Medicine & Science in Sports & Exercise Publish Ahead of Print

DOI: 10.1249/MSS.0000000000004212

Predicting Military Training Outcome: A Prospective Cohort Study of Supervised Statistical Learning Classification Methods in British Army Recruits

Thomas J. O'Leary[1], Alexander T. Carswell[2], Sarah Jackson[1], Rachel M. Izard[3], Jonathan C. Y. Tang[4], Samuel J. Oliver[5], Neil P. Walsh[6], and Julie P. Greeves[1,4]

[1]Army Health and Performance Research, Army Headquarters, Andover, UNITED KINGDOM; [2]Norwich Medical School, University of East Anglia, Norwich, UNITED KINGDOM; [3]Defence Science and Technology, Ministry of Defence, Porton Down, UNITED KINGDOM; [4]School of Genetics and Cancer, University of Edinburgh, Edinburgh, UNITED KINGDOM; [5]Institute for Applied Human Physiology, College of Medicine and Health, Bangor University, Bangor, UNITED KINGDOM; [6]School of Sport and Exercise Sciences, Liverpool John Moores University, Liverpool, UNITED KINGDOM

Running Title: PREDICTING MILITARY TRAINING OUTCOME

Address for correspondence: Thomas J O'Leary PhD, Army Health and Performance Research, Army Headquarters, Andover, Hampshire, SP11 8HT, United Kingdom. Email: thomas.oleary100@mod.gov.uk

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

ABSTRACT

Purpose: Investigate predictors of failure and injury during British Army basic training and test the performance of statistical learning models to predict these outcomes. **Methods:** 968 male and 450 female recruits had demographics, lifestyle habits, mood, physical performance, body composition, micronutrient status, and bone metabolism measured at week 1 of training. Associations between predictors and training outcome (failure; lower limb overuse musculoskeletal injury) were assessed by logistic regression. Training outcome was predicted by logistic regression, elastic net, random forests, and gradient boosting. **Results:** 66% of men and 76% of women passed with 20% and 34% developing an injury. For men, slower 2.4 km run time (odds ratio [OR]=1.004, p=0.011) and higher symptoms of depression (OR=1.173, p=0.001) were associated with higher risk of failing, and lower areal bone mineral density (OR=0.008, p=0.002), slower 2.4 km run time (OR=1.006, p=0.011), previous stress fracture (OR=3.267, p=0.002), lower parathyroid hormone (OR=0.783, p=0.009), and lower c-telopeptide cross-links of type I collagen (βCTX) (OR=0.255, p=0.041) were associated with higher risk of injury. For women, slower 2.4 km run time (OR=1.009, p=0.009) and higher total 25-hydroxyvitamin-D (OR=1.013, p=0.001) were associated with higher risk of failing, and slower 2.4 km run time (OR=1.007, p=0.009), higher alcohol intake (OR=1.006, p=0.017), and lower βCTX (OR=0.057, p=0.026) were associated with higher risk of injury. Training outcome was poorly predicted in men and women (area under the receiver operating characteristic curve≤0.68). All models had low sensitivity and high specificity, but application of an outcome-specific threshold determined using the Youden Index improved sensitivity. **Conclusions:** Interventions to improve military training outcome should address aerobic fitness prior to starting training. Prediction models may be useful in identifying some ‘at-risk’ recruits. **Key**

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

Words: ENDURANCE, LOAD CARRIAGE, MACHINE LEARNING, MUSCULOSKELETAL INJURY, MUSCLE STRENGTH

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

INTRODUCTION

British Army basic training is physically (1–8) and psychologically (9, 10) demanding. The arduous nature of training means many recruits fail to complete the course due to voluntary withdrawal, poor course performance, or because they develop musculoskeletal injuries that result in rehabilitation or medical discharge (1, 5, 11–13). High numbers of recruits failing training impose a significant operational and financial burden from not having enough trained military personnel to meet military objectives and high costs associated with training each recruit. Better understanding the reasons for, and who is at risk of, failing military training can support military selection and training practices.

Principal pre-training risk factors associated with developing a musculoskeletal injury during military training have been well studied and include being female, low aerobic fitness, low levels of previous physical activity, various components of body composition, and poorer micronutrient status, amongst others (14–17). There are comparatively fewer data on risk factors associated with failing military training courses (from basic training to advanced selection courses) (18–28), which is arguably the most important outcome for military planners. Some studies have tested the performance of classification models to estimate individual musculoskeletal injury risk (22, 29–36) or training course outcome (19, 22–24, 26, 28) in military populations, however, most studies either included small samples, did not validate performance on new observations—inflating model performance, limited analyses to a single regression method (*i.e.*, logistic regression), and did not explore sex-specific risk factors. Compared with men during military training, women experience higher physiological stress (1, 37), and a higher incidence of musculoskeletal and bone stress injuries (13), endocrine disturbances (10), and micronutrient deficiencies (38), likely in part, due to having lower aerobic

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

fitness, less muscle mass, lower strength, and more body fat (1, 7, 37). Additionally, exploring a range of increasingly complex and flexible statistical learning classification methods may improve prediction accuracy for training outcome (26). We are unaware of any study exploring prediction models for British Army basic training outcome, one of the most important times in a career where there is a high risk of musculoskeletal injury and failure.

The primary aim of this study was to examine factors measured at the start of training that were associated with British Army basic training outcome (failure and developing a lower limb overuse musculoskeletal injury) in men and women. Secondary aims were to develop and evaluate predictive models to identify individual risk for the same outcomes. We hypothesised that lower aerobic fitness, lower lift strength, and lower mood would be associated with failing basic military training, and lower aerobic fitness and lower lift strength would be associated with developing a musculoskeletal injury. We anticipated that these factors could be used to accurately classify individual risk of training failure or lower limb overuse musculoskeletal injury. Better predicting individual risk of adverse training outcome in men and women may be useful in assigning limited resources to support those most in need before or during military training.

METHODS

Participants

The study was advertised to new British Army male infantry recruits and female soldier recruits during week one of their basic training courses. The male recruits were training to be line infantry soldiers at the Infantry Training Centre, Catterick, UK and the female recruits were training to be non-infantry soldiers at the Army Training Centre, Pirbright, UK. Recruitment was open from January 2014 to July 2017, before women were permitted to join the infantry. British

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

Army infantry training is 26-weeks split into two phases. The first 14-weeks teaches basic military skills (field exercise, weapon handling), and involves periods of aerobic endurance training, strength and conditioning, military-specific fitness training (obstacle course, circuit training), military drill, and progressive loaded marching. The final 12-weeks teaches infantry specific specialist skills. British Army soldier (non-infantry) basic training involves 14-weeks of teaching basic military skills and developing physical fitness. The first 14 weeks of each British Army basic training course are similar across training sites. All participants had passed an initial military medical assessment and were confirmed to be injury free and not have any medical condition that precluded military service. This study was approved by the Ministry of Defence Research Ethics Committee (ref: 165/Gen/10). Each participant had the study procedures and risks fully explained verbally and in writing, and provided written informed consent.

Experimental Design

This study was an observational, prospective cohort study. These data were secondary analyses from a larger study exploring micronutrient deficiencies and health and performance outcomes (38–45). Demographics, lifestyle and physical activity habits, mood, physical performance, body composition, and biochemical markers of micronutrient status and bone metabolism were measured at week 1 before military training had commenced. Participants self-reported their exercise history, sleep habits, smoking habits, alcohol intake, mood, and stress fracture history using questionnaires. Physical performance was assessed by a maximal effort 2.4 km run, a maximal lift test, and a vertical jump. Venous blood samples were drawn for the analysis of biochemical markers of iron status, vitamin D status, and biochemical markers of bone metabolism. Body mass and height were measured (SECA, Hamburg, Germany), and whole-body lean mass, fat mass, and areal bone mineral density were measured by dual-energy

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

X-ray absorptiometry (DXA). Participants were then followed prospectively for 14 weeks of basic military training for course outcome and grouped as: i) passed; ii) failed for non-medical reasons [voluntary withdrawal, poor course performance, unsuitable for Army service]; iii) failed and entered rehabilitation; iv) failed and medically discharged. Participants' medical records were used to record clinician-diagnosed lower limb overuse musculoskeletal injuries and lower limb stress fractures (including hip/pelvis) during the first 14 weeks of training. Medical records were reviewed at the end of the study by a single military occupational medicine physician according to ICD-9 codes. A diagnostic code was assigned on first presentation with clear symptoms but changed if imaging confirmed a different diagnosis. All medical encounters were recorded and each participant with one or more injury was coded as 'injured.'

