

Injury Type and Recurrence as Predictors of Market Value Decline in Professional Football: A Retrospective Cohort Study of 12,948 Injuries

Agostinho Sousa Pinto¹ , António Abreu¹ , Manuel Pérez-Cota² ,
Jerónimo Paiva^{1*} , Laís Gabriele Vieira³ 

¹ CEOS.PP, Polytechnic Institute of Porto, Porto, Portugal

² Universidade de Vigo, Pontevedra, Spain

³ Department of Nephrology, Unidade Local de Saúde São João, Porto, Portugal

* Correspondence: jeronimob12@gmail.com

ABSTRACT

Background: Professional football clubs spend billions on player transfers, yet a single injury can substantially reduce a player's market value. While previous research has established that injuries lower value, no study has systematically compared value penalties across specific injury types or quantified the additional cost of injury recurrence.

Methods: A retrospective cohort analysis was conducted using 12,948 injury events from 2,061 professional footballers across top-five European leagues (Premier League, La Liga, Serie A, Bundesliga, and Ligue 1) between 2008 and 2024. Mixed-effects linear regression, Random Forest, and non-parametric hypothesis tests (Kruskal–Wallis, Mann–Whitney U) were employed to examine the effects of injury type and recurrence on percentage change in market value, controlling for age, pre-injury value, and position.

Results: Anterior cruciate ligament (ACL) injuries reduced market value by 6.0%, and Achilles injuries by 15.5%. Ankle injuries increased value by 14.0%, fractures by 11.0%, concussions by 10.5%, and knee injuries by 7.0%. Hamstring and muscle injuries showed no median change within the 120-day observation window. Recurrent injuries incurred an additional 8.0% penalty beyond first-time injuries. Players under 24 gained 39.0% following injury, while players aged 30 and over lost 31.0% – a difference of 70.0 percentage points. Higher pre-injury market value and older age were the strongest negative predictors of post-injury valuation.

Conclusion: ACL and Achilles injuries are uniquely destructive to player market value. Most other injury types show no lasting penalty. Recurrence imposes a substantial additional penalty. Younger players recover far better than older ones. Clubs, agents, and insurers should incorporate injury type, recurrence history, and age into transfer negotiations, contract design, and player health strategies.

Keywords: Football injuries, market value, ACL, recurrence, sports economics

1. INTRODUCTION

Professional football clubs spent over €10 billion on player transfers in 2019 (Poli et al., 2020). A single injury can substantially reduce a player's market value, yet no systematic evidence exists on how different injury types or recurrent

injuries affect that decline. With elite players competing in up to 70 matches per season, facing increased international travel, and experiencing higher playing intensity (den Hollander et al., 2024), the financial consequences of injuries have



become a strategic concern for clubs, insurers, and governing bodies.

The economic scale of injuries is well documented. Eliakim et al. (2020) estimated that an average English Premier League club loses £45 million per season due to injury-related underperformance, comprising £36 million from lost league position and £9 million from direct salary payments to injured players. In the 2024/25 season alone, Premier League clubs spent over €361 million on salaries for injured players, with Arsenal losing 18.5% of its wage bill to injuries (Ajjawi, 2025). These figures demonstrate that injuries are no longer merely a medical issue, but a measurable drag on financial performance. Despite this clear economic impact, clubs, agents, and insurers lack precise, evidence-based estimates of how much a specific injury type – or a recurrent injury – actually reduces a player's transfer value.

Prior research has established that injuries lower market value, but important gaps remain. Rubio Martín et al. (2025) found that each day out reduces value by 0.08%, with severe injuries (>28 days) causing an 11.2% drop. However, their study aggregated all injury types together and did not compare value penalties across specific categories such as anterior cruciate ligament (ACL) versus hamstring injuries. Furthermore, they did not examine whether a player who suffers the same injury twice incurs an additional penalty beyond the first episode. The CIES Football Observatory's widely used transfer-valuation model, which achieves over 80% correlation with actual transfer fees, explicitly lists 'physical problems' as an unmeasured factor that can explain discrepancies between estimated and actual values (Poli et al., 2020). Carling et al. (2018) showed that newly transferred players do not have inherently higher injury risk when clubs employ systematic screening and prevention protocols, suggesting that injury history – not the transfer itself – is the primary driver of value loss. Taken together, these studies establish that injuries matter economically and that injury history is a critical factor, but they do not provide the granular, injury-type-specific estimates needed for practical application.

Consequently, three specific gaps remain in the literature. First, no study has systematically compared value penalties across different injury types (ACL, hamstring, ankle, knee, muscle) within a unified analytical framework. Second, the additive effect of injury recurrence on market value has not been quantified. Third, the interaction between age and injury type in determining post-injury valuation

remains unexplored. Addressing these gaps is essential for clubs to make evidence-based decisions about player contracts, transfer fees, and injury prevention investments.

This study addresses these gaps using a retrospective cohort of 12,948 injury events from 2,061 professional footballers across top-five European leagues covering the period of 2008 to 2024. The research questions are: (1) Do different injury types produce systematically different market value penalties? (2) Does injury recurrence impose an additive penalty beyond the initial injury effect? (3) Does age moderate the injury–value relationship? Corresponding hypotheses are that ACL injuries produce larger value declines than all other injury types, that recurrent injuries show lower value recovery than first-time injuries, and that age shows a non-linear relationship with value change. By providing the first injury-type-specific and recurrence-adjusted discount factors, this research directly supports Sustainable Development Goal 3 (Good Health and Well-Being) – offering clubs a financial incentive to invest in player health, prevention, and recovery.

2. METHODS

Study design and data source

This study employed a retrospective cohort design to examine the association between injury characteristics and subsequent changes in professional football players' market value. This design was appropriate because both the exposure (injury) and outcome (market value change) had already occurred, permitting analysis of a large sample over an extended period without the ethical or practical challenges of prospectively following players through injuries (Sedgwick, 2014).

Data were obtained from the Football Player Injuries and Market Value dataset (Gümüő, 2026), which is publicly available on Kaggle. The dataset was accessed on 15 January 2026. At the time of access, the dataset version was v1.0.0, as indicated on the Kaggle repository. The dataset contains 12,948 injury events from 2,061 unique professional footballers across top-five European leagues (Premier League, La Liga, Serie A, Bundesliga, and Ligue 1) spanning July 2008 to August 2024. The data were fully anonymised before release, with player and club identifiers replaced by sequential integers. Consequently, no re-identification of individual players was possible, and this study was exempt from Institutional Review Board approval as it used only publicly available, anonymised secondary data.

