



# Development and validation of a machine learning-based risk model for metastatic disease in nmCRPC patients: a tumor marker prognostic study

Xudong Ni, MD<sup>a,b,c</sup>, Ziyun Wang, MD<sup>a,b,c</sup>, Xiaomeng Li, MD<sup>a,b,c</sup>, Jixinnan Sui, MD<sup>a,b,c</sup>, Weiwei Ma, MD<sup>a,b,c</sup>, Jian Pan, MD<sup>a,b,c</sup>, Dingwei Ye, MD<sup>a,b,c</sup>, Yao Zhu, MD, PhD<sup>a,b,c,\*</sup>

**Background:** Nonmetastatic castration-resistant prostate cancer (nmCRPC) is a clinical challenge due to the high progression rate to metastasis and mortality. To date, no prognostic model has been developed to predict the metastatic probability for nmCRPC patients. In this study, we developed and externally validated a machine-learning model capable of calculating risk scores and predicting the likelihood of metastasis in nmCRPC patients.

**Patients and methods:** A total of 2716 nmCRPC patients were included in this study. The training and testing datasets were derived from SPARTAN (NCT01946204) and ARAMIS (NCT02200614), respectively. Regarding metastasis-free survival as the endpoint, we subjected 13 clinical features to 10 machine-learning models and their combinations to predict metastasis. Model performance was assessed through accuracy (AUC), calibration (slope and intercept), and clinical utility (DCA). The risk score calculated by the model and risk factors based on eight identified variates were used for metastatic risk stratification.

**Results:** The final prognostic model included eight prognostic factors, including novel hormone therapy application, Gleason score, previous treatments received (both surgery and radiotherapy, or neither), Race (White), PSA doubling time (PSADT), hemoglobin (HGB), and IgPSA. The prognostic model resulted in a C-index of 0.724 (95% CI 0.700–0.747) in internal validation and relatively good performance through tAUC (>0.70 at 3-month intervals between 6 and 39 months) in external validation. In the risk score stratifying strategy, compared with the low-risk group, the metastasis HRs for medium- and high-risk groups were 1.72 (95% CI 1.39–2.12) and 4.43 (95% CI 3.66–5.38); as for risk factor count, the HRs are 1.98 (95% CI 1.50–2.61) and 4.17 (95% CI 3.16–5.52), respectively.

**Conclusions:** In this study, we developed and validated a machine learning prognostic model to predict the risk of metastasis in nmCRPC patients. This model can assist in the risk stratification of nmCRPC patients, guide follow-up strategies, and aid in selecting personalized treatment intensities.

**Keywords:** machine learning, metastasis-free survival, nonmetastatic castration-resistant prostate cancer, prognostic model, prostate cancer

<sup>a</sup>Department of Urology, Fudan University Shanghai Cancer Center, Shanghai, China, <sup>b</sup>Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China and <sup>c</sup>Shanghai Genitourinary Cancer Institute, Shanghai, China

Xudong Ni, Ziyun Wang, Xiaomeng Li, and Jixinnan Sui contributed equally and are co-first authors.

Dingwei Ye and Yao Zhu contributed equally and are co-senior authors.

\*Corresponding author. Address: Department of Urology, Fudan University Shanghai Cancer Center, 270 Dong'an Road, Shanghai 200032, China. Tel.: +86 21 64175590. E-mail: zhuyao@fudan.edu.cn (Y. Zhu).

Copyright © 2025 The Author(s). Published by Wolters Kluwer Health, Inc. This is an open access article distributed under the Creative Commons Attribution License 4.0 (CCBY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

International Journal of Surgery (2025) 111:3331–3341

Received 9 August 2024; Accepted 5 November 2025

Supplemental Digital Content is available for this article. Direct URL citations are provided in the HTML and PDF versions of this article on the journal's website, [www.ijso.com/international-journal-of-surgery](http://www.ijso.com/international-journal-of-surgery).

