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.

Prediction of Vertebral Recollapse After Percutaneous Pedicle Screw Fixation for Single-Level Traumatic Thoracolumbar Fractures: Model Development and Internal Validation

Xin Xu
Affiliated 960th Hospital of PLA, Shandong Second Medical University
Linyang Song
Qilu Hospital of Shandong University
Junlin Han
Affiliated 960th Hospital of PLA, Shandong Second Medical University
Ang Li
Affiliated 960th Hospital of PLA, Shandong Second Medical University
Ruoxian Song
ruoxiansong@sina.com
Affiliated 960th Hospital of PLA, Shandong Second Medical University

Research Article

Keywords: thoracolumbar fracture, percutaneous pedicle screw fixation, vertebral recollapse, prediction model, Hounsfield units, facet joint violation

Posted Date: November 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11271359/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

Progressive loss of restored vertebral height can occur after percutaneous pedicle screw fixation (PPSF) for traumatic thoracolumbar fractures, yet individualized prediction of vertebral recollapse remains limited. We aimed to develop and internally validate an early postoperative prediction model for vertebral recollapse after short-segment PPSF.

Methods

We retrospectively included 420 neurologically intact patients with single-level T11–L2 traumatic fractures treated with a uniform four-screw, two-rod PPSF construct. Recollapse was defined as a > 20-percentage-point increase in normalized anterior vertebral height loss (ΔHL) from the early postoperative assessment to follow-up before planned implant removal. Twelve candidate predictors were evaluated using least absolute shrinkage and selection operator (LASSO) regression followed by multivariable logistic regression. Internal validation used 1,000 bootstrap resamples. Post hoc comparisons of four nested models used a common stratified 10-fold cross-validation partition. Sensitivity analyses examined alternative recollapse thresholds and continuous ΔHL using multivariable linear regression with HC3 robust standard errors.

Results

Recollapse occurred in 98 patients (23.3%). Six predictors were retained: age, vertebral compression degree, injured-vertebral Hounsfield unit (HU) value, screw-to-endplate distance, facet joint violation (FJV) grade, and lower endplate destruction. The apparent and optimism-corrected areas under the receiver operating characteristic curve (AUCs) were 0.851 and 0.829, respectively; the optimism-corrected Brier score was 0.135 and the calibration slope was 0.883. In nested comparisons, cross-validated AUC increased from 0.734 for age and compression to 0.834 for the full model, while Brier score decreased from 0.158 to 0.132. The largest incremental gain occurred after addition of the postoperative imaging block, whereas HU added little discrimination. Continuous-outcome analysis showed concordant association directions, with residual uncertainty for HU.

Conclusions

The six-predictor model showed promising internally validated performance for early postoperative prediction of vertebral recollapse. Post hoc comparisons suggested additional predictive information from endplate integrity and the combined postoperative imaging variables, whereas the observed

incremental contribution of HU was small. External validation and prospective clinical-impact assessment are required before clinical implementation.

1 Introduction

The thoracolumbar junction is a biomechanically vulnerable transition between the relatively rigid thoracic spine and the more mobile lumbar spine and is therefore particularly susceptible to traumatic injury. In patients requiring surgical stabilization without open decompression, percutaneous pedicle screw fixation (PPSF) can provide indirect reduction and posterior stabilization while reducing paraspinal muscle dissection, intraoperative blood loss, and soft-tissue injury compared with conventional open fixation [1]. The clinical effectiveness and feasibility of minimally invasive percutaneous fixation for traumatic thoracolumbar fractures have also been supported by systematic reviews [2]. Nevertheless, satisfactory early postoperative reduction does not necessarily ensure durable restoration of vertebral morphology. Progressive vertebral-height loss, loss of kyphotic correction, and vertebral recollapse may develop during follow-up despite initially satisfactory fixation [3].

Maintenance of vertebral height after posterior fixation depends in part on the residual structural integrity of the injured anterior column. Endplate and intervertebral disc injuries have been associated with postoperative correction loss after short-segment fixation of traumatic thoracolumbar fractures [4], and disruption of the disc-endplate complex has similarly been linked to poorer maintenance of vertebral height and sagittal alignment [5]. Improved restoration of the injured endplate may contribute to better preservation of vertebral height after short-segment fixation [6]. Greater preoperative vertebral compression has also been associated with subsequent vertebral-height loss after pedicle screw fixation [7], while recent evidence suggests that fracture severity and bone quality may jointly influence maintenance of vertebral height and kyphotic correction [8]. These findings suggest that postoperative recollapse reflects the combined effects of initial structural damage and residual load-bearing capacity rather than a single radiographic feature.

Local bone quality may further influence postoperative mechanical stability. Hounsfield unit (HU) values obtained from routine computed tomography (CT) provide an opportunistic quantitative measure of cancellous bone attenuation and correlate with vertebral bone strength and other biomechanical indicators of bone quality [9, 10]. More recent evidence has linked lower CT-derived HU values to an increased risk of mechanical complications after spinal instrumentation [11]. In traumatic thoracolumbar fractures, HU measurements may therefore complement fracture morphology, particularly because clinically relevant variation in local bone quality can persist even among patients without typical osteoporotic or fragility fractures.

Technical characteristics of posterior instrumentation may also affect maintenance of fracture reduction. Pedicle screw depth and position can influence screw-bone anchorage and postoperative vertebral-height restoration [12]. Facet joint violation (FJV) is another relevant postoperative imaging feature of percutaneous pedicle screw placement [13], and its occurrence may vary with screw-

placement technique and local anatomy in thoracolumbar fracture fixation [14]. Screw-related imaging features may therefore provide prognostic information beyond preoperative patient and fracture characteristics. Integrating patient-related factors, fracture severity, local bone quality, endplate integrity, and postoperative screw-related characteristics may offer a more comprehensive framework for estimating subsequent recollapse risk.

Individualized prediction of vertebral recollapse after thoracolumbar fracture fixation has received increasing attention. Xiong et al. reported a nomogram incorporating age, injury severity, screw distribution, and preoperative vertebral compression [15], and machine-learning methods have also been explored for treatment failure after short-segment posterior fixation [16]. However, published studies differ substantially in patient selection, fixation strategy, predictor definition, and radiographic outcome criteria. These differences limit direct comparison and may reduce transportability across clinical settings. A model developed within a relatively homogeneous fixation strategy and evaluated at a clearly defined postoperative prediction time point may therefore provide a more interpretable estimate of recollapse risk.

Accordingly, we aimed to develop and internally validate a prediction model for vertebral recollapse after short-segment PPSF in neurologically intact patients with single-level traumatic thoracolumbar fractures. The intended prediction time point was after routine early postoperative CT assessment and before hospital discharge. The primary outcome was progression in normalized anterior vertebral height loss from the early postoperative assessment to follow-up before planned implant removal. We additionally examined whether sequentially adding endplate integrity, HU, and postoperative screw-related imaging features improved prediction beyond age and vertebral compression, and whether the observed associations were consistent when height-loss progression was analyzed as a continuous outcome.

2 Materials and methods

2.1 Study design and participants

This retrospective cohort study consecutively included neurologically intact patients with single-level traumatic thoracolumbar fractures at T11–L2 who underwent posterior PPSF at the 960th Hospital of the Joint Logistics Support Force, affiliated with Shandong Second Medical University, between December 2022 and December 2024. The patient-selection and study workflow is shown in Fig. 1. Clinical data were de-identified using anonymous case codes before analysis.

Patients were eligible if they met the following criteria: (1) an acute single-level traumatic fracture involving T11–L2 confirmed by imaging; (2) no neurological deficit; (3) treatment with posterior short-segment PPSF; (4) availability of the clinical, preoperative imaging, and early postoperative imaging data required for model development; and (5) radiographic follow-up at the time of planned implant removal. Exclusion criteria were multilevel spinal fractures, pathological or osteoporotic/fragility fractures, neurological deficits or a requirement for decompression or open surgery, previous surgery involving the

fractured or adjacent levels, severe associated injuries, inadequate radiographic follow-up, or unavailable data required to determine the primary outcome.

All eligible patients were included in the model-development cohort, and internal validation was performed by bootstrap resampling.

The study was approved by the institutional ethics committee (No. 2024009), and the requirement for informed consent was waived because of the retrospective design. The study was conducted in accordance with the Declaration of Helsinki and reported in accordance with the TRIPOD statement [17].

2.2 Surgical procedure and postoperative management

All procedures were performed under general anesthesia with the patient in the prone position. Pedicle screws were inserted percutaneously under C-arm fluoroscopic guidance. A uniform short-segment construct was used, with bilateral pedicle screws placed one vertebra above and one vertebra below the fractured level, resulting in a four-screw, two-rod construct. No screws were inserted into the fractured vertebra.

Fracture reduction was achieved primarily by prone positioning and ligamentotaxis. Moderate distraction or compression was applied through the screw-rod construct when necessary to restore vertebral height and local alignment. Routine postoperative management included analgesia, thromboprophylaxis, wound care, brace use, and progressive ambulation according to each patient's clinical condition.

2.3 Prediction time point and primary outcome

The intended prediction time point was after the early postoperative CT assessment and before hospital discharge.

Early postoperative imaging was routinely obtained during the initial postoperative hospitalization, generally around postoperative day 3. Follow-up radiographs used for the final vertebral-height assessment were obtained at the time of planned implant removal. Exact patient-level intervals between surgery and the imaging examinations were not systematically extracted. The primary outcome was vertebral recollapse assessed on these follow-up radiographs before implant removal. To reduce the influence of differences in radiographic magnification between examinations, anterior vertebral height was normalized separately at the early postoperative and follow-up assessments using the immediately adjacent cranial and caudal vertebrae as internal references.

Normalized anterior vertebral height loss was calculated as:

$$HL(t)\ (\%) = [1 - H_F(t) / \{(H_{cranial}(t) + H_{caudal}(t))/2\}] \times 100$$

where $H_F(t)$, $H_{cranial}(t)$, and $H_{caudal}(t)$ denote the anterior heights of the fractured, cranial adjacent, and caudal adjacent vertebrae at assessment t, respectively. HU is reserved for CT attenuation and is not used as a height symbol.

ΔHL = HL(follow-up) − HL(early postoperative)

Previous studies have evaluated postoperative structural deterioration using vertebral-height loss, correction loss, or vertebral recollapse, with heterogeneous outcome definitions [4, 7, 8, 15]. In the present study, ΔHL > 20 percentage points was used as an author-defined operational radiographic criterion for vertebral recollapse. The threshold was not optimized by maximizing model performance and has not been established as a clinically validated cutoff. Representative measurements and the outcome calculation are shown in Fig. 2F.

Alternative-threshold analyses and continuous ΔHL analyses were used to examine the robustness of the findings to outcome specification. These analyses were not intended to establish the clinical validity or optimality of the > 20-percentage-point criterion.

