fitness Index (FI)	0.90	0.72	1.12	846	140	84	56	7.0
All	Freshness Index (FR)	1.10	0.88	1.37	846	140	84	56	7.0
All	Match-day m·min⁻¹ (MD)	1.24	0.99	1.55	846	140	84	56	7.0
All	FI × FR	1.02	0.83	1.25	846	140	84	56	7.0
All	FI × MD	0.98	0.80	1.21	846	140	84	56	7.0
All	FR × MD	0.96	0.79	1.16	846	140	84	56	7.0
All	FI × FR × MD	0.99	0.82	1.20	846	140	84	56	7.0
U21	Fitness Index (FI)	0.94	0.60	1.48	171	24	13	11	1.4
U21	Freshness Index (FR)	1.16	0.73	1.85	171	24	13	11	1.4
U21	Match-day m·min⁻¹ (MD)	1.37	0.82	2.29	171	24	13	11	1.4
U21	FI × FR	1.01	0.65	1.58	171	24	13	11	1.4
U21	FI × MD	0.89	0.57	1.38	171	24	13	11	1.4
U21	FR × MD	1.08	0.68	1.71	171	24	13	11	1.4
U21	FI × FR × MD	1.17	0.64	2.12	171	24	13	11	1.4
U18	Fitness Index (FI)	0.85	0.55	1.31	162	24	16	8	1.0
U18	Freshness Index (FR)	1.21	0.75	1.97	162	24	16	8	1.0
U18	Match-day m·min⁻¹ (MD)	1.83	1.03	3.25	162	24	16	8	1.0
U18	FI × FR	1.04	0.59	1.84	162	24	16	8	1.0
U18	FI × MD	0.96	0.60	1.54	162	24	16	8	1.0
U18	FR × MD	0.95	0.63	1.42	162	24	16	8	1.0
U18	FI × FR × MD	0.88	0.58	1.32	162	24	16	8	1.0
U16	Fitness Index (FI)	0.77	0.45	1.29	134	23	11	12	1.4
U16	Freshness Index (FR)	1.01	0.58	1.75	134	23	11	12	1.4
U16	Match-day m·min⁻¹ (MD)	1.25	0.68	2.29	134	23	11	12	1.4
U16	FI × FR	0.94	0.53	1.66	134	23	11	12	1.4
U16	FI × MD	1.07	0.60	1.92	134	23	11	12	1.4
U16	FR × MD	0.94	0.51	1.76	134	23	11	12	1.4
U16	FI × FR × MD	1.03	0.54	1.99	134	23	11	12	1.4
U15	Fitness Index (FI)	0.83	0.45	1.55	121	20	14	6	0.8
U15	Freshness Index (FR)	1.09	0.59	2.03	121	20	14	6	0.8
U15	Match-day m·min⁻¹ (MD)	1.44	0.65	3.18	121	20	14	6	0.8
U15	FI × FR	1.02	0.56	1.88	121	20	14	6	0.8
U15	FI × MD	0.98	0.47	2.04	121	20	14	6	0.8
U15	FR × MD	0.88	0.42	1.82	121	20	14	6	0.8
U15	FI × FR × MD	1.05	0.48	2.30	121	20	14	6	0.8
U14	Fitness Index (FI)	0.90	0.52	1.55	128	25	14	11	1.4
U14	Freshness Index (FR)	0.97	0.55	1.71	128	25	14	11	1.4
U14	Match-day m·min⁻¹ (MD)	1.01	0.57	1.80	128	25	14	11	1.4
U14	FI × FR	0.99	0.50	1.95	128	25	14	11	1.4
U14	FI × MD	1.00	0.55	1.84	128	25	14	11	1.4
U14	FR × MD	1.00	0.54	1.84	128	25	14	11	1.4
U14	FI × FR × MD	1.00	0.41	2.42	128	25	14	11	1.4
U13	Fitness Index (FI)	0.77	0.44	1.33	130	24	16	8	1.0
U13	Freshness Index (FR)	1.11	0.63	1.97	130	24	16	8	1.0
U13	Match-day m·min⁻¹ (MD)	0.89	0.51	1.53	130	24	16	8	1.0
U13	FI × FR	0.93	0.49	1.73	130	24	16	8	1.0
U13	FI × MD	1.00	0.52	1.94	130	24	16	8	1.0
U13	FR × MD	0.93	0.49	1.75	130	24	16	8	1.0
U13	FI × FR × MD	1.05	0.52	2.10	130	24	16	8	1.0