Lifestyle, Sleep, Physical Activity Habits, and Mood

Adolescent exercise volume (min·wk^{-1}) was determined from the question '*During your teenage years (between the ages of 12 and 16), what type of exercise did you do during your leisure time, for how many minutes each session, and how many times in a typical week?'* Exercise during adolescence is important for developing musculoskeletal tissues, particularly bone (46), and could be important for protecting against musculoskeletal injuries in military training (46–48). Participants were asked to report the time they went to sleep and awoke on a typical night before training started to allow the calculation of habitual sleep time (h·night^{-1}). Smoking status (Yes / No / Previous) was determined from the question '*Do you smoke tobacco (cigarettes, roll-ups, cigars, or pipe) (Yes / I occasionally smoke / I used to smoke regularly / I used to smoke occasionally / I have never smoked)*? Alcohol intake (units per month) was calculated from the following questions: '*Do you drink alcohol? (Yes / No)*', '*Generally, how many times per month do you drink alcohol? (Once / 2 to 4 / 5 to 12 / 13 +)'*, and '*Generally,*

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

how many units do you drink on each occasion at home and when socialising (one pint of beer = 3 units / one glass of wine = 2 units / one shot of spirits = 1 unit / one bottle of alcopop = 1 unit)? (1 to 5 / 6 to 10 / 11 to 15 / 16 to 20 / 20 +)'. History of stress fracture (Yes / No / Unsure) was determined from the question *'Have you ever been told by a doctor that you have suffered a stress fracture (an injury to a bone caused from over exercising)?'* Anger, confusion, depression, fatigue, tension, and vigour were determined from the Brunel Mood Scale (49).

Physical Performance

Aerobic fitness was assessed as the time to complete a maximal effort 2.4 km run on a standardised running course. Participants completed an 800 m warm-up, and completion time was recorded to the nearest second. The time to complete a 2.4 km run is indicative of maximal aerobic capacity (50) and is a military field test assessed during selection, training, and throughout a military career. All participants were accustomed to performing this test from selection, before commencing military training. Maximal lift strength was assessed with a custom-made incremental lift machine that simulates a power clean. The devices were custom made for British Army assessment centres for use as a selection test. The device consists of a vertically moving carriage with handgrips positioned approximately 30 cm above the floor. The vertically moving carriage is loaded with a weight stack and pin that allows selection of the load in 5 kg increments. The handles must be lifted to a height of 1.45 m from the ground, which replicates the height of a British Army 4-tonne truck. Participants started by holding the handles whilst squatted down before lifting the bar in a power clean motion by extended the knees and hips, finishing with a press of the bar. Correct technique was first demonstrated. Participants attempted the first lift at 20 kg, and the weight was increased by 5 kg after each successful attempt. The test was terminated when participants failed to lift the weight to the required height

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

on their second attempt. Maximal dynamic lift strength has a test re-test reliability of r = 0.91 (51, 52), is a representative military task that is used as a selection test (53), and has been used widely in military human performance research (51, 52). Peak leg power was determined from a vertical jump test using a digital jump meter (Takei Scientific Instruments, Niigata, Japan). A belt was fitted around the waist of each participant and secured to a rubber mat. Participants were instructed to descend and then immediately jump as high as possible with their hands placed on their hips. Vertical jump height was recorded as the highest score achieved out of three attempts. Power was calculated from jump height and body mass using an equation that reflects instantaneous power output and has been validated with the use of a force platform (54): *Peak power (W) = ([51.9 × jump height (cm)] + [48.9 × body mass (kg)]) −2007*. Vertical jump height and peak power have been shown to be important predictors of military performance (55) and have good test-retest reliability (56).

Body Composition

Whole-body lean mass, fat mass, and areal bone mineral density (aBMD) were assessed by DXA (GE Healthcare, Buckinghamshire, UK) whilst wearing light clothing.

Blood Collection and Biochemical Analyses

A venous blood sample was collected either in the morning (~0900 to 1100 h) after breakfast (0600 to 0700 h), or early afternoon (~1300 to 1500 h) after lunch (1200 to 1300h). Venous blood was withdrawn from a vein in the antecubital fossa and collected in serum and EDTA BD Vacutainer® tubes (Becton Dickinson, New Jersey, USA). Serum samples were left to clot for 1 hour at room temperature. Blood samples were centrifuged at 1500 *g* and 4°C for 10 min before serum and plasma were separated into universal tubes and stored at −80°C until analysis. Plasma procollagen type 1 N-terminal propeptide (PINP), c-telopeptide cross-links of

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

type 1 collagen (βCTX), intact parathyroid hormone (PTH), and serum ferritin were analysed by electro-chemiluminescence immunoassays on the COBAS c601 (Roche Diagnostics, Mannheim, Germany) platform. PINP inter-assay coefficient of variation (CV) was < 3% between 20 and 600 $\mu g \cdot L^{-1}$ with a sensitivity of 8 $\mu g \cdot L^{-1}$. βCTX inter-assay CV was < 3% between 0.20 and 1.50 $\mu g \cdot L^{-1}$ with a sensitivity of 0.01 $\mu g \cdot L^{-1}$. PTH inter-assay CV was < 3.8% between 0.1 and 530.0 $pmol \cdot L^{-1}$. Ferritin inter-assay CV was < 4.2% between 0.5 and 2000.0 $\mu g \cdot L^{-1}$. Serum samples were analysed for total 25-hydroxyvitamin-D (25(OH)D) by liquid chromatography tandem mass spectrometry (57). Total 25(OH)D was calculated from the sum of $25(OH)D_3$ and $25(OH)D_2$. The $25(OH)D_2$ and $25(OH)D_3$ assays were calibrated using the National Institute of Science and Technology standard reference material SRM972a. Total 25(OH)D inter-assay CV was < 8.5% between 0.1 and 200.0 $nmol \cdot L^{-1}$. All biochemical analyses were undertaken by the Good Clinical Laboratory Practice and Vitamin D External Quality Assessment Scheme (DEQAS) certified Bioanalytical Facility at the University of East Anglia, Norwich, UK.

Statistical Analyses

These data were secondary analyses (38–45) and so no *a priori* sample size was calculated. All data were analysed using the R programming language (*v.*4.5.1). Participant demographics and anthropometrics at week 1 were analysed descriptively. Inferential statistics were used to test associations between our predictors and four binary outcomes, resulting in four models: 1) training failure (any reason) (0 = pass *vs* 1 = fail for any reason; 2) training failure (medical) (0 = pass *vs* 1 = fail for medical reason); 3) lower limb overuse musculoskeletal injury (0 = no *vs* 1 = yes); 4) lower limb stress fracture (0 = no *vs* 1 = yes) using binary logistic regression (*glm* function). The lower limb stress fracture model produced implausibly large estimates and standard errors, likely due to the low incidence of stress fractures leading to sparse

Copyright © 2026 by the American College of Sports Medicine

data bias; therefore, stress fracture models were not reported or taken further. Stress fractures were also included as a lower limb overuse musculoskeletal injury in analyses. Those who had failed training for non-medical reasons were excluded from models 2 to 3. The predictors for this study were: age [y], height [m], body mass [kg], lean mass [kg], fat mass [kg], body fat [%], aBMD [$g \cdot cm^{-2}$], 2.4 km run time [s], lift strength [kg], vertical jump power [W], smoking status [no *vs* previous *vs* yes], alcohol intake [$units \cdot month^{-1}$], habitual sleep [$h \cdot night^{-1}$], adolescent exercise volume [$min \cdot wk^{-1}$], ferritin [$\mu g \cdot L^{-1}$], total 25(OH)D [$nmol \cdot L^{-1}$], PTH ($pmol \cdot L^{-1}$), βCTX [$\mu g \cdot L^{-1}$], PINP [$\mu g \cdot L^{-1}$], mood scores [anger, confusion, depression, fatigue, tension, vigour], and stress fracture history [never had a stress fracture *vs* unsure *vs* had a previous stress fracture]. Mood scores were only included in training failure models (models 1 and 2). PTH, βCTX, PINP, and previous stress fracture were only included in the musculoskeletal injury model (model 3). All numeric predictors were correlated before building models and fat mass, body mass, and vertical jump power were removed from the male and female models due to having a high correlation with other predictors (> 0.80). Due to the large number of predictors, multiple imputation ($m = 50$) using predictive mean matching (*mice* package, *v*.3.19.0) was used to minimise case-wise deletion. Multiple imputation was used for inferential analysis to allow uncertainty due to missing data to be incorporated into the coefficient estimates (*i.e.*, data were imputed multiple times, and the coefficients were pooled across these imputed data sets) (58). Mean and maximum generalised variance inflation factors were checked for each model (*car* package), and all models had mean and maximum generalised variance inflation factors ≤ 2.2. Odds ratios were calculated from the logistic regression by exponentiating the coefficients. Significance of the marginal effects were taken as $p < 0.05$.