2.4 Candidate predictors and imaging assessment

Twelve candidate predictors were selected for model development based on clinical relevance, previous evidence, routine availability, and availability at the intended prediction time point: age, body mass index (BMI), injured vertebral compression degree, injured vertebral Hounsfield unit (HU) value, screw-to-endplate distance, facet joint violation (FJV) grade, maximum screw breach distance, sex, current smoking, upper endplate destruction, lower endplate destruction, and adjacent intervertebral disc injury. Definitions and model specifications are summarized in Supplementary Table S1.

All imaging measurements were performed using the institutional picture archiving and communication system. Representative measurement methods are shown in Fig. 2.

The injured-vertebral HU value was measured on preoperative CT using an elliptical region of interest placed within cancellous bone while avoiding cortical bone, fracture lines, focal sclerosis, and the posterior venous plexus (Fig. 2A).

Vertebral compression degree was calculated from the anterior heights of the cranial adjacent vertebra (b1), fractured vertebra (b2), and caudal adjacent vertebra (b3):

Compression (%) = [1 − b2 / ((b1 + b3)/2)] × 100

Screw-to-endplate distance was defined as the shortest distance from each screw tip to the nearest vertebral endplate on early postoperative sagittal CT; the minimum value among the four screws was used as the patient-level measurement (Fig. 2C).

FJV was graded from 0 to 3 according to the classification described by Babu et al. [13], and the higher bilateral grade was recorded (Fig. 2D). Upper and lower endplate destruction were assessed separately according to cortical discontinuity, focal collapse, step-off deformity, or fracture-line involvement (Fig. 2E). All patients underwent preoperative CT and MRI, and adjacent intervertebral disc injury was assessed using both modalities. Maximum screw breach distance was defined as the greatest cortical breach measured among the four pedicle screws.

2.5 Interobserver reliability and missing data

One trained spine surgeon performed all imaging measurements and classifications for the full cohort. For the interobserver reliability assessment, a second trained spine surgeon independently repeated the same measurements and classifications in a randomly selected subset of 50 patients. Continuous imaging variables were assessed using a two-way random-effects, absolute-agreement, single-measure intraclass correlation coefficient [ICC(2,1)]. FJV grade was evaluated using quadratic-weighted Cohen's kappa, whereas binary imaging variables were assessed using unweighted Cohen's kappa.

The analytic dataset was checked for missingness before model development. No missing values were present for the primary outcome or any of the 12 candidate predictors; therefore, statistical imputation was not required.

2.6 Sample-size considerations

The sample size was determined by the available eligible cohort. There were 98 outcome events for 12 candidate predictor parameters. Sample-size adequacy was considered in light of contemporary prediction-model guidance [18]; no prospective sample-size calculation determined the cohort size. A post hoc calculation based on model fit in the development dataset cannot establish that the model is free from overfitting.

2.7 Statistical analysis and model development

Continuous variables were summarized as mean ± standard deviation or median (interquartile range), as appropriate, and categorical variables as number (percentage). Baseline characteristics were descriptively compared between patients with and without vertebral recollapse using appropriate parametric, nonparametric, or categorical tests. These comparisons were not used for predictor screening.

All 12 candidate predictors were entered simultaneously into a least absolute shrinkage and selection operator (LASSO) logistic regression model [19]. Continuous predictors were standardized before penalized regression. The inverse regularization strength parameter (C) was tuned over a logarithmic grid from 0.001 to 100 using 10-fold stratified cross-validation, and the 1-standard-error rule was applied for predictor selection. Predictors retaining nonzero coefficients were then entered simultaneously into an unpenalized multivariable logistic regression model to obtain interpretable regression coefficients, odds ratios (ORs), and 95% confidence intervals (CIs).

No additional predictor elimination based on univariable or multivariable P values, Akaike information criterion, or Bayesian information criterion was performed after LASSO selection. Continuous predictors remained continuous, and ORs were expressed using clinically interpretable increments where appropriate.

Potential nonlinear associations for the four continuous predictors retained in the final model were assessed using restricted cubic splines with five prespecified knots placed at the 5th, 27.5th, 50th, 72.5th, and 95th percentiles of each predictor distribution. Likelihood-ratio tests compared each spline specification with the corresponding linear specification, with 3 degrees of freedom assigned to the nonlinear component. The ordinal FJV term was additionally compared with a categorical specification to assess the adequacy of a linear grade effect. Variance inflation factors were calculated to assess multicollinearity. The final logistic regression equation was used to estimate individual probabilities of vertebral recollapse, and a nomogram was constructed for graphical risk estimation.

2.8 Internal validation and model performance

Internal validation was performed using 1,000 patient-level bootstrap resamples. The entire model-development procedure was repeated within each resample, including predictor standardization, LASSO penalty tuning, predictor selection, and unpenalized logistic model refitting. Five-fold stratified cross-validation was used for LASSO tuning within each bootstrap sample. Optimism was estimated from the difference between bootstrap-sample and original-sample performance and used to derive optimism-corrected estimates.

Model discrimination was quantified using the area under the receiver operating characteristic curve (AUC), and overall predictive accuracy was evaluated using the Brier score. Calibration was assessed using the calibration intercept, calibration slope, and graphical comparison of predicted and observed probabilities; ideal intercept and slope values are 0 and 1, respectively [20].

Predictor-selection stability was assessed as the proportion of bootstrap resamples in which each of the 12 candidate predictors retained a nonzero LASSO coefficient.

For descriptive classification performance, the probability cutoff that maximized the Youden index was calculated, and sensitivity, specificity, positive predictive value, negative predictive value, and accuracy were reported. This cutoff was considered exploratory and was not interpreted as a validated clinical decision or treatment threshold.

Decision curve analysis was used to evaluate model net benefit across threshold probabilities relative to treat-all and treat-none strategies [21]. Patients were also ranked by predicted probability and divided into deciles to compare mean predicted probabilities with observed recollapse rates. These displays were descriptive assessments within the development cohort and were not used to define clinical intervention thresholds.

All statistical tests were two-sided, with $P < 0.05$ used as the conventional reference for statistical significance. Statistical significance was not used as a criterion for predictor inclusion or exclusion from the prediction model.

2.9 Incremental predictive performance

A post hoc analysis compared four nested logistic regression models to describe the additional predictive information associated with successive predictor blocks. Model 1 included age and vertebral compression degree. Model 2 added lower endplate destruction; Model 3 additionally included injured-vertebral HU value; and Model 4 added screw-to-endplate distance and FJV grade, yielding the final six-predictor model. Coding and scaling were identical to those used in the final model, and no interactions were added.

All four models were assessed using the same stratified 10-fold partition (random seed 42). Within each training set, the specified unpenalized logistic regression model was refitted and predictions were generated for the held-out patients. AUC and Brier score were calculated from pooled out-of-fold predictions. The cross-validated calibration slope was estimated by regressing the binary outcome on the logit of these predictions. Apparent performance was also reported, and the change in cross-validated AUC was calculated relative to the preceding model.

The predictor sets were fixed after development of the main model, and LASSO selection was not repeated within this comparison. These estimates therefore characterize the specified model structures conditional on prior selection and do not replace bootstrap assessment of the development procedure. Differences were interpreted descriptively; no confidence intervals or hypothesis tests for between-model differences were calculated.

2.10 Sensitivity analyses

The fixed six-predictor logistic model was assessed using recollapse thresholds of > 15, >17.5, > 20, >22.5, and > 25 percentage points. Coefficients were refitted within the training folds, and 10-fold cross-validated AUC and Brier score were reported for each outcome definition. For the primary > 20-percentage-point outcome, the values correspond to Model 4 in the incremental analysis. Performance across thresholds was interpreted descriptively because both outcome definition and event prevalence changed.

To examine associations without dichotomizing height-loss progression, ΔHL was also analyzed as a continuous outcome using ordinary least-squares regression with the same six predictors entered simultaneously. Negative values were retained. Coefficients represent differences in ΔHL, in percentage points, using the same predictor increments as in the logistic model. HC3 heteroskedasticity-robust standard errors were used, with 95% confidence intervals and two-sided P values based on a t distribution with 413 residual degrees of freedom. R^2 and adjusted R^2 were reported. This analysis was complementary and used the same cohort; it was not an independent validation or an assessment of the optimal binary threshold.

Restricted cubic spline analyses and figure generation were performed using R version 4.5.1 (R Foundation for Statistical Computing, Vienna, Austria). The nested-model and continuous-outcome analyses were implemented in Python using scikit-learn, NumPy, and SciPy.

3 Results

3.1 Study population and baseline characteristics

Of 616 patients assessed for eligibility, 196 were excluded and 420 eligible patients were included in the analysis (Fig. 1). Among the 420 included patients, 98 (23.3%) developed vertebral recollapse and 322 (76.7%) did not. The mean age was 52.45 ± 10.51 years, 294 patients (70.0%) were male, and the mean BMI was 24.31 ± 3.28 kg/m².

Compared with patients without recollapse, those who developed recollapse were older (55.99 ± 10.35 vs. 51.37 ± 10.33 years, P < 0.001), had greater preoperative vertebral compression (49.88 ± 12.17% vs. 41.22 ± 11.49%, P < 0.001), lower injured-vertebral HU values (113.03 ± 22.67 vs. 121.35 ± 27.16, P = 0.003), and greater screw-to-endplate distances (5.38 ± 1.48 vs. 4.60 ± 1.56 mm, P < 0.001). Higher FJV grades and lower endplate destruction were also more frequent among patients with recollapse. Other baseline characteristics are summarized in Table 1.

Table 1
Baseline and perioperative characteristics according to vertebral recollapse status

Characteristic	Overall (n = 420)	No recollapse (n = 322)	Recollapse (n = 98)	P value	SMD
Continuous characteristics, mean ± SD					
Age, years	52.45 ± 10.51	51.37 ± 10.33	55.99 ± 10.35	< 0.001	0.447
BMI, kg/m²	24.31 ± 3.28	24.34 ± 3.19	24.20 ± 3.55	0.716	0.044
Highest VAS score within 0–8 h postoperatively	3.23 ± 1.07	3.23 ± 1.04	3.26 ± 1.15	0.800	0.031
Injury-to-surgery interval, days	3.02 ± 1.89	3.10 ± 1.96	2.77 ± 1.64	0.096	0.175
Operative time, min	116.07 ± 29.69	116.23 ± 30.36	115.53 ± 27.49	0.829	0.024
Intraoperative blood loss, mL	114.60 ± 75.52	113.49 ± 73.64	118.23 ± 81.66	0.608	0.063
Time to first postoperative ambulation, days	5.52 ± 2.96	5.57 ± 2.90	5.36 ± 3.14	0.559	0.071
Injured vertebral compression degree, %	43.24 ± 12.20	41.22 ± 11.49	49.88 ± 12.17	< 0.001	0.743
Injured vertebral HU value	119.41 ± 26.39	121.35 ± 27.16	113.03 ± 22.67	0.003	0.318
Screw-to-endplate distance, mm	4.78 ± 1.57	4.60 ± 1.56	5.38 ± 1.48	< 0.001	0.506
Upper-level mean sagittal screw angle, °	7.68 ± 3.75	7.66 ± 3.82	7.74 ± 3.55	0.851	0.021
Lower-level mean sagittal screw angle, °	7.18 ± 4.12	7.20 ± 4.06	7.09 ± 4.33	0.832	0.025
Upper-level mean axial convergence angle, °	30.62 ± 4.57	30.56 ± 4.58	30.82 ± 4.58	0.621	0.057

Note: Data are presented as mean ± standard deviation, median (interquartile range), or number (%), as appropriate. P values were calculated using Welch's t test, the Mann–Whitney U test, Pearson's χ^2 test, or Fisher's exact test, as appropriate, and were considered descriptive only. Standardized mean differences were additionally reported to quantify between-group differences. For nominal variables with more than two levels, the reported SMD represents the maximum absolute level-specific SMD.