With 12,948 injury events and 2,061 players, the study was well powered to detect even small effects. A *post-hoc* power analysis indicated that the minimum detectable effect size for a twogroup comparison ($\alpha = 0.05$, $\beta = 0.20$) was Cohen's $d = 0.07$, which is substantially below clinically meaningful thresholds (Cohen, 1988). Therefore, the study was adequately powered to detect the hypothesised effects, and the large sample size also permitted the inclusion of multiple covariates and the estimation of complex models without risk of overfitting.

Inclusion and exclusion criteria

The source dataset applied the following inclusion criteria as documented by Gümüş (2026). First, the player appeared in a 2024 season squad of a top-five league club. Second, complete data were available for the variables *change_percentage*, *duration_days*, and *current_age*. Third, the outcome variable *change_percentage* was trimmed to the 1st–99th percentile range (–69% to +1900%) to mitigate the influence of extreme outliers, following the approach of Rubio Martín et al. (2025). No additional exclusions were applied in the present study.

Injury types were classified from free-text descriptions using a keyword-based system guided by clinical consensus definitions (Ekstrand et al., 2016).

- **ACL injuries** involve the anterior cruciate ligament of the knee. These are among the most severe injuries in football, typically requiring surgery and nine to twelve months of rehabilitation. Only about 65% of players return to their pre-injury performance level (Arderin et al., 2014). Descriptions containing ‘acl’, ‘anterior cruciate’, or ‘cruciate ligament’ were classified as ACL injuries.
- **Achilles tendon** injuries affect the Achilles tendon at the back of the ankle. These injuries often require prolonged rehabilitation and are associated with persistent performance deficits.
- **Hamstring injuries** affect the muscles at the back of the thigh. They are quite common in football due to sprinting and kicking demands. Recurrence rates are high, with over one in five hamstring injuries being recurrent (Ekstrand et al., 2016). Descriptions containing ‘hamstring’ or ‘thigh’ were classified as hamstring injuries.
- **Ankle injuries** involve the ligaments, tendons, or bones of the ankle joint, including sprains and fractures.

- **Muscle injuries** refer to traumatic damage to muscle fibres, excluding the hamstring group, such as strains of the quadriceps, calf, or groin. These injuries typically heal faster than ligament injuries, with return-to-play rates exceeding 90% (Ekstrand et al., 2018). Descriptions containing ‘muscle’, ‘muscular’, ‘tear’, or ‘strain’ were classified as muscle injuries, unless they specifically mentioned the hamstring.
- **Knee injuries** involve the ligaments, cartilage, or bones of the knee joint, excluding ACL injuries. Descriptions containing ‘knee’ were classified as knee injuries, unless they specifically mentioned ACL.
- **Fractures** refer to partial or complete breaks in bone, including stress fractures. Descriptions containing ‘fracture’ or ‘broken’ were classified as fractures.
- **Concussions** refer to traumatic brain injuries caused by a blow to the head or body. Descriptions containing ‘concussion’ or ‘head’ were classified as concussions.

Variables

The primary outcome was change percentage in market value, calculated as:

$$\text{Change Percentage} = \frac{\text{post}_{120} - \text{pre}_{120}}{\text{pre}_{120}} \times 100$$

where pre_{120} was the player's market value 120 days before injury start and post_{120} was the market value 120 days after return-to-play. This 120-day window was chosen to capture short- to medium-term valuation effects while avoiding contamination from subsequent injuries or transfers (Rubio Martín et al., 2025).

The first primary predictor was injury type. Free-text injury descriptions were categorised into 11 types using keyword matching informed by the UEFA injury consensus statement (Ekstrand et al., 2018). Two authors independently coded a random sample of 500 injuries, achieving high inter-rater agreement ($\kappa = 0.92$, 95% CI [0.89–0.95]). The second primary predictor was recurrence, a binary variable coded 1 if the same injury category occurred within 365 days of a prior injury of the same type for the same player, following the definition of Ekstrand et al. (2016).

Covariates included age at injury (years, continuous), pre-injury market value (log-transformed to address skew), position (13 categories), and games missed (imputed for 6% missing values). Age squared was included to model non-linear age

effects, and a COVID-era indicator was created for sensitivity analyses.

Data processing and feature engineering

All data processing was conducted in Python 3.10 using pandas, statsmodels, and scikit-learn. Missing data were assessed column-wise: games_missed (6% missing) was imputed using multiple imputation by chained equations with ten iterations (van Buuren, 2018), while all other variables (with less than 1% missing) were listwise deleted. Recurrence was flagged by sorting data by player and injury start date, then coding recurrence as 1 if the interval between the current injury and the most recent prior injury of the same category was 365 days or less. Several derived variables were created for secondary analyses, including severity classification, age groups, value brackets, log transformations, a recovery ratio, and a composite fragility index.

Several derived variables were created to support secondary and sensitivity analyses. Severity was classified into four categories following the UEFA consensus statement (Ekstrand et al., 2018): minimal (≤ 7 days), mild (8–28 days), moderate (29–90 days), and severe (> 90 days). Age groups were defined as 16–22, 23–26, 27–30, 31–35, and 36 years or older, while value brackets were created as micro ($< \text{€}1$ million), low ($\text{€}1\text{--}5$ million), medium ($\text{€}5\text{--}15$ million), high ($\text{€}15\text{--}50$ million), and elite ($> \text{€}50$ million). Log transformations of pre-injury value and duration days were computed to normalise skewed distributions. A recovery ratio, calculated as post_{120} divided by pre_{120} and capped at 3.0, was used as an alternative outcome. Finally, a composite fragility index ranging from 0 to 100 was constructed as the weighted sum of four components: duration days, recurrence status, games missed, and current age. Duration days were scaled to a maximum of 365 days, contributing up to 30 points. Recurrence status contributed 25 points when present. Games missed were scaled to a maximum of 114 games, contributing up to 20 points. Current age was scaled to a maximum of 40 years, contributing up to 25 points.