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

Predictive analyses were performed using the same predictors and binary outcomes with the correlated predictors excluded. All analyses were performed using the *tidymodels* packages (*v*.1.4.1). The following supervised statistical learning classification models were evaluated: i) logistic regression; ii) elastic net; iii) random forest, and; iv) gradient boosting. Logistic regression models the probability of a binary outcome from the linear combination of predictors using a logistic link function and elastic net builds on logistic regression by using a shrinkage method to constrain the coefficients (L1 "LASSO' penalty and L2 "Ridge" penalty) towards 0 to reduce overfitting and manage correlated predictors (58, 59). Random forests is an ensemble method that constructs multiple decision trees using a random subset of predictors and combining their predictions, and gradient boosting is a method where decision trees are built sequentially using information from previous tress to improve predictive performance (59). These models were selected based on previous use in military musculoskeletal injury prediction studies (15, 26, 30, 36). Missing predictor data were imputed using k-nearest neighbours (k = 5) within the tidymodels preprocessing workflow. K-nearest neighbours was used for predictive analyses to produce one imputed data set based on the training data within each cross-fold validation, thereby preventing held-out data from contributing to the imputation. Internal validation was performed using five-fold cross-validation (15), whereby the data were randomly split into five equal folds. During each iteration, models were trained using four folds and evaluated on the remaining hold-out-fold, such that each participant was included in a hold-out fold one. Folds were stratified by the binary outcomes being predicted to ensure balance in the outcome across folds. Elastic net, random forest, and gradient boosting models were tuned using grid search with five-fold cross-validation, and the optimal hyperparameters were selected according to the highest receiver operating characteristic area under the curve (ROC AUC). For

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

each held-out participant, the fitted model generated a predicted probability of the outcome ranging from 0 to 1. Binary classifications were generated using both the conventional probability threshold of > 0.5 (original model) and an outcome-specific threshold determined using the Youden Index, which optimises the sum of sensitivity and specificity. Models were evaluated using accuracy, ROC AUC, sensitivity, specificity, positive predictive value, and F1 calculated from the pooled out-of-fold predictions obtained from the five cross-validation folds. The ROC AUC were categorised as poor (≤ 0.70), fair (0.71 to 0.80), good (0.81 to 0.90), and excellent (0.91 to 1.00) model performance (19, 36). Final models were fitted on the entire data set and are provided in Supplemental Data (Supplemental Digital Content 1, https://links.lww.com/MSS/D523).

RESULTS

Participants

A total of 968 male and 450 female recruits volunteered to participate (Table 1). Of the male recruits, 634 participants passed the course (65.5%), 185 were diagnosed with a lower limb overuse musculoskeletal injury (20.4%) and 20 had a stress fracture (2.1%). Of the female recruits, 344 passed the course (76.4 %), 143 were diagnosed with a lower limb overuse musculoskeletal injury (33.7%) and 12 had a stress fracture (2.7%).

Associations Between Predictors and Training Outcome

Associations between each predictor variable and each training outcome for men are presented in Table 2. Slower 2.4 km run time (OR [95%CI] = 1.004 [1.001, 1.008], p = 0.011) and higher reported symptoms of depression (OR [95%CI] = 1.173 [1.065, 1.292], p = 0.001) were associated with higher risk of failing training when controlling for all other predictors. No factor was associated with failing training for a medical reason. Lower whole-body aBMD (OR

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

[95%CI] = 0.008 [0.000, 0.160], p = 0.002), slower 2.4 km run time (OR [95%CI] = 1.006 [1.001, 1.010], p = 0.011), having had a previous stress fracture (OR [95%CI] = 3.267 [1.607, 8.189], p = 0.002), lower PTH (OR [95%CI] = 0.783 [0.652, 0.940], p = 0.009), and lower βCTX (OR [95%CI] = 0.255 [0.069, 0.944], p = 0.041) were associated with higher risk of developing a lower limb overuse musculoskeletal injury when controlling for all other predictors.

Associations between each predictor variable and each training outcome for women are presented in Table 3. Slower 2.4 km run time (OR [95%CI] = 1.009 [1.002, 1.015], p = 0.009) and higher total 25(OH)D (OR [95%CI] = 1.013 [1.005, 1.021], p = 0.001) were associated with higher risk of failing training when controlling for all other predictors. Lower whole-body aBMD (OR [95%CI] = 0.003 [0.000, 0.608], p = 0.033) and slower 2.4 km run time (OR [95%CI] = 1.009 [1.001, 1.017], p = 0.029) were associated with higher risk of failing training for medical reasons when controlling for all other predictors. Slower 2.4 km run time (OR [95%CI] = 1.007 [1.002, 1.012], p = 0.009), higher habitual alcohol intake (OR [95%CI] = 1.006 [1.001, 1.012], p = 0.017), and lower βCTX (OR [95%CI] = 0.057 [0.005, 0.708], p = 0.026) were associated with higher risk of developing a lower limb overuse musculoskeletal injury when controlling for all other predictors.

Training Outcome Prediction

Model performances for each outcome are shown in Figure 1 and Tables 4 and 5 and Supplemental Table 1 (Supplemental Digital Content 2, https://links.lww.com/MSS/D524). Final fitted models are provided in Supplemental Data (Supplemental Digital Content 1, https://links.lww.com/MSS/D523) For predicting training failure (any reason), training failure (medical), and lower limb overuse musculoskeletal injury in men, all of the four modelling methods performed similarly (training failure [any reason], ROC AUC = 0.56 to 0.59; training

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

failure [medical], ROC AUC = 0.57 to 0.60; lower limb overuse musculoskeletal injury, ROC AUC = 0.61 to 0.63) within outcomes, generally with high specificity but poor sensitivity. Model discriminatory performance was generally best for predicting lower limb overuse musculoskeletal injury compared with training failure outcomes. Application of the Youden Index markedly improved sensitivity across all models at the expense of specificity. For predicting training failure (any reason) and training failure (medical) in women, logistic regression generally performed poorer (ROC AUC = 0.61 and 0.53) than elastic net (ROC AUC = 0.63 and 0.54), random forests (ROC AUC = 0.65 and 0.61), and gradient boosting (ROC AUC = 0.64 and 0.64). There was similar performance across the four modelling methods for predicting lower limb overuse musculoskeletal injury (ROC AUC = 0.65 to 0.68). All models had high specificity and poor sensitivity, but application of the Youden Index markedly improved sensitivity across all models at the expense of specificity. Model discriminatory performance was generally better in women than men across outcomes, and best when predicting training failure (any reason) and lower limb overuse musculoskeletal injury compared with training failure (medical) in women.

DISCUSSION

This study explored factors associated with British Army basic training outcome (failure and lower limb overuse musculoskeletal injury) in men and women and whether supervised statistical learning classification models—with a range of flexibilities and interpretabilities—could accurately predict outcome. Slower 2.4 km run time at the start of training was consistently associated with poorer training outcome (failure and lower limb overuse musculoskeletal injury) in men and women. Higher symptoms of depression were associated with training failure, and several bone outcomes (lower aBMD, previous stress fracture, lower PTH, and lower βCTX)

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

were associated with developing a lower limb overuse musculoskeletal injury in men. In women, several bone-related outcomes (higher total 25(OH)D and lower aBMD and βCTX) were also associated with poorer training outcome (failure or lower limb overuse musculoskeletal injury). Statistical learning models provided poor discriminatory performance but may have most utility in identifying at-risk women. Military training represents an excellent opportunity to study training outcomes in a large cohort of young men and women undergoing similar arduous training, and here we provide novel insight into factors associated with training performance and musculoskeletal health, as well as the utility of statistical learning methods.

Associations Between Predictors and Training Outcome

Slower 2.4 km run time was associated with higher risk of failing training for any reason in men and women, and with higher risk of failing training for a medical reason in women only. These findings are in agreement with previous studies showing that lower aerobic performance is a predictor of training failure (19, 20). Best effort 2.4 km run time is related to maximal oxygen uptake (50) and higher maximal oxygen uptake (60–63) and faster 2.4 km run time (37, 64) are associated with better performance and / or lower physiological strain in key military tasks, including load carriage. Therefore, slower 2.4 km run time could increase risk of training failure due to an inability to reach performance standards in key tests like load carriage tests. British Army basic military training courses are physically demanding and involve high volumes of aerobic exercise (1, 8), and slower 2.4 km run time was associated with higher risk of developing a lower limb overuse musculoskeletal injury. Low aerobic fitness one of the most consistent risk factors for musculoskeletal injury (14, 16, 17, 65, 66) and musculoskeletal injuries are the most common cause of entering rehabilitation and medical discharge in basic military training (67). It is unclear why 2.4 km run time was not associated with failing training for a medical reason in

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

men, but men had a higher rate of training failure and medical discharge than women despite a lower incidence of lower limb overuse musculoskeletal injury. Accordingly, other factors than musculoskeletal injury—such as voluntary withdrawal or common mental disorders—could have contributed more training failures in men.

Higher symptoms of depression at the start of training were associated with training failure in men but not women. More depressive symptoms (18) and higher reported occupational stress (68) have previously been associated with training failure in basic military training. British Army basic training is psychologically demanding, and symptoms of stress and depression increase during training (9, 10). Accordingly, starting training with symptoms of depression could increase risk of training failure due to an inability to cope with these psychological demands. Alternatively, higher symptoms of depression during week 1 of training could be an early sign of failing to adjust to Army life. Data from British Army Officer training suggest women have a heightened hypothalamic pituitary adrenal response compared with men in the same basic military training course (10), but men in the present study were completing infantry training—which is the most arduous and psychologically demanding form of British Army soldier training—whereas women were completing soldier (non-infantry) training (13). Therefore, symptoms of depression may be more important for course outcome in infantry vs non-infantry training. Comparisons between sexes should be made with caution as we did not explore the interaction between sex and predictor, men and women completed different training courses, our sample size for women was smaller than for men, and men and women had different incidences of outcomes, which will impact our ability to detect an effect within sex.