Abbreviations: BMI, body mass index; FJV, facet joint violation; HU, Hounsfield unit; IQR, interquartile range; SD, standard deviation; SMD, standardized mean difference; VAS, visual analog scale.

Characteristic	Overall (n = 420)	No recollapse (n = 322)	Recollapse (n = 98)	P value	SMD
Continuous characteristics, mean ± SD					
Lower-level mean axial convergence angle, °	31.81 ± 4.51	31.77 ± 4.55	31.96 ± 4.42	0.720	0.041
Ordinal or non-normally distributed characteristics, median (IQR)					
FJV grade (0–3)	0.00 (0.00–1.00)	0.00 (0.00–1.00)	1.00 (0.00–2.00)	< 0.001	0.837
Grade 0	249 (59.3%)	219 (68.0%)	30 (30.6%)		
Grade 1	113 (26.9%)	72 (22.4%)	41 (41.8%)		
Grade 2	47 (11.2%)	28 (8.7%)	19 (19.4%)		
Grade 3	11 (2.6%)	3 (0.9%)	8 (8.2%)		
Maximum screw breach distance, mm	0.80 (0.49–1.17)	0.80 (0.49–1.20)	0.81 (0.43–1.11)	0.783	0.016
Categorical characteristics, n (%)					
Male sex	294 (70.0%)	223 (69.3%)	71 (72.4%)	0.632	0.070
Current smoking	110 (26.2%)	87 (27.0%)	23 (23.5%)	0.570	0.082
Major medical comorbidity	43 (10.2%)	37 (11.5%)	6 (6.1%)	0.179	0.190
Injury mechanism				0.858	0.151
Fall	176 (41.9%)	132 (41.0%)	44 (44.9%)		
Fall from height	100 (23.8%)	75 (23.3%)	25 (25.5%)		
Traffic accident	83 (19.8%)	68 (21.1%)	15 (15.3%)		
Training-related injury	25 (6.0%)	19 (5.9%)	6 (6.1%)		
Crush injury by heavy object	20 (4.8%)	15 (4.7%)	5 (5.1%)		
Sprain	16 (3.8%)	13 (4.0%)	3 (3.1%)		

Note: Data are presented as mean ± standard deviation, median (interquartile range), or number (%), as appropriate. P values were calculated using Welch's t test, the Mann–Whitney U test, Pearson's χ^2 test, or Fisher's exact test, as appropriate, and were considered descriptive only. Standardized mean differences were additionally reported to quantify between-group differences. For nominal variables with more than two levels, the reported SMD represents the maximum absolute level-specific SMD.

Abbreviations: BMI, body mass index; FJV, facet joint violation; HU, Hounsfield unit; IQR, interquartile range; SD, standard deviation; SMD, standardized mean difference; VAS, visual analog scale.

Characteristic	Overall (n = 420)	No recollapse (n = 322)	Recollapse (n = 98)	P value	SMD
Continuous characteristics, mean ± SD					
Injured vertebral level				0.961	0.059
T11	36 (8.6%)	27 (8.4%)	9 (9.2%)		
T12	89 (21.2%)	67 (20.8%)	22 (22.4%)		
L1	198 (47.1%)	154 (47.8%)	44 (44.9%)		
L2	97 (23.1%)	74 (23.0%)	23 (23.5%)		
Brace compliance	340 (81.0%)	261 (81.1%)	79 (80.6%)	1.000	0.011
Upper endplate destruction	161 (38.3%)	119 (37.0%)	42 (42.9%)	0.351	0.121
Lower endplate destruction	165 (39.3%)	109 (33.9%)	56 (57.1%)	< 0.001	0.481
Adjacent intervertebral disc injury	202 (48.1%)	156 (48.4%)	46 (46.9%)	0.884	0.030
Hazardous screw breach (≥ 2 mm)	45 (10.7%)	35 (10.9%)	10 (10.2%)	1.000	0.022
Screw-tail facet-capsule impingement	84 (20.0%)	65 (20.2%)	19 (19.4%)	0.977	0.020

Note: Data are presented as mean ± standard deviation, median (interquartile range), or number (%), as appropriate. P values were calculated using Welch's t test, the Mann–Whitney U test, Pearson's χ^2 test, or Fisher's exact test, as appropriate, and were considered descriptive only. Standardized mean differences were additionally reported to quantify between-group differences. For nominal variables with more than two levels, the reported SMD represents the maximum absolute level-specific SMD.

Abbreviations: BMI, body mass index; FJV, facet joint violation; HU, Hounsfield unit; IQR, interquartile range; SD, standard deviation; SMD, standardized mean difference; VAS, visual analog scale.

Interobserver agreement was good to excellent for the imaging measurements and classifications, with ICCs ranging from 0.870 to 0.968 for continuous variables and kappa statistics ranging from 0.700 to 0.908 for categorical variables (Supplementary Table S2).

3.2 Predictor selection and final model

All 12 candidate predictors were entered into the LASSO logistic regression model. Using 10-fold stratified cross-validation and the 1-standard-error rule, six predictors retained nonzero coefficients: age, injured vertebral compression degree, injured vertebral HU value, screw-to-endplate distance, FJV grade, and lower endplate destruction (Fig. 3 and Supplementary Table S3).

These six predictors were subsequently entered into the multivariable logistic regression model (Table 2). Greater age was associated with higher predicted risk of vertebral recollapse (OR per 10-year increase, 1.620; 95% CI, 1.237–2.121; P < 0.001). Each 10-percentage-point increase in vertebral compression was associated with approximately twofold higher odds of recollapse (OR, 2.002; 95% CI, 1.568–2.556; P < 0.001).

Table 2
Final six-predictor multivariable logistic regression model for vertebral recollapse

Predictor	Unit/comparison	β	SE	Adjusted OR	95% CI	P value
Intercept	—	−8.65983	—	—	—	—
Age	Per 10-year increase	0.48221	0.13762	1.620	1.237–2.121	< 0.001
Injured vertebral compression degree	Per 10-percentage-point increase	0.69404	0.12474	2.002	1.568–2.556	< 0.001
Injured vertebral HU value	Per 50-HU increase	−0.47676	0.27010	0.621	0.366–1.054	0.078
Screw-to-endplate distance	Per 1-mm increase	0.30333	0.09164	1.354	1.132–1.621	< 0.001
FJV grade	Per 1-grade increase	1.06277	0.17705	2.894	2.046–4.095	< 0.001
Lower endplate destruction	Yes vs no	1.20997	0.28853	3.353	1.905–5.903	< 0.001

Note: Vertebral recollapse was defined as > 20-percentage-point progression in normalized anterior vertebral height loss from the early postoperative assessment to follow-up before planned implant removal. All six LASSO-selected predictors were entered simultaneously into an unpenalized multivariable logistic regression model. No additional elimination based on P values or information criteria was performed. Odds ratios for continuous predictors are reported using clinically interpretable increments.

Abbreviations: β, regression coefficient; CI, confidence interval; FJV, facet joint violation; HU, Hounsfield unit; OR, odds ratio; SE, standard error.

Greater screw-to-endplate distance was also associated with higher risk (OR per 1-mm increase, 1.354; 95% CI, 1.132–1.621; P < 0.001), as was increasing FJV grade (OR per grade, 2.894; 95% CI, 2.046–4.095; P < 0.001). Lower endplate destruction was associated with substantially higher odds of recollapse (OR, 3.353; 95% CI, 1.905–5.903; P < 0.001). Higher HU values were associated with lower predicted risk (OR per 50-HU increase, 0.621; 95% CI, 0.366–1.054), although the individual coefficient did not reach the conventional level of statistical significance (P = 0.078).

The final prediction equation was:

LP = − 8.65983 + 0.48221 × (Age/10) + 0.69404 × (Compression/10) − 0.47676 × (HU/50) + 0.30333 × Distance + 1.06277 × FJV + 1.20997 × LowerEndplate

where lower endplate destruction was coded as 1 when present and 0 when absent. The predicted probability was calculated as:

P = 1 / (1 + e^(− LP))

A graphical representation of the final logistic regression model was generated to facilitate interpretation of individual risk estimates (Fig. 4); the regression equation remains the primary specification of the model.

3.3 Model diagnostics and predictor stability

Spline analyses using the prespecified five-knot specification did not provide strong evidence of nonlinearity for age (P = 0.847), vertebral compression degree (P = 0.518), HU value (P = 0.555), or screw-to-endplate distance (P = 0.113) (Supplementary Figure S1). Modeling FJV as a categorical variable did not significantly improve model fit compared with treating it as an ordinal linear term (P = 0.140). Variance inflation factors ranged from approximately 1.01 to 1.02, indicating no relevant multicollinearity (Supplementary Table S4).

Bootstrap selection stability was high for all six predictors retained in the final model. Selection frequencies were 99.4% for age, 100.0% for vertebral compression degree, 86.7% for HU value, 99.7% for screw-to-endplate distance, 100.0% for FJV grade, and 99.9% for lower endplate destruction. In contrast, the remaining candidate predictors were selected in only 29.6%–37.5% of bootstrap resamples (Supplementary Table S5 and Supplementary Figure S2).

3.4 Model performance and internal validation

The final model yielded an apparent AUC of 0.851 (95% CI, 0.808–0.890). After full-pipeline bootstrap correction, the optimism-corrected AUC was 0.829. The apparent and optimism-corrected Brier scores were 0.125 and 0.135, respectively (Table 3 and Fig. 5A).

Table 3
Performance of the final prediction model and bootstrap internal validation

Performance measure	Apparent estimate	Bootstrap-corrected estimate / classification result
Area under the ROC curve	0.851 (0.808–0.890)	0.829
Brier score	0.125 (0.106–0.145)	0.135
Calibration intercept	0.000	−0.087
Calibration slope	1.000	0.883
Cox–Snell R^2	0.267	—
Nagelkerke R^2	0.404	—
Exploratory probability cutoff	—	0.272
Sensitivity	—	76.5%
Specificity	—	81.4%
Positive predictive value	—	55.6%
Negative predictive value	—	91.9%
Accuracy	—	80.2%

Note: Apparent estimates were obtained from the full development cohort. Optimism-corrected AUC, Brier score, calibration intercept, and calibration slope were obtained from 1,000 patient-level bootstrap resamples in which the full model-development pipeline—including predictor standardization, LASSO penalty tuning, predictor selection, and unpenalized logistic regression refitting—was repeated. Values in parentheses indicate 95% confidence intervals where available. Ideal calibration intercept and slope values are 0 and 1, respectively. The Youden-derived cutoff of 0.272 was used only for descriptive classification performance and does not represent a validated clinical decision or treatment threshold.