Published online 26 March 2025

<http://dx.doi.org/10.1097/JS9.0000000000002321>

## HIGHLIGHTS

- This study introduces a machine learning-based model to predict metastasis risk in non-metastatic castration-resistant prostate cancer (nmCRPC) patients, addressing a critical need in the field.
- The model achieved a time-dependent area under the curve greater than 0.7 at 3-month intervals between 6 and 36 months during external validation, demonstrating its robustness and reliability.
- The model allows for effective risk stratification by assigning risk scores or counting risk factors, significantly improving personalized treatment planning and clinical decision-making for nmCRPC patients.

## Introduction

Nonmetastatic castration-resistant prostate cancer (nmCRPC) is characterized by increasing levels of PSA despite androgen

deprivation therapy (ADT), without evidence of metastasis on conventional imaging modalities such as computed tomography or bone scintigraphy<sup>[1]</sup>. As an advanced stage of prostate cancer, the annual incidence of nmCRPC in the United States was estimated to be approximately 60 000 cases in 2020, with a 34% annual progression rate to metastatic disease and a 16% yearly overall mortality<sup>[2]</sup>. Most prostate cancer-related deaths are attributed to widespread metastasis, and a recent study has validated that metastasis-free survival (MFS) can serve as a surrogate endpoint for overall survival (OS) in localized CRPC patients<sup>[3,4]</sup>. Therefore, it is necessary to accurately identify nmCRPC patients at high risk of developing metastasis and estimate the time to progression to mCRPC to establish realistic patient expectations and enable precise implementation of treatment and follow-up by healthcare providers.

Currently, the prognosis stratification of nmCRPC primarily relies on the PSA level and the change rate, as reported by PSADT<sup>[5]</sup>. Evidence suggests that a shorter PSADT indicates a higher risk of progression<sup>[6,7]</sup>, and a PSADT <6 months is sometimes suggested as a cutoff to identify the need for more aggressive therapy<sup>[8]</sup>. However, relying on a single binary indicator can only roughly categorize nmCRPC into high-risk and low-risk groups, making it challenging to provide individualized assessments of MFS for patients, limiting its guidance on patient prognosis. In contrast to the lack of well-established prognostic models for nmCRPC patients, high-quality clinical multivariable prognostic models have been established for metastatic castration-resistant prostate cancer (mCRPC) patients, both for chemotherapy-naïve mCRPC and post-chemotherapy mCRPC<sup>[9,10]</sup>. In 2023, Halabi *et al* externally validated a previously established mCRPC prognostic model based on the CALGB 90401 Phase III clinical trial ( $n = 1050$ ), with a tAUC (time-dependent area under the curve) of 0.74 (95% CI, 0.73–0.75) in the validation dataset<sup>[11]</sup>. Given the current lack of effective and externally validated prognostic models for nmCRPC patients, we aim to leverage data from the latest Phase III clinical trials (SPARTAN and ARAMIS) to identify prognostic and predictive variables for the prognosis of nmCRPC patients and establish a prognostic model tailored explicitly to nmCRPC patients.

In this study, we utilized machine learning methodologies to develop a prognostic model using MFS as the primary endpoint for nmCRPC patients. Considering that novel hormone therapy (NHT) such as enzalutamide, apalutamide, and darolutamide have become the first-line treatment for nmCRPC<sup>[12,13]</sup>, we incorporated the prognostic value of NHT application and other clinical data, including baseline PSA level, PSADT, previous treatments received, Gleason score, race, and laboratory indicators, which are related to disease progression, as model parameters. The objective of this study was to create and externally validate a model that calculates risk scores and predicts the probability of metastasis in nmCRPC patients, providing urologists with a valuable tool for clinical practice.

## Methods

### Derivation and validation cohorts

The training dataset was derived from the SPARTAN clinical trial (NCT01946204)<sup>[14]</sup>, while the validation dataset was obtained from the ARAMIS clinical trial (NCT02200614)<sup>[15]</sup>. These trials shared similar study designs, as well as inclusion and

exclusion criteria. ARAMIS enrolled 1509 nmCRPC patients (955 randomized to darolutamide; 554 to placebo), and SPARTAN enrolled 1207 patients (806 randomized to apalutamide; 401 to placebo).