Higher total 25(OH)D was associated with failing training in women. There is some evidence that musculoskeletal injury rates are higher in the summer (69) when total 25(OH)D is

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

also higher (42), which may explain a higher failure rate with higher total 25(OH)D. Total 25(OH)D was not associated with musculoskeletal injury in this study, but low total 25(OH)D has been associated with increased risk of injury (14, 15), however, the relationships between the vitamin D metabolites and injury risk (42) and bone structure (70) are complex. Lower whole-body aBMD was associated with failing training in women and developing a musculoskeletal injury in men. Whole-body aBMD is not a clear risk factor for lower limb bone stress injuries, with lower limb bone geometry likely more important (71, 72), but whole-body aBMD can increase in response to military training (12) and has previously been associated with increased risk of musculoskeletal injury (14, 15). Therefore, whole-body aBMD may represent musculoskeletal adaptation to previous training, although we controlled for adolescent exercise volume and lean mass in our analyses. Lower PTH and / or βCTX were associated with higher risk of developing a musculoskeletal injury in men and women. Parathyroid hormone maintains circulating calcium, in part, by mobilising calcium from bone (73), and βCTX is a marker of bone resorption. There is no clear relationship between PTH, βCTX, or other bone metabolism markers and bone adaptation or bone stress injury risk in military training, and no clear pattern in their response to military training (71). It is not clear why lower PTH and / or βCTX was associated with increased risk of musculoskeletal injury but exercise can acutely increase PTH and βCTX (73, 74); although speculative, lower PTH and βCTX could represent lower levels of bone remodelling, which might be important in promoting bone adaptation to exercise (46).

Training Outcome Prediction

All prediction models generally produced poor discriminatory performance (ROC AUC ≤ 0.68). These performances are similar to those previously reported in military populations (ROC AUC ≤ 0.69) for predicting musculoskeletal injury (29, 32, 36) or failing basic military training

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

(22, 26). Better model performance (ROC AUC ≥ 0.75) has been reported in other studies when predicting training outcomes (19, 28), musculoskeletal injury in service (31, 35), or bone stress injuries in military training (33, 34), but most of these studies did not test performance on naïve data, which overestimates model performance (36). Our inferential analyses showed limited association between our predictors and outcomes, which likely contributed to the poor predictive performance of the models. Military training outcomes are multifactorial and highly complex; training failure can be caused by a range of reasons that likely have different risk factors (*e.g.*, developing a musculoskeletal injury *vs* dissatisfaction with military life) that may interact and different musculoskeletal injuries have different pathophysiology. The more complex and flexible models sometimes performed better (ROC AUC) than logistic regression for women (*e.g.*, training failure) but often did not. The relatively comparable performance between the more complex models and logistic regression could be because the relationship between the predictors and outcome were linear and additive, the limited number of events for some outcomes limited the ability to estimate complex relationships reliably, or because of the limited association between predictors and outcome. Although predictive performance was poor, some models may have practical utility when the objective is early identification of at-risk individuals. For example, screening new recruits for targeted interventions before training (*e.g.*, a training course to improving run time) or extra monitoring during training. Application of the Youden index improved the ability to detect 'at-risk' individuals with several models producing sensitivities of 61 to 74% whilst maintaining specificity of 53 to 63%, mainly for women. We propose predicting course failure—provided any follow-up intervention is supportive, does no harm, and is inexpensive—generally favours a model that produces more false positives than false negatives.

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

When predicting training failure (any reason), no model produced a ROC AUC above 0.59 for men, but the random forest model produced an ROC AUC of 0.65 for women with a sensitivity of 68% and a specificity 59% when classifying based on the Youden Index. The prevalence of training failure (any reason) for women was 23.6%, and so for 1000 new female recruits, this model would correctly identify 160 of the 236 failures (true positives) and incorrectly identify 313 of the 764 passes as potential failures (false positives). Similarly, the elastic net model for predicting lower limb overuse musculoskeletal injuries in women (ROC AUC = 0.68, sensitivity = 64%, specificity = 63%) would correctly identify 216 of the 337 injuries (true positives) and incorrectly flag 245 of the 663 injury free women (false positives). It is unclear why model prediction performance was slightly better for women than men, but as previously proposed, the higher lower limb overuse musculoskeletal injury incidence in women could mean a musculoskeletal injury contributed to training failure—either through medical discharge, voluntary withdrawal, or through impairing performance—more in women than men. Women can also have more severe injuries in some military training courses, such as stress fractures (13), and present with more risk factors for injuries such as lower aerobic fitness. Training failure can be for a variety of reasons—including inability to learn military skills, disliking military service, common mental disorders, musculoskeletal injury—and so is likely much harder to predict than lower limb overuse musculoskeletal injuries, which are a more homogenous grouping of outcomes with some established risk factors. Previous studies including both musculoskeletal injury and training outcome support musculoskeletal injuries are generally better predicted (22) with some studies showing very high classification accuracy (ROC AUC = 0.98) when predicting a single injury outcome (medial tibial stress syndrome) (30). We did not attempt to predict stress fracture due to the low number of events, but future work might benefit

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

from predicting more specific outcomes. Nevertheless, implementing effective interventions for those identified by the best performing models here, may improve training outcomes for some individuals.

Limitations

Our data are limited by the small sample size, particularly for women, and where we have outcomes with low prevalence and a low number of events per variable. This low sample size limited our ability to robustly estimate coefficients in our inferential analyses and produced more unstable prediction models and performance estimates in our prediction analyses. Men and women completed different training courses in this study, and so we developed separate models for men and women; however, sex could be a powerful predictive factor in men and women attempting the same course. We also did not formally assess differences between men and women, and any sex comparison should be made with caution. Our analyses also failed to measure several predictors that may explain more variance in our outcome, for example more detailed psychological assessments. We also did not separate non-medical training failures out any further (*e.g.,* failing a fitness test *vs* dissatisfaction with military life) or aim to predict several other relevant outcomes including acute musculoskeletal injuries or mental health outcomes, which could be an area for future investigation. Although all musculoskeletal injuries were recorded at a single medical facility and were classified by a single military clinician using military medical records, each musculoskeletal injury was recorded by a different treating clinician at the point of injury. Accordingly, there will be some variability in how musculoskeletal injuries were diagnosed and recorded on the medical records at the point of medical encounter. Some musculoskeletal injuries may have also not been reported to medical professionals. Other standardised methods for musculoskeletal injury classification from medical

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

records are available (e.g., (75)), which would have likely resulted in different classifications of lower limb overuse musculoskeletal injuries—and therefore effected estimates of injury prevalence, risk factors, and prediction accuracy—and contributed to improved reproducibility of injury classification. Our data are several years old but the findings are still relevant as the demographics of the recruits in this study are similar to recently published data (76) and the predictors of musculoskeletal injury (*e.g.,* run time) have been consistent over several decades (14). Several of our predictor variables were self-reported, including injury history, which is likely to be less reliable than medical records. Our findings will likely not generalise to other military populations or training environments as the reasons for failure and injury are likely site and population specific.

CONCLUSIONS

Slower 2.4 km run time at the start of training was consistently associated with poorer training outcome (failure or lower limb overuse musculoskeletal injury) in men and women in British Army basic training. A series of prediction models could not predict training outcome accurately in either men or women but may be useful in identifying some at risk female recruits. Assessment of aerobic fitness before basic military training might remain the most important method for assessing risk of poor training outcome. Interventions to increase basic training pass rates and reduce lower limb overuse musculoskeletal injury risk should explore improving aerobic fitness prior to starting military training.

Acknowledgements

The authors thank Dr Daniel Kashi, Xin Hui Aw Yong, Mark Ward, Claire Potter, Dr Laurel Wentz, and Miss Rebecca Double for their assistance with data collection. The authors would also like to thank the help of Professor William Fraser for study design and conduct. This study

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

was funded by the UK Ministry of Defence (Army). The authors have no conflicts of interests to declare. The results of the present study do not constitute endorsement by the American College of Sports Medicine. The results of this study are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation. SJ, RMI, SJO, NPW, and JPG designed the study. TJO, ATC, SJ, SJO, and NPW collected the data. JCYT analysed the biochemical samples. TJO produced the manuscript and performed the data analyses. All authors edited the manuscript and approved the final version. All data are available from the corresponding author pending approval of public release from the UK Ministry of Defence.

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

REFERENCES

1. O'Leary TJ, Saunders SC, McGuire SJ, Venables MC, Izard RM. Sex Differences in Training Loads during British Army Basic Training. *Med Sci Sports Exerc*. 2018;50(12):2565–74.
2. O'Leary TJ, Coombs CV, Edwards VC, et al. The effect of sex and protein supplementation on bone metabolism during a 36-h military field exercise in energy deficit. *J Appl Physiol*. 2023;134(6):1481–95.
3. Richmond VL, Carter JM, Wilkinson DM, et al. Comparison of the physical demands of single-sex training for male and female recruits in the British Army. *Mil Med*. 2012;177(6):709–15.
4. Blacker SD, Wilkinson DM, Rayson MP. Gender differences in the physical demands of British Army recruit training. *Mil Med*. 2009;174(8):811–6.
5. Wilkinson DM, Rayson MP, Bilzon JL. A physical demands analysis of the 24-week British Army Parachute Regiment recruit training syllabus. *Ergonomics*. 2008;51(5):649–62.
6. Gifford RM, Greeves JP, Wardle SL, et al. Measuring the Exercise Component of Energy Availability during Arduous Training in Women. *Med Sci Sports Exerc*. 2021;53(4):860–8.
7. O'Leary TJ, Gifford RM, Knight RL, et al. Sex differences in energy balance, body composition, and metabolic and endocrine markers during prolonged arduous military training. *J Appl Physiol*. 2024;136(4):938–48.
8. Powell SD, Siddall AG, Needham-Beck SC, et al. Association between external training loads and injury incidence during 44 weeks of military training. *Scand J Med Sci Sports*. 2023;33(7):1211–20.