Abbreviations: AUC, area under the receiver operating characteristic curve; NPV, negative predictive value; PPV, positive predictive value; ROC, receiver operating characteristic.

The apparent calibration intercept was approximately 0 and the slope was 1.000, as expected when an unpenalized logistic model is evaluated in its development data. After bootstrap correction, the intercept was − 0.087 and the slope was 0.883, indicating optimism and some overfitting (Table 3). Apparent grouped calibration is shown in Fig. 5B, together with the bootstrap-corrected summary calibration metrics. The apparent Cox–Snell and Nagelkerke R^2 values were 0.267 and 0.404, respectively.

At the exploratory Youden-derived probability cutoff of 0.272, sensitivity was 76.5%, specificity was 81.4%, positive predictive value was 55.6%, negative predictive value was 91.9%, and overall accuracy was 80.2%. This cutoff was used only to describe classification performance and was not considered a clinical intervention threshold.

Within the development cohort, decision curve analysis suggested greater net benefit for the model than for the treat-all and treat-none strategies over portions of the displayed threshold range (Fig. 5C). Observed recollapse rates generally increased across predicted-risk deciles (Fig. 5D). These findings describe internal performance and do not establish clinical effectiveness.

3.5 Incremental predictive performance

The nested-model comparison is presented in Table 4 and Fig. 6. The model containing age and vertebral compression alone had an apparent AUC of 0.742, a cross-validated AUC of 0.734, and a cross-validated Brier score of 0.158. Adding lower endplate destruction increased cross-validated AUC to 0.769 (ΔAUC, 0.035) and reduced the Brier score to 0.152. Adding HU produced a cross-validated AUC of 0.770 (ΔAUC, 0.001) and a Brier score of 0.151.

Table 4
Incremental predictive performance of progressively extended prediction models

Model	Predictors	Apparent AUC	10-fold CV AUC	ΔCV-AUC	Apparent Brier score	10-fold CV Brier score	CV calibration slope
Model 1	Age + compression	0.742	0.734	—	0.156	0.158	0.930
Model 2	Model 1 + lower endplate destruction	0.781	0.769	+ 0.035	0.148	0.152	0.924
Model 3	Model 2 + injured vertebral HU value	0.786	0.770	+ 0.001	0.146	0.151	0.901
Model 4	Model 3 + screw-to-endplate distance + FJV grade	0.851	0.834	+ 0.065	0.125	0.132	0.905

Note: All models used the same 420 observations, including 98 recollapse events defined by > 20-percentage-point height-loss progression. Continuous predictors were retained as continuous variables; lower endplate destruction was binary and FJV grade was ordinal (0–3). Identical stratified 10-fold splits were used for all models (random seed 42). Unpenalized logistic regression was refitted within each training set, and cross-validated AUC and Brier score were calculated from pooled out-of-fold predictions. The cross-validated calibration slope was estimated by logistic regression of the outcome on the logit of the out-of-fold probabilities. ΔCV-AUC is the difference from the immediately preceding model, calculated before rounding. Predictor sets were fixed for this post hoc comparison and selection was not repeated within folds; therefore, these estimates are conditional on the specified predictor sets and are not equivalent to full-pipeline bootstrap correction. Differences are descriptive, and no formal test of incremental performance is reported.

Abbreviations: AUC, area under the receiver operating characteristic curve; CV, cross-validation; FJV, facet joint violation; HU, Hounsfield unit.

Adding screw-to-endplate distance and FJV grade together yielded a cross-validated AUC of 0.834 (ΔAUC, 0.065) and a Brier score of 0.132. Cross-validated calibration slopes for Models 1–4 were 0.930, 0.924, 0.901, and 0.905, respectively. Thus, the largest observed increase between successive models occurred after addition of the combined postoperative imaging block, whereas the additional discrimination associated with HU at the preceding step was small. These descriptive comparisons do not establish statistical significance of the increments or isolate the contribution of either postoperative variable.

3.6 Alternative recollapse thresholds

Sensitivity analyses were performed using alternative thresholds for progression in normalized anterior vertebral height loss (Supplementary Table S6 and Supplementary Figure S3).

The number of recollapse events decreased as the threshold increased, from 132 events (31.4%) at > 15 percentage points to 59 events (14.0%) at > 25 percentage points. The corresponding 10-fold cross-validated AUCs were 0.748, 0.804, 0.834, 0.825, and 0.769 for thresholds of > 15, >17.5, > 20, >22.5, and > 25 percentage points, respectively.

The corresponding cross-validated Brier scores were 0.181, 0.150, 0.132, 0.116, and 0.106. AUC exceeded 0.80 for the > 17.5-, > 20-, and > 22.5-percentage-point outcomes but was lower for the > 15- and > 25-percentage-point outcomes. Because event prevalence differed across definitions, lower Brier scores at higher thresholds should not be interpreted as evidence of better performance for the same prediction task. These analyses do not establish an optimal recollapse threshold.

3.7 Continuous height-loss progression

In the continuous-outcome analysis of all 420 patients, greater age, vertebral compression, screw-to-endplate distance, FJV grade, and lower endplate destruction were associated with greater ΔHL (Supplementary Table S7 and Supplementary Figure S4). Adjusted coefficients were 1.068 percentage points per 10 years of age (95% CI, 0.297–1.839), 1.702 per 10-percentage-point increase in compression (95% CI, 1.023–2.382), 0.750 per 1-mm increase in distance (95% CI, 0.213–1.287), and 3.091 per FJV grade (95% CI, 2.027–4.156). Lower endplate destruction was associated with a 3.338-percentage-point greater ΔHL (95% CI, 1.582–5.093).

The coefficient for HU was negative but uncertain (− 1.099 percentage points per 50 HU; 95% CI, − 2.511 to 0.313; P = 0.127). R^2 was 0.216 and adjusted R^2 was 0.205. Association directions were concordant with those in the logistic model, although most variation in continuous height-loss progression remained unexplained.

4 Discussion

4.1 Principal findings

We developed and internally validated a six-predictor model for vertebral recollapse after short-segment PPSF in neurologically intact patients with single-level traumatic thoracolumbar fractures. The retained predictors—age, injured vertebral compression degree, injured-vertebral HU value, screw-to-endplate distance, FJV grade, and lower endplate destruction—represent complementary domains of patient susceptibility, fracture severity, local bone quality, anterior-column integrity, and postoperative technical imaging characteristics.

The model had an apparent AUC of 0.851 and an optimism-corrected AUC of 0.829. The corrected calibration slope of 0.883 indicated some overfitting. In post hoc nested comparisons, cross-validated AUC increased from 0.734 for age and compression to 0.834 for the full model, with the largest stepwise increase occurring when screw-to-endplate distance and FJV were added together. Continuous ΔHL analysis showed concordant association directions. These findings support further evaluation of the model but do not establish external validity.

Vertebral recollapse was characterized by progression in normalized anterior vertebral height loss. Previous studies have used vertebral-height loss, correction loss, kyphotic progression, or recollapse with heterogeneous measurement approaches [4, 5, 7, 8, 15, 22]. Normalization to the adjacent vertebrae at each time point may reduce the influence of radiographic magnification while focusing the outcome on relative loss of restored morphology. Nevertheless, the threshold used here remains an operational radiographic definition, and its relationship with pain, function, and sagittal alignment requires further study.

4.2 Fracture severity, age, local bone quality, and endplate integrity

Preoperative vertebral compression was one of the strongest predictors. Greater compression may reflect more extensive cancellous disruption and reduced residual load-bearing capacity of the anterior column. This interpretation is consistent with the Load Sharing Classification [23] and with studies linking fracture severity and vertebral compression to correction loss or fixation failure after short-segment posterior fixation [7, 8, 24–27].

Age was also retained as a stable predictor. Increasing age may capture differences in trabecular architecture, tissue quality, healing capacity, and spinal biomechanics rather than osteoporosis alone, particularly because typical osteoporotic or fragility fractures were excluded. Age has also appeared in previous models of vertebral recollapse and correction loss [15, 28].

The estimated association of higher injured-vertebral HU with lower recollapse risk was uncertain (OR per 50 HU, 0.621; 95% CI, 0.366–1.054; P = 0.078). The continuous-outcome estimate was likewise negative, with a confidence interval spanning zero. Moreover, adding HU to age, compression, and lower endplate destruction increased cross-validated AUC by only 0.001. Thus, the present analyses provide limited evidence of a substantial incremental contribution from HU within this cohort. HU was retained because it was selected by the prespecified LASSO procedure, and no subsequent P-value-based

elimination was performed. This does not demonstrate that HU is essential to a transportable model. CT-derived attenuation remains a biologically plausible marker given its reported relationships with bone quality, strength, and mechanical complications [9–11, 29–32].

Treatment selection may partly explain the limited incremental contribution of HU in this cohort. In routine practice at our center, neurologically intact older patients with compression fractures and poor bone quality are more often considered for vertebral augmentation, whereas PPSF is preferentially used in relatively younger patients when posterior stabilization is indicated. Consequently, patients selected for PPSF may represent a comparatively younger population with a narrower and generally more preserved range of bone quality. Together with the exclusion of typical osteoporotic or fragility fractures, this treatment pathway may have reduced between-patient variability in HU and attenuated its independent prognostic contribution after adjustment for fracture severity and postoperative mechanical factors. Because treatment selection and HU distributions among patients receiving alternative procedures were not directly analyzed, this explanation should be considered hypothesis-generating rather than definitive.

Lower endplate destruction was associated with substantially higher recollapse risk. The vertebral endplate contributes to load transfer between the disc and vertebral body, and endplate or disc-endplate injury has previously been linked to loss of correction after thoracolumbar fracture fixation [4–6, 28]. Early restoration of vertebral height may therefore be less durable when the injured endplate and underlying anterior column remain structurally compromised.

4.3 Screw-related characteristics and facet joint violation

Screw-to-endplate distance emerged as a postoperative technical imaging marker. In the standardized four-screw construct used here, this measure reflects screw position within the adjacent instrumented vertebrae rather than within the fractured vertebra itself.

Screw depth, trajectory, and surrounding bone quality influence the screw-bone interface [12, 32–34]. A shorter screw-to-endplate distance could reflect deeper purchase or greater engagement of subendplate bone, although this interpretation remains mechanistic. The present data do not define an optimal screw position and should not be interpreted as evidence that screws should deliberately be placed at a specific endplate distance.