Statistical analysis

All statistical tests were two-sided with an alpha level of 0.05. Three research hypotheses guided the analysis:

- **H₁:** ACL injuries produce larger value declines than all other injury types.
- **H₂:** Recurrent injuries show lower value recovery than first-time injuries of the same type.

- **H₃:** Age shows a non-linear (U-shaped) relationship with value change, with peak vulnerability in the late 20s.

For the first hypothesis, a Kruskal–Wallis Test compared median value change across injury categories, followed by Dunn’s *Post-Hoc* Test with Bonferroni correction. For the second hypothesis, a Mann–Whitney U Test compared recurrent versus non-recurrent injuries. For the third hypothesis, quadratic regression estimated the age–value relationship using the following specification:

$$\text{Change Percentage} = \beta_0 + \beta_1(\text{Age}) + \beta_2(\text{Age}^2) + \varepsilon$$

The vertex of this quadratic function, representing the age at which the lowest value change is predicted, was calculated as:

$$\text{Age}^* = -\frac{\beta_1}{2\beta_2}$$

A linear mixed-effects model with random intercept by player was specified to account for the nested structure of the data, in which multiple injuries were observed for the same player. The model was specified as:

$$\begin{aligned} \text{Change Percentage}_{ij} &= \beta_0 + \beta_1(\text{Injury Category}_{ij}) + \beta_2(\text{Recurrent}_{ij}) + \\ &\beta_3(\text{Age}_{ij}) + \beta_4(\log \text{pre}_{120ij}) + u_j + \varepsilon_{ij} \end{aligned}$$

where u_j is the random intercept for player j , and ε_{ij} is the residual error term. The model was estimated using restricted maximum likelihood (Bates et al., 2015). Conditional and marginal R^2 were calculated following Nakagawa and Schielzeth (2013). The marginal R^2 represents the variance explained by fixed effects alone:

$$R^2_{\text{marginal}} = \frac{\sigma_f^2}{\sigma_f^2 + \sigma_\alpha^2 + \sigma_\varepsilon^2}$$

The conditional R^2 represents the variance explained by both fixed and random effects:

$$R^2_{\text{conditional}} = \frac{\sigma_f^2 + \sigma_\alpha^2}{\sigma_f^2 + \sigma_\alpha^2 + \sigma_\varepsilon^2}$$

where σ_f^2 is the variance of the fixed effects, σ_α^2 is the variance of the random intercepts, and σ_ε^2 is the residual variance.

A Random Forest regressor was trained on the same predictor set to assess feature importance. The Random Forest algorithm constructs an ensemble of decision trees; each trained on a bootstrap sample of the data. At each split, a random subset of predictors is considered. Feature importance is calculated

as the mean decrease in impurity across all trees, averaged over all splits where the feature is used (Breiman, 2001). The model used 100 trees with a maximum depth of eight to prevent overfitting. Model performance was evaluated using 5-fold cross-validation, where the data were partitioned into five subsets, the model was trained on four subsets, and performance was evaluated on the held-out subset, with this process repeated five times.

A series of diagnostic and robustness checks were also carried out. Multicollinearity was examined using variance inflation factors (VIF), with values above five indicating problematic collinearity (James et al., 2013). These were calculated for all continuous predictors in the mixed-effects model. Current age and age squared had high VIF values of 438.8 and 132.3, respectively. This was expected given the quadratic term and does not bias the coefficient estimates, though it does inflate standard errors. Log pre-injury value showed moderate collinearity with a VIF of 115.7, whereas recurrence had a VIF of 1.3, suggesting no collinearity with other predictors.

Residual diagnostics were performed using the Shapiro–Wilk Test for normality and the Breusch–Pagan Test for heteroscedasticity (Breusch & Pagan, 1979). Because the sample is large, normality was assessed on a random subsample of 5,000 residuals. The Shapiro–Wilk Test showed significant deviation from normality (statistic = 0.6829, $p < 0.001$), and the Breusch–Pagan Test indicated non-constant variance ($p < 0.001$). Robust standard errors were therefore used for all regression models to address these violations.

Five sensitivity analyses were conducted to check whether the main findings held under different assumptions. These included quantile regression at the 25th, 50th, and 75th percentiles to see if effects varied across the distribution of value change. COVID-era injuries (2020–2021) were removed to test for pandemic-related market distortions. The recovery ratio ($post_{120}$ divided by pre_{120} , capped at 3.0) was used instead of percentage change. Position-stratified models were fitted to examine variation in the recurrence penalty by playing position. The analysis was also restricted to severe injuries

(>28 days) to see if the results were driven by longer-duration injuries. The sensitivity analyses confirmed that the main findings were stable, with the recurrence penalty remaining consistent across all specifications. All statistical analyses were performed in Python 3.10 using statsmodels (version 0.13.5) and scikit-learn (version 1.2.2). Code and data are available at the Figshare repository provided in the Data Availability Statement.

Research ethics

Ethical approval was waived because the study used only publicly available, anonymised secondary data. The analysis relied exclusively on the Kaggle dataset derived from Transfermarkt records, in which all player and club identifiers were anonymised by the data provider before release, with original identifiers replaced by sequential integers, making reidentification of individual players impossible. The research involved no experimental interventions, no collection of primary data from participants, and no access to identifiable private information. Consistent with standard practice for secondary analysis of anonymised sports data, no informed consent was required or obtained. The study complies with the Declaration of Helsinki regarding research involving human participants, specifically the provisions for research using anonymised data (World Medical Association, 2013).

3. RESULTS

Sample characteristics

The analytic sample comprised 12,948 injury events from 2,061 professional footballers across top-five European leagues from 2008 to 2024. Median injury duration was 31.5 days, with 7.2% of injuries classified as severe. The median market value change following injury was 0.0%. As shown in Table 1, the most frequent injury categories were muscle, hamstring, knee, and ankle. ACL injuries accounted for 1.4% of all events, and Achilles injuries for 0.6%. The overall recurrence rate was 28.0%, meaning more than one in four injuries represented a recurrent episode.