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

9. Gifford RM, O'Leary TJ, Double RL, et al. Positive adaptation of HPA axis function in women during 44 weeks of infantry-based military training. *Psychoneuroendocrinology*. 2019;110:104432.
10. Gifford RM, O'Leary TJ, Knight RL, et al. Sex-related hypothalamic-pituitary-gonadal and hypothalamic-pituitary-adrenal axis adaptation during military training. *J Appl Physiol*. 2025;138(1):13–21.
11. Blacker SD, Wilkinson DM, Bilzon JL, Rayson MP. Risk factors for training injuries among British Army recruits. *Mil Med*. 2008;173(3):278–86.
12. O'Leary TJ, Izard RM, Walsh NP, Tang JCY, Fraser WD, Greeves JP. Skeletal macro- and microstructure adaptations in men undergoing arduous military training. *Bone*. 2019;125:54–60.
13. O'Leary TJ, Wardle SL, Rawcliffe AJ, Chapman S, Mole J, Greeves JP. Understanding the musculoskeletal injury risk of women in combat: the effect of infantry training and sex on musculoskeletal injury incidence during British Army basic training. *BMJ Mil Health*. 2023;169(1):57–61.
14. Sammito S, Hadzic V, Karakolis T, et al. Risk factors for musculoskeletal injuries in the military: a qualitative systematic review of the literature from the past two decades and a new prioritizing injury model. *Mil Med Res*. 2021;8(1):66.
15. Foulis SA, Proctor SP, Spiering BA, et al. Model for Musculoskeletal Injury Risk Factors Among US Army Basic Combat Trainees. *JAMA Netw Open*. 2025;8(6):e2513177.
16. Rhon DI, Molloy JM, Monnier A, Hando BR, Newman PM. Much work remains to reach consensus on musculoskeletal injury risk in military service members: A systematic review with meta-analysis. *Eur J Sport Sci*. 2022;22(1):16–34.

ACCEPTED
Copyright © 2026 by the American College of Sports Medicine

17. Dos Santos Bunn P, de Oliveira Meireles F, de Souza Sodre R, Rodrigues AI, da Silva EB. Risk factors for musculoskeletal injuries in military personnel: a systematic review with meta-analysis. *Int Arch Occup Environ Health*. 2021;94(6):1173–89.
18. Tait JL, Drain JR, Bulmer S, Gastin PB, Main LC. Factors Predicting Training Delays and Attrition of Recruits during Basic Military Training. *Int J Environ Res Public Health*. 2022;19(12):7271.
19. Moran DS, Evans RK, Arbel Y, Hadid A, Laor A, Fuks Y. Prediction model for attrition from a combat unit training program. *J Strength Cond Res*. 2011;25(11):2963–70.
20. Pope RP, Herbert R, Kirwan JD, Graham BJ. Predicting attrition in basic military training. *Mil Med*. 1999;164(10):710–4.
21. Vaara JP, Eränen L, Ojanen T, et al. Can Physiological and Psychological Factors Predict Dropout from Intense 10-Day Winter Military Survival Training? *Int J Environ Res Public Health*. 2020;17(23):9064.
22. Orr RM, Cohen BS, Allison SC, Bulathsinhala L, Zambraski EJ, Jaffrey M. Models to predict injury, physical fitness failure and attrition in recruit training: a retrospective cohort study. *Mil Med Res*. 2020;7(1):26.
23. Hunt AP, Orr RM, Billing DC. Developing Physical Capability Standards That are Predictive of Success on Special Forces Selection Courses. *Mil Med*. 2013;178(6):619–24.
24. Kalns J, Baskin J, Reinert A, et al. Predicting Success in the Tactical Air Combat Party Training Pipeline. *Mil Med*. 2011;176(4):431–7.
25. Talcott GW, Haddock CK, Klesges RC, Lando H, Fiedler E. Prevalence and predictors of discharge in United States Air Force Basic Military Training. *Mil Med*. 1999;164(4):269–74.

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

26. Alzate Vanegas JM, Wine W, Drasgow F. Predictions of attrition among US Marine Corps: Comparison of four predictive methods. *Mil Psychol*. 2022;34(2):147–66.

27. O'Neil Snoddy R, Henderson JM. Predictors of Basic Infantry Training Success. *Mil Med*. 1994;159(9):616–22.

28. Forse JN, Koltun KJ, Bird MB, et al. Low psychological resilience and physical fitness predict attrition from US Marine Corps Officer Candidate School training. *Mil Psychol*. 2025;37(6):517–26.

29. Eckard TG, Marshall SW, Kucera KL, Wikstrom EA, Cameron KL, Padua DA. Development of a Pragmatic Prediction Model for Bone Stress Injuries in First-Year US Military Academy Cadets. *J Athl Train*. 2026;61(6):401–7.

30. Shaw A, Newman P, Witchalls J, Hedger T. Externally validated machine learning algorithm accurately predicts medial tibial stress syndrome in military trainees: a multicohort study. *BMJ Open Sport Exerc Med*. 2023;9(2):e001566.

31. Teyhen DS, Shaffer SW, Goffar SL, et al. Identification of Risk Factors Prospectively Associated With Musculoskeletal Injury in a Warrior Athlete Population. *Sports Health*. 2020;12(6):564–72.

32. Teyhen DS, Shaffer SW, Butler RJ, et al. What Risk Factors Are Associated With Musculoskeletal Injury in US Army Rangers? A Prospective Prognostic Study. *Clin Orthop Relat Res*. 2015;473(9):2948–58.

33. Moran DS, Israeli E, Evans RK, et al. Prediction model for stress fracture in young female recruits during basic training. *Med Sci Sports Exerc*. 2008;40(11 Suppl):S636-44.

34. Moran DS, Finestone AS, Arbel Y, Shabshin N, Laor A. A Simplified Model to Predict Stress Fracture in Young Elite Combat Recruits. *J Strength Cond Res*. 2012;26(9):2585–92.

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

35. Rhon DI, Teyhen DS, Collins GS, Bullock GS. Predictive models for musculoskeletal injury risk: why statistical approach makes all the difference. *BMJ Open Sport Exerc Med*. 2022;8(4):e001388.

36. Bird MB, Koltun KJ, Mi Q, et al. Predictive utility of commercial grade technologies for assessing musculoskeletal injury risk in US Marine Corps Officer candidates. *Front Physiol*. 2023;14:1088813.

37. O'Leary TJ, Saunders SC, McGuire SJ, Izard RM. Sex differences in neuromuscular fatigability in response to load carriage in the field in British Army recruits. *J Sci Med Sport*. 2018;21(6):591–5.

38. O'Leary TJ, Jackson S, Izard RM, et al. Sex differences in iron status during military training: a prospective cohort study of longitudinal changes and associations with endurance performance and musculoskeletal outcomes. *Br J Nutr*. 2024;131(4):581-92.

39. Carswell AT, Jackson S, Swinton P, et al. Vitamin D Metabolites are Associated with Physical Performance in Young Healthy Adults. *Med Sci Sports Exerc*. 2022;54(11):1982-89.

40. Kashi DS, Oliver SJ, Wentz LM, et al. Vitamin D and the hepatitis B vaccine response: a prospective cohort study and a randomized, placebo-controlled oral vitamin D3 and simulated sunlight supplementation trial in healthy adults. *Eur J Nutr*. 2021;60(1):475–91.

41. Harrison SE, Oliver SJ, Kashi DS, et al. Influence of Vitamin D Supplementation by Simulated Sunlight or Oral D3 on Respiratory Infection during Military Training. *Med Sci Sports Exerc*. 2021;53(7):1505–16.

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

42. Carswell AT, O'Leary TJ, Swinton P, et al. Vitamin D Metabolites Are Associated With Musculoskeletal Injury in Young Adults: A Prospective Cohort Study. *J Bone Min Res*. 2023;38(10):1453-64.

43. O'Leary TJ, Evans HA, Close M-EO, et al. Hormonal Contraceptive Use and Physical Performance, Body Composition, and Musculoskeletal Injuries during Military Training. *Med Sci Sports Exerc*. 2025;57(3):613–24.

44. O'Leary TJ, Waters HL, Coombs CV, et al. Folate and B12 status in women during British Army basic training: a prospective cohort study exploring implications for endurance performance and bone density and metabolism. *Br J Nutr*. 2026;doi:10.1017/S0007114526106515.

45. Carswell AT, Oliver SJ, Wentz LM, et al. Influence of Vitamin D Supplementation by Sunlight or Oral D3 on Exercise Performance. *Med Sci Sports Exerc*. 2018;50(12):2555–64.

46. Hughes JM, O'Leary TJ, Koltun KJ, Greeves JP. Promoting adaptive bone formation to prevent stress fractures in military personnel. *Eur J Sport Sci*. 2022;22(1):4-15.

47. Fallowfield JL, Leiper RG, Shaw AM, et al. Risk of Injury in Royal Air Force Training: Does Sex Really Matter? *Mil Med*. 2018;185(1-2):170-77.