The relevance of adjacent-level screw anchorage may also depend on construct design. Meta-analyses suggest that fracture-level screws can improve radiographic maintenance in selected patients [35–37], whereas other evidence shows heterogeneity across fracture morphology, fixation strategy, and load-sharing characteristics [38, 39]. The present model should therefore be viewed as specific to the four-screw, two-rod construct in which it was developed.

FJV showed a strong graded association with vertebral recollapse. Although FJV has been studied mainly in relation to adjacent-segment pathology and longer-term outcomes [13, 14, 40–43], altered segmental mechanics provide a plausible explanation for its prognostic relevance. Alternatively, FJV may

act as a marker of challenging anatomy or suboptimal trajectory rather than as a direct cause of recollapse.

This distinction is important because the present data establish predictive association rather than causation. The high bootstrap selection frequency of FJV supports its prognostic stability within this dataset, but three-dimensional trajectory analysis and biomechanical studies are needed to determine whether FJV is causally related to maintenance of fracture reduction.

4.4 Incremental prediction and clinical interpretation

Previous prediction studies have used nomograms or machine-learning methods to estimate vertebral recollapse or treatment failure after thoracolumbar fixation [15, 16], while related studies have examined risk factors for correction loss and fixation failure [24–27]. The present model differs by focusing on a homogeneous single-level T11–L2 cohort treated with a uniform PPSF construct and by combining preoperative characteristics with early postoperative screw-related imaging features.

The nested comparison assessed whether the complete model contained predictive information beyond a simpler model based on age and vertebral compression. The observed improvement was concentrated in lower endplate destruction and the combined postoperative imaging block, with little additional discrimination from HU at the tested step. Incremental contributions depend on the order of variable entry and correlations among predictors. The 0.065 AUC increase after adding distance and FJV is therefore a joint contribution and cannot be attributed to either variable separately. Because the candidate model structures were specified after predictor selection, these comparisons should be viewed as exploratory rather than as an unbiased validation of the full selection process.

The optimism-corrected AUC of 0.829 and the fixed-model cross-validated AUC of 0.834 arose from different validation procedures and should be interpreted separately. Bootstrap assessment evaluated the development procedure, whereas the nested comparison refitted fixed predictor sets. The calibration slope of 0.883 indicates residual overfitting. Decision curve analysis suggested potential net benefit within the development cohort, but neither the DCA nor the exploratory Youden cutoff establishes a validated clinical decision rule.

The most plausible near-term role of the model is risk-informed follow-up rather than treatment selection. In future prospective studies, patients with higher predicted risk could be considered for closer radiographic surveillance and more structured assessment of postoperative loading and rehabilitation. Whether model-guided follow-up improves patient-centered outcomes remains unknown.

Screw-to-endplate distance and FJV are potentially modifiable postoperative imaging characteristics, but the present findings do not establish that changing either variable will reduce recollapse. They should therefore be considered candidate technical quality markers for external validation rather than immediate operative targets.

4.5 Interpretation of outcome sensitivity analyses

Agreement in the directions of association between the binary and continuous analyses suggests that the main findings were not confined to dichotomization at exactly 20 percentage points. However, the continuous model explained approximately 22% of outcome variation, indicating substantial unexplained heterogeneity. These analyses reused the same cohort and predictors and do not constitute independent replication.

Discrimination also varied with the binary outcome definition, with lower AUCs at the most permissive and strictest thresholds. The > 20-percentage-point definition remains an operational radiographic endpoint. Neither the threshold analysis nor the continuous-outcome analysis establishes its clinical importance, identifies an optimal cutoff, or demonstrates correspondence with pain, disability, or a need for further intervention.

4.6 Strengths, limitations, and future directions

Strengths of this study include the uniform fracture-level range and fixation construct, assessment of predictors available at a defined early postoperative time point, retention of continuous predictors in continuous form, and evaluation of calibration and predictor-selection stability. The added nested-model comparison and continuous-outcome analysis further clarify the observed contributions of predictor groups and the dependence of the findings on outcome definition.

Several limitations should be considered. First, the retrospective single-center design remains susceptible to selection bias, measurement error, and residual confounding. Exact screw dimensions, three-dimensional trajectory, reduction quality, residual vertebral voids, surgeon effects, rehabilitation intensity, and postoperative loading could not be measured uniformly.

Second, the primary outcome was radiographic and its clinical importance remains uncertain. Adjacent-threshold and continuous-outcome analyses provide complementary descriptions but cannot validate the > 20-percentage-point cutoff. Future studies should assess relationships with pain, function, sagittal alignment, and subsequent intervention. Although imaging was obtained according to a relatively consistent institutional follow-up protocol, exact patient-level imaging intervals were not available for analysis, and minor variation in assessment timing may have influenced the magnitude of observed vertebral-height loss.

Third, HU measurements may vary with scanner calibration, acquisition parameters, reconstruction algorithms, acute fracture-related trabecular impaction, and region-of-interest placement. Interobserver reliability was high in the present study, but reproducibility across institutions and imaging platforms remains to be established.

Fourth, all validation was internal. The nested-model comparisons were post hoc, conditional on prior predictor selection, and based on a single 10-fold partition. Uncertainty intervals for performance differences were not estimated. The continuous analysis likewise reused the selected variables and should be regarded as exploratory. Independent external validation is required before absolute risk estimates or incremental performance can be considered transportable.

Finally, the model applies specifically to neurologically intact patients with single-level traumatic T11–L2 fractures treated with a four-screw, two-rod PPSF construct without fracture-level screws and excludes typical osteoporotic or fragility fractures. Generalizability to other levels, fracture patterns, osteoporotic fractures, open or long-segment procedures, and constructs incorporating fracture-level screws is unknown [37, 38, 44]. Longer follow-up is also needed because radiographic deterioration may continue beyond the observation period assessed here [45].

Independent multicenter external validation should therefore be the next step. If transportability is confirmed, prospective impact studies should evaluate whether model-informed surveillance or technical optimization improves radiographic and patient-centered outcomes.

5 Conclusions

A six-predictor model showed promising internally validated performance for early postoperative prediction of vertebral recollapse after four-screw, two-rod PPSF in neurologically intact patients with single-level traumatic T11–L2 fractures. Post hoc comparisons suggested additional predictive information from endplate integrity and the combined postoperative imaging variables, while the observed incremental contribution of HU was small. Continuous-outcome analysis showed concordant association directions but did not validate the binary threshold. Multicenter external validation and prospective clinical-impact assessment are required before clinical use.

Abbreviations

One-standard-error rule (1-SE)

Area under the receiver operating characteristic curve (AUC)

Body mass index (BMI)

Confidence interval (CI)

Computed tomography (CT)

Cross-validation (CV)

Decision curve analysis (DCA)

Facet joint violation (FJV)

Hounsfield unit (HU)

Intraclass correlation coefficient (ICC)

Interquartile range (IQR)

Least absolute shrinkage and selection operator (LASSO)

Linear predictor (LP)

Likelihood ratio (LR)

Magnetic resonance imaging (MRI)

Negative predictive value (NPV)

Odds ratio (OR)

Percutaneous pedicle screw fixation (PPSF)

Positive predictive value (PPV)

Receiver operating characteristic (ROC)

Standard deviation (SD)

Standard error (SE)

Standardized mean difference (SMD)

Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD)

Variance inflation factor (VIF)

Visual analog scale (VAS)

Declarations

Ethics approval and consent to participate

This retrospective study was approved by the Ethics Committee of the Affiliated 960th Hospital of PLA, Shandong Second Medical University (Approval No. 2024009). The requirement for written informed consent was waived by the ethics committee because of the retrospective design. All procedures were conducted in accordance with the Declaration of Helsinki and applicable institutional ethical standards.

Clinical trial number: not applicable.

Consent for publication

Not applicable. All clinical images included in the manuscript were fully anonymized and contained no identifiable personal information.

Competing interests

The authors declare no competing interests.

Availability of data and materials

The datasets generated and/or analysed during the current study are not publicly available because they contain potentially identifiable clinical information, but they are available from the corresponding author on reasonable request and subject to applicable institutional requirements.

Funding

This study received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author contributions

X.X. and R.S. conceived and designed the study. X.X. performed the statistical analyses and drafted the manuscript. L.S. and J.H. contributed to clinical data acquisition and imaging assessment. A.L. contributed to data management, figure preparation, and manuscript organization. R.S. supervised the study and critically revised the manuscript. All authors contributed to interpretation of the results, reviewed the manuscript, and approved the final version.

Acknowledgements

We thank Lili Yu for statistical consultation.

References

1. Tian F, Tu LY, Gu WF, et al. Percutaneous versus open pedicle screw instrumentation in treatment of thoracic and lumbar spine fractures: a systematic review and meta-analysis. Med (Baltim). 2018;97(41):e12535. 10.1097/MD.0000000000012535.
2. Pannu CD, Farooque K, Sharma V, Singal D. Minimally invasive spine surgeries for treatment of thoracolumbar fractures of spine: a systematic review. J Clin Orthop Trauma. 2019;10(Suppl 1):S147–55. 10.1016/j.jcot.2019.04.012.

3. Oh HS, Seo HY. Percutaneous Pedicle Screw Fixation in Thoracolumbar Fractures: Comparison of Results According to Implant Removal Time. Clin Orthop Surg. 2019;11(3):291–6. 10.4055/cios.2019.11.3.291.
4. Hashimura T, Onishi E, Ota S, et al. Correction loss following short-segment posterior fixation for traumatic thoracolumbar burst fractures related to endplate and intervertebral disc destruction. BMC Musculoskelet Disord. 2023;24:174. 10.1186/s12891-023-06288-y.
5. Lee KY, Kim MW, Seok SY, Kim DR, Im CS. The Relationship between Superior Disc-Endplate Complex Injury and Correction Loss in Young Adult Patients with Thoracolumbar Stable Burst Fracture. Clin Orthop Surg. 2017;9(4):465–71. 10.4055/cios.2017.9.4.465.
6. Hou J, Ren D, Chen Y, et al. Effectiveness of the Endplate Reduction Technique Combined With Bone Grafting for the Treatment of Thoracolumbar Fractures by Using Posterior Short-Segment Fixation. Neurospine. 2023;20(1):353–64. 10.14245/ns.2244980.490.
7. Shi Z, Zhang C, Wang B, et al. Analysis of Influencing Factors of Vertebral Height Loss After Pedicle Screw Fixation of Thoracolumbar Fracture. Clin Spine Surg. 2023;36(8):E362–8. 10.1097/BSD.0000000000001467.
8. Zhang J, Ye Z, Mao Y. Factors associated with loss of vertebral height and kyphosis correction after intermediate screws in short segment pedicular fixation for type-A fractures of the thoracolumbar spine: a retrospective study. Med (Baltim). 2024;103(22):e38343. 10.1097/MD.0000000000038343.
9. Schreiber JJ, Anderson PA, Rosas HG, Buchholz AL, Au AG. Hounsfield units for assessing bone mineral density and strength: a tool for osteoporosis management. J Bone Joint Surg Am. 2011;93(11):1057–63. 10.2106/JBJS.J.00160.
10. Martini ML, Mikula AL, Lakomkin N, et al. Opportunistic CT-Based Hounsfield Units Strongly Correlate with Biomechanical CT Measurements in the Thoracolumbar Spine. Spine (Phila Pa 1976). 2024;49(14):1021–8. 10.1097/BRS.0000000000004822.
11. Egu C, Najjar E, Komaitis S, et al. Evaluating the prognostic role of computed tomography Hounsfield units in anticipating spinal outcomes post-instrumentation: a systematic review and meta-analysis. Eur Spine J. 2025;34(4):1420–32. 10.1007/s00586-025-08737-9.
12. Wang J, Ren D, Geng L, et al. Effect of screw insertion depth into fractured vertebrae in the treatment of thoracolumbar fractures. J Orthop Surg Res. 2024;19(1):665. 10.1186/s13018-024-05026-x.
13. Babu R, Park JG, Mehta AI, et al. Comparison of superior-level facet joint violations during open and percutaneous pedicle screw placement. Neurosurgery. 2012;71(5):962–70. 10.1227/NEU.0b013e31826a88c8.
14. Zhang RJ, Zhou LP, Zhang HQ, et al. Rates and risk factors of intrapedicular accuracy and cranial facet joint violation among robot-assisted, fluoroscopy-guided percutaneous, and freehand techniques in pedicle screw fixation of thoracolumbar fractures: a comparative cohort study. BMC Surg. 2022;22(1):52. 10.1186/s12893-022-01502-5.
15. Xiong X, Chen WW, Zhou WH, et al. A Prediction Nomogram for Fractured Vertebra Recollapse After Posterior Reduction and Pedicle Screw Fixation in Thoracolumbar Fractures. World Neurosurg.