Note. 90% CIs from match-clustered standard errors, pooled across 50 imputed datasets. One interval excluded 1. Records: median across imputed datasets; Won and Not won: matches; EPV, events per parameter.

Analysis with players identified by season (post hoc sensitivity)

Player names were written differently in 2024 and 2025, so before players were linked across seasons each player was identified separately in each season, and within-player z-scores were computed within player and season. The whole analysis was repeated with the linked players; the earlier version, with players identified by season (committed outputs, version a28e1f7), is shown here as a sensitivity analysis.

 © 2027 ARD Asociación Española

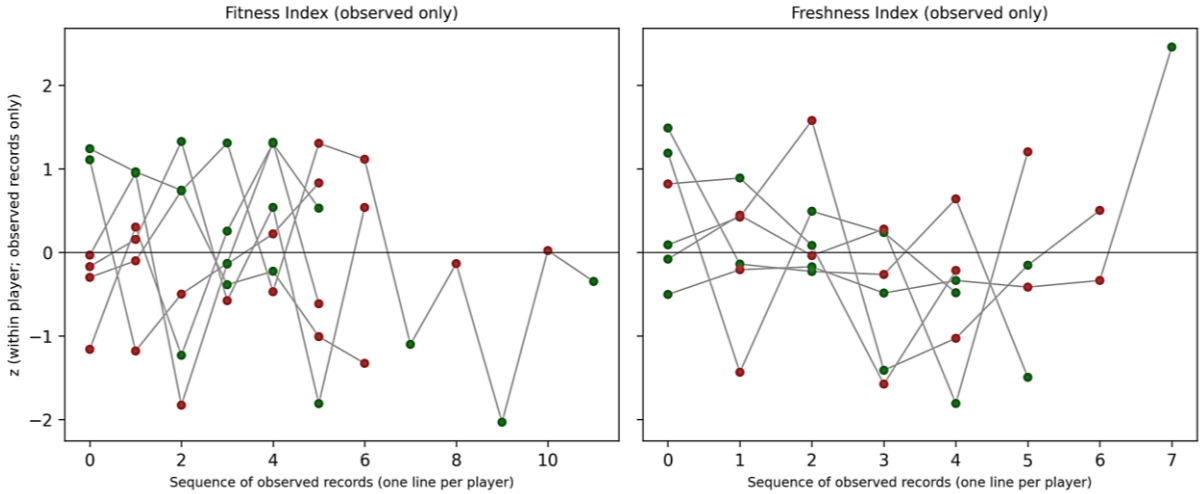
Table S12. main results with players identified by season and with players linked across seasons.

Quantity	Players identified by season	*LL*	*UL*	Players linked across seasons (primary)	*LL*	*UL*
Players in the eligible records	208	—	—	146	—	—
Imputation method selected by the protocol	chained equations with Bayesian regression	—	—	chained equations with predictive mean matching	—	—
Records excluded for undefined within-player z (median per dataset)	41	—	—	27	—	—
Original model: terms with 95% CI excluding 1 (of 49)	0	—	—	0	—	—
Main-effects model: terms with 95% CI excluding 1 (of 21)	0	—	—	0	—	—
Category × index interactions, joint test p	> .999	—	—	> .999	—	—
Home venue (venue-adjusted model), OR and 95% CI	2.14	1.01	4.51	2.11	0.999	4.46
Match-day running, match-clustered SE, OR and 95% CI	1.16	0.89	1.52	1.24	0.95	1.61
Match-day running, GLMM (player and category), OR and 95% credible interval	1.17	0.99	1.39	1.28	1.07	1.52
Match-day running, GLMM with opponent club, OR and 95% credible interval	1.26	1.00	1.58	1.29	1.02	1.64