Both ARAMIS and SPARTAN studies were conducted by the Declaration of Helsinki, the International Conference on Harmonisation for Good Clinical Practice, and local regulations<sup>[16–18]</sup>. The study protocols were approved by the respective authorities and ethics committees, and signed informed consent was obtained from all patients. Informed consent for constructing and validating the nmCRPC model was optimal due to the de-identified nature of the individual patient data from ARAMIS and SPARTAN. The patient-level data from the SPARTAN clinical trial were obtained through the Yale University Open Data Access Project, while the ARAMIS data were provided by Bayer and made available through Vivli, Inc. The present work has been reported in compliance with REporting recommendations for tumor MARKer prognostic studies (REMARK) criteria and guidelines<sup>[19]</sup>.

### Study population

Our study initially included all patients from two clinical cohorts and applied the following criteria for exclusion. Exclusion criteria are as follows: (1) lack of the endpoint result MFS; (2) presence of baseline data missing. Forty-seven patients from SPARTAN and 127 patients from ARAMIS were excluded. Finally, 1160 nmCRPC patients from the Clinical Trial A database and 1247 patients were included in the analysis.

### Study outcome and candidate predictor variables

The study's outcome was MFS, defined as the time from randomization to evidence of metastasis or death from any cause in both SPARTAN and ARAMIS. A comprehensive literature review identified Candidate predictor variables for potential inclusion in our models. In the data set, 13 variables (NHT application, baseline PSA level, PSADT, previous treatments received, Gleason score, race, and laboratory indicators) were available for analysis. The basic clinical information of the included patients is shown in Table 1.

Binary variables included: NHT application (NHT drugs Darolutamide or Apalutamide or placebo), N stage (N0 or N1), bone-sparing agent (yes or no), ECOG (0 or 1), Gleason grade (<8 or ≥8). Multi-class variables included previous treatments (surgery, radiotherapy, neither, or both applications), race (White, Asian, Black, or Other), and age (<65, 65–74, or ≥74). The remaining parameters were included as continuous variables. To prevent the parameter estimation from being influenced by individual extreme outliers, we performed the subsequent analysis using the log10-transformed PSA values.

To conform to the input format required by all models, multi-class variables are converted into binary variables. Variable selection in the training set was performed using Univariate Cox proportional hazards models. Categorical variables with a  $P$  value of <0.1 in the univariate analysis were included in the subsequent model building. Additionally, split multi-class variables, with any of the binary variables'  $P$  value <0.1, were included in the following procedure.

**Table 1**  
**Baseline characteristics of patients in the development and validation cohort**

Variable	Development cohort (SPARTAN; n = 1160)	Validation cohort (ARAMIS; n = 1247)
Treatment arm		
Apalutamide or Darolutamide	776(66.9%)	799 (64.1%)
Placebo	384(33.1%)	448 (35.9%)
N stage		
N0	969(83.5%)	617(49.5%)
N1	191(16.5%)	206(16.5%)
NX	/	424(34.0%)
Bone-sparing agent		
N	1044(90.0%)	1192(95.6%)
Y	116(9.98%)	55(4.41%)
ECOG		
0	883(76.1%)	866(69.4%)
1	227(23.9%)	381(30.6%)
Gleason grade		
<8	654(56.4%)	734(58.9%)
≥8	506(43.6%)	513(41.1%)
Therapy		
Both	349(30.1%)	50(4.01%)
Neither	315(27.2%)	587(47.1%)
Radiotherapy	368(31.7%)	563(45.1%)
Surgery	128(11.0%)	47(3.77%)
Race		
White	770(66.4%)	1016(81.5%)
Asian	129(11.1%)	172(13.8%)
Black	66(5.69%)	47(3.77%)
Others	195(16.8%)	12(0.96%)
Age		
<65	357(30.8%)	168(13.5%)
65–74	508(43.8%)	480(38.5%)
≥75	295(25.4%)	599(48.0%)
PSA double time (months)	4.73 (2.29)	4.93(2.33)
HGB (g/dL)	13.3 (1.22)	13.1(1.19)
ALP (U/L)	80.5(25.1)	79.7(24.0)
ALB (g/dL)	4.37 (0.25)	4.26(0.29)
IgPSA	0.80(0.44)	0.96(0.42)