2024;189:e807–13. 10.1016/j.wneu.2024.07.009.

16. Khaledian N, Bagheri SR, Sharifi H, Alimohammadi E. The efficacy of machine learning models in forecasting treatment failure in thoracolumbar burst fractures treated with short-segment posterior spinal fixation. J Orthop Surg Res. 2024;19(1):211. 10.1186/s13018-024-04690-3.
17. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med. 2015;162(1):55–63. 10.7326/M14-0697.
18. Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. BMJ. 2020;368:m441. 10.1136/bmj.m441.
19. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Ser B Stat Methodol. 1996;58(1):267–88.
20. Van Calster B, McLernon DJ, van Smeden M, et al. Calibration: the Achilles heel of predictive analytics. BMC Med. 2019;17(1):230. 10.1186/s12916-019-1466-7.
21. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Mak. 2006;26(6):565–74. 10.1177/0272989X06295361.
22. Perna A, Santagada DA, Bocchi MB, et al. Early loss of angular kyphosis correction in patients with thoracolumbar vertebral burst (A3–A4) fractures who underwent percutaneous pedicle screws fixation. J Orthop. 2021;24:77–81. 10.1016/j.jor.2021.02.029.
23. McCormack T, Karaikovic E, Gaines RW. The load sharing classification of spine fractures. Spine (Phila Pa 1976). 1994;19(15):1741–4. 10.1097/00007632-199408000-00014.
24. Javadnia P, Ramazani Y, Moradinazar M, et al. Predictors of failure in posterior short segment instrumentation for thoracolumbar burst fractures: a systematic review and meta-analysis. J Orthop Surg Res. 2025;20(1):736. 10.1186/s13018-025-06149-5.
25. Alimohammadi E, Bagheri SR, Joseph B, et al. Analysis of factors associated with the failure of treatment in thoracolumbar burst fractures treated with short-segment posterior spinal fixation. J Orthop Surg Res. 2023;18(1):690. 10.1186/s13018-023-04190-w.
26. Tanaka M, Thakur A, Rahman MA, et al. Risk Factors of Correction Loss After Percutaneous Reduction and Fixation for Thoracolumbar Burst Fracture: A One-Year Follow-Up Study. Cureus. 2025;17(5):e84961. 10.7759/cureus.84961.
27. Vilela A, Couto B, Ferreira D, et al. Risk factors for failure in short segment pedicle instrumentation in thoracolumbar fractures. Brain Spine. 2025;5:104266. 10.1016/j.bas.2025.104266.
28. Ge T, Li W, Wu J, et al. Impact of traumatic intervertebral disc injury on loss of correction following pedicle screw fixation for thoracolumbar fractures. BMC Musculoskelet Disord. 2025;26(1):507. 10.1186/s12891-025-08759-w.
29. Xu F, Zou D, Li W, et al. Hounsfield units of the vertebral body and pedicle as predictors of pedicle screw loosening after degenerative lumbar spine surgery. Neurosurg Focus. 2020;49(2):E10. 10.3171/2020.5.FOCUS20249.

30. Singh M, Kim J, Nassar J, et al. Generation and validation of a Hounsfield unit (HU) threshold for predicting mechanical and junctional complications following thoracolumbar or lumbar fusion: a systematic review and meta-analysis. Eur Spine J. 2026;35(4):1527–38. 10.1007/s00586-025-09236-7.
31. Akabane T, Suzuki T, Konno Y, et al. Bone density as a risk of early loss of correction after percutaneous posterior spinal fixation for traumatic thoracolumbar fracture: a study on the usefulness of Hounsfield unit values on computed tomography scan. Eur Spine J. 2025;34(2):498–505. 10.1007/s00586-024-08508-y.
32. Lee SJ, Lee JH, Lee HJ, Oh JW, Park IH. Pullout strength of pedicle screws using cadaveric vertebrae with or without artificial demineralization. Spine J. 2021;21(9):1580–6. 10.1016/j.spinee.2021.04.010.
33. Wadhwa RK, Thakur JD, Khan IS, et al. Adjustment of suboptimally placed lumbar pedicle screws decreases pullout strength and alters biomechanics of the construct: a pilot cadaveric study. World Neurosurg. 2015;83(3):368–75. 10.1016/j.wneu.2014.04.065.
34. Goda Y, Higashino K, Toki S, et al. The Pullout Strength of Pedicle Screws Following Redirection After Lateral Wall Breach or End-plate Breach. Spine (Phila Pa 1976). 2016;41(15):1218–23. 10.1097/BRS.0000000000001600.
35. Kapoen C, Liu Y, Bloemers FW, Deunk J. Pedicle screw fixation of thoracolumbar fractures: conventional short segment versus short segment with intermediate screws at the fracture level-a systematic review and meta-analysis. Eur Spine J. 2020;29(10):2491–504. 10.1007/s00586-020-06479-4.
36. Li K, Zhang W, Liu D, et al. Pedicle screw fixation combined with intermediate screw at the fracture level for treatment of thoracolumbar fractures: a meta-analysis. Med (Baltim). 2016;95(33):e4574. 10.1097/MD.0000000000004574.
37. Perna A, Franchini A, Rovere G, et al. Do More Screws Mean Better Stability? Four (4S) vs. Six (6S) Screws for Short-Segment Fixation in Thoracolumbar Fractures-A Systematic Review and Meta-Analysis. J Clin Med. 2025;14(16):5672. 10.3390/jcm14165672.
38. Grin A, Karanadze V, Lvov I, et al. Effective method of pedicle screw fixation in patients with neurologically intact thoracolumbar burst fractures: a systematic review of studies published over the last 20 years. Neurocirugia (Engl Ed). 2024;35(6):299–310. 10.1016/j.neucie.2024.07.009.
39. Chen Y, Zhang W, Ma H, et al. Whether patients with load sharing classification (LSC) score ≤ 6 need the additional screws in the fractured vertebra? a prospective randomized study with 2 years of follow up. Eur Spine J. 2026;35(2):913–21. 10.1007/s00586-025-09688-x.
40. Liu PC, Lu Y, Lin HH, et al. Superior facet joint violation between open and minimally invasive techniques in lumbar fusion surgery: an updated systematic review and meta-analysis. J Chin Med Assoc. 2023;86(1):113–21. 10.1097/JCMA.0000000000000788.
41. Oh HS, Seo HY. The Relationship between Adjacent Segment Pathology and Facet Joint Violation by Pedicle Screw after Posterior Lumbar Instrumentation Surgery. J Clin Med. 2021;10(13):2911.

10.3390/jcm10132911.

42. Sinz NJ, Hwang RW, Lee GH, et al. Pedicle Screw-Associated Violation of the Adjacent Unfused Facet Joint: Clinical Outcomes and Fusion Rates. Global Spine J. 2024;14(5):1477–84. 10.1177/21925682221145651.
43. McNamee C, McDonnell JM, Kelly D, et al. Superior facet joint violation after lumbar pedicle screw placement: a scoping review of prevalence, biomechanics, and implications for adjacent segment disease. Asian Spine J. 2025;19(6):1032–44. 10.31616/asj.2025.0143.
44. Perna A, Franchini A, Gorgoglione FL, et al. Short-segment percutaneous fusion versus open posterior fusion with screw in the fractured vertebra for thoracolumbar junction burst vertebral fracture treatment. J Neurosci Rural Pract. 2024;15(1):34–41. 10.25259/JNRP_370_2023.
45. Kultur Y, Sarikaya İ, Ozsahin MK, Davulcu CD, Aydingoz O. Twenty year outcomes following short-segment posterior instrumentation and fusion for thoracolumbar burst fractures: A retrospective observational study. Med (Baltim). 2024;103(46):e40579. 10.1097/MD.0000000000040579.