48. Sanchez-Santos MT, Davey T, Leyland KM, et al. Development of a Prediction Model for Stress Fracture During an Intensive Physical Training Program: The Royal Marines Commandos. *Orthop J Sports Med*. 2017;5(7):2325967117716381.

49. Terry PC, Lane AM, Fogarty GJ. Construct validity of the Profile of Mood States — Adolescents for use with adults. *Psychol Sport Exerc*. 2003;4(2):125–39.

50. Burger SC, Bertram SR, Stewart RI. Assessment of the 2.4 km run as a predictor of aerobic capacity. *S Afr Med J*. 1990;78(6):327–9.

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

51. Nindl BC, Friedl KE, Frykman PN, Marchitelli LJ, Shippee RL, Patton JF. Physical performance and metabolic recovery among lean, healthy men following a prolonged energy deficit. *Int J Sports Med*. 1997;18(5):317–24.

52. Nindl BC, Barnes BR, Alemany JA, Frykman PN, Shippee RL, Friedl KE. Physiological consequences of U.S. Army Ranger training. *Med Sci Sports Exerc*. 2007;39(8):1380–7.

53. Rayson M, Holliman D, Belyavin A. Development of physical selection procedures for the British Army. Phase 2: Relationship between physical performance tests and criterion tasks. *Ergonomics*. 2000;43(1):73–105.

54. Sayers SP, Harackiewicz DV, Harman EA, Frykman PN, Rosenstein MT. Cross-validation of three jump power equations. *Med Sci Sports Exerc*. 1999;31(4):572–7.

55. Harman EA, Gutekunst DJ, Frykman PN, et al. Prediction of Simulated Battlefield Physical Performance from Field-Expedient Tests. *Mil Med*. 2008;173(1):36–41.

56. Fortes MB, Diment BC, Greeves JP, Casey A, Izard R, Walsh NP. Effects of a daily mixed nutritional supplement on physical performance, body composition, and circulating anabolic hormones during 8 weeks of arduous military training. *Appl Physiol Nutr Meab*. 2011;36(6):967–75.

57. Tang JCY, Nicholls H, Piec I, et al. Reference intervals for serum 24,25-dihydroxyvitamin D and the ratio with 25-hydroxyvitamin D established using a newly developed LC-MS/MS method. *J Nutr Biochem*. 2017;46:21–9.

58. Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. 2nd ed. Cham (Switzerland): Springer; 2019.

59. James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning: with Applications in R*. 2nd ed. New York (USA): Springer; 2021.

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

60. Vaara JP, Groeller H, Drain J, et al. Physical training considerations for optimizing performance in essential military tasks. *Eur J Sport Sci*. 2022;22(1):43–57.

61. Friedl KE, Knapik JJ, Häkkinen K, et al. Perspectives on Aerobic and Strength Influences on Military Physical Readiness: Report of an International Military Physiology Roundtable. *J Strength Cond Res*. 2015;29(Supplement 11):S10–23.

62. Feigel ED, Sterczala AJ, Krajewski KT, et al. Physiological characteristics predictive of passing military physical employment standard tasks for ground close combat occupations in men and women. *Eur J Sport Sci*. 2024;24(9):1247–59.

63. Robertson EE, Coombs CV, Halsall EB, et al. Determinants of physiological demands, physiological resilience, and muscle fatigability during load carriage in women: an observational trial. *Eur J Appl Physiol*. 2026;doi:10.1007/s00421-026-06382-8.

64. Coakley SL, Myers SD, Walker EF, et al. 1.5 mile run time and body mass predict 8 mile loaded march performance, irrespective of sex. *J Sci Med Sport*. 2019;22(2):217–21.

65. Lovalekar M, Hauret K, Roy T, et al. Musculoskeletal injuries in military personnel-Descriptive epidemiology, risk factor identification, and prevention. *J Sci Med Sport*. 2021;24(10):963-69.

66. Robinson M, Siddall A, Bilzon J, et al. Low fitness, low body mass and prior injury predict injury risk during military recruit training: a prospective cohort study in the British Army. *BMJ Open Sport Exerc Med*. 2016;2(1):e000100.

67. Sharma J, Greeves JP, Byers M, Bennett AN, Spears IR. Musculoskeletal injuries in British Army recruits: a prospective study of diagnosis-specific incidence and rehabilitation times. *BMC Musculoskelet Disord*. 2015;16:106.

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

68. Jackson S, Agius R, Bridger R, Richards P. Occupational stress and the outcome of basic military training. *Occup Med*. 2011;61(4):253–8.

69. Knapik JJ, Canham-Chervak M, Hauret K, et al. Seasonal variations in injury rates during US Army Basic Combat Training. *Ann Occup Hyg*. 2002;46(1):15–23.

70. O'Leary TJ, Izard RM, Jackson S, et al. Vitamin D and tibial bone density, geometry, and microarchitecture in male military recruits: an observational study and randomized controlled trial. *J Bone Min Res*. 2025;40(7):847–59.

71. Greeves JP, Beck B, Nindl BC, O'Leary TJ. Current risks factors and emerging biomarkers for bone stress injuries in military personnel. *J Sci Med Sport*. 2023;26:S14–21.

72. O'Leary TJ, Rice HM, Greeves JP. Biomechanical Basis of Predicting and Preventing Lower Limb Stress Fractures During Arduous Training. *Curr Osteoporos Rep*. 2021;19(3):308-17.

73. Wherry SJ, Swanson CM, Kohrt WM. Acute catabolic bone metabolism response to exercise in young and older adults: A narrative review. *Exp Gerontol*. 2022;157:111633.

74. Coombs CV, Greeves JP, Young CD, et al. The effect of calcium supplementation on bone calcium balance and calcium and bone metabolism during load carriage in women: a randomized controlled crossover trial. *J Bone Min Res*. 2025;40(6):753–65.

75. Roy TC, Richardson MD, Ritland BM, Cushing RE, Nguyen VT. The Occupational Military Neuromusculoskeletal Injury Matrix. *Mil Med*. 2022;187(7–8):e889–97.

76. Armstrong NC, Bailey SJ, Osofa J, et al. Characterizing the respiratory muscle strength of male British Army infantry recruits. *Physiol Rep*. 2026;14(11):e70833.

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

FIGURE CAPTIONS

Figure 1. Overall discriminatory performance of each model for each outcome for men and women.

SUPPLEMENTAL DIGITAL CONTENT

SDC 1: Supplementary Data.html

SDC 2: Supplementary Table 1.docx

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

Table 1. Participant characteristics (n = 1418). Data are mean ± standard deviation, median [25th percentile, 75th percentile], or n (%).

	Male Recruits (n = 968)	**Female Recruits (n = 450)**
Age (years)	22 ± 11	22 ± 3
Height (m)	1.77 ± 0.06	1.65 ± 0.06
Body Mass (kg)	75.4 ± 10.2	64.7 ± 8.2
Lean Mass (kg)	56.8 ± 6.4	42.0 ± 4.8
Body Fat (%)	19.8 ± 5.5	30.3 ± 4.9
Whole-Body aBMD ($g \cdot cm^2$)	1.23 ± 0.10	1.15 ± 0.09
2.4 km Run Time (s)	631 ± 53	717 ± 70
Maximum Lift Strength (kg)	71 ± 13	42 ± 8
Peak Power Output (W)	3,829 ± 567	2,771 ± 454
Total 25(OH)D ($nmol \cdot L^{-1}$)	56.6 ± 28.8	68.3 ± 32.3
Ferritin ($ug \cdot L^{-1}$)	92.7 ± 55.0	49.3 ± 38.2
PTH ($pmol \cdot L^{-1}$)	3.8 ± 1.2	3.5 ± 1.0
βCTX ($ug \cdot L^{-1}$)	0.51 ± 0.22	0.34 ± 0.15
PINP ($ug \cdot L^{-1}$)	101.9 ± 43.5	70.2 ± 28.1
Sleep ($h \cdot night^{-1}$)	8 ± 2	9 ± 1
Adolescent Exercise Volume ($min \cdot wk^{-1}$)	400 [220, 693]	320 [180, 540]
Previous Stress Fracture		
No	856 (91.8%)	424 (94.6%)
Unsure	41 (4.4%)	9 (2.0%)

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

	Male Recruits (n = 968)	Female Recruits (n = 450)
Yes	35 (3.8%)	15 (3.3%)
Smoker		
No	327 (34.9%)	246 (54.8%)
Previous	186 (19.8%)	96 (21.4%)
Yes	425 (45.3%)	107 (23.8%)
Alcohol (unit·month^{-1})	39 [13, 78]	23 [9, 48]
Anger (au)	0 [0, 1]	0 [0, 1]
Confusion (au)	2 [0, 4]	1 [0, 3]
Depression (au)	0 [0, 1]	0 [0, 1]
Fatigue (au)	4 [2, 6]	4 [3, 6]
Tension (au)	3 [1, 6]	5 [3, 8]
Vigour (au)	9 [7, 11]	8 [6, 10]
Training Outcomes		
Pass	634 (65.5%)	344 (76.4%)
Fail	197 (20%)	68 (15%)
Medically Discharged	63 (6.5%)	10 (2.2%)
Rehabilitation	74 (7.6%)	28 (6.2%)
Lower Limb Overuse Musculoskeletal Injury	185 (20.4%)	143 (33.7%)
Time to First Lower Limb Overuse Musculoskeletal Injury (Days)	37 [15, 59]	27 [15, 65]
Stress Fracture	20 (2.1%)	12 (2.7%)
Time to First Stress Fracture (Days)	38 [15, 62]	62 [43, 86]

aBMD, areal bone mineral density; βCTX, c-telopeptide cross-links of type 1 collagen; PINP, procollagen type 1

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

	Male Recruits (n = 968)	Female Recruits (n = 450)

N-terminal propeptide; PTH, intact parathyroid hormone; total 25(OH)D, total 25-hydroxyvitamin-D.