Note. LL and UL apply only to the rows with an OR. In both versions, none of the 49 terms of the original model had a 95% CI excluding 1, so the conclusions for the indices are the same. Outside the indices, the venue interval excluded 1 with players identified by season (OR 2.14 [1.01, 4.51]) but included 1 with linked players (OR 2.11 [0.999, 4.46]), and the match-day running intervals of the two mixed models excluded 1 only with linked players (GLMM: OR 1.17 [0.99, 1.39] vs. 1.28 [1.07, 1.52]; GLMM with opponent club: OR 1.26 [1.00, 1.58] vs. 1.29 [1.02, 1.64]). With match-clustered standard errors, the match-day running interval included 1 in both versions (OR 1.16 [0.89, 1.52] and 1.24 [0.95, 1.61]). The mixed models do not account for clustering by match. The version with players identified by season also expressed the indices as observed − expected, which only inverts their odds ratios; venue and match-day running are not affected.

Individual trajectories with observed values (post hoc)

The Fitness and Freshness indices were observed in only 27.7% and 21.3% of eligible records, so trajectories built from the imputed data would be dominated by imputed values. This analysis uses only observed values (not imputed), within-player standardised from those observed values alone: players with fewer than 4 observed records, or without within-player variability among them, are excluded (27 players for the Fitness Index, 18 for the Freshness Index). It is purely descriptive.



Note. Green markers: match won; red markers: match not won. Descriptive only, small samples; not used for inference.

Figure S1. Individual trajectories (within-player z, observed values only) of the 6 players with the most observed records of each index.

Table S13. Frequency of extreme within-player deviations (|z| > 1.5, observed values only) before wins and non-wins

Index	Outcome	*n*	Extreme deviations (|z| > 1.5)	%
Fitness Index	Win	78	5	6.4
Fitness Index	Non-win	69	2	2.9
Freshness Index	Win	45	2	4.4
Freshness Index	Non-win	43	3	7.0

Note. Descriptive only; samples too small for a formal comparison between outcomes.

Imputation: sensitivity to the method

Table S14. Terms with a 95% CI excluding 1 by imputation method (post hoc except the primary method).

Imputation	Model	Records (all)	Models estimable (of 7)	Terms estimable	Terms excluding 1	Category of term excluding 1	Term excluding 1	*OR*	95% CI *LL*	95% CI *UL*
Chained equations, predictive mean matching (primary)	Original (7 terms)	846	7	49	0	—	None	—	—	—
Chained equations, predictive mean matching (primary)	Main effects (3 terms)	846	7	21	0	—	None	—	—	—
Chained equations, Bayesian regression (post hoc sensitivity)	Original (7 terms)	846	7	49	0	—	None	—	—	—
Chained equations, Bayesian regression (post hoc sensitivity)	Main effects (3 terms)	846	7	21	0	—	None	—	—	—
Category × season mean	Original (7 terms)	701	7	49	3	U16	FI × FR × MD	2.27	1.28	4.02
Category × season mean	Original (7 terms)	701	7	49	3	U13	FI	0.42	0.22	0.80
Category × season mean	Original (7 terms)	701	7	49	3	U13	FI × FR	0.50	0.26	0.97
Category × season mean	Main effects (3 terms)	701	7	21	1	U16	FI	0.62	0.40	0.96
Category mean, original convention (undefined z set to 0; reference)	Original (7 terms)	873	7	49	4	U16	FI	0.55	0.32	0.96
Category mean, original convention (undefined z set to 0; reference)	Original (7 terms)	873	7	49	4	U16	FI × FR × MD	2.61	1.06	6.44
Category mean, original convention (undefined z set to 0; reference)	Original (7 terms)	873	7	49	4	U14	FI × FR × MD	2.39	1.01	5.67
Category mean, original convention (undefined z set to 0; reference)	Original (7 terms)	873	7	49	4	U13	FI	0.45	0.26	0.78
Category mean, original convention (undefined z set to 0; reference)	Main effects (3 terms)	873	7	21	1	U16	FI	0.60	0.37	0.97
k-nearest neighbours (k = 10)	Original (7 terms)	846	7	49	3	U14	FR	0.55	0.31	0.99
k-nearest neighbours (k = 10)	Original (7 terms)	846	7	49	3	U14	FI × FR	1.68	1.14	2.49
k-nearest neighbours (k = 10)	Original (7 terms)	846	7	49	3	U13	MD	0.59	0.35	0.97
k-nearest neighbours (k = 10)	Main effects (3 terms)	846	7	21	2	U14	FR	0.57	0.33	0.98
k-nearest neighbours (k = 10)	Main effects (3 terms)	846	7	21	2	U13	MD	0.59	0.34	1.000
Player's own mean	Original (7 terms)	468	6	42	1	U18	FR	1.65	1.07	2.54
Player's own mean	Main effects (3 terms)	468	7	21	2	U18	FR	1.60	1.10	2.32
Player's own mean	Main effects (3 terms)	468	7	21	2	U16	FR	0.54	0.30	0.99
Complete cases (no imputation)	Original (7 terms)	34	2	14	1	U18	FI × MD	0.20	0.05	0.79