Figures

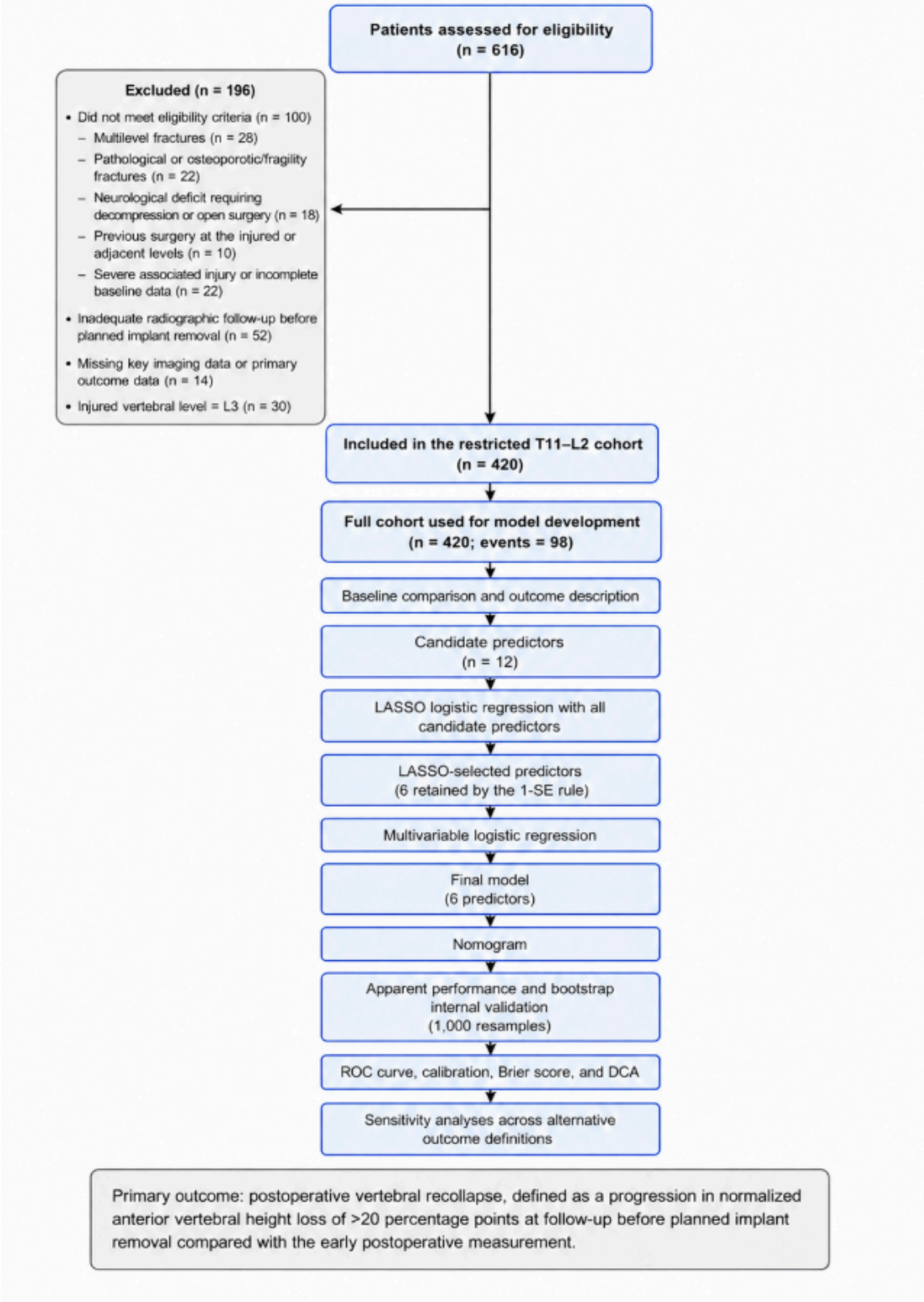


Figure 1

Patient selection and model-development workflow

Of 616 patients assessed for eligibility, 196 were excluded and 420 were included; 98 met the primary recollapse outcome. Twelve candidate predictors were considered, six were retained by LASSO, and the final logistic model was internally assessed using bootstrap resampling. The primary outcome was >20-

percentage-point progression in normalized anterior vertebral height loss from the early postoperative assessment to follow-up before planned implant removal.

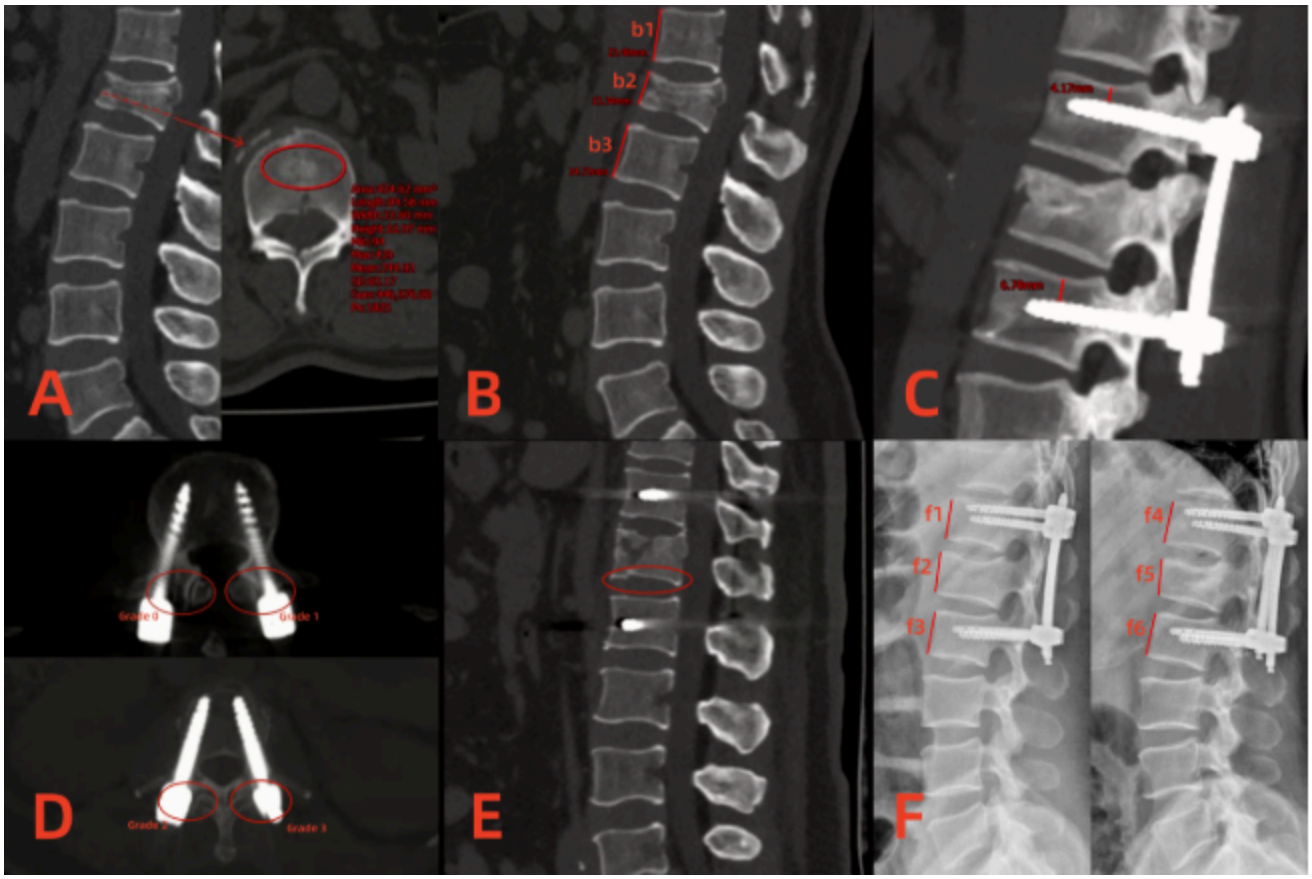


Figure 2

Imaging assessment and definition of vertebral recollapse

(A) Injured-vertebral HU measurement on preoperative CT. (B) Anterior heights of the cranial adjacent vertebra (b1), fractured vertebra (b2), and caudal adjacent vertebra (b3), used to calculate compression degree. (C) Screw-to-endplate distance on early postoperative sagittal CT; the minimum among the four screws was recorded. (D) FJV grades 0–3. (E) Lower endplate destruction. (F) Early postoperative and follow-up standing lateral radiographs obtained before planned implant removal. The labels f1/f2/f3 and f4/f5/f6 denote the cranial adjacent, fractured, and caudal adjacent anterior vertebral heights at the two assessments. At each assessment, HL (%) = $[1 - H_F/\{(H_{cranial} + H_{caudal})/2\}] \times 100$. ΔHL is the follow-up value minus the early postoperative value; recollapse is defined as ΔHL >20 percentage points. CT, computed tomography; FJV, facet joint violation; HU, Hounsfield unit.

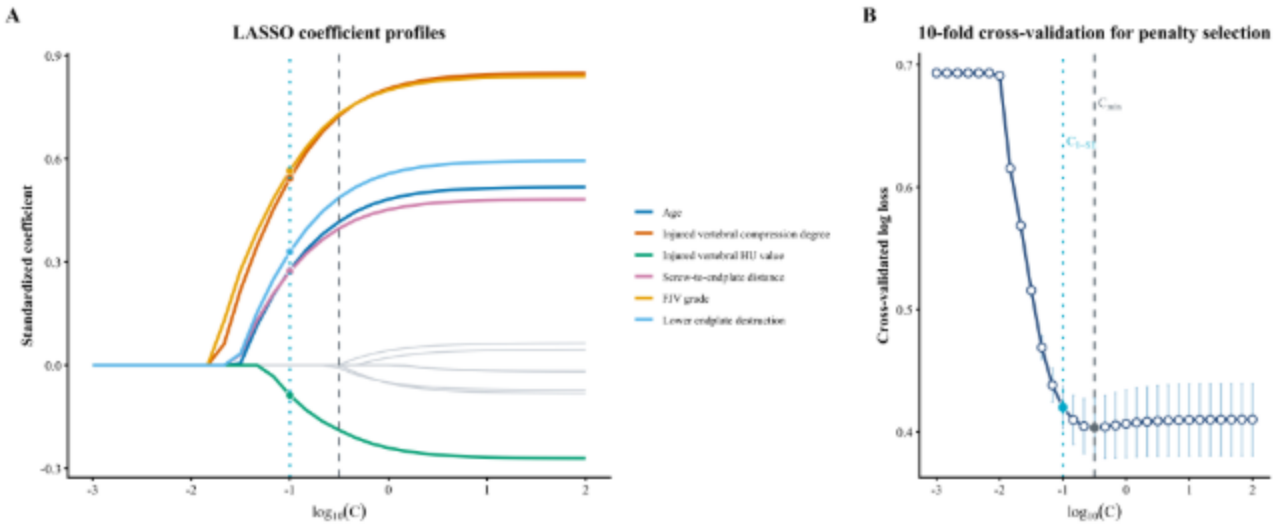


Figure 3

LASSO predictor selection

(A) Standardized coefficient profiles across inverse regularization strength C. (B) Mean 10-fold cross-validated log loss with one-standard-error bars. The 1-standard-error rule selected C = 0.100 and retained age, compression degree, HU, screw-to-endplate distance, FJV grade, and lower endplate destruction. LASSO, least absolute shrinkage and selection operator.