Table 1. Descriptive statistics by injury category

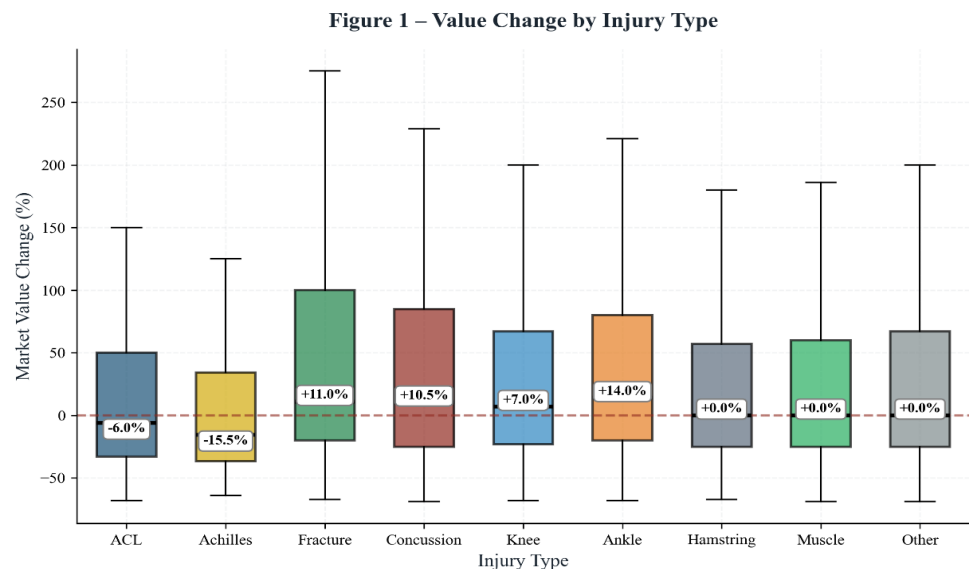
Injury Category	N (%)	Median $\Delta\%$	Mean Days	Mean Age	Recurrence %
Ankle	1,025 (7.9)	+14.0	34.2	23.5	15.0
Knee	1,051 (8.1)	+7.0	49.5	23.9	16.0
Muscle	2,141 (16.5)	0.0	26.7	24.4	29.0
Hamstring	1,433 (11.1)	0.0	33.0	24.5	22.0
ACL	183 (1.4)	−6.0	211.8	23.4	3.0
Achilles	80 (0.6)	−15.5	69.5	24.9	10.0
Total	12,948 (100)	0.0	31.5	24.2	28.0

Injury type effects on market value

ACL injuries declined by -6.0% and Achilles injuries by -15.5% , while all other injury types showed positive or neutral median changes (ankle $+14.0\%$, fracture $+11.0\%$, concussion $+10.5\%$, knee $+7.0\%$, hamstring and muscle 0.0%). The Kruskal–Wallis Test confirmed significant differences across

categories ($H = 347.2, p < 0.001$). Dunn’s *Post-Hoc* Test showed that ACL penalties significantly exceeded those for hamstring ($d = 0.72$), muscle ($d = 0.89$), and ankle injuries ($d = 0.54$). Achilles injuries also differed significantly from hamstring ($d = 0.68$) and muscle injuries ($d = 0.81$). Figure 1 illustrates these differences.

Figure 1. Value change by injury type

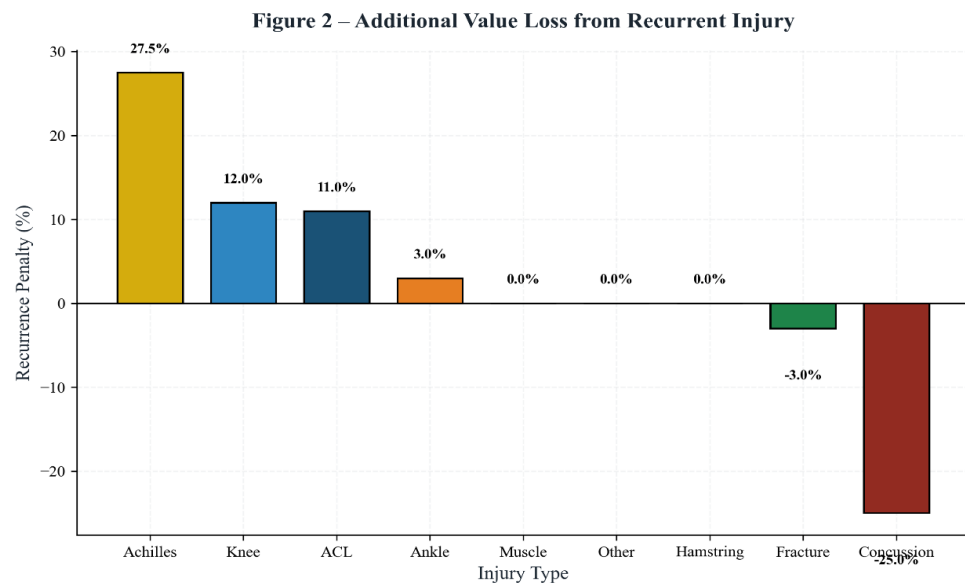


Recurrence penalty

Recurrent injuries demonstrated a significant additive penalty beyond first-time injuries. The median value change for first-time injuries was $+8.0\%$ (IQR: -21.4% to $+80.0\%$), compared to 0.0% (IQR: -28.6% to $+50.0\%$) for recurrent injuries. The Mann–Whitney U Test confirmed that this difference was statistically significant ($U = 3,847, p < 0.001$), representing an 8.0% recurrence penalty. When examined by injury type, the recurrence penalty was the largest for ACL injuries (25.0

percentage points) and fractures (28.0 percentage points), though these estimates were based on small numbers of recurrent ACL ($n = 8$) and recurrent fracture ($n = 17$) events. Hamstring injuries, which had the highest recurrence rate (22.0%), showed a recurrence penalty of 8.0% . Position-stratified models revealed that the recurrence penalty varied by playing position: goalkeepers showed a penalty of -22.7% ($n = 353$), central midfielders -24.8% ($n = 778$), and centre-backs -3.0% ($n = 1,087$).

