Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Repeated Measures of Burnout and Well-Being as Early Indicators of Later Athlete Burnout

Nathanael Adu
Mampong Technical College of Education
Kofi Nyantakyi Appiah
nyantakyi.appiah@gmail.com
Wesley College of Education
Neelam Sharma
Lovely Professional University

Research Article

Keywords: athlete burnout, risk prediction, machine learning, longitudinal study, sport psychology

Posted Date: April 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9441649/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background: This study examined whether baseline levels and short-term changes in burnout, well-being, sleep, and health symptoms were associated with later high athlete burnout, with the practical aim of identifying early psychological indicators of risk.

Method: We conducted a prediction-focused secondary analysis of a three-wave longitudinal dataset of 267 competitive athletes. Burnout, depressive symptoms, sleep quality, life satisfaction, physical symptoms, and illness were assessed at baseline (T1), three months (T2), and six months (T3). High T3 burnout was defined as scoring at or above the 75th percentile of the burnout distribution. Multivariable logistic regression, lasso-regularized logistic regression, and random forest models were fit with repeated 10×10-fold cross-validation. Performance metrics included area under the receiver operating characteristic curve (AUC), Brier score, and calibration; sensitivity analyses and falsification tests assessed robustness.

Results: The lasso model achieved the best internal performance (AUC ≈ .82; Brier ≈ .15). Baseline burnout and early increases in burnout were the dominant predictors; depressive symptoms, sleep, physical symptoms, and demographics added limited incremental value. The random forest did not outperform penalized logistic regression. Sensitivity and falsification analyses supported robustness of findings.

Conclusion: Repeated burnout assessments may help identify athletes whose psychological experience is shifting toward higher burnout risk, supporting earlier monitoring and intervention in applied sport settings.

Trial registration: Not applicable

Background

Athlete burnout is a pervasive concern in competitive sport and has been linked to impaired performance, mental health problems, and dropout from sport [1, 2]. Conceptually, burnout reflects a multidimensional syndrome of emotional and physical exhaustion, reduced sense of accomplishment, and sport devaluation that develops in response to chronic stress and insufficient recovery [3]. Empirical work over the past two decades has documented high rates of burnout indicators among athletes across sports and competitive levels and has associated burnout with elevated depressive symptoms, sleep disturbance, and lower well-being [1, 4]. Despite this, relatively little research has examined which early psychological changes signal that athletes are moving toward higher burnout over time [5].

Most longitudinal studies in sport psychology have emphasized explanatory questions such as whether stress, coping, motivation, or interpersonal factors predict changes in burnout using regression, structural equation modeling, or cross-lagged panel designs [6, 7]. These approaches are invaluable for theory testing but are not optimized for answering applied questions about who, among a squad of

athletes, is at highest risk of severe burnout [8]. In contrast, a predictive approach can show which combinations of burnout, well-being, sleep, and health symptoms most clearly track later psychological risk in athletes [9–11]. In clinical medicine and psychiatry, multivariable prediction models are increasingly used to support screening and risk stratification; however, comparable tools for athlete burnout risk are largely absent [12].

Several psychological and health indicators have been linked to burnout and may help identify athletes who are on a worsening trajectory. Cross-sectional and longitudinal research has linked higher burnout to greater depressive symptoms, poorer sleep quality, more physical complaints and illness, and lower life satisfaction or well-being [13–16]. Demographic and sport-context factors such as age, sex, sport type (team vs. individual), injury, and training load also appear to relate to burnout risk, although findings are mixed [17, 18]. Importantly, most prior work has examined these variables individually or in small sets, without asking how they jointly perform in multivariable, out-of-sample prediction or whether more complex machine-learning methods meaningfully outperform traditional regression in this context.

At the same time, researchers in psychology and sport science have increasingly emphasized the value of distinguishing mechanisms that explain burnout from measures that can identify athletes at rising risk, to use appropriate internal validation, and to report performance metrics such as discrimination and calibration when developing prediction models [19–21]. Regularized regression approaches such as the lasso and elastic net can handle correlated predictors and reduce overfitting in modest samples, while tree-based ensemble methods such as random forests can capture non-linearities and interactions, albeit often at the cost of interpretability [22, 23]. There is increasing recognition that in many psychological datasets, carefully tuned regularized generalized linear models may perform as well as or better than more complex machine-learning algorithms [24, 25]. However, these insights have rarely been applied systematically to athlete burnout.

The present study addresses these gaps by conducting a prediction-focused secondary analysis of a three-wave longitudinal dataset of competitive athletes, originally collected to examine reciprocal relations between burnout and health-related outcomes [14]. The parent study used a random-intercept cross-lagged panel model to show that increases in emotional and physical exhaustion were associated with later depressive symptoms and decrements in well-being [14]. Building on this foundation, we shift the focus from population-level associations to individual-level risk stratification by asking: to what extent can baseline levels and short-term changes in burnout, depressive symptoms, sleep, life satisfaction, and physical symptoms predict which athletes will experience high burnout six months later?

We examined which early psychological indicators were most strongly associated with later high burnout, using repeated k-fold cross-validation to obtain honest estimates of out-of-sample performance [26, 27]. By evaluating discrimination (area under the receiver operating characteristic curve), calibration, and overall prediction error, we aim to determine whether more complex machine-learning models offer meaningful advantages over regularized logistic regression in this context [28, 29]. In addition, we

conduct planned sensitivity analyses (baseline-only models, alternative high-burnout thresholds, depressive-symptom outcome) and falsification tests (time-reversed and shuffled outcomes) to assess robustness and guard against overfitting [30, 31].

This work makes three primary contributions. First, it provides one of the first longitudinal examinations of how repeated burnout and well-being measures relate to later high burnout in competitive athletes, using a rich longitudinal dataset with repeated measures of burnout and key psychological and health indicators [14]. Second, it illustrates how careful longitudinal modeling can be used to identify which early psychological changes matter most for later burnout in sport-psychology research to complement traditional explanatory analyses [9, 32]. Third, by directly comparing logistic regression, lasso, and random forest models, the study informs ongoing debates about the relative value of interpretable versus more complex algorithms for applied decision-making in sport settings [33].

Based on prior evidence that burnout levels and trajectories are closely tied to later ill-being, we hypothesized that (a) models including both baseline burnout and early changes in burnout would achieve at least acceptable discrimination (AUC ≥ .75) for predicting high T3 burnout, (b) penalized logistic regression would perform as well as or better than standard logistic regression and at least competitively with random forests, and (c) additional predictors (depressive symptoms, sleep, physical symptoms, life satisfaction, and basic demographics) would add only modest incremental prediction beyond the burnout variables themselves [34–36]. By testing these hypotheses, we aim to provide practically useful insights into how simple, repeated questionnaire measures can be leveraged to identify athletes at greatest risk of progressing to high burnout, thereby supporting early, targeted intervention in applied sport environments.

Methods

Study design and data source

This study used a secondary analysis of a pre-existing three-wave longitudinal dataset of competitive athletes that is publicly available via PsychArchives under a Scientific Use License [37]. The parent study followed 267 athletes across approximately six months with assessments at baseline (T1), three months (T2), and six months (T3) to examine longitudinal associations between athlete burnout and mental and physical health indicators [14]. The original investigation employed a random-intercept cross-lagged panel model to disaggregate within- and between-person effects of burnout and health outcomes over time [14].

Participants

The original sample comprised 267 competitive athletes aged 18 years and older who were recruited via sport organizations and clubs and who provided informed consent to participate [14]. Athletes were currently engaged in organized training and competition and represented both team and individual

sports across a range of performance levels, consistent with recommendations for ecologically valid sport psychology research [38]. Inclusion criteria in the parent project were being at least 18 years old, currently participating in competitive sport, and sufficient proficiency in the study language to complete online questionnaires; athletes were excluded if they reported conditions that precluded questionnaire completion or if they withdrew consent at any wave [14]. For the present secondary analysis, we included all athletes with available data on the T3 burnout outcome and at least baseline burnout, depressive symptoms, sleep, life satisfaction, and covariates; complete-case samples for specific models are reported in the Statistical Analysis subsection.