 © 2027 ARD Asociación Española

Imputation	Model	Records (all)	Models estimable (of 7)	Terms estimable	Terms excluding 1	Category of term excluding 1	Term excluding 1	*OR*	95% CI *LL*	95% CI *UL*
Complete cases (no imputation)	Main effects (3 terms)	34	2	6	0	—	None	—	—	—
Multilevel chained equations (post hoc)	Original (7 terms)	846	7	49	0	—	None	—	—	—
Multilevel chained equations (post hoc)	Main effects (3 terms)	846	7	21	0	—	None	—	—	—

Note. One row per term excluding 1 (None when no term did). A model is estimable if it has more than 20 rows and converges. Single imputation methods (category mean, player mean, k-nearest neighbours) understate uncertainty. The original convention sets undefined within-player z-scores to 0 and keeps all 873 records; without it, the category mean leaves many players without variation. Complete cases: 53 records with the three variables observed, 34 after excluding undefined z-scores.

Table S15. Original model, aggregated: OR and 95% CI by imputation method.

Imputation	Term	*OR*	95% CI *LL*	95% CI *UL*
Chained equations, predictive mean matching (primary)	FI	0.90	0.69	1.16
Chained equations, predictive mean matching (primary)	FR	1.10	0.84	1.43
Chained equations, predictive mean matching (primary)	MD	1.24	0.95	1.61
Chained equations, predictive mean matching (primary)	FI × FR	1.02	0.80	1.30
Chained equations, predictive mean matching (primary)	FI × MD	0.98	0.77	1.26
Chained equations, predictive mean matching (primary)	FR × MD	0.96	0.76	1.20
Chained equations, predictive mean matching (primary)	FI × FR × MD	0.99	0.79	1.25
Chained equations, Bayesian regression (post hoc sensitivity)	FI	0.88	0.71	1.08
Chained equations, Bayesian regression (post hoc sensitivity)	FR	1.07	0.85	1.34
Chained equations, Bayesian regression (post hoc sensitivity)	MD	1.23	0.95	1.59
Chained equations, Bayesian regression (post hoc sensitivity)	FI × FR	1.02	0.80	1.30
Chained equations, Bayesian regression (post hoc sensitivity)	FI × MD	0.98	0.79	1.22
Chained equations, Bayesian regression (post hoc sensitivity)	FR × MD	0.99	0.79	1.23
Chained equations, Bayesian regression (post hoc sensitivity)	FI × FR × MD	0.97	0.79	1.20
Category × season mean	FI	0.99	0.77	1.28
Category × season mean	FR	0.89	0.69	1.15
Category × season mean	MD	1.17	0.90	1.53
Category × season mean	FI × FR	1.21	0.995	1.47
Category × season mean	FI × MD	0.93	0.76	1.14
Category × season mean	FR × MD	0.96	0.79	1.18
Category × season mean	FI × FR × MD	1.05	0.92	1.19
Category mean, original convention (undefined z set to 0; reference)	FI	1.00	0.78	1.28
Category mean, original convention (undefined z set to 0; reference)	FR	0.90	0.70	1.15
Category mean, original convention (undefined z set to 0; reference)	MD	1.19	0.92	1.55
Category mean, original convention (undefined z set to 0; reference)	FI × FR	1.21	0.997	1.48
Category mean, original convention (undefined z set to 0; reference)	FI × MD	0.94	0.77	1.15
Category mean, original convention (undefined z set to 0; reference)	FR × MD	0.97	0.80	1.18
Category mean, original convention (undefined z set to 0; reference)	FI × FR × MD	1.04	0.92	1.19
k-nearest neighbours (k = 10)	FI	0.96	0.81	1.12
k-nearest neighbours (k = 10)	FR	0.92	0.77	1.11
k-nearest neighbours (k = 10)	MD	1.14	0.88	1.48
k-nearest neighbours (k = 10)	FI × FR	1.03	0.88	1.20
k-nearest neighbours (k = 10)	FI × MD	0.94	0.79	1.12
k-nearest neighbours (k = 10)	FR × MD	1.01	0.85	1.19
k-nearest neighbours (k = 10)	FI × FR × MD	0.97	0.84	1.13
Player's own mean	FI	1.03	0.83	1.29
Player's own mean	FR	0.97	0.77	1.20
Player's own mean	MD	1.15	0.84	1.56
Player's own mean	FI × FR	0.95	0.73	1.23
Player's own mean	FI × MD	0.86	0.68	1.08
Player's own mean	FR × MD	1.08	0.90	1.29
Player's own mean	FI × FR × MD	0.92	0.77	1.11