**Model derivation**

After data analysis variable selection, we preliminarily assessed the performance of 10 machine learning models allowing the use of survival data [Least absolute shrinkage and selection operator (LASSO), stepwise Cox proportional hazards (Stepwise Cox), partial least squares proportional hazard regression (plsRCox), CoxBoost, Efficient Neural Network (Enet), gradient boosting machine (GBM), random survival forest (RSF), Survival Support Vector Analysis (survivalSVM), and Supervised Principal Components (superPC)] and their combinations. Further details regarding the machine learning methods have been published previously<sup>[20]</sup>.

After comprehensively evaluating all the machine learning models, we selected five models with centered and scaled continuous variables due to their varying scales. These models include a Stepwise Cox model<sup>[21]</sup>, a LASSO model<sup>[22]</sup>, a LASSO and Stepwise Cox combined model<sup>[23]</sup>, a CoxBoost model<sup>[24]</sup>, and an RSF model<sup>[25]</sup>. The overall framework of machine-learning model development is illustrated in Fig. 1.

**Prediction model performance**

To assess the accuracy of these five models, we evaluated the model performance through internal and external validation. C-indexes and 95% CI were calculated using bootstrapping (500 times) to evaluate the models' accuracy. In the validation cohort, predictive accuracy was acquired by tAUC assessed by plotting the time-AUC curve at 3-month intervals between 6 and 36 months. The hazard ratio (HR) and 95% confidence interval (CI) were computed for each potential predictor. In addition to accuracy, it is recommended to assess calibration and clinical utility in developing a prognostic model. Therefore, we employed calibration curves and decision curve analysis to evaluate these aspects in the training and testing cohorts. Multicollinearity was calculated using Spearman correlation and variance inflation factor (VIF).

To further support the clinical use of the model, a risk score was developed from the multivariable model in both the training and testing sets. We stratified the patients into three risk groups according to the risk score trisection value calculated in the training set.

Moreover, we built another three-risk prognostic group based on the number of risk factors converted from model predictors. Table S1 (available at: <http://links.lww.com/JS9/E13>) defines these risk factors. To achieve a balanced distribution, patients were stratified into low-risk (0-2 risk factors), medium-risk (3-4 risk factors), or high-risk (>4 risk factors) groups.

The Kaplan–Meier curve was applied to estimate MFS and OS distribution by prognostic risk groups with and without placebo patients, and the log-rank statistic was applied to assess the differences in MFS among three-risk groups. Median MFS, metastasis-free survival probability, and hazard ratio with 95% CI were calculated for the three risk groups. The statistical analysis was performed using R Studio, version 4.0.2. Statistical significance was defined as a two-sided *P* value of less than 0.05 (*P* < 0.05).