Demographic and sport-related characteristics used as covariates included biological sex (1 = male, 2 = female), age in years, sport type (1 = individual, 2 = team), years of competitive experience, current injury status, current use of pain medication, and self-reported history of mental disorder [14]. These variables were derived directly from the archived dataset and treated as manifest indicators.

Measures

Burnout

Athlete burnout was assessed at each wave using a multidimensional inventory with subscales capturing emotional and physical exhaustion, sport devaluation, and reduced sense of accomplishment [3]. For each assessment, we used the total burnout score and the exhaustion, devaluation, and reduced accomplishment subscale scores (t1BURN, t1EXH, t1DEV, t1RSA; and corresponding T2 and T3 variables). Higher scores indicated greater burnout. Prior work has shown acceptable internal consistency ($\alpha \approx .80–.90$) and construct validity for this measure in athlete samples [3, 39].

Depressive symptoms

Depressive symptoms were measured at each wave using a self-report scale indexing the frequency of core depressive experiences over the past two weeks (t1DS, t2DS, t3DS), with higher scores reflecting greater symptom severity [40]. The underlying instrument has demonstrated good reliability and validity in both athlete and general populations.

Physical symptoms and illnes **s**

Physical symptom burden and illness were assessed with a physical symptoms scale (t1PS, t2PS, t3PS) and a cold/illness symptom inventory based on the Wisconsin Upper Respiratory Symptom Survey (t1WURSSg, t2WURSSg, t3WURSSg, and brief cold severity items; [41]. These measures capture the severity and impact of somatic complaints and upper respiratory symptoms over the preceding days and have shown acceptable psychometric properties in prior research [41].

Sleep

Sleep disturbances were measured with the Pittsburgh Sleep Quality Index (PSQI) components at each wave (PSQI components 1–7 and the global and graded scores; e.g., t1PSQI_c1–c7, t1PSQIg, t1PSQIgr; [42]. Higher scores indicate poorer sleep quality. The PSQI is widely used in sport and clinical populations and has demonstrated good internal consistency, test-retest reliability, and convergent validity [43, 44].

Life satisfaction

Life satisfaction was assessed with a multi-item scale at each wave (t1LS, t2LS, t3LS); higher scores indicate greater satisfaction with life. The items correspond to a standardized measure of global life satisfaction with well-documented reliability and validity [45, 46].

Training load and medication

At each wave, athletes reported their average weekly training load (e.g., hours or sessions per week; t1trainingload, t2trainingload, t3trainingload) and current use of medication (t1medication, t2medication, t3medication). These variables are commonly used contextual indicators in sport-science research to characterize athlete workload and health status [47–49]. Training load and medication were used descriptively and in sensitivity analyses, but not as primary predictors in the final models.

Derived variables

To capture early changes in key constructs, we computed change scores between T1 and T2 for burnout, depressive symptoms, sleep, and life satisfaction. Specifically, we calculated Δburnout (d_BURN_12 = t2BURN − t1BURN), Δdepressive symptoms (d_DS_12 = t2DS − t1DS), Δsleep disturbance (d_PSQI_12 = t2PSQIg − t1PSQIg), and Δlife satisfaction (d_LS_12 = t2LS − t1LS). Change scores are a straightforward way to index short-term within-person change in longitudinal studies when there are few time points [50]. Analogous change scores from T2 to T3 were created for use in sensitivity and placebo analyses.

To operationalize clinically relevant high burnout at T3, we defined a binary outcome variable indicating whether an athlete scored at or above the 75th percentile of the T3 burnout distribution (t3_high_burn = 1) versus lower burnout (t3_high_burn = 0). Upper-quartile thresholds are commonly used in risk-stratification analyses when validated cut-points are unavailable [51].

Design and analytic framework

The present study was explicitly framed as a prediction-focused, longitudinal observational analysis rather than a causal inference study, consistent with recommendations to distinguish predictive from explanatory aims in psychological and clinical research [52, 53]. We sought to evaluate the extent to which baseline levels and early changes in burnout, depressive symptoms, sleep, life satisfaction, and physical symptoms predict high burnout at six months, and whether more complex machine-learning

models outperform traditional logistic regression while maintaining interpretability [14]. Our analytic framework followed best-practice guidelines for developing and evaluating clinical prediction models, including explicit outcome definition, clear predictor specification, appropriate handling of missing data, internal cross-validation, and comprehensive reporting of discrimination and calibration [11, 54].

Statistical analysis

All analyses were conducted in R (version 4.5.2) using standard packages for regression, machine learning, cross-validation, and imputation [55]. We began by summarizing the distributions of all variables and examining missing-data patterns across waves [56]. The primary analysis sample comprised athletes with non-missing T3 burnout scores and complete data on all selected baseline predictors and early change scores; this resulted in 137 complete cases for the main multivariable models, which is acceptable for prediction-oriented logistic models with regularization [57].

For the primary prediction model, we used penalized logistic regression via glmnet with lasso regularization, and we fit standard multivariable logistic regression and a non-linear random forest classifier as comparative models [58]. Penalized regression is recommended for prediction models with multiple correlated predictors to reduce overfitting and improve generalizability [59]. All models included baseline burnout, depressive symptoms, global sleep disturbance, life satisfaction, physical symptoms, upper respiratory symptoms, early change scores in burnout, sleep, and life satisfaction, and the demographic covariates biological sex, age, sport type, competitive experience, injury status, and history of mental disorder.

To avoid optimistic performance estimates and provide a fair comparison of algorithms, we used repeated 10-fold cross-validation (10 folds, 10 repeats) for internal model tuning and performance evaluation, as recommended for internal validation of prediction models [60, 61]. For the glmnet models, we tuned the mixing parameter (α) and penalty (λ) over a grid and selected the combination that maximized mean cross-validated area under the receiver operating characteristic curve (AUC). For random forests, we tuned the number of variables randomly sampled at each split (mtry) while fixing the number of trees at 1,000, consistent with common practice in classification tasks. Logistic regression was fit without penalization, and its cross-validated performance was computed within the same resampling framework using the caret infrastructure [62].

Discrimination was quantified using AUC, averaged across all resamples [63]. For interpretability, we also reported mean sensitivity and specificity at the default 0.5 probability threshold. To evaluate calibration and overall predictive accuracy, we obtained cross-validated predicted probabilities for the final glmnet model and computed the Brier score (mean squared error between predicted probabilities and observed outcomes) and visualized calibration by plotting observed versus predicted event rates across deciles of predicted risk [64, 65].

We conducted several planned sensitivity analyses. First, we estimated a baseline-only glmnet model excluding all early change scores to quantify the incremental value of longitudinal information [66].

Second, we redefined the high-burnout outcome using an alternative cut-off (70th percentile) and refit the glmnet models to examine robustness to threshold choice, as recommended when clinically validated cut-points are not available [51]. Third, we repeated models using high depressive symptoms at T3 as an alternative outcome to gauge whether predictor patterns generalized across mental-health endpoints in athletes.

To probe potential overfitting and model specificity, we performed two falsification (negative-control) tests. In a time-reversal placebo model, we attempted to predict high burnout at T1 from T3 burnout, T3 sleep, and T3 life satisfaction; discrimination in this model was expected to approximate chance if temporal ordering is meaningful [67]. In a shuffled-outcome model, we randomly permuted the high-burnout labels and refit the glmnet models; AUCs near 0.5 in this analysis would support that the original predictive performance derived from genuine predictor-outcome structure rather than artefacts.

Missing data were handled in two complementary ways. The primary analyses used complete-case datasets within each model, with the extent and pattern of missingness described in the Results [68]. In a robustness check, we implemented multiple imputation by chained equations (20 imputations) including all predictors and the T3 burnout outcome in the imputation model and refit the main glmnet model on imputed datasets [69]. Coefficient patterns and AUCs were compared qualitatively to the complete-case results to ensure that conclusions were not driven by selective attrition [70].

All statistical tests were two-sided with an alpha level of .05 for logistic regression coefficients. For prediction performance, we emphasized effect sizes (AUC, Brier score, sensitivity, specificity) and their uncertainty rather than null-hypothesis significance testing, in line with current recommendations for clinical prediction research and modern reporting standards in psychology [71, 72]. Analytic code and variable definitions will be made available in an online repository upon publication, and we will provide links to the original data DOI to enable independent replication, consistent with transparency and openness principles [73].