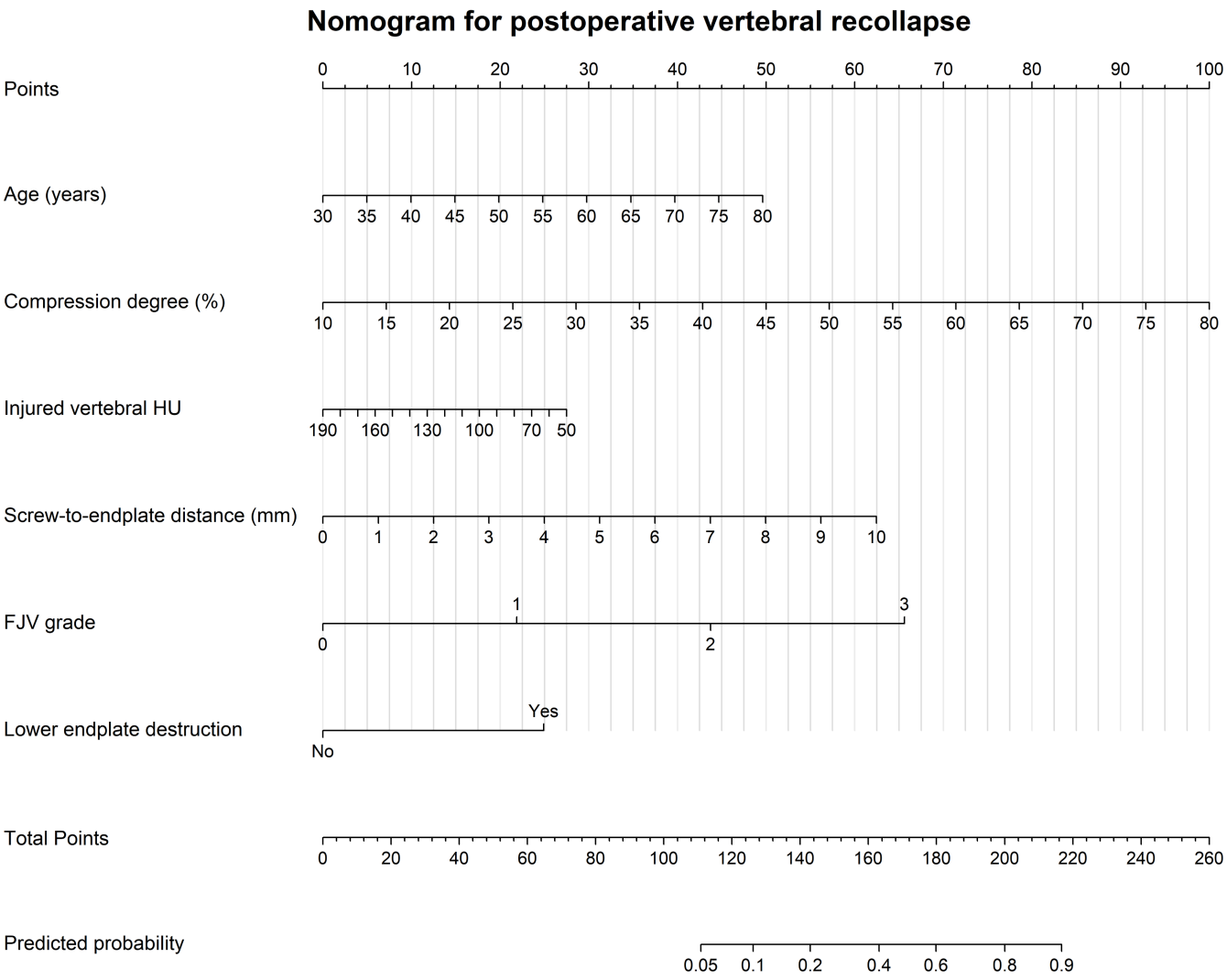


Figure 4

Graphical representation of the final prediction model

The nomogram maps the six predictor values to an estimated probability of vertebral recollapse. The regression equation in Section 3.2 is the primary model specification. This is an early postoperative prediction model and not a preoperative treatment-selection rule.

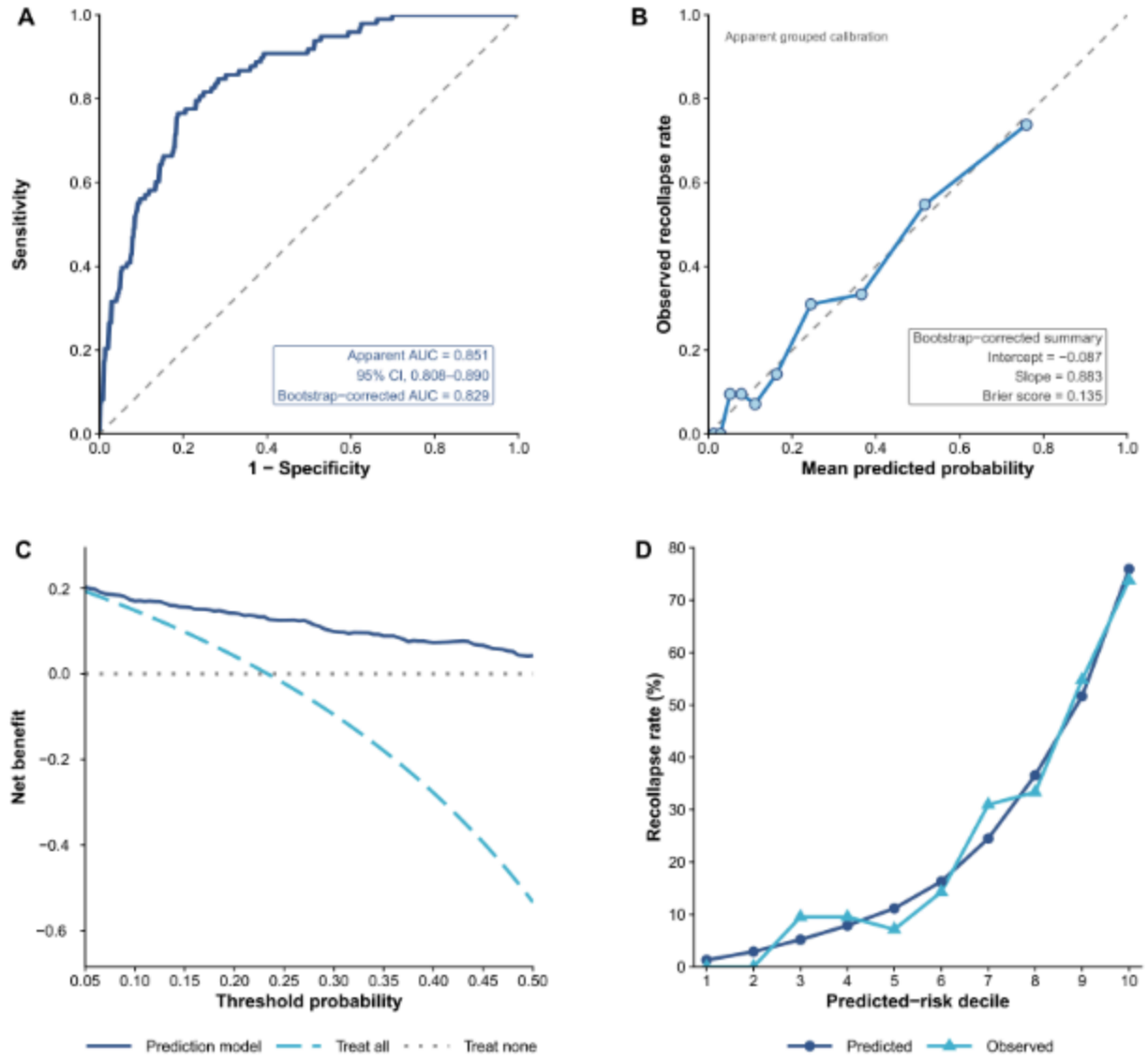


Figure 5

Model performance and clinical utility

(A) Receiver operating characteristic curve of the final six-predictor model. The apparent AUC was 0.851 (95% CI, 0.808–0.890), and the optimism-corrected AUC obtained from full-pipeline bootstrap validation was 0.829. (B) Apparent grouped calibration based on deciles of predicted probability. Bootstrap-corrected summary estimates were a calibration intercept of −0.087, calibration slope of 0.883, and Brier score of 0.135; the plotted grouped calibration curve itself represents apparent development-cohort calibration. (C) Decision curve analysis comparing the prediction model with treat-all and treat-none strategies. Threshold probabilities from 0.05 to 0.50 are displayed for visualization and should not be interpreted as validated clinical intervention thresholds. (D) Mean predicted probability and observed recollapse rate across deciles of predicted risk. The decision-curve and risk-decile analyses are descriptive assessments within the development cohort and do not establish clinical effectiveness or validated risk categories. AUC, area under the ROC curve; CI, confidence interval; ROC, receiver operating characteristic.

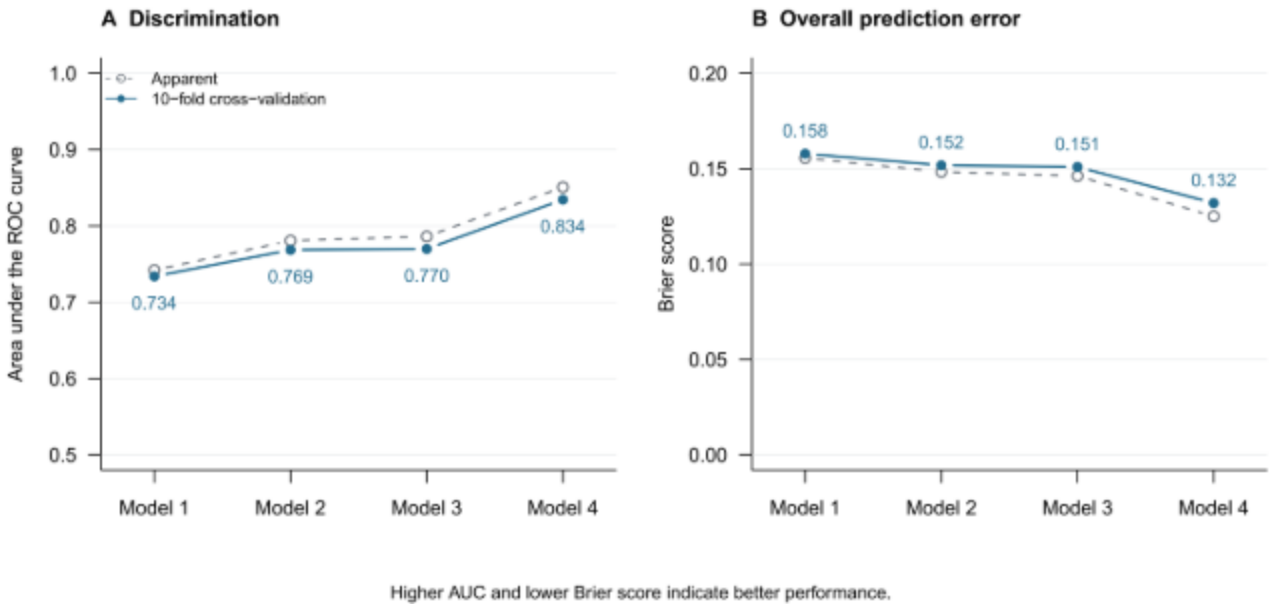


Figure 6

Incremental predictive performance across progressively extended models

(A) Apparent and cross-validated AUC. (B) Apparent and cross-validated Brier score. Model 1 included age and compression; Model 2 added lower endplate destruction; Model 3 added HU; and Model 4 added screw-to-endplate distance and FJV grade. Cross-validated estimates used pooled out-of-fold predictions from a common stratified 10-fold partition. Predictor sets were fixed, and LASSO selection was not repeated. Comparisons are descriptive and conditional on prior selection.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTables17.docx
- FigureS2.png
- FigureS3.png
- FigureS3.png
- FigureS4.png
- FigureS1.png
- Supplementarytabletitlesandnotes.docx