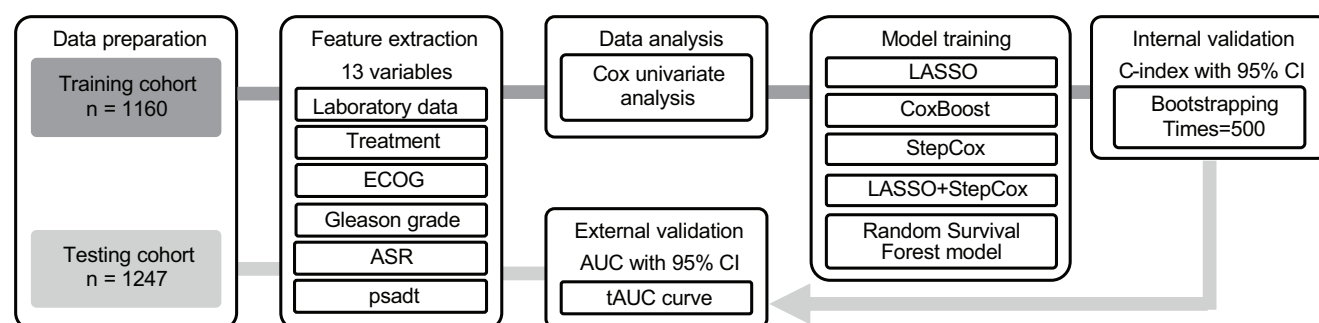
**Results**

**Characteristics of the cohorts**

A total of 1160 participants were included in the derivation cohort (SPARTAN) and 1247 in the external validation cohort (ARAMIS). Based on our exclusion criteria, 47 patients from the full SPARTAN population (*n* = 1207) and 262 patients from the full ARAMIS population (*n* = 1509) were excluded from the analysis. The baseline characteristics of patients in the derivation and validation cohort are shown in Table 1. Dating to the cutoff date, the MFS times distribution of the development and validation cohorts were presented in Figure S1 (available at: <http://links.lww.com/JS9/E13>).

**Prognostic factors influencing MFS**

Before conducting the univariate analysis, we converted multi-class variables (such as previous therapy received, race, and age) into binary variables to facilitate the following model-building step. We used the univariate Cox proportional hazards model to analyze the relationship between the following parameters and MFS (Table 2): treatment, N stage, Bone-sparing agent use, ECOG score, Gleason score, previous therapy (surgery, radiotherapy, both, or neither), race (White, Asian, Black, and



**Figure 1.** The framework of machine-learning model development for MFS prediction.

others), age (<65, 65–74, and ≥75), PSADT, HGB, alkaline phosphatase (ALP), albumin (ALB), and IgPSA. Among the parameters associated with MFS, NHT, PSADT, and IgPSA application exhibited strong correlations with MFS prognosis, previously reported as OS risk factors in nmCRPC patients<sup>[26]</sup>.

It was noted that the common prognostic parameter lactate dehydrogenase (LDH) was not included due to a high number of missing values. The categorical variables (N stage, bone-sparing agent use, and age) with  $P$  value >0.1 were excluded in the following steps, while all the continuous variables were preserved. Additionally, we observed that white people had an HR of 1.229 (95% CI: 1.012–1.494;  $P = 0.038$ ). In order to make this model applicable to all ethnicities, the race variable was preserved.

**Table 2**  
**Univariate analysis of patients' characteristics and MFS in the development cohort**

Variable	HR	95% CI	$P$ value
Treatment arm <sup>a</sup>	0.276	(0.230, 0.332)	<0.001
N stage	1.197	(0.945, 1.515)	0.136
Bone-sparing agent	1.005	(0.748, 1.350)	0.975
ECOG	1.282	(1.046, 1.572)	0.017
Gleason grade <sup>b</sup>	1.203	(1.004, 1.440)	0.045
Therapy			
Both	1.223	(1.011, 1.479)	0.038
Surgery	0.876	(0.650, 1.180)	0.383
Radiotherapy	1.123	(0.930, 1.355)	0.229
Neither	0.737	(0.593, 0.916)	0.006
Race			
White	1.229	(1.012, 1.494)	0.038
Asian	0.850	(0.612, 1.181)	0.333
Black	0.932	(0.628, 1.384)	0.726
Others	0.824	(0.648, 1.048)	0.115
Age			
<65	0.959	(0.790, 1.164)	0.672
65–74	0.996	(0.831, 1.194)	0.967
≥75	1.055	(0.858, 1.298)	0.610
PSA double time	0.881	(0.845, 0.918)	<0.001
HGB	0.919	(0.852, 0.991)	0.028
ALP	1.002	(0.998, 1.005)	0.345
ALB	0.898	(0.631, 1.278)	0.551
IgPSA	1.593	(1.302, 1.949)	<0.001

<sup>a</sup>NHT (apalutamide) vs. placebo.