Results

Participant flow and analytic samples

Of the 267 athletes who completed the baseline survey, 171 provided data at T2 and 171 at T3 [14]. After restricting to cases with non-missing T3 burnout and complete data on all baseline predictors and early change scores, 137 athletes (51.3%) were available for the primary prediction models [68]. For baseline-only models (no change scores), 158 athletes had complete data. Multiple-imputation analyses used all 267 athletes with at least partial data on outcomes and predictors [74, 75].

Descriptive analysis

At baseline, athletes reported moderate burnout (M = 1.22, SD = 0.50; possible range 0–4) and depressive symptoms (M = 14.62, SD = 7.54; range 2–45) [14]. Sleep quality was generally poor (PSQI

global M = 8.77, SD = 2.53; range 3–15), whereas life satisfaction was moderately high (M = 23.64, SD = 5.57; range 5–35). Early changes from T1 to T2 were small on average but heterogeneous: Δburnout (d_BURN_12) had M = 0.18 (SD = 0.57; range − 1.27 to 2.47), Δsleep disturbance (d_PSQI_12) had M = 0.26 (SD = 2.19; range − 7 to 8), and Δlife satisfaction (d_LS_12) had M = − 0.22 (SD = 4.73; range − 24 to 19).

High T3 burnout was defined as scoring at or above the 75th percentile of the T3 total burnout distribution (cut-off ≈ 1.83), yielding 27 high-burnout athletes (19.7%) and 110 lower-burnout athletes (80.3%) in the complete-case sample, a distribution suitable for risk prediction modeling.

Table 1 displays descriptive statistics for all predictors and the high-burnout outcome in the complete-case sample.

Table 1
Descriptive statistics for predictors and T3 high burnout (complete-case sample, n = 137)

Variable	M	SD	Min	Max
Baseline burnout (t1BURN)	1.22	0.50	0.00	3.27
Baseline depressive symptoms (t1DS)	14.62	7.54	2	45
Baseline PSQI global (t1PSQIg)	8.77	2.53	3	15
Baseline life satisfaction (t1LS)	23.64	5.57	5	35
Baseline physical symptoms (t1PS)	39.29	15.80	18	112
Baseline WURSS global (t1WURSSg)	18.37	14.62	0	70
Δburnout T1–T2 (d_BURN_12)	0.18	0.57	−1.27	2.47
ΔPSQI global T1–T2 (d_PSQI_12)	0.26	2.19	−7	8
Δlife satisfaction T1–T2 (d_LS_12)	−0.22	4.73	−24	19
Age (years)	20.87	3.47	18	57
Competitive experience (years)	8.80	6.81	0	40
T3 high burnout (≥ 75th percentile)	0.20	—	0	1

Note. T3 high burnout coded 1 = high burnout, 0 = lower burnout.

Primary multivariable logistic regression

The full multivariable logistic regression including all baseline predictors, change scores, and covariates fit well (residual deviance = 104.04, df = 122). Baseline burnout (b = 3.16, SE = 0.74, p < .001) and early increase in burnout from T1 to T2 (b = 1.73, SE = 0.66, p = .009) were statistically significant predictors of

high T3 burnout. Baseline depressive symptoms, sleep, life satisfaction, physical/illness symptoms, and demographic covariates were not significant when modelled jointly.

Using repeated 10-fold cross-validation (10 repeats); this logistic regression achieved a mean AUC of 0.78, with high sensitivity (0.86) but modest specificity (0.51) at the 0.50 probability threshold.

Penalized regression and random forest models

Penalized logistic regression via glmnet produced slightly better discrimination and more parsimonious models. Cross-validated AUCs across α–λ combinations ranged from approximately 0.78 to 0.82. The optimal model selected by caret used α = 1 (lasso) and λ = 0.044, yielding a cross-validated AUC of 0.817 and Brier score of 0.147 based on pooled out-of-fold predictions, indicating good discrimination and acceptable calibration [76].

In this final lasso model, non-zero coefficients were retained for baseline burnout, early change in burnout, baseline life satisfaction, and competitive experience; all other predictors were shrunk to zero, underscoring that burnout trajectory and life satisfaction carry most of the predictive information.

The random forest classifier (1,000 trees; tuned mtry) achieved an out-of-bag error of 19.7%, corresponding to an AUC of approximately 0.74. Variable importance indices (mean decrease in accuracy and Gini) were highest for baseline burnout, followed by early change in burnout, baseline life satisfaction, depressive symptoms, and competitive experience. However, the random forest underperformed both logistic and lasso models in AUC and showed high misclassification for high-burnout cases (class error ≈ 0.54), suggesting that additional model complexity did not translate into better prediction in this sample [24].

Table 2 summarizes the cross-validated performance of the primary models.

Table 2
Cross validated performance of primary prediction models for high T3 burnout

Model	AUC (CV)	Brier (CV)	Sensitivity	Specificity
Logistic regression	0.78	0.16	0.86	0.51
Penalized logistic (glmnet, final)	0.82	0.15	0.88	0.52
Random forest (1,000 trees)	0.74	0.17	0.77	0.60

Note. Values are approximate averages across 10×10-fold cross-validation (10 repeats). AUC = area under the ROC curve.

A calibration plot for the final lasso model (Fig. 1) showed that observed event rates tracked predicted probabilities closely across most risk deciles, with slight over-prediction in the highest-risk group.

To illustrate discrimination, Fig. 2 presents cross-validated ROC curves for the logistic, lasso, and random forest models, demonstrating the superior performance of the penalized model across the range

of thresholds.

Sensitivity and alternative-specification analyses

Baseline-only penalized models

When change scores were excluded and only baseline predictors were entered into glmnet, the best baseline-only model (n = 158) achieved a cross-validated AUC around 0.80. This represents a modest decrease relative to the full lasso model (AUC 0.82); indicating that early change in burnout and life satisfaction adds incremental predictive value beyond baseline measures alone. In the baseline-only solution, baseline burnout remained the dominant predictor, with smaller contributions from depressive symptoms and life satisfaction.

Alternative high-burnout threshold

Redefining high T3 burnout at the 70th percentile cut-off produced similar results. Penalized models continued to show AUCs in the 0.80–0.82 range, and baseline burnout plus early change in burnout remained the strongest predictors. Predictor importance rankings and coefficient patterns were highly stable across thresholds, suggesting robustness of the main findings to the choice of cut-point [77].

Depressive-symptom outcome

Using high T3 depressive symptoms (upper quartile of t3DS) as the outcome yielded slightly lower performance (AUC ≈ 0.75–0.77), with baseline depressive symptoms and early change in depressive symptoms emerging as primary predictors, alongside burnout indicators. This pattern is consistent with prior evidence that burnout and depression are related but distinct constructs in athletes [78, 79].

Falsification and placebo tests

The time-reversed placebo model, in which T3 burnout, sleep, and life satisfaction were used to predict high burnout at T1, produced cross-validated AUCs close to chance (≈ 0.47–0.50) with high sensitivity and very low specificity. This supports that the main predictive direction (early → later burnout) reflects meaningful temporal ordering rather than symmetric prediction.

In the shuffled-outcome lasso model, where high-burnout labels were randomly permuted, cross-validated AUC collapsed to ≈ 0.49, effectively chance-level performance. This falsification result indicates that the strong AUC in the main models arises from genuine structure in the data rather than overfitting to noise.

Table 3 summarizes AUC values from key sensitivity and falsification analyses.

Table 3
AUCs from sensitivity and falsification models

Model / specification	AUC (CV)
Full lasso (all predictors, change scores)	0.82
Baseline-only lasso	0.80
High burnout defined at 70th percentile	0.80–0.82
High T3 depressive symptoms (lasso)	0.75–0.77
Placebo (T3 predictors → T1 high burnout)	≈ 0.47–0.50
Shuffled-outcome lasso	≈ 0.49
Note. Values are approximate cross-validated AUCs across 10×10-fold CV.	