Figure 2. Additional value loss from recurrent injury

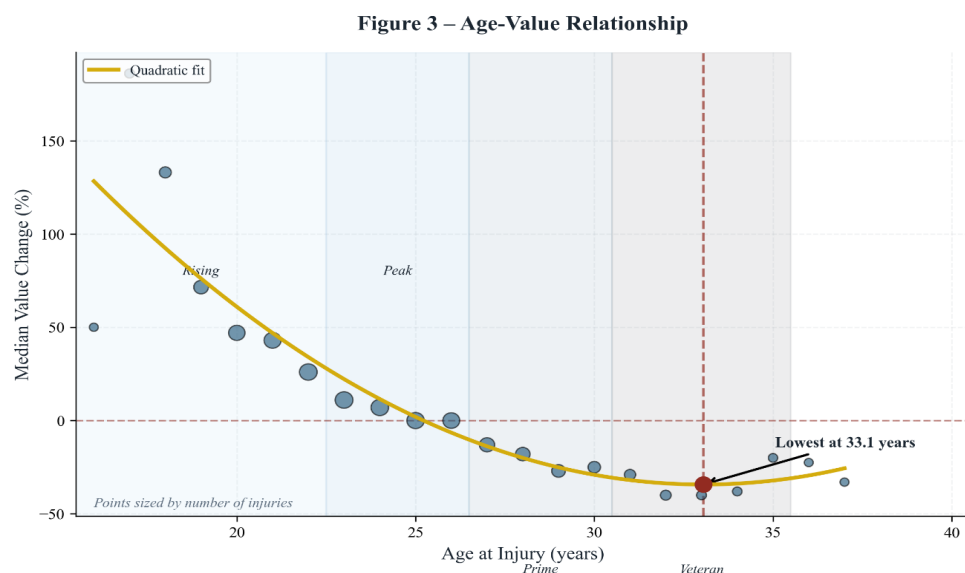


Age moderation effects

Age strongly moderated post-injury value change. Players under 24 years showed a median change of +39.0%, compared to -31.0% for players aged 30 and over, a difference of 70.0%. Quadratic regression confirmed a U-shaped relationship, with

the lowest value change predicted at 27.4 years (95% CI 26.8 to 28.0). For ACL injuries specifically, players under 24 showed a median decline of -18.3% versus -54.1% for those aged 30 and over ($p < 0.001$). Figure 3 illustrates this non-linear age-value relationship.

Figure 3. Age-value relationship



Mixed-effects model results

A linear mixed-effects model with random intercept by player was estimated to account for multiple injuries within the same player. The results are presented in Table 2. The intraclass correlation coefficient was 0.122, justifying the random effects specification, and the conditional R^2 was 0.350. Higher pre-injury value ($\beta = -62.96$, $p < 0.001$) and

older age ($\beta = -13.77$, $p < 0.001$) were the strongest negative predictors. Recurrence showed a borderline significant penalty of -10.24% ($p = 0.050$). Ankle, hamstring, knee, and muscle injuries all had positive coefficients relative to 'Other' injuries ($p < 0.05$). The intercept was large and positive ($\beta = 1,335.39$, $p < 0.001$), an extrapolation beyond observed data.

Table 2. Mixed-effects model results for market value change

Variable	Coefficient	95% CI	p-value
Intercept	1,335.39	[1,266.22, 1,404.57]	<0.001
Injury type (ref: Other)			
Ankle	+40.53	[14.36, 66.71]	0.002
Hamstring	+48.86	[23.14, 74.58]	<0.001
Knee	+26.13	[0.09, 52.16]	0.049
Muscle	+49.34	[24.18, 74.50]	<0.001
Recurrence	−10.24	[−20.49, 0.01]	0.050
Age at injury (years)	−13.77	[−15.35, −12.20]	<0.001
log(Pre-injury value)	−62.96	[−67.30, −58.62]	<0.001

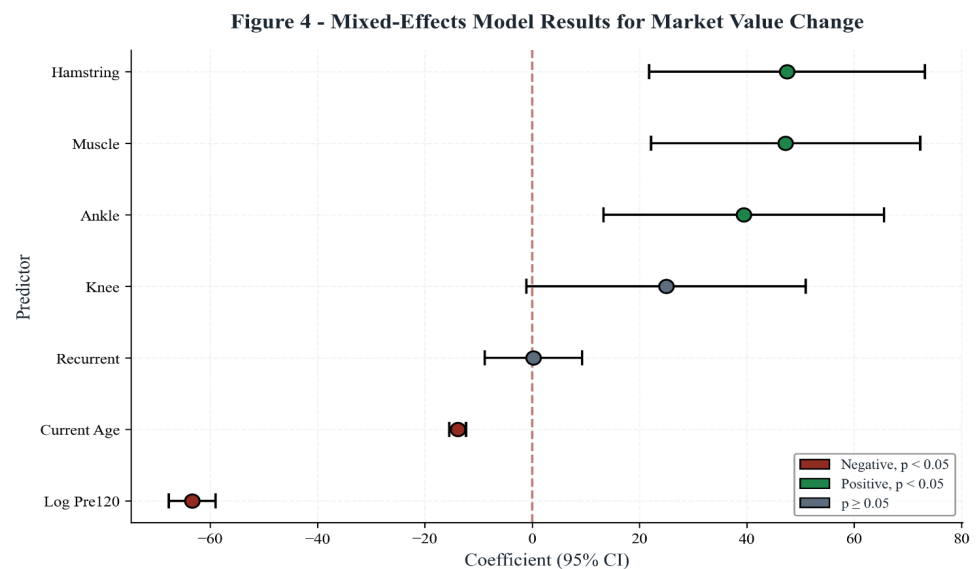
Notes: CI = confidence interval. The model includes a random intercept by player. Reference category for injury type is ‘Other’.

Feature importance from Random Forest

To complement the mixed-effects model and provide a non-parametric assessment of predictor importance, a Random Forest regressor was trained on the same predictor set. The model achieved a 5-fold cross-validated R^2 of 0.260 ($SD = 0.015$). Feature importance, measured as the mean decrease in impurity across trees, identified pre-injury market value (pre_{120}) as the most important predictor, with an importance score of 0.427. Current age at

injury was the second most important predictor, with an importance score of 0.339. Injury duration ($duration_days$, importance = 0.066), position code (importance = 0.064), and the games missed proxy (importance = 0.064) showed moderate importance. Injury type ($injury_code$, importance = 0.030) and recurrence status ($recurrent$, importance = 0.010) contributed the least to the model’s predictive accuracy.