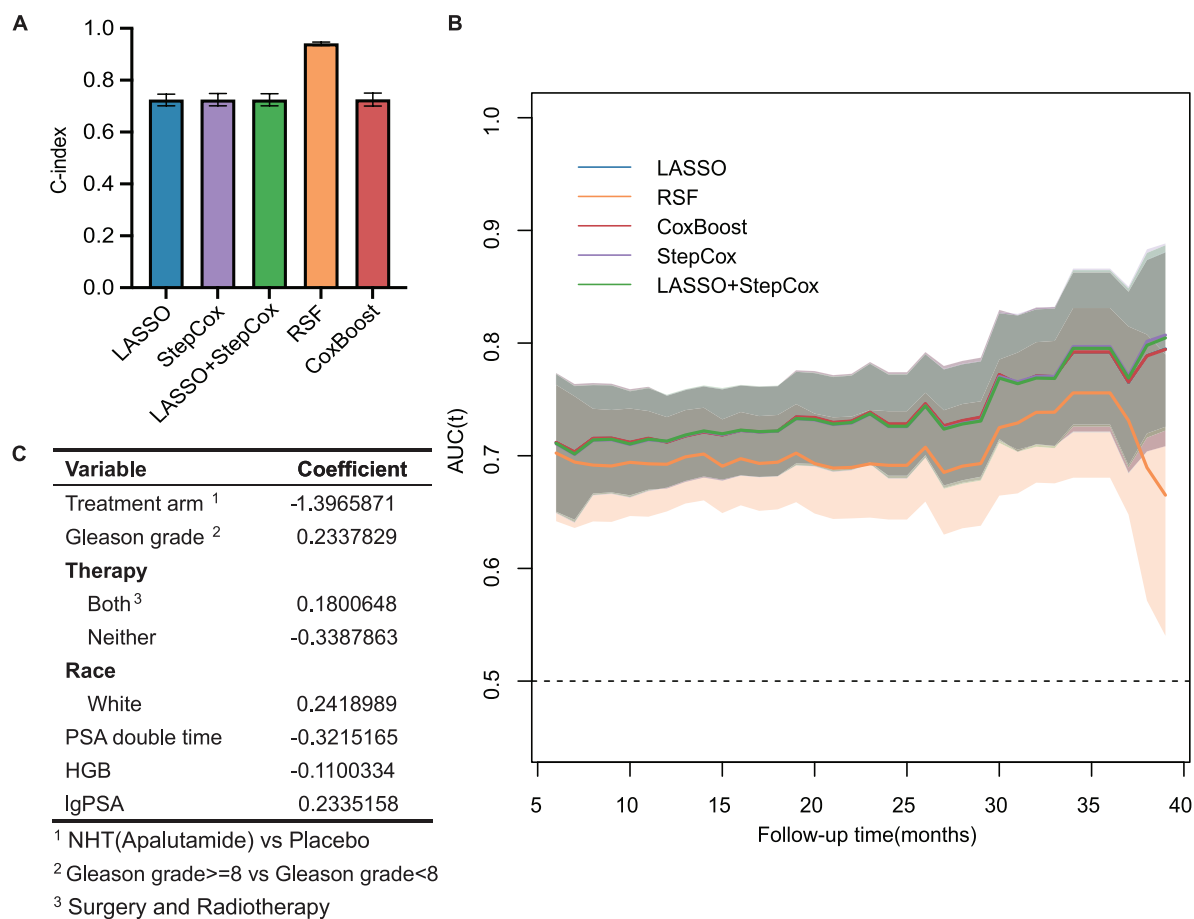
<sup>b</sup>Gleason grade ≥8 vs. Gleason grade <8.