Multiple-imputation robustness

Multiple imputation with 20 datasets yielded lasso models with cross-validated AUCs between 0.81 and 0.83 and Brier scores around 0.15–0.16, closely mirroring the complete-case results. We compared coefficient patterns and AUCs qualitatively to the complete-case model, and the similarity of these results supports the robustness of our primary complete-case analysis [80]. In all imputed models, baseline burnout and early increase in burnout remained the strongest predictors, with baseline life satisfaction and competitive experience providing smaller but consistent contributions. These findings suggest that missing data and attrition did not materially alter the substantive conclusions.

Summary of main findings

Across traditional regression, penalized logistic models, random forests, sensitivity analyses, falsification tests, and multiple imputation, results converged on a consistent pattern: baseline burnout and early increases in burnout over three months are the central predictors of high burnout six months later, while adding numerous additional self-report indicators (sleep, physical symptoms, illness, life satisfaction, depressive symptoms, demographics) yields only modest incremental gains in predictive performance. This pattern supports the practical value of simple repeated burnout screening for identifying high-risk athletes in applied sport settings.

Discussion

Summary of main findings

This secondary analysis of a three-wave dataset of competitive athletes examined whether baseline levels and early changes in psychological and health variables can predict high burnout six months later using modern prediction methods. Consistent across traditional logistic regression, penalized regression, and random forest models, baseline burnout and early increases in burnout between T1 and

T2 emerged as the dominant predictors of high T3 burnout, whereas depressive symptoms, sleep, physical symptoms, and demographic factors added only modest incremental information when modelled jointly [14]. The main substantive finding was that baseline burnout and early increases in burnout were the clearest indicators of later high burnout, and this pattern was robust across analytic approaches while remaining parsimonious [65, 81]. Sensitivity analyses (baseline-only models, alternative cut-offs, depressive-symptom outcome), falsification tests (time-reversed and shuffled outcomes), and multiple imputation all supported the robustness of these findings [67].

Relation to previous research

Our results extend existing work on burnout and mental health in athletes in several ways. Prior longitudinal analyses using random-intercept cross-lagged panel models in this dataset showed that burnout was associated with a range of health outcomes at the between-person level and that within-person increases in exhaustion predicted later depressive symptoms [14]. The present study shifts the focus from explanation to individual-level risk prediction, asking how well simple repeated measures can identify athletes at risk of later high burnout.

The central role of baseline burnout and its short-term worsening aligns with conceptualizations of burnout as a cumulative, chronic response to prolonged stress in sport [3, 82]. That depressive symptoms sleep problems, and physical complaints provided relatively little incremental predictive value once burnout trajectory was accounted for suggests that these correlates may be more useful as co-occurring consequences or correlates of burnout than as early screening markers beyond the burnout measure itself [1]. The modest additional contribution of life satisfaction is consistent with evidence that higher global well-being protects against stress-related outcomes in athletes and other populations [83, 84].

Methodologically, our finding that a relatively simple penalized logistic model slightly outperformed a more complex random forest is in line with work showing that, in moderate-sized psychological datasets with mostly linear relations, regularized generalized linear models often match or exceed the performance of black-box machine-learning algorithms while offering greater interpretability [85, 86]. The strong performance of lasso with only a handful of non-zero coefficients underscores that careful feature selection and regularization can be more beneficial than escalating model complexity in typical sport-psychology samples [24].

Practical implications for sport settings

From an applied perspective, the results suggest that brief, repeated burnout screening may help coaches and sport psychologists notice when athletes are beginning to move toward a more concerning psychological profile over a six-month period. In practice, sport organizations could implement brief, periodic burnout questionnaires (e.g., every 2–3 months) and track both current level and short-term change, flagging athletes whose scores are high and/or increasing for further assessment or preventive intervention [82]. Because our models indicate only modest incremental gains from adding sleep,

physical symptoms, and depressive-symptom measures, practitioners facing time or resource constraints might prioritize burnout monitoring while considering broader assessments where feasible.

At the same time, the predictive role of life satisfaction suggests that supporting global well-being through interventions targeting autonomy, competence, relatedness, and life balance could have protective effects against progression to severe burnout [84, 87]. For example, counsellors or sport psychologists might use changes in burnout and life satisfaction scores to tailor feedback: athletes with rising burnout and declining life satisfaction may benefit from workload adjustments, recovery planning, or psychological skills training focused on values and meaning in sport [82].

The main model showed fairly strong internal discrimination, but because it has only internal validation, it should be read as an encouraging indicator rather than a ready-to-use screening tool in psychiatry and behavioral medicine, where AUCs in the .70–.80 range are typically considered acceptable to good [57, 88]. Nevertheless, given the relatively low base rate of high burnout in this sample (~ 20%), even well-performing models will yield false positives and negatives, so predictions should be used as one component of a broader clinical and performance-support process rather than as deterministic labels [11].

Methodological contributions

This study illustrates how transparent, prediction-focused secondary analyses of existing sport-psychology datasets can complement traditional structural modeling approaches. We followed several best-practice recommendations for prediction modeling, including clear outcome specification, internal validation via repeated cross-validation, penalization to reduce overfitting, multi-metric evaluation (AUC, Brier score, calibration), prespecified sensitivity analyses, and falsification tests [88, 89]. The falsification analyses showing that time-reversed and shuffled-label models performed at chance helped demonstrate that our main results reflect genuine temporal and structural signal rather than artefacts or model overfitting [67, 89].

Importantly, our findings reinforce the value of penalized regression in sport-psychology research. Lasso reduced a relatively large predictor set (burnout dimensions, symptoms, sleep, life satisfaction, illness, demographics) to a small, interpretable subset without sacrificing predictive accuracy, offering a principled alternative to stepwise selection or arbitrary p-value thresholds [90]. This approach may be particularly attractive for applied researchers who require both predictive performance and clear communication of key risk indicators to coaches and support staff.

Limitations and directions for future research

Several limitations warrant consideration. First, although the original sample of 267 athletes is substantial for a longitudinal sport-psychology study, the effective sample size for prediction (n = 137 complete cases for the main models) is modest by machine-learning standards, which may limit model

stability and external generalizability [57, 91]. Our use of repeated cross-validation and multiple imputation mitigates but does not eliminate this concern. External validation in independent athlete samples and contexts (e.g., different countries, sports, or performance levels) is needed before the models can be considered ready for implementation.

Second, all predictors and outcomes were self-report questionnaire measures, which are vulnerable to common-method bias and socially desirable responding [92]. Future work could integrate objective or contextual indicators (e.g., training-load metrics, competition schedules, injury severity, performance data) and ecologically valid digital markers (e.g., sleep-tracking data) to test whether they add incremental predictive value beyond self-reported burnout trajectory.

Third, although we leveraged longitudinal data and conducted falsification tests, the analyses remain predictive, not causal. We cannot conclude that modifying burnout or life satisfaction at T1–T2 would necessarily change the probability of high burnout at T3, as unmeasured confounding and reverse causality are possible [8]. Future studies could combine prediction models with quasi-experimental or intervention designs, or with modern causal-inference approaches (e.g., marginal structural models), to better understand how modifying specific risk factors might prevent burnout progression [93].

Fourth, our operationalization of high burnout used sample-based percentiles rather than clinically validated cut-points. While this is common in early prediction work, it complicates direct comparison across samples and may not align perfectly with clinically meaningful impairment [57, 88]. Work to establish clinically anchored thresholds for athlete burnout, perhaps drawing on functional outcomes or expert consensus, would enhance the applied interpretability of prediction models.

Finally, we focused on a relatively short six-month window with three measurement occasions. Burnout may develop over longer time scales and in response to cumulative training loads, life events, and career transitions [82]. Studies with more frequent assessments and longer follow-up could use dynamic, time-updated prediction models (e.g., joint modeling, dynamic logistic regression) to capture how risk evolves across seasons and career stages [94].