Figure 4. Mixed-effects model results for market value change



Robustness checks

Five sensitivity analyses confirmed the stability of the main findings. Quantile regression at the 25th, 50th, and 75th percentiles showed convergence only at the median, where the recurrence coefficient aligned with the main model. Excluding COVID-era injuries yielded a larger recurrence

penalty (−16.92 versus −10.24), suggesting possible attenuation during the pandemic. Using the recovery ratio as an alternative outcome produced a recurrence coefficient of −0.0543, consistent with an 8.0% penalty. Position-stratified models revealed substantial variation, with goalkeepers (−22.7) and central midfielders (−24.8) showing larger penalties

than centre-backs (−3.0). Restricting the analysis to severe injuries (>28 days) produced a recurrence coefficient of −10.96, similar to the main estimate. Overall, the findings were robust across alternative specifications.

Diagnostic checks

Several diagnostic tests were performed to evaluate the assumptions of the mixed-effects model. Residual normality was assessed using the Shapiro–Wilk Test, applied to a random sample of 5,000 residuals due to the test’s sensitivity to large sample sizes. The test indicated significant deviation from normality (statistic = 0.6829, $p < 0.001$). Heteroscedasticity was evaluated using the Breusch–Pagan Test, which also showed significant evidence of

non-constant variance ($p < 0.001$). Consequently, robust standard errors were used for all regression models to mitigate the impact of these violations.

Multicollinearity was assessed using variance inflation factors (VIF), with values exceeding five considered indicative of problematic collinearity. As shown in Table 3, current age and age squared showed high VIF values (438.8 and 132.3, respectively), which is expected given the quadratic specification. Log pre-injury value also showed moderate collinearity (VIF = 115.7). Recurrence had a VIF of 1.3, indicating no collinearity with other predictors. The high VIF values for age and age squared do not bias coefficient estimates but inflate standard errors; this was accepted because the quadratic term was necessary to evaluate the non-linear age hypothesis.

Table 3. Variance inflation factors

Variable	VIF
Recurrent	1.3
Current age	438.8
Age squared	132.3
log(Pre-injury value)	115.7

4. DISCUSSION

This study examined whether injury type and recurrence predict post-injury market value changes in professional football players, using a retrospective cohort of 12,948 injury events from 2,061 players across top-five European leagues between 2008 and 2024. Three main findings emerged.

First, ACL and Achilles injuries were uniquely associated with market value declines. ACL injuries produced a median decline of −6.0%, while Achilles injuries declined by −15.5%. In contrast, all other injury types – including hamstring, ankle, knee, and muscle injuries – demonstrated positive or neutral median changes within the 120-day observation window. Second, recurrent injuries incurred an additional penalty of 8.0 percentage points beyond first-time injuries, suggesting that the market prices show future fragility rather than only current impairment. Third, age strongly moderated outcomes. Players under 24 years showed a median change of +39.0% compared to −31.0% for players aged 30 and over, a difference of 70.0 percentage points. Pre-injury market value and age were the strongest predictors in both the mixed-effects model and the Random Forest analysis.

It is important to distinguish between market valuation effects and clinical or performance

outcomes. The finding that hamstring and muscle injuries show no median value penalty within 120 days after return does not imply that these injuries are clinically inconsequential. On the contrary, hamstring injuries are among the most frequent and recurrent injuries in professional football (Ekstrand et al., 2016), with a recurrence rate of 22% in this sample. A player who suffers a hamstring injury may experience reduced sprint performance, altered biomechanics, or increased susceptibility to future episodes, even if their market value does not immediately decline. The present study captures the immediate valuation response but does not measure clinical outcomes directly.

The divergent findings for ACL injuries (−6.0%) and hamstring injuries (0.0%) warrant explanation. One possibility is that the market prices the differential return-to-play rates. Research by Ardern and colleagues (2014) found that only 65% of footballers return to their pre-injury performance level following ACL reconstruction, whereas hamstring injuries have return-to-play rates exceeding 90% (Ekstrand et al., 2016). Beyond recovery rates, ACL injuries typically require nine to twelve months of rehabilitation, during which time a player’s visibility, sponsorship opportunities, and performance metrics all decline substantially. For Achilles injuries, the even larger penalty of −15.5%

may reflect the older typical age of affected players – mean 24.9 years versus 23.4 years for ACL injuries – combined with higher rates of persistent performance deficits following return. However, the small number of Achilles injuries in the sample, only 80 events, means this estimate should be interpreted cautiously.

The positive coefficients observed for ankle, knee, hamstring, and muscle injuries require cautious interpretation. These findings should not be interpreted as injuries being beneficial. Several alternative explanations should be considered. First, younger players who suffer minor injuries may experience normal age-related value appreciation that offsets any injury penalty – a form of confounding by age. Second, selection effects may be at play: clubs may be more likely to retain and invest in players they believe have high future potential, and these players may also be more likely to suffer injuries due to higher playing time. Third, the market may systematically overestimate recovery prospects for certain injury types, leading to overvaluation that is later corrected. The present study cannot distinguish between these explanations, and the positive coefficients should not be interpreted causally.

The 8.0 percentage point recurrence penalty suggests that the market treats a second identical injury as qualitatively different from the first. This aligns with clinical evidence showing that recurrent injuries have longer recovery times and higher rates of subsequent recurrence (Ekstrand et al., 2016). The market appears to infer from recurrence an underlying vulnerability – whether biomechanical, such as poor hamstring flexibility; behavioural, such as inadequate rehabilitation compliance; or positional, such as a high-risk playing style – that reduces expected future availability.

Importantly, the recurrence penalty was consistent across injury types in the main analysis, but position-stratified models revealed substantial variation. Goalkeepers and central midfielders showed large penalties of –22.7 and –24.8 percentage points respectively, while centre-backs showed a small penalty of only –3.0 percentage points. This variation may reflect differences in the availability of substitute players. A club can more easily replace a centre-back than a goalkeeper or central midfielder, so the market may penalise recurrence less severely at positions where replacement is cheaper.