Men missing data: age, n = 1; height, n = 9; body mass, n = 9; lean mass, n = 239; body fat, n = 239; whole-body aBMD, n = 239; 2.4 km run time, n = 204; maximum lift strength, n = 345; peak power output, n = 337; total 25(OH)D, n = 5; ferritin, n = 4; PTH, n = 8; βCTX, n = 8; PINP, n = 8; sleep, n = 115; previous stress fracture, n = 36; smoker, n = 30; alcohol, n = 24; anger, n = 125; confusion, n = 123; depression, n = 117; fatigue, n = 120; tension, n = 118; vigour, n = 127; lower limb overuse musculoskeletal injury, n = 59.

Women missing data: height, n = 3; body mass, n = 5; lean mass, n = 60; body fat, n = 60; whole-body aBMD, n = 60; 2.4 km run time, n = 144; maximum lift strength, n = 154; peak power output, n = 109; PTH, n = 2; βCTX, n = 2; PINP, n = 2; sleep, n = 62; exercise, n = 1; previous stress fracture, n = 2; smoker, n = 1; alcohol, n = 68; anger, n = 64; confusion, n = 67; depression, n = 64; fatigue, n = 67; tension, n = 64; vigour, n = 63; lower limb overuse musculoskeletal injury, n = 26.

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

Table 2. Associations between predictors and training outcomes in male recruits (n = 968).

	Training Failure (Any Reason)		**Training Failure (Medical)**		**Lower Limb Musculoskeletal Injury**	
	Odds Ratio [95% CI]	p	Odds Ratio [95% CI]	p	Odds Ratio [95% CI]	p
Age (y)	0.985 [0.969, 1.001]	0.072	0.986 [0.965, 1.009]	0.228	1.005 [0.990, 1.021]	0.505
Height (m)	1.000 [0.051, 19.479]	1.000	1.256 [0.021, 75.527]	0.913	0.979 [0.018, 54.318]	0.992
Lean Mass (kg)	1.000 [1.000, 1.000]	0.170	1.000 [1.000, 1.000]	0.221	1.000 [1.000, 1.000]	0.279
Body Fat (%)	0.978 [0.944, 1.012]	0.200	0.991 [0.946, 1.039]	0.719	1.001 [0.956, 1.048]	0.977
aBMD ($g \cdot cm^2$)	0.704 [0.092, 5.402]	0.735	0.091 [0.006, 1.438]	0.088	0.008 [0.000, 0.160]	**0.002**
2.4 km Run Time (per 10 s)	1.045 [1.010, 1.080]	**0.011**	1.030 [0.986, 1.077]	0.183	1.058 [1.013, 1.105]	**0.011**
Lift Strength (kg)	0.996 [0.977, 1.016]	0.691	0.996 [0.972, 1.022]	0.775	0.990 [0.966, 1.014]	0.395
Total 25(OH)D ($nmol \cdot L^{-1}$)	1.001 [0.996, 1.006]	0.816	1.001 [0.994, 1.008]	0.843	0.999 [0.991, 1.006]	0.701
Ferritin ($ug \cdot L^{-1}$)	1.000 [0.997, 1.003]	0.997	1.000 [0.996, 1.004]	0.958	0.998 [0.994, 1.002]	0.306
Sleep ($h \cdot night^{-1}$)	1.032 [0.948, 1.123]	0.467	1.022 [0.908, 1.150]	0.724	1.061 [0.939, 1.200]	0.341
Adolescent Exercise Volume ($min \cdot wk^{-1}$)	1.000 [1.000, 1.000]	0.287	1.000 [1.000, 1.000]	0.257	1.000 [1.000, 1.000]	0.894
Smoking (Previous Smoker[a])	1.070 [0.717, 1.597]	0.740	1.255 [0.709, 2.224]	0.435	0.777 [0.427, 1.415]	0.409

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

	Training Failure (Any Reason)		Training Failure (Medical)		Lower Limb Musculoskeletal Injury	
	Odds Ratio [95% CI]	p	Odds Ratio [95% CI]	p	Odds Ratio [95% CI]	p
Smoking (Smoker[a])	0.971 [0.700, 1.346]	0.859	1.226 [0.769, 1.955]	0.392	1.527 [0.983, 2.374]	0.060
Alcohol (unit·month^{-1})	1.001 [0.999, 1.003]	0.307	1.001 [0.999, 1.004]	0.274	1.000 [0.998, 1.003]	0.855
Anger (au)	0.995 [0.905, 1.094]	0.923	1.055 [0.930, 1.196]	0.406	—	—
Confusion (au)	1.030 [0.949, 1.119]	0.477	0.991 [0.886, 1.107]	0.867	—	—
Depression (au)	1.173 [1.065, 1.292]	**0.001**	1.103 [0.972, 1.250]	0.127	—	—
Fatigue (au)	1.011 [0.960, 1.065]	0.674	1.015 [0.943, 1.092]	0.691	—	—
Tension (au)	0.950 [0.892, 1.011]	0.107	1.002 [0.920, 1.091]	0.967	—	—
Vigour (au)	0.975 [0.929, 1.022]	0.294	0.983 [0.919, 1.052]	0.624	—	—
Previous Stress Fracture (Unsure[b])	—		—		0.628 [0.229, 1.720]	0.365
Previous Stress Fracture (Yes[b])	—		—		3.627 [1.607, 8.189]	**0.002**
PTH (pmol·L^{-1})	—		—		0.783 [0.652, 0.940]	**0.009**
βCTX (ug·L^{-1})	—		—		0.255 [0.069, 0.944]	**0.041**
PINP (ug·L^{-1})	—		—		1.000 [0.994, 1.006]	0.944

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

[a]*vs* never smoked; [b]*vs* never had a stress fracture.

aBMD, whole-body areal bone mineral density; au, arbitrary unit; βCTX, c-telopeptide cross-links of type 1 collagen; CI, confidence interval; PINP, procollagen type 1 N-terminal propeptide; PTH, intact parathyroid hormone; total 25(OH)D, total 25-hydroxyvitamin-D.

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

Table 3. Associations between predictors and training outcomes in female recruits (n = 450).

	Training Failure (Any Reason)		**Training Failure (Medical)**		**Lower Limb Musculoskeletal Injury**	
	Odds Ratio [95% CI]	p	Odds Ratio [95% CI]	p	Odds Ratio [95% CI]	p
Age (y)	1.029 [0.950, 1.115]	0.481	1.074 [0.961, 1.201]	0.208	0.972 [0.891, 1.061]	0.523
Height (m)	0.046 [0.000, 23.812]	0.333	0.003 [0.000, 35.913]	0.229	10.952 [0.036, 3325.313]	0.411
Lean Mass (kg)	1.000 [1.000, 1.000]	0.960	1.000 [1.000, 1.000]	0.617	1.000 [1.000, 1.000]	0.335
Body Fat (%)	0.994 [0.929, 1.063]	0.854	1.036 [0.936, 1.147]	0.489	1.039 [0.972, 1.109]	0.259
aBMD ($g \cdot cm^2$)	0.383 [0.009, 15.574]	0.610	0.003 [0.000, 0.608]	**0.033**	0.031 [0.001, 1.055]	0.054
2.4 km Run Time (per 10 s)	1.090 [1.022, 1.162]	**0.009**	1.095 [1.009, 1.187]	**0.029**	1.069 [1.017, 1.125]	**0.009**
Lift Strength (kg)	0.964 [0.910, 1.020]	0.201	0.943 [0.880, 1.011]	0.100	0.988 [0.947, 1.031]	0.586
Total 25(OH)D ($nmol \cdot L^{-1}$)	1.013 [1.005, 1.021]	**0.001**	0.998 [0.985, 1.011]	0.733	0.994 [0.986, 1.003]	0.179
Ferritin ($ug \cdot L^{-1}$)	0.996 [0.988, 1.004]	0.295	1.002 [0.993, 1.012]	0.672	0.993 [0.985, 1.001]	0.101
Sleep ($h \cdot night^{-1}$)	0.987 [0.802, 1.214]	0.900	1.086 [0.785, 1.504]	0.617	0.939 [0.761, 1.158]	0.555
Adolescent Exercise Volume	1.000 [0.999, 1.000]	0.716	1.000 [1.000, 1.001]	0.290	1.000 [1.000, 1.001]	0.409