Conclusion

This prediction-focused secondary analysis of a three-wave athlete sample demonstrates that baseline burnout and short-term increases in burnout are the central indicators of who will experience high burnout six months later, with only modest incremental value from depressive symptoms, sleep, physical health, and demographic variables when modelled jointly. A penalized logistic regression model achieved good internal discrimination and calibration and slightly outperformed both standard logistic regression and a random forest classifier, underscoring the utility of regularized, interpretable models in applied sport-psychology contexts. Practically, the findings support the implementation of simple, repeated burnout screening as a feasible approach to identifying athletes at elevated risk, while highlighting the potential protective role of life satisfaction and overall well-being. Further research should externally validate these models, test their utility over longer periods and in diverse athlete populations, and

integrate causal-inference and intervention designs to determine how best to translate predictive signals into effective prevention and support for athletes.

Abbreviations

AUC	area under the receiver operating characteristic curve
Brier	Brier score (mean squared prediction error)
CV	cross-validation
df	degrees of freedom
DOI	Digital Object Identifier
DS	depressive symptoms
LS	life satisfaction
M	mean
PS	physical symptoms
PSQI	Pittsburgh Sleep Quality Index
PSQIg	Pittsburgh Sleep Quality Index global score
PSQIgr	Pittsburgh Sleep Quality Index graded score
ROC	receiver operating characteristic
RSA	reduced sense of accomplishment
SD	standard deviation
SE	standard error
WURSS	Wisconsin Upper Respiratory Symptom Survey

Declarations

Ethics approval and consent to participate

Not applicable. This study is a secondary analysis of de-identified, publicly archived data (Glandorf et al., 2024; DOI: 10.23668/PSYCHARCHIVES.14066). The original parent study received ethical approval from the host institution's research ethics committee, and all participants provided informed consent prior to participation. As our analysis involved only previously collected, anonymized data with no direct contact with human participants, no additional ethical approval was required for this secondary analysis.

Consent for publication

Not applicable.

Availability of data and materials

The datasets analyzed during the current study are available in the PsychArchives repository under a Scientific Use License https://doi.org/10.23668/PSYCHARCHIVES.14066 (Glandorf et al., 2024). All analysis scripts, derived variables, supplementary materials, and complete project documentation are publicly available on the Open Science Framework https://doi.org/10.17605/OSF.IO/K45VY. The OSF project is licensed under CC BY Attribution 4.0 International.

Competing interests

The authors declare that they have no competing interests.

Funding

No funding was received for this secondary analysis. The original data collection was funded by the host institution of the parent study authors, but the funder had no role in the conceptualization, design, data collection, analysis, decision to publish, or preparation of this manuscript.

Authors' contributions

NA contributed to the conceptualization of the prediction framework, methodology design, and drafting of the introduction and discussion sections. KNA performed the statistical analyses (including logistic regression, lasso regularization, random forest models, cross-validation, sensitivity analyses, and falsification tests), contributed to the methodology and results sections, and served as the corresponding author. NS contributed to the interpretation of findings, critical revision of the manuscript for important intellectual content, and supervision of the analytic approach. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Glandorf HL, Madigan DJ, Kavanagh O, Mallinson-Howard SH. Mental and physical health outcomes of burnout in athletes: a systematic review and meta-analysis. Int Rev Sport Exerc Psychol. 2025;18:372–416. https://doi.org/10.1080/1750984X.2023.2225187
2. Isoard-Gautheur S, Guillet-Descas E, Gustafsson H. Athlete Burnout and the Risk of Dropout Among Young Elite Handball Players. Sport Psychol. 2016;30:123–30. https://doi.org/10.1123/tsp.2014-0140

3. Raedeke TD, Smith AL. Development and Preliminary Validation of an Athlete Burnout Measure. J Sport Exerc Psychol. 2001;23:281–306. https://doi.org/10.1123/jsep.23.4.281
4. Eklund RC, Defreese JD. Burnout in Sport and Performance. Oxf Res Encycl Psychol [Internet]. Oxford University Press; 2017 [cited 2026 Apr 16]. https://doi.org/10.1093/acrefore/9780190236557.013.165
5. Nixdorf I, Beckmann J, Nixdorf R. Psychological Predictors for Depression and Burnout Among German Junior Elite Athletes. Front Psychol. 2020;11:601. https://doi.org/10.3389/fpsyg.2020.00601
6. Madigan DJ, Stoeber J, Passfield L. Perfectionism and Burnout in Junior Athletes: A Three-Month Longitudinal Study. J Sport Exerc Psychol. 2015;37:305–15. https://doi.org/10.1123/jsep.2014-0266
7. Madigan DJ, Stoeber J, Passfield L. Motivation Mediates the Perfectionism–Burnout Relationship: A Three-Wave Longitudinal Study With Junior Athletes. J Sport Exerc Psychol. 2016;38:341–54. https://doi.org/10.1123/jsep.2015-0238
8. Shmueli G. To Explain or to Predict? Stat Sci [Internet]. 2010 [cited 2026 Apr 16];25. https://doi.org/10.1214/10-STS330
9. Moons KGM, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. Heart. 2012;98:683–90. https://doi.org/10.1136/heartjnl-2011-301246
10. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. Ann Intern Med. 2015;162:W1–73. https://doi.org/10.7326/M14-0698
11. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J. 2014;35:1925–31. https://doi.org/10.1093/eurheartj/ehu207
12. Bernardini F, Attademo L, Cleary SD, Luther C, Shim RS, Quartesan R, et al. Risk Prediction Models in Psychiatry: Toward a New Frontier for the Prevention of Mental Illnesses. J Clin Psychiatry. 2017;78:572–83. https://doi.org/10.4088/JCP.15r10003
13. Ahola K, Hakanen J. Job strain, burnout, and depressive symptoms: A prospective study among dentists. J Affect Disord. 2007;104:103–10. https://doi.org/10.1016/j.jad.2007.03.004
14. Glandorf HL, Madigan DJ, Kavanagh O, Mallinson-Howard SH, Donachie TC, Olsson LF, et al. Athlete burnout and mental and physical health: A three-wave longitudinal study of direct and reciprocal effects. Sport Exerc Perform Psychol. 2024;13:412–31. https://doi.org/10.1037/spy0000355
15. Kim H, Ji J, Kao D. Burnout and Physical Health among Social Workers: A Three-Year Longitudinal Study. Soc Work. 2011;56:258–68. https://doi.org/10.1093/sw/56.3.258
16. Rothe N, Schulze J, Kirschbaum C, Buske-Kirschbaum A, Penz M, Wekenborg MK, et al. Sleep disturbances in major depressive and burnout syndrome: A longitudinal analysis. Psychiatry Res. 2020;286:112868. https://doi.org/10.1016/j.psychres.2020.112868