### Selection of prognostic models

After screening the predictors through univariate analysis, we fitted 101 machine learning models in the training cohort. We calculated the AUC between training and testing cohorts using the method published by the previous study<sup>[20]</sup>. As shown in Table S2 (available at: <http://links.lww.com/JS9/E13>), we chose the RSF model because its mean AUC was superior to the other models, while a significant gap existed between the training and testing cohort attributed to the overfitting probability. Subsequently, we selected four other representative models, including the Stepwise Cox model, the LASSO, the LASSO + Stepwise Cox combined model, and the CoxBoost model. They presented relatively excellent predictive efficiency (AUC>0.7), performance consistency (the AUC discrepancy between two cohorts), model simplicity (fewer variables), as well as wide usage in prognostic model development (high level of acceptance and replication convenience).

This study recommended using all available data in the dataset to develop the model and estimate its performance<sup>[27]</sup>. Thus we took the bootstrapping resampling approach instead of randomly splitting the training cohort to acquire C-index and 95% CI in five models. Random forest model exhibited highest C-index 0.941 (95% CI 0.934–0.947), while other models' C-index were approximately the same [C-index with 95% CI: LASSO 0.724 (0.701–0.750), CoxBoost 0.725 (0.700–0.749), LASSO + Stepwise Cox 0.724(0.700–0.747), and Stepwise Cox 0.725 (0.702–0.749)] (Fig. 2A).

In the following, we further evaluated the performance of these models in the external dataset ARAMIS. Despite the excellent performance in internal validation, the RSF model showed significantly and consistently poor performance compared to the other four models. In all five models constructed, the LASSO + Stepwise Cox model presented the highest prediction accuracy (Fig. 2B). And the number of parameters (eight) chosen by these models is minimum (Table S2, available at: <http://links.lww.com/JS9/E13>). Given the model's simplicity and broad applicability, the LASSO + Stepwise Cox model was selected as the optimal prognostic model for subsequent analysis. The detailed model establishment procedure (LASSO for variable screening, multivariate Cox model analysis, multicollinearity measurement through Spearman correlation and VIF) was presented in Figure S2 (available at: <http://links.lww.com/JS9/E13>). Meanwhile, the model presented excellent calibration ability in two cohorts because their calibration curves were close to the diagonal line (Fig. 3A and B). Furthermore, we adopted decision



**Figure 2.** Internal and external evaluation of machine-learning models. (A) C-indices of models calculated by bootstrapping (times = 500) in the training cohort. (B) Time-dependent AUC curve of models in the testing cohort at 3-month intervals between 6 and 39 months. (C) The parameters and coefficients of the LASSO + Stepwise Cox model.

curve analysis to assess our model’s clinical utility. In the training cohort, DCA curves (Decision Curve Analysis curves) showed that if the threshold probability of patients is 5–40%, 10–74%, and 12–89%, using nomogram in the current study to predict metastasis risk respectively in 6, 12, and 24 months could add more benefit (Fig. 3C–E). As for the testing cohort, the DCA curves illustrated that the threshold probability is 4–18%, 10–42%, and 12–63%, corresponding to 6, 12, and 24 months (Fig. 3F–H).