Imputation	Term	*OR*	95% CI *LL*	95% CI *UL*
Complete cases (no imputation)	FI	1.51	0.41	5.59
Complete cases (no imputation)	FR	1.30	0.37	4.53
Complete cases (no imputation)	MD	1.24	0.35	4.40
Complete cases (no imputation)	FI × FR	0.67	0.21	2.17
Complete cases (no imputation)	FI × MD	0.43	0.10	1.75
Complete cases (no imputation)	FR × MD	0.45	0.10	2.06
Complete cases (no imputation)	FI × FR × MD	0.36	0.05	2.62
Multilevel chained equations (post hoc)	FI	0.85	0.56	1.28
Multilevel chained equations (post hoc)	FR	1.14	0.83	1.55
Multilevel chained equations (post hoc)	MD	1.23	0.93	1.62
Multilevel chained equations (post hoc)	FI × FR	1.00	0.78	1.27
Multilevel chained equations (post hoc)	FI × MD	1.00	0.78	1.27
Multilevel chained equations (post hoc)	FR × MD	1.01	0.79	1.28
Multilevel chained equations (post hoc)	FI × FR × MD	0.96	0.77	1.20

Stability: model fit and main-effects model

Table S16. Model fit (medians across imputations) and post hoc joint Wald Test of the seven terms.

Model	Log-likelihood	Null log-likelihood	Pseudo-R^2	LR test *p*	Joint Wald *p* (post hoc)	Imputations estimable	Total imputations
All	−571.0	−580.2	.016	.012	.895	50	50
U21	−113.7	−118.3	.039	.238	.982	50	50
U18	−99.8	−108.7	.082	.013	.868	50	50
U16	−86.8	−90.7	.043	.347	.994	50	50
U15	−71.8	−76.8	.065	.191	.997	50	50
U14	−85.0	−88.0	.033	.568	> .999	50	50
U13	−83.3	−87.1	.040	.427	.997	50	50

Note. The likelihood-ratio (LR) test, stated in the original Methods, assumes independent records; records of the same match share the outcome, so it overstates the evidence. The joint Wald test (post hoc) uses match-clustered standard errors and is combined across imputations with the D1 statistic.

Table S17. Main-effects model (three terms).