17. Montejo-Lambaren VH, Ramirez-Hernandez S, Gaytan-Gonzalez A, Gonzalez-Mendoza RG, Lopez-Taylor JR. Indicators Of Burnout In College Athletes: Proportion Comparisons By Sex And Sport Type 1357 Board #165 May 31 9 00 AM - 10 30 AM. Med Sci Sports Exerc. 2018;50:325. https://doi.org/10.1249/01.mss.0000536151.79507.2d
18. Reche C, De Francisco C, Martínez-Rodríguez A, Ros-Martínez A. Relationship among sociodemographic and sport variables, exercise dependence, and burnout: a preliminary study in athletes. An Psicol. 2018;34:398. https://doi.org/10.6018/analesps.34.2.289861
19. Alba AC, Agoritsas T, Walsh M, Hanna S, Iorio A, Devereaux PJ, et al. Discrimination and Calibration of Clinical Prediction Models: Users' Guides to the Medical Literature. JAMA. 2017;318:1377. https://doi.org/10.1001/jama.2017.12126
20. Binuya MAE, Engelhardt EG, Schats W, Schmidt MK, Steyerberg EW. Methodological guidance for the evaluation and updating of clinical prediction models: a systematic review. BMC Med Res Methodol. 2022;22:316. https://doi.org/10.1186/s12874-022-01801-8
21. Steyerberg EW, Harrell FE. Prediction models need appropriate internal, internal–external, and external validation. J Clin Epidemiol. 2016;69:245–7. https://doi.org/10.1016/j.jclinepi.2015.04.005
22. Tibshirani R. Regression Shrinkage and Selection Via the Lasso. J R Stat Soc Ser B Stat Methodol. 1996;58:267–88. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x
23. Tutz G, Ulbricht J. Penalized regression with correlation-based penalty. Stat Comput. 2009;19:239–53. https://doi.org/10.1007/s11222-008-9088-5
24. Jacobucci R, Littlefield AK, Millner AJ, Kleiman EM, Steinley D. Evidence of Inflated Prediction Performance: A Commentary on Machine Learning and Suicide Research. Clin Psychol Sci. 2021;9:129–34. https://doi.org/10.1177/2167702620954216
25. Jacobucci R, Grimm KJ. Machine Learning and Psychological Research: The Unexplored Effect of Measurement. Perspect Psychol Sci. 2020;15:809–16. https://doi.org/10.1177/1745691620902467
26. Austin PC, Harrell FE, Steyerberg EW. Predictive performance of machine and statistical learning methods: Impact of data-generating processes on external validity in the "large N, small p" setting. Stat Methods Med Res. 2021;30:1465–83. https://doi.org/10.1177/09622802211002867
27. Brester C, Voutilainen A, Tuomainen T-P, Kauhanen J, Kolehmainen M. Epidemiological predictive modeling: lessons learned from the Kuopio ischemic heart disease risk factor study. Ann Epidemiol. 2022;70:1–8. https://doi.org/10.1016/j.annepidem.2022.03.010
28. Senders JT, Staples PC, Karhade AV, Zaki MM, Gormley WB, Broekman MLD, et al. Machine Learning and Neurosurgical Outcome Prediction: A Systematic Review. World Neurosurg. 2018;109:476-486.e1. https://doi.org/10.1016/j.wneu.2017.09.149
29. Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? Liu B, editor. PLOS ONE. 2017;12:e0174944. https://doi.org/10.1371/journal.pone.0174944
30. Bijlsma MJ, Vansteelandt S, Janssen F, Hak E. The effect of adherence to statin therapy on cardiovascular mortality: quantification of unmeasured bias using falsification end-points. BMC

Public Health. 2016;16:303. https://doi.org/10.1186/s12889-016-2986-0

31. Pizer SD. Falsification Testing of Instrumental Variables Methods for Comparative Effectiveness Research. Health Serv Res. 2016;51:790–811. https://doi.org/10.1111/1475-6773.12355
32. Moons KGM, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: II. External validation, model updating, and impact assessment. Heart. 2012;98:691–8. https://doi.org/10.1136/heartjnl-2011-301247
33. Baker RE, Kwartler T. Sport Analytics: Using Open Source Logistic Regression Software to Classify Upcoming Play Type in the NFL. J Appl Sport Manag [Internet]. 2015 [cited 2026 Apr 15];7. https://doi.org/10.7290/jasm07c14e
34. Hagström D, Sundström A, Hakelind C, Nordin S, Nordin M. Mental and Somatic Ill-Health as Long-Term Predictors of the Burnout Symptoms Cognitive Weariness, Exhaustion, and Lack of Vitality in a General Adult Population. Stress Health. 2025;41:e70119. https://doi.org/10.1002/smi.70119
35. Shi H, Liu J, Yang C, Shang J, Zeng Y. Machine learning for predicting burnout among healthcare workers: a systematic review and meta-analysis. Contemp Nurse. 2025;1–21. https://doi.org/10.1080/10376178.2025.2593294
36. Stylianou N, Akbarov A, Kontopantelis E, Buchan I, Dunn KW. Mortality risk prediction in burn injury: Comparison of logistic regression with machine learning approaches. Burns. 2015;41:925–34. https://doi.org/10.1016/j.burns.2015.03.016
37. Glandorf H, Madigan D, Kavanagh O, Mallinson-Howard S, Donachie T, Olsson L, et al. Data for: Athlete burnout and mental and physical health outcomes: A three-wave longitudinal study [Internet]. PsychArchives; 2024 [cited 2026 Apr 15]. https://doi.org/10.23668/PSYCHARCHIVES.14066
38. Araújo D, Davids K, Passos P. Ecological Validity, Representative Design, and Correspondence Between Experimental Task Constraints and Behavioral Setting: Comment on Rogers, Kadar, and Costall (2005). Ecol Psychol. 2007;19:69–78. https://doi.org/10.1080/10407410709336951
39. Isoard-Gautheur S, Oger M, Guillet E, Martin-Krumm C. Validation of a French Version of the Athlete Burnout Questionnaire (ABQ): In Competitive Sport and Physical Education Context. Eur J Psychol Assess. 2010;26:203–11. https://doi.org/10.1027/1015-5759/a000027
40. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: Validity of a brief depression severity measure. J Gen Intern Med. 2001;16:606–13. https://doi.org/10.1046/j.1525-1497.2001.016009606.x
41. Barrett B, Brown R, Mundt M, Safdar N, Dye L, Maberry R, et al. The Wisconsin Upper Respiratory Symptom Survey is responsive, reliable, and valid. J Clin Epidemiol. 2005;58:609–17. https://doi.org/10.1016/j.jclinepi.2004.11.019
42. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research. Psychiatry Res. 1989;28:193–213. https://doi.org/10.1016/0165-1781(89)90047-4
43. Carpi M. The Pittsburgh Sleep Quality Index: a brief review. Occup Med. 2025;75:14–5. https://doi.org/10.1093/occmed/kqae121

44. Ørskov PT, Norup A. Validity and reliability of the Danish version of the Pittsburgh Sleep Quality Index. Scand J Public Health. 2023;51:1182–8. https://doi.org/10.1177/14034948221100099
45. Akyurek G, Efe A, Aki E. Psychometric properties of the Temporal Satisfaction with Life Scale among Turkish people. Psychiatry Clin Psychopharmacol. 2019;29:204–11. https://doi.org/10.1080/24750573.2018.1445898
46. Dirzyte A, Perminas A, Biliuniene E. Psychometric Properties of Satisfaction with Life Scale (SWLS) and Psychological Capital Questionnaire (PCQ-24) in the Lithuanian Population. Int J Environ Res Public Health. 2021;18:2608. https://doi.org/10.3390/ijerph18052608
47. Halson SL. Monitoring Training Load to Understand Fatigue in Athletes. Sports Med. 2014;44:139–47. https://doi.org/10.1007/s40279-014-0253-z
48. Jaspers A, Brink MS, Probst SGM, Frencken WGP, Helsen WF. Relationships Between Training Load Indicators and Training Outcomes in Professional Soccer. Sports Med. 2017;47:533–44. https://doi.org/10.1007/s40279-016-0591-0
49. Jones CM, Griffiths PC, Mellalieu SD. Training Load and Fatigue Marker Associations with Injury and Illness: A Systematic Review of Longitudinal Studies. Sports Med. 2017;47:943–74. https://doi.org/10.1007/s40279-016-0619-5
50. Cronbach LJ, Furby L. How we should measure 'change': Or should we? Psychol Bull. 1970;74:68–80. https://doi.org/10.1037/h0029382
51. Hewitt DB, Ellis RJ, Hu Y-Y, Cheung EO, Moskowitz JT, Agarwal G, et al. Evaluating the Association of Multiple Burnout Definitions and Thresholds With Prevalence and Outcomes. JAMA Surg. 2020;155:1043. https://doi.org/10.1001/jamasurg.2020.3351
52. Nielson DM, Keren H, O'Callaghan G, Jackson SM, Douka I, Vidal-Ribas P, et al. Great Expectations: A Critical Review of and Suggestions for the Study of Reward Processing as a Cause and Predictor of Depression. Biol Psychiatry. 2021;89:134–43. https://doi.org/10.1016/j.biopsych.2020.06.012
53. Rohrer JM, Murayama K. These Are Not the Effects You Are Looking for: Causality and the Within-/Between-Persons Distinction in Longitudinal Data Analysis. Adv Methods Pract Psychol Sci. 2023;6:25152459221140842. https://doi.org/10.1177/25152459221140842
54. Efthimiou O, Seo M, Chalkou K, Debray T, Egger M, Salanti G. Developing clinical prediction models: a step-by-step guide. BMJ. 2024;386:e078276. https://doi.org/10.1136/bmj-2023-078276
55. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; 2025.
56. Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G, editors. Incomplete data: Introduction and overview. Longitud Data Anal [Internet]. 0 edn Chapman and Hall/CRC; 2008 [cited 2026 Apr 15]. p. 409–22. https://doi.org/10.1201/9781420011579-29
57. Riley RD, Snell KI, Ensor J, Burke DL, Harrell Jr FE, Moons KG, et al. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. Stat Med. 2019;38:1276–96. https://doi.org/10.1002/sim.7992