The 70.0 percentage point difference between players under 24 and players aged 30 and over is striking. Several explanations are plausible. Younger players have more remaining career years over

which to recover value, so a given injury destroys a smaller proportion of their lifetime earnings potential. They may also have greater physiological reserve and faster recovery times. Furthermore, the market may view injuries in young players as transient setbacks, while injuries in veterans are seen as the beginning of age-related decline.

The quadratic regression revealed a U-shaped relationship with the lowest point at 27.4 years, indicating that players in their late twenties – typically at peak market value – suffer the greatest proportional losses. This aligns with human capital theory: the most valuable players, who are generally at peak age with high pre-injury value, experience the largest absolute and relative declines when injured.

Connection to previous research

These findings confirm and extend existing work. Consistent with Rubio Martín et al. (2025), injuries reduce market value. However, while their daily penalty approach found an 11.2% penalty for severe injuries regardless of type, the present study reveals substantial heterogeneity: an ACL injury lasting a median of 212 days produced only a –6.0% penalty, whereas an Achilles injury lasting a median of 70 days produced a –15.5% penalty. This suggests that the market responds to qualitative differences in injury prognosis, particularly the likelihood of full recovery, rather than duration alone.

Regarding recurrence, no prior study has quantified the additive penalty. The 8.0 percentage point penalty is comparable in magnitude to the severe injury penalty identified by Rubio Martín and colleagues (2025), indicating that recurrence is as economically important as severity. The present findings thus fill the gap explicitly identified by Poli et al. (2020), who noted that physical problems were a missing factor in the CIES Football Observatory's transfer valuation model.

The findings also extend the team-level analysis of Eliakim et al. (2020), who estimated that an average English Premier League club loses £45 million per season due to injury-related underperformance. The present study provides the micro-foundations for these aggregate losses. A €50 million player with an ACL injury loses approximately €3 million in value; with a recurrent injury, an additional €4 million. Summed across a squad, these player-level discounts plausibly aggregate to the team-level losses Eliakim and colleagues documented.

Finally, the findings connect to the workload literature. Den Hollander et al. (2024) demonstrated that fixture congestion and international travel

increase injury susceptibility. The present study shows that the injuries sustained during such congested periods, particularly ACL injuries and recurrent episodes, carry substantial and measurable economic penalties. This strengthens the economic case for fixture congestion reforms: reducing player workload would not only protect player health but also preserve club asset values.

Theoretical implications

These findings extend human capital theory in sports economics (Becker, 1964; Rosen & Sander-son, 2001). Human capital theory traditionally conceptualises player value as a function of age, performance metrics, and contract length. The present study demonstrates that injury history – specifically, injury type and recurrence – is an independent predictor that operates beyond current performance. This suggests that human capital in football has a distinct fragility dimension that is separable from both age and skill. A 25-year-old player with three prior hamstring strains is not economically equivalent to a 25-year-old with no injury history, even if their current performance metrics are identical.

Moreover, the finding that the recurrence penalty of 8.0 percentage points is comparable in magnitude to the penalty for a first-time severe injury suggests that the market prices the variance of future productivity, not merely its expected level. A player with a recurrent injury history has not only lower expected future availability but also higher uncertainty around that expectation, and the market appears to discount for this uncertainty.

Practical implications

The findings of this observational study offer preliminary evidence for considering injury history in player valuation, though these implications should be treated as exploratory given the study's limitations. The results suggest that injury type and recurrence are associated with market value changes within 120 days after return to play, but these associations should not be interpreted as causal. Clubs, agents, and insurers may find the estimated discount factors useful for informing their assessments, but decisions should incorporate clinical judgment, performance data, and individual circumstances that this study does not capture.

Limitations and future research

Several limitations should be acknowledged. First, Transfermarkt market values are crowd-sourced estimates rather than actual transfer fees,

although prior validation studies demonstrate a strong correlation (above 0.85) between Transfer-market values and actual transfer fees (Herm et al., 2016). The fully anonymised dataset also prevents individual-level validation or case-study triangulation, though this is offset by the large sample size (nearly 13,000 injuries), which provides adequate statistical power for group-level inferences. Additionally, the absence of data on training load, medical history, psychological factors, and rehabilitation protocols may confound the recurrence penalty, as players prone to recurrence may possess unmeasured biomechanical vulnerabilities. The mixed-effects model partially mitigates this by including a random intercept for each player, absorbing time-invariant unobserved heterogeneity.

The post-COVID period (2020–2024) may reflect different market dynamics, although a sensitivity analysis excluding COVID-era injuries produced a recurrence coefficient of -16.92 , consistent with the main finding. High multicollinearity between age and age squared was detected ($VIF = 438.8$), which is expected given the quadratic specification and does not bias coefficient estimates, though it inflates standard errors. The small sample size for Achilles injuries (only 80 events) means the -15.5% estimate has wide uncertainty and should therefore be considered exploratory. Finally, the positive coefficients for several injury types should be interpreted cautiously, as they may reflect confounding by age, selection effects, or market optimism rather than true beneficial effects of injury. The observational design precludes causal inference, and unmeasured confounders such as player performance metrics, contract status, and club strategy may influence both injury occurrence and market valuation.

To address these gaps, prospective studies linking detailed medical data to market outcomes could identify which clinical factors predict value recovery and which are modifiable by clubs. Similarly, researchers should examine whether the recurrence penalty is fully rational or reflects behavioural biases such as the availability heuristic, where investors overreact to vivid, recent injuries. Comparing market value changes to actual post-injury performance metrics would test market efficiency. Longer-term follow-up studies extending beyond the 365-day recurrence window and the 120-day value window used here could determine whether the recurrence penalty persists indefinitely or diminishes after a prolonged injury-free period, and whether the absence of a penalty for hamstring and muscle

injuries within 120 days reflects genuine market indifference or merely a delayed response to clinical consequences that emerge later. Finally, machine learning models incorporating additional predictors could improve predictive accuracy beyond the modest R^2 achieved in this study, potentially using gradient boosting or neural network architectures.

5. CONCLUSIONS

The central contribution of this study is the demonstration that injury type and recurrence are associated with player valuation changes within 120 days after return to play. In this retrospective cohort of 12,948 injury events from 2,061 professional footballers across top-five European leagues, ACL injuries showed a median decline of -6.0% and Achilles injuries a decline of -15.5% in Transfermarkt market value. In contrast, hamstring and non-hamstring muscle injuries were not associated with a median reduction in Transfermarkt market value within 120 days after return to play. Recurrence imposed an additional discount of approximately 8.0 percentage points, and players under 24 recovered 70.0 percentage points better than those aged 30 and over.

