SUPPLEMENTARY METHODS

Study Population and Design

This retrospective cohort study included consecutive patients with histologically confirmed head and neck squamous cell carcinoma (HNSCC) who received definitive or postoperative/adjuvant radiotherapy or chemoradiotherapy at a single tertiary oncology center between December 2010 and October 2023. Patients with missing survival data or zero follow-up duration were excluded. The study was approved by the institutional ethics committee, and the requirement for informed consent was waived owing to the retrospective design.

Outcome Definition

The primary endpoint was progression-free survival (PFS), defined as the interval from the start of treatment to documented disease progression (local, regional, or distant) or death from any cause, whichever occurred first. Patients alive without documented progression at the last follow-up were censored at that date. Three patients recorded as deceased with a PFS event code of zero were reclassified as events to maintain the composite endpoint definition.

Candidate Predictors

Nine candidate predictors were selected a priori on the basis of established prognostic relevance in head and neck cancer and the criterion that all variables must be obtainable from routine assessments before the start of radiation treatment. Treatment-related variables (treatment modality, prior surgery, chemotherapy regimen, and feeding gastrostomy) were deliberately excluded to avoid indication bias and to ensure the model reflects baseline biological risk rather than treatment allocation patterns. The candidate set comprised: age (continuous), Eastern Cooperative Oncology Group performance status (ECOG PS; ordinal 0-3), AJCC tumor stage (I, II, III, IVA, IVB; modeled as an ordinal variable, 8th edition), neutrophil-to-lymphocyte ratio (NLR; continuous), C-reactive protein (CRP; continuous, mg/L), serum albumin (continuous, g/L), body mass index category (1 = underweight, 2 = normal, 3 = overweight; ordinal), hemoglobin (continuous, g/dL), and lactate dehydrogenase (LDH; continuous, U/L). These predictors span five distinct pathophysiological axes: demographics (age), functional status (ECOGPS), tumor burden (stage, LDH), systemic inflammation (NLR, CRP), and nutritional status (albumin, BMI, hemoglobin). Composite inflammatory indices derived from the same complete blood count components (systemic immune-inflammation index, pan-immune-inflammation value, lymphocyte-to-monocyte ratio, platelet-to-lymphocyte ratio) were not included because of high collinearity with NLR (Spearman's r = 0.72-0.92). Prognostic nutritional index was excluded because of near-perfect correlation with albumin (r = 1.00). A single albumin value of 0.9 g/L, judged to be a data entry error, was set to missing prior to imputation.

Missing Data Handling

Missing predictor values were handled by multiple imputation using chained equations (MICE) via the miceforest library. Twenty imputed datasets were generated with five iterations per dataset. The imputation model included all nine candidate predictors and the outcome variables (PFS time and event status) to preserve the outcome-covariate association. Imputed distributions were compared visually with the observed data to verify plausibility.

Variable Selection: LASSO Cox Regression

The least absolute shrinkage and selection operator (LASSO) penalized Cox proportional hazards model was applied for simultaneous variable selection and coefficient shrinkage. All predictor values were standardized (mean = 0, SD = 1) before model fitting. For each of the 20 imputed datasets, the optimal regularization parameter (λ) was determined by 5-fold cross-validated concordance index (C-index), and the value maximizing the mean C-index (λmin rule) was selected. The λmin strategy was preferred overthe more conservative one-standard-error rule because the events-per-variable ratio (EPV) of the final model exceeded 7, indicating a borderline risk of overfitting. Penalized regression coefficients were averagedacross the 20 imputed datasets; because LASSO yields biased coefficient estimates with non-standard sampling distributions, full Rubin's rules (which require unbiased estimates and within/betweenvariance components) were not applied, and corresponding 95% confidence intervals are not reported. A predictor was classified as stable if its coefficient was nonzero in at least 50% of the imputed datasets; an adaptive threshold was applied whereby the selection cut-off was raised in 5-percentage-point increments until no more than eight predictors remained, to preserve EPV ≥7. In this cohort, the final threshold settled above 70%, retaining the eight predictors listed in Table 2.

Final Model Fitting and Internal Validation

Rather than fitting the final model on a single imputed dataset, a pooled prediction approach was adopted. The LASSO Cox model was fitted separately on each of the 20 imputed datasets using the pooled mean λ and the set of stable predictors. Linear predictor scores and survival functions were averaged across the 20 models to produce pooled predicted probabilities at each time horizon. Internal validation was performed using two complementary approaches; both should be interpreted as approximate internal validation because the variable selection step was not fully nested within each resample:

Bootstrap Optimism Correction

Two hundred bootstrap resamples were drawn with replacement from the first imputed dataset. For each resample, a model was fitted with the previously selected stable predictors and pooled λ, and its apparent C-index was recorded; the same model was then evaluated on the out-of-bag observations. The mean difference between the apparent and out-of-bag C-index (optimism) was subtracted from the pooled apparent C-index to yield an optimism-corrected estimate. We acknowledge that this strategy does not repeat the imputation and LASSO variable selection within each bootstrap resample, which may underestimate true optimism; a fully nested implementation is computationally prohibitive in this setting.

Temporal Split Validation

Patients were divided into a training set (diagnosis date ≤ median) and a test set (diagnosis date > median). Because variable selection was performed on the entire cohort prior to the temporal split, the test set is not strictly independent of the model development process; consequently, the temporal test C-index represents an approximation of external generalizability rather than fully held-out validation.

Model Performance Assessment

Discrimination was quantified by the optimism-corrected Harrell C-index and by time-dependent receiver operating characteristic (ROC) curves at 12, 24, and 36 months, constructed using the cumulative/dynamic case-control definition. Calibration was assessed by dividing patients into risk quartiles and plotting Kaplan-Meier-observed PFS probabilities against pooled model-predicted probabilities at 12, 24, and 36 months. Overall prediction accuracy was summarized by the integrated Brier score (IBS).

Risk Stratification

The pooled linear predictor served as a continuous risk score. Two dichotomization strategies were evaluated: a median split (two equal-sized groups) and a 67th-percentile cut-point, classifying patients into a high-risk group (upper third) and a low-moderate-risk group (lower two-thirds). Group separationwas tested with the log-rank test. The median and 67th-percentile cut-points were evaluated for exploratory risk stratification. A Harrell-type nomogram was constructed to translate the model into a point-based clinical tool.

Software

All analyses were performed in Python 3.10 using scikit-survival (LASSO Cox, C-index, time-dependent AUC, Brier score), lifelines (Kaplan-Meier curves, log-rank test), miceforest (MICE), and scikit-learn (StandardScaler, cross-validation). A two-sided α of 0.05 was used for all significance tests.