58. Yan Y, Yang Z, Semenkovich TR, Kozower BD, Meyers BF, Nava RG, et al. Comparison of standard and penalized logistic regression in risk model development. JTCVS Open. 2022;9:303–16. https://doi.org/10.1016/j.xjon.2022.01.016
59. Pavlou M, Ambler G, Seaman S, De Iorio M, Omar RZ. Review and evaluation of penalised regression methods for risk prediction in low-dimensional data with few events. Stat Med. 2016;35:1159–77. https://doi.org/10.1002/sim.6782
60. Ali A, Gravino C. An empirical comparison of validation methods for software prediction models. J Softw Evol Process. 2021;33:e2367. https://doi.org/10.1002/smr.2367
61. Zhang C, Yan R, Liu X, Nie X, Peng Y, Peng X. Empirical simulation of internal validation methods for prediction models: comparing k-fold cross-validation with bootstrap-based optimism correction. J Clin Epidemiol. 2026;190:112101. https://doi.org/10.1016/j.jclinepi.2025.112101
62. Lewis MJ, Spiliopoulou A, Goldmann K, Pitzalis C, McKeigue P, Barnes MR. nestedcv: an R package for fast implementation of nested cross-validation with embedded feature selection designed for transcriptomics and high-dimensional data. Lengauer T, editor. Bioinforma Adv. 2023;3:vbad048. https://doi.org/10.1093/bioadv/vbad048
63. Xiang Liu PX. Feature Selection using Bootstrapped ROC Curves. J Proteomics Bioinform [Internet]. 2014 [cited 2026 Apr 15];s9. https://doi.org/10.4172/jpb.S9-006
64. Simon RM, Subramanian J, Li M-C, Menezes S. Using cross-validation to evaluate predictive accuracy of survival risk classifiers based on high-dimensional data. Brief Bioinform. 2011;12:203–14. https://doi.org/10.1093/bib/bbr001
65. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the Performance of Prediction Models: A Framework for Traditional and Novel Measures. Epidemiology. 2010;21:128–38. https://doi.org/10.1097/EDE.0b013e3181c30fb2
66. Wang S, Gao W, Ngwa J, Allard C, Liu C-T, Cupples LA. Comparing baseline and longitudinal measures in association studies. BMC Proc. 2014;8:S84. https://doi.org/10.1186/1753-6561-8-S1-S84
67. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative Controls: A Tool for Detecting Confounding and Bias in Observational Studies. Epidemiology. 2010;21:383–8. https://doi.org/10.1097/EDE.0b013e3181d61eeb
68. Eekhout I, De Boer RM, Twisk JWR, De Vet HCW, Heymans MW. Missing Data: A Systematic Review of How They Are Reported and Handled. Epidemiology. 2012;23:729–32. https://doi.org/10.1097/EDE.0b013e3182576cdb
69. Royston P. Multiple Imputation of Missing Values. Stata J Promot Commun Stat Stata. 2004;4:227–41. https://doi.org/10.1177/1536867X0400400301
70. Alderman H, Behrman JR, Kohler H-P, Maluccio JA, Cotts Watkins S. Attrition in Longitudinal Household Survey Data: Some Tests for Three Developing-Country Samples [Internet]. World Bank, Washington, DC; 2000 [cited 2026 Apr 15]. https://doi.org/10.1596/1813-9450-2447

71. Pepe MS, Kerr KF, Longton G, Wang Z. Testing for improvement in prediction model performance. Stat Med. 2013;32:1467–82. https://doi.org/10.1002/sim.5727
72. Steyerberg EW, Van Calster B. Redefining significance and reproducibility for medical research: A plea for higher *P* -value thresholds for diagnostic and prognostic models. Eur J Clin Invest. 2020;50:e13229. https://doi.org/10.1111/eci.13229
73. Aalbersberg IjJ, Appleyard T, Brookhart S, Carpenter T, Clarke M, Curry S, et al. Making Science Transparent By Default; Introducing the TOP Statement [Internet]. Open Science Framework; 2018 [cited 2026 Apr 15]. https://doi.org/10.31219/osf.io/sm78t
74. Austin PC, White IR, Lee DS, Van Buuren S. Missing Data in Clinical Research: A Tutorial on Multiple Imputation. Can J Cardiol. 2021;37:1322–31. https://doi.org/10.1016/j.cjca.2020.11.010
75. Li P, Stuart EA, Allison DB. Multiple Imputation: A Flexible Tool for Handling Missing Data. JAMA. 2015;314:1966. https://doi.org/10.1001/jama.2015.15281
76. Matheny ME, Ohno-Machado L, Resnic FS. Discrimination and calibration of mortality risk prediction models in interventional cardiology. J Biomed Inform. 2005;38:367–75. https://doi.org/10.1016/j.jbi.2005.02.007
77. Di Gennaro P, Fordellone M, Nicolao G, Schiattarella P, Smimmo A, Speranza T, et al. Assessing Methods for Predictive Cut-Point Estimation: A Simulation-Based Comparison. Epidemiol Biostat Public Health [Internet]. 2025 [cited 2026 Apr 15]; https://doi.org/10.54103/2282-0930/29367
78. Cresswell SL, Eklund RC. The convergent and discriminant validity of burnout measures in sport: A multi-trait/multi-method analysis. J Sports Sci. 2006;24:209–20. https://doi.org/10.1080/02640410500131431
79. Frank R, Nixdorf I, Beckmann J. Analyzing the Relationship Between Burnout and Depression in Junior Elite Athletes. J Clin Sport Psychol. 2017;11:287–303. https://doi.org/10.1123/JCSP.2017-0008
80. Austin PC, Lee DS, Ko DT, White IR. Effect of Variable Selection Strategy on the Performance of Prognostic Models When Using Multiple Imputation. Circ Cardiovasc Qual Outcomes. 2019;12:e005927. https://doi.org/10.1161/CIRCOUTCOMES.119.005927
81. Rufibach K. Use of Brier score to assess binary predictions. J Clin Epidemiol. 2010;63:938–9. https://doi.org/10.1016/j.jclinepi.2009.11.009
82. Gustafsson H, DeFreese J, Madigan DJ. Athlete burnout: review and recommendations. Curr Opin Psychol. 2017;16:109–13. https://doi.org/10.1016/j.copsyc.2017.05.002
83. Diener E, Suh EM, Lucas RE, Smith HL. Subjective well-being: Three decades of progress. Psychol Bull. 1999;125:276–302. https://doi.org/10.1037/0033-2909.125.2.276
84. Lundqvist C. Well-being in competitive sports—The feel-good factor? A review of conceptual considerations of well-being. Int Rev Sport Exerc Psychol. 2011;4:109–27. https://doi.org/10.1080/1750984X.2011.584067
85. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. Nat Methods. 2018;15:233–4. https://doi.org/10.1038/nmeth.4642