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

	Training Failure (Any Reason)		Training Failure (Medical)		Lower Limb Musculoskeletal Injury	
	Odds Ratio [95% CI]	p	Odds Ratio [95% CI]	p	Odds Ratio [95% CI]	p
(min·wk^{-1})						
Smoking (Previous Smoker[a])	0.812 [0.423, 1.559]	0.530	0.752 [0.283, 2.000]	0.567	1.087 [0.585, 2.019]	0.791
Smoking (Smoker[a])	1.014 [0.543, 1.893]	0.965	0.579 [0.211, 1.587]	0.287	1.107 [0.596, 2.054]	0.748
Alcohol (unit·month^{-1})	0.997 [0.991, 1.003]	0.299	0.998 [0.990, 1.007]	0.646	1.006 [1.001, 1.012]	**0.017**
Anger (au)	1.076 [0.896, 1.292]	0.435	1.195 [0.905, 1.578]	0.209	—	—
Confusion (au)	0.989 [0.847, 1.154]	0.885	1.067 [0.863, 1.319]	0.547	—	—
Depression (au)	1.098 [0.876, 1.377]	0.415	0.865 [0.614, 1.220]	0.408	—	—
Fatigue (au)	1.083 [0.978, 1.199]	0.124	1.020 [0.870, 1.195]	0.808	—	—
Tension (au)	0.915 [0.827, 1.011]	0.082	0.969 [0.847, 1.108]	0.642	—	—
Vigour (au)	0.947 [0.864, 1.039]	0.249	0.956 [0.826, 1.106]	0.544	—	—
Previous Stress Fracture (Unsure[b])	—	—	—	—	5.229 [0.393, 69.580]	0.210
Previous Stress Fracture	—	—	—	—	0.500 [0.115, 2.179]	0.355

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

	Training Failure (Any Reason)		Training Failure (Medical)		Lower Limb Musculoskeletal Injury	
	Odds Ratio [95% CI]	p	Odds Ratio [95% CI]	p	Odds Ratio [95% CI]	p
(Yes[b])						
PTH ($pmol \cdot L^{-1}$)	—	—	—	—	1.224 [0.950, 1.576]	0.117
βCTX ($ug \cdot L^{-1}$)	—	—	—	—	0.057 [0.005, 0.708]	**0.026**
PINP ($ug \cdot L^{-1}$)	—	—	—	—	1.009 [0.997, 1.022]	0.132

[a]*vs* never smoked; [b]*vs* never had a stress fracture.

aBMD, whole-body areal bone mineral density; au, arbitrary unit; βCTX, c-telopeptide cross-links of type I collagen; CI, confidence interval; PINP, procollagen type 1 N-terminal propeptide; PTH, intact parathyroid hormone; total 25(OH)D, total 25-hydroxyvitamin-D.

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

Table 4. Performance of prediction models for training outcomes in male recruits (n = 968).

	Training Failure (Any Reason)				Training Failure (Medical)				Lower Limb Musculoskeletal Injury			
	Logistic Regression	Elastic Net	Random Forest	Gradient Boosting	Logistic Regression	Elastic Net	Random Forest	Gradient Boosting	Logistic Regression	Elastic Net	Random Forest	Gradient Boosting
Original	0.66	0.66	0.66	0.65	0.81	0.82	0.82	0.81	0.79	0.80	0.80	0.79
Accuracy	0.56	0.59	0.57	0.58	0.57	0.60	0.59	0.58	0.61	0.63	0.62	0.60
ROC AUC	0.54	0.50	0.50	0.47	0.11	0.00	0.00	0.17	0.48	0.50	0.00	0.36
PPV	0.14	0.00	0.12	0.02	0.01	0.00	0.00	0.02	0.08	0.01	0.00	0.03
Sensitivity	0.94	1.00	0.94	0.99	0.99	1.00	1.00	0.98	0.98	1.00	1.00	0.99
Specificity	0.22	0.06	0.19	0.04	0.01	0.00	0.00	0.04	0.13	0.02	0.00	0.05
F1	0.38	0.34	0.39	0.31	0.14	0.17	0.18	0.12	0.16	0.24	0.24	0.20
Youden	0.59	0.59	0.60	0.51	0.47	0.51	0.55	0.46	0.53	0.70	0.72	0.65
Threshold	0.41	0.42	0.43	0.39	0.22	0.23	0.23	0.21	0.26	0.33	0.35	0.30
Accuracy	0.40	0.56	0.46	0.78	0.76	0.74	0.66	0.75	0.69	0.44	0.44	0.53
PPV	0.69	0.60	0.68	0.36	0.41	0.46	0.53	0.40	0.49	0.77	0.79	0.68
Sensitivity	0.40	0.48	0.44	0.52	0.34	0.35	0.35	0.33	0.38	0.37	0.39	0.38
Specificity												
F1												

F1, F score; ROC AUC, receiver operating characteristic area under the curve; PPV, positive predictive value.

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

Table 5. Performance of prediction models for training outcomes in female recruits (n = 450).

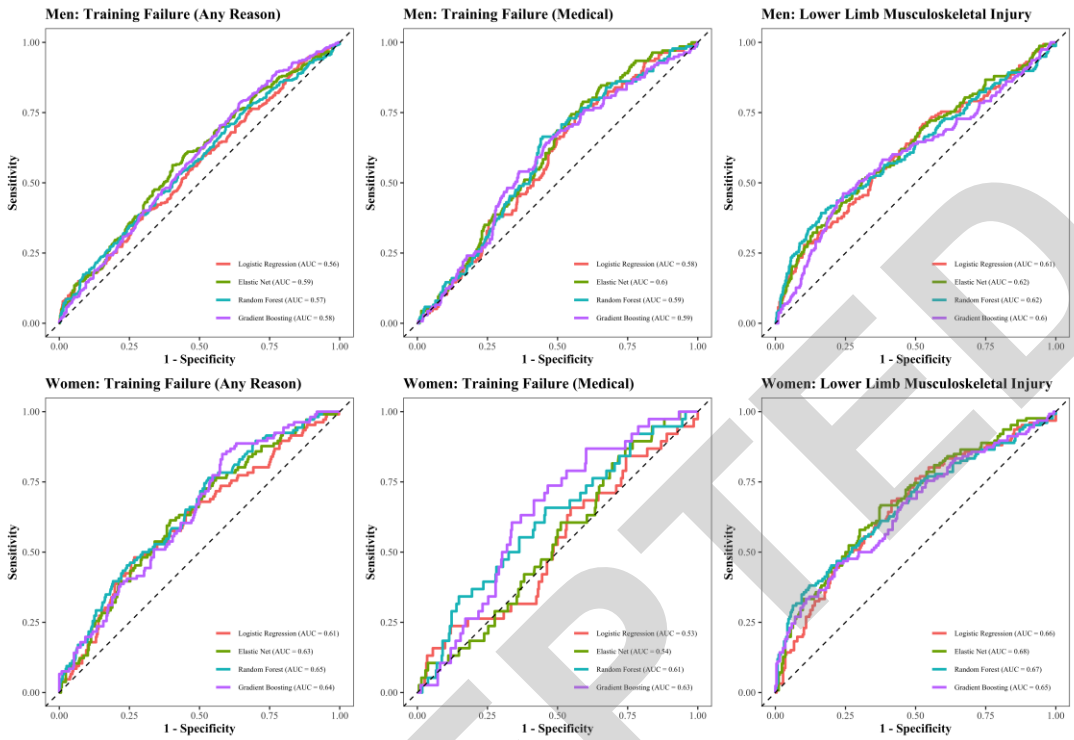
	Training Failure (Any Reason)				Training Failure (Medical)				Lower Limb Musculoskeletal Injury			
	Logistic Regression	Elastic Net	Random Forest	Gradient Boosting	Logistic Regression	Elastic Net	Random Forest	Gradient Boosting	Logistic Regression	Elastic Net	Random Forest	Gradient Boosting
Original	0.73	0.76	0.76	0.77	0.89	0.90	0.90	0.88	0.68	0.67	0.68	0.70
Accuracy	0.61	0.63	0.65	0.64	0.53	0.54	0.61	0.64	0.66	0.68	0.67	0.65
ROC AUC	0.33	0.33	0.46	0.62	0.13	0.00	0.00	0.11	0.53	1.00	0.71	0.60
PPV	0.13	0.01	0.09	0.08	0.03	0.00	0.00	0.03	0.32	0.01	0.04	0.31
Sensitivity	0.92	0.99	0.97	0.99	0.98	1.00	1.00	0.98	0.86	1.00	0.99	0.90
Specificity	0.19	0.02	0.16	0.13	0.04	0.00	0.00	0.04	0.40	0.02	0.08	0.41
F1	0.27	0.23	0.24	0.16	0.22	0.08	0.10	0.02	0.29	0.33	0.40	0.29
Youden	0.67	0.61	0.58	0.52	0.83	0.23	0.59	0.56	0.62	0.64	0.33	0.60
Threshold	0.36	0.32	0.32	0.31	0.16	0.11	0.15	0.15	0.45	0.47	0.33	0.43
Accuracy	0.48	0.61	0.68	0.85	0.16	0.90	0.66	0.74	0.68	0.67	1.00	0.69
PPV	0.73	0.61	0.59	0.42	0.91	0.23	0.59	0.54	0.59	0.63	0.00	0.55
Sensitivity	0.41	0.42	0.44	0.46	0.16	0.20	0.24	0.25	0.54	0.55	0.50	0.53
Specificity												
F1												

F1, F score; ROC AUC, receiver operating characteristic area under the curve; PPV, positive predictive value.

ACCEPTED

Copyright © 2026 by the American College of Sports Medicine

Figure 1



ACCEPTED

Copyright © 2026 by the American College of Sports Medicine