SUPPLEMENTARY MATERIALS

METHODS

Construction of prediction model and risk score

To develop a novel prediction model, our objective was to pinpoint the most critical prognostic factors associated with overall survival (OS) in patients with early-stage hepatocellular carcinoma (HCC). The initial step involved employing univariate Cox proportional hazard models. Variables that demonstrated statistical significance (*P*-value<0.05) in these univariate analyses were then considered as potential candidates for inclusion in the multivariable Cox proportional hazard model. The construction of the multivariable model followed a stepwise selection process, carefully evaluating and incorporating variables that significantly contribute to predicting OS. The selection of the final model was guided by the Akaike Information Criterion (AIC), where models with lower AIC values were preferred, as they indicate a better fit to the data. Additionally, residual analysis was conducted to assess the goodness-of-fit of the model. This analysis helps in verifying whether the model adequately captures the underlying patterns in the data, ensuring that the final risk equation is both robust and reliable in predicting OS in early-stage HCC patients.

The Cox-based risk score was calculated for each identified prognostic risk factor, utilizing the coefficients derived from the final model. This score was then standardized on a scale ranging from 0 to 100. To determine the total risk score for each patient, the individual scores for all relevant prognostic factors were summed. Subsequently, these total scores were categorized into three distinct risk groups: high, medium, and low risk. The categorization was based on the 33rd and 66th percentiles, effectively dividing the patient cohort into three roughly equal groups according to their calculated risk levels. This stratification aids in identifying patients with varying levels of risk and can be instrumental in guiding clinical decision-making and treatment planning.

Given that the prediction model incorporates variables related to initial treatment, tumor characteristics, and biomarkers, there is a potential for multicollinearity, particularly between treatment variables and biomarkers. Additionally, the inclusion of numerous variables raises the concern of overfitting in the predictive model. To address these issues, the least absolute shrinkage and selection operator (LASSO) method was chosen. The LASSO method is particularly effective in situations where multicollinearity and overfitting are concerns, as it can shrink the coefficients of less important variables to zero, effectively performing variable selection and regularization simultaneously. The variables retained in the LASSO-based Cox model were then used to develop the machine learning (ML)-based risk score.

To evaluate the efficacy of the LASSO-based Cox regression model, a comparison was made with a standard Cox regression model. The LASSO model, which was fitted using Cox regression after variable selection through LASSO Cox regression, was compared against the COX model, fitted directly by Cox regression without the LASSO variable selection step. This comparison aimed to determine whether the LASSO approach provided a better fit and more accurate predictions, thereby validating the effectiveness of the LASSO method in this context.

Validation of risk score

The ML-based risk score using LASSO-based Cox regression and the conventional Cox-based CATS-IF score using stepwise Cox analysis was first internally validated in the training cohort and then externally validated in the validation cohorts. A three-step validation process was conducted for both the conventional Cox-based risk score and the ML-based risk score in the validation cohort. First, the predictive performance of the conventional Cox-based risk score and the ML-based risk score for OS in patients with early-stage HCC was evaluated by comparing the homogeneity, AIC, and the area under ROC (AUROC) among different models. Second, calibration plots of the conventional Cox-based risk score and the ML-based risk score for OS were generated to compare the predicted survival probability with the observed survival.

The standard ROC curve analysis considers event status and maker value (i.e., risk score in this study) for patients as fixed over time. However, both disease status and survival time change over time in this study. Patients who have received treatment ear-

lier may die later due to longer study follow-up, and their risk score may also change from baseline during follow-up. Thus, we used the time-dependent ROC curve of conventional Cox-based risk scores and ML-based risk scores and compared the prognostic performance of the two risk scores at 1, 2, 3, and 5 years after HCC diagnosis.