Model	Term	*OR*	95% CI *LL*	95% CI *UL*	Records	Matches	Won	Not won	EPV
All	Fitness Index (FI)	0.90	0.70	1.16	846	140	84	56	14.0
All	Freshness Index (FR)	1.10	0.85	1.43	846	140	84	56	14.0
All	Match-day m·min^{-1} (MD)	1.23	0.95	1.59	846	140	84	56	14.0
U21	Fitness Index (FI)	0.92	0.55	1.54	171	24	13	11	2.8
U21	Freshness Index (FR)	1.16	0.69	1.97	171	24	13	11	2.8
U21	Match-day m·min^{-1} (MD)	1.36	0.76	2.44	171	24	13	11	2.8
U18	Fitness Index (FI)	0.89	0.55	1.46	162	24	16	8	2.0
U18	Freshness Index (FR)	1.28	0.76	2.15	162	24	16	8	2.0
U18	Match-day m·min^{-1} (MD)	1.72	0.93	3.21	162	24	16	8	2.0
U16	Fitness Index (FI)	0.79	0.46	1.38	134	23	11	12	2.8
U16	Freshness Index (FR)	1.02	0.56	1.85	134	23	11	12	2.8
U16	Match-day m·min^{-1} (MD)	1.20	0.63	2.31	134	23	11	12	2.8
U15	Fitness Index (FI)	0.85	0.44	1.63	121	20	14	6	1.5
U15	Freshness Index (FR)	1.10	0.59	2.06	121	20	14	6	1.5
U15	Match-day m·min^{-1} (MD)	1.41	0.57	3.49	121	20	14	6	1.5
U14	Fitness Index (FI)	0.90	0.52	1.58	128	25	14	11	2.8
U14	Freshness Index (FR)	0.98	0.52	1.85	128	25	14	11	2.8
U14	Match-day m·min^{-1} (MD)	1.01	0.54	1.88	128	25	14	11	2.8
U13	Fitness Index (FI)	0.77	0.44	1.36	130	24	16	8	2.0
U13	Freshness Index (FR)	1.09	0.59	2.02	130	24	16	8	2.0
U13	Match-day m·min^{-1} (MD)	0.89	0.50	1.60	130	24	16	8	2.0

Note. Records: median across imputed datasets; Won and Not won: matches; EPV, events per parameter.

 © 2027 ARD Asociación Española

Adjustment for the opponent

Table S18. Original model (aggregated) with fixed effects for category and opponent club and match-clustered standard errors (post hoc).

Term	*OR*	95% CI *LL*	95% CI *UL*	*p*	Imputations estimable	Total imputations
Fitness Index (FI)	0.88	0.62	1.25	.459	50	50
Freshness Index (FR)	1.12	0.80	1.59	.504	50	50
Match-day m·min^{-1} (MD)	1.27	0.92	1.76	.140	50	50
FI × FR	1.02	0.72	1.42	.928	50	50
FI × MD	0.97	0.73	1.30	.852	50	50
FR × MD	1.01	0.76	1.35	.933	50	50
FI × FR × MD	0.98	0.72	1.33	.877	50	50

Note. Opponent clubs never beaten (2) or always beaten (8) carry no within-opponent information and produce separation, so they were excluded: 97 of 140 matches and 624 records remained, with 18 fixed effects.

Developmental stages: Main-effects model (post hoc)

Table S19. Main-effects model by developmental stage.

Stage	Term	*OR*	95% CI *LL*	95% CI *UL*
U13–U14	Fitness Index (FI)	0.87	0.58	1.31
U13–U14	Freshness Index (FR)	1.02	0.67	1.57
U13–U14	Match-day m·min^{-1} (MD)	0.94	0.61	1.43
U15–U16	Fitness Index (FI)	0.91	0.61	1.35
U15–U16	Freshness Index (FR)	1.02	0.66	1.57
U15–U16	Match-day m·min^{-1} (MD)	1.24	0.77	2.01
U18–U21	Fitness Index (FI)	0.91	0.62	1.33
U18–U21	Freshness Index (FR)	1.22	0.85	1.77
U18–U21	Match-day m·min^{-1} (MD)	1.50	0.98	2.29

This work is licensed under a Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0 DEED).