86. Yarkoni T, Westfall J. Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning. Perspect Psychol Sci. 2017;12:1100–22. https://doi.org/10.1177/1745691617693393
87. Deci EL, Ryan RM. The 'What' and 'Why' of Goal Pursuits: Human Needs and the Self-Determination of Behavior. Psychol Inq. 2000;11:227–68. https://doi.org/10.1207/S15327965PLI1104_01
88. Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating [Internet]. Cham: Springer International Publishing; 2019 [cited 2026 Apr 16]. https://doi.org/10.1007/978-3-030-16399-0
89. Goldstein BA, Pencina MJ. Testing Clinical Prediction Models. JAMA. 2020;324:1998. https://doi.org/10.1001/jama.2020.19392
90. Heinze G, Wallisch C, Dunkler D. Variable selection – A review and recommendations for the practicing statistician. Biom J. 2018;60:431–49. https://doi.org/10.1002/bimj.201700067
91. Van Smeden M, Moons KG, De Groot JA, Collins GS, Altman DG, Eijkemans MJ, et al. Sample size for binary logistic prediction models: Beyond events per variable criteria. Stat Methods Med Res. 2019;28:2455–74. https://doi.org/10.1177/0962280218784726
92. Podsakoff PM, MacKenzie SB, Lee J-Y, Podsakoff NP. Common method biases in behavioral research: A critical review of the literature and recommended remedies. J Appl Psychol. 2003;88:879–903. https://doi.org/10.1037/0021-9010.88.5.879
93. VanderWeele TJ. Explanation in causal inference: developments in mediation and interaction. Int J Epidemiol. 2016;dyw277. https://doi.org/10.1093/ije/dyw277
94. Rizopoulos D. Joint Models for Longitudinal and Time-to-Event Data [Internet]. 0 edn. Chapman and Hall/CRC; 2012 [cited 2026 Apr 16]. https://doi.org/10.1201/b12208

Figures

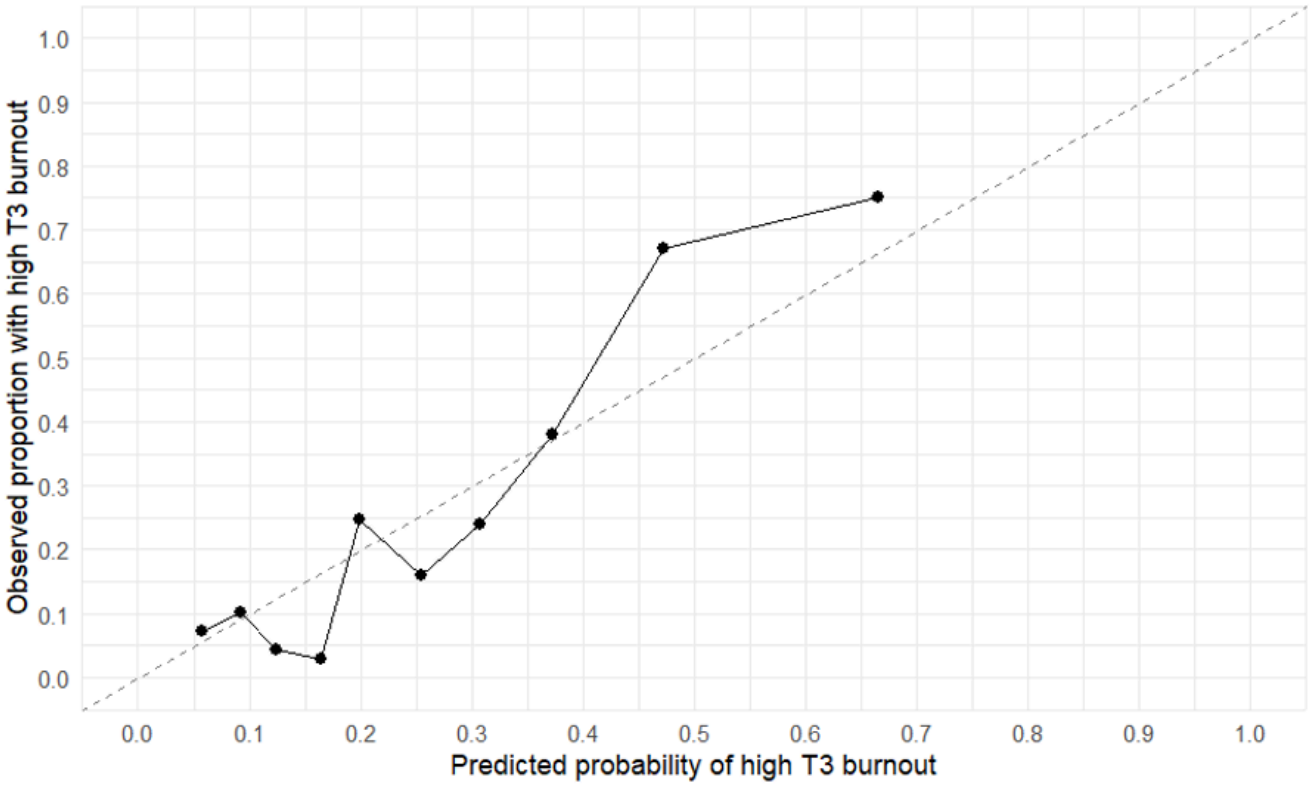


Figure 1

Calibration plot for lasso logistic regression predicting high T3 athlete burnout (observed vs predicted probability across deciles).

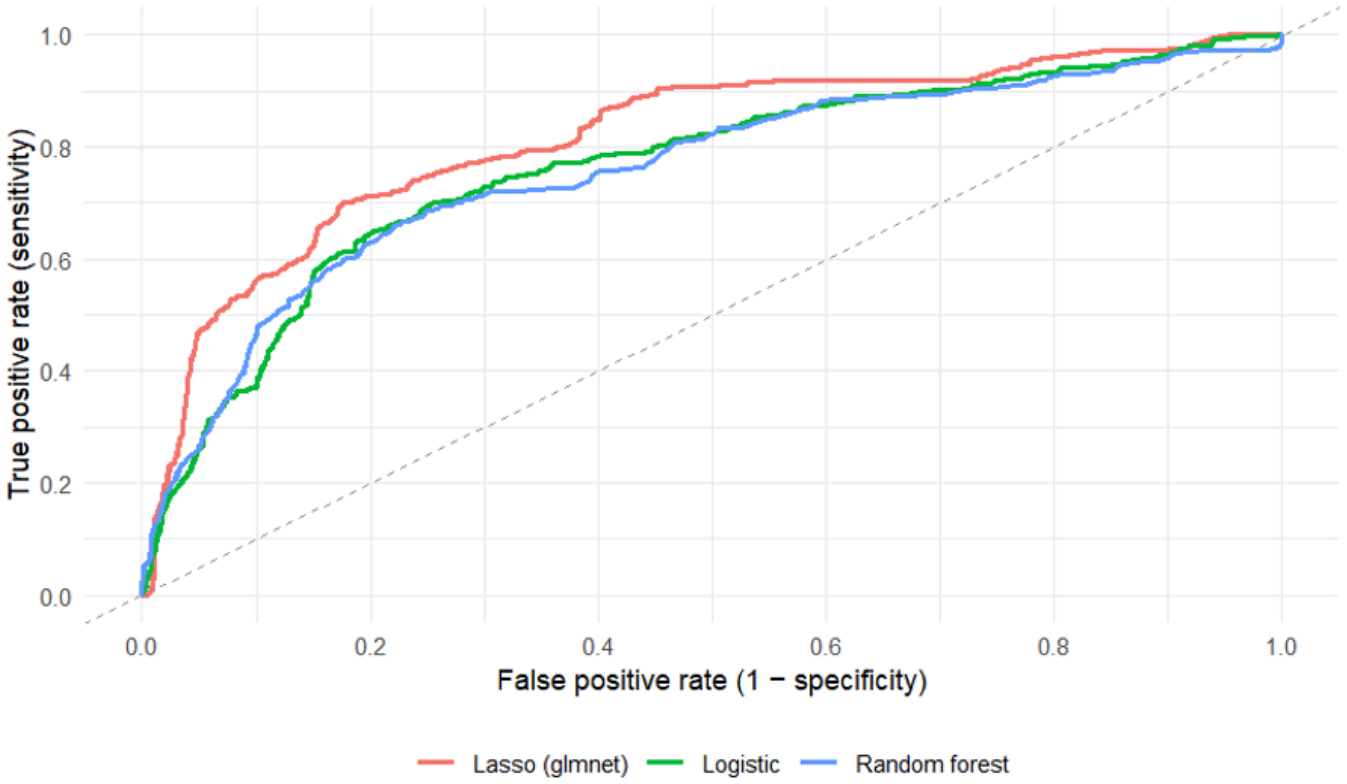


Figure 2

Cross-validated ROC curves for logistic regression, lasso, and random forest models predicting high T3 athlete burnout.