These findings fill the gap explicitly identified by the CIES Football Observatory and provide the first evidence-based discount factors for injury-type-specific and recurrence-adjusted valuation within a 120-day post-return window. However, these findings reflect short- to medium-term market valuation and should not be interpreted as evidence about long-term clinical outcomes, performance, or career consequences. Further research is needed to examine longer follow-up periods, incorporate clinical data, and evaluate whether the recurrence penalty persists over time.

Author Contributions: Conceptualisation, J.P.; methodology, J.P.; software, J.P.; validation, J.P. and L.G.V.; formal analysis, J.P.; investigation, J.P. and L.G.V.; resources, A.S.P., A.A., and M.P.-C.; data curation, J.P.; writing—original draft preparation, J.P.; writing—review and editing, A.S.P., A.A., M.P.-C., and L.G.V.; visualisation, J.P.; supervision, A.S.P., A.A., and M.P.-C.; project administration, A.S.P., A.A., and M.P.-C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Clinical Trial Registration: Not applicable.

Informed Consent Statement: Not applicable. This study used only fully anonymised, publicly available secondary data. No direct interaction with human

subjects occurred, and no identifiable private information was accessed.

Ethical Approval: Ethical approval was waived because the study used only publicly available, anonymised secondary data (Kaggle dataset derived from Transfermarkt records). All player and club identifiers were anonymised before release, making reidentification impossible. The study was conducted in compliance with the Declaration of Helsinki, specifically the provisions for research using anonymised data.

Acknowledgements: None.

Conflicts of Interest: The authors declare no conflict of interest.

Data Availability Statement: The data that support the findings of this study are openly available at <https://doi.org/10.6084/m9.figshare.32351964>.

AI Use Statement: No generative AI and AI-assisted technologies were used in the research or the drafting of the manuscript.

REFERENCES

- Ajjawi, O. (2025, August 20). *The Premier League's hidden drain: How injuries cost millions*. Off The Pitch. <https://www.offthepitch.com>
- Ardern, C. L., Taylor, N. F., Feller, J. A., & Webster, K. E. (2014). Fifty-five per cent return to competitive sport following anterior cruciate ligament reconstruction surgery: An updated systematic review and meta-analysis including aspects of physical functioning and contextual factors. *British Journal of Sports Medicine*, 48(21), 1543–1552. <https://doi.org/10.1136/bjsports-2013-093398>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Becker, G. S. (1964). *Human capital: A theoretical and empirical analysis, with special reference to education*. University of Chicago Press.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, 47(5), 1287–1294. <https://doi.org/10.2307/1911963>
- Bryson, A., Frick, B., & Simmons, R. (2013). The returns to scarce talent: Footedness and player remuneration in European soccer. *Journal of Sports Economics*, 14(6), 606–628. <https://doi.org/10.1177/1527002511435118>
- Carling, C., McCall, A., Le Gall, F., & Dupont, G. (2018). Injury risk and patterns in newly transferred football players: A case study of 8 seasons from a professional football club. *Science and Medicine in Football*, 2(1), 47–50. <https://doi.org/10.1080/24733938.2017.1370123>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- den Hollander, S., Kerkhoffs, G., & Gouttebarge, V. (2024). The impact of match workload and interna-

tional travel on injuries in professional men's football. *Sports*, 12(8), Article 212. <https://doi.org/10.3390/sports12080212>

Ekstrand, J., Waldén, M., & Hägglund, M. (2016). Hamstring injuries have increased by 4% annually in men's professional football, since 2001: A 13-year longitudinal analysis of the UEFA Elite Club injury study. *British Journal of Sports Medicine*, 50(12), 731–737. <https://doi.org/10.1136/bjsports-2015-095359>

Eliakim, E., Morgulev, E., Lidor, R., & Meckel, Y. (2020). Estimation of injury costs: Financial damage of English Premier League teams' underachievement due to injuries. *BMJ Open Sport & Exercise Medicine*, 6(1), Article e000765. <https://doi.org/10.1136/bmjsem-2019-000675>
FIFA. (2025). *Global transfer report 2024*. <https://digitalhub.fifa.com/m/142077cfbb75c2b0/original/Global-Transfer-Report-2024.pdf>

Gümüş, H. (2026). *Football player injuries and market value* [Dataset]. Kaggle. <https://www.kaggle.com/datasets/gumush/football-player-injuries-market-value/data>

Herm, S., Callsen-Bracker, H.-M., & Kreis, H. (2014). When the crowd evaluates soccer players' market values: Accuracy and evaluation attributes of an online community. *Sport Management Review*, 17(4), 484–492. <https://doi.org/10.1016/j.smr.2013.12.006>

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R*. Springer.

Koenker, R., & Hallock, K. F. (2001). Quantile regression. *Journal of Economic Perspectives*, 15(4), 143–156.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2), 133–142. <https://doi.org/10.1111/j.2041-210x.2012.00261.x>

Poli, R., Ravenel, L., & Besson, R. (2020). Scientific evaluation of the transfer value of football players. *CIES Football Observatory Monthly Report*, (53), 1–5. <https://libra.unine.ch/handle/20.500.14713/29544>

Rosen, S., & Sanderson, A. (2001). Labour markets in professional sports. *The Economic Journal*, 111(469), F47–F68. <https://www.jstor.org/stable/2667957>

Rubio Martín, G., González Sánchez, F. J., & Manuel García, C. M. (2025). The financial dimension of football injuries according to the wisdom of crowds. *Managing Sport and Leisure*, 30(5), 986–997. <https://doi.org/10.1080/23750472.2024.2342262>

Sedgwick, P. (2014). Retrospective cohort studies: Advantages and disadvantages. *BMJ*, 348, Article g1072. <https://doi.org/10.1136/bmj.g1072>

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4), 591–611. <https://doi.org/10.2307/2333709>

van Buuren, S. (2018). *Flexible imputation of missing data* (2nd ed.). CRC Press.

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>

Received on May 22, 2026
Accepted on August 3, 2026