**Clinical utility of the prediction model**

**Risk groups based on the risk score**

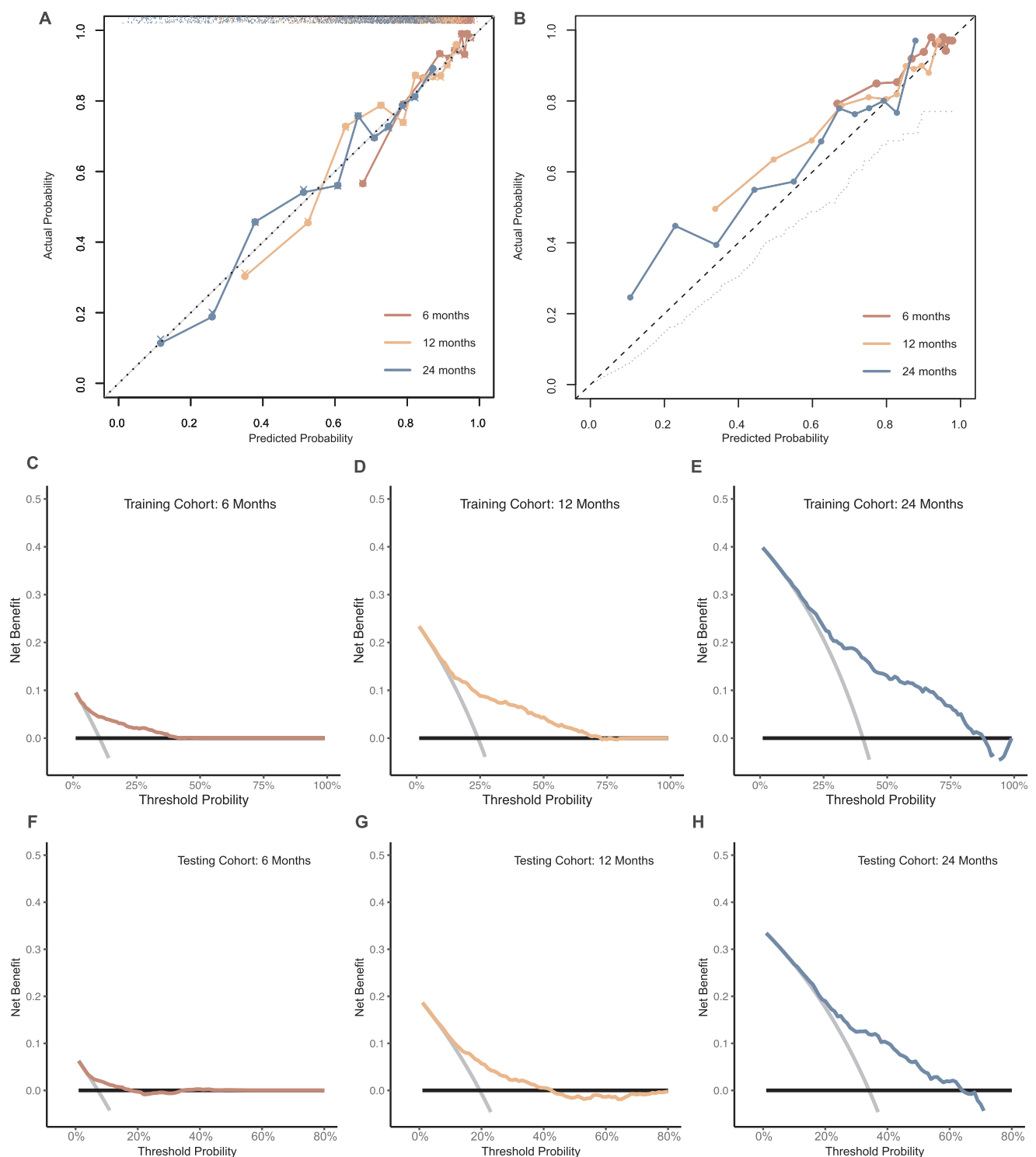
The risk score acquired from the model can serve as the basis for metastatic risk stratification in clinical practice or for the randomization of patients in clinical trials. For three-risk groups, through getting cutoff value using the training data, there were respectively 781 patients (32.4%), 786 patients (32.6%), and 840 patients (34.9%) in the low-, medium-, and high-risk groups with unreached median MFS, median MFS times of 41 months (95% CI 34 months, not reached) and 18 months (95% CI 15–19, log-rank test  $P < 1 \times 10^{-4}$ ; Fig 4A and B). The nonmetastasis probability decreases over time, with the high-

and medium-risk group substantially higher than the low-risk group in 12, 34, and 36 months (Fig. 4C). Compared with the low-risk group, the metastasis HR for medium- and high-risk groups were 1.72 (95% CI 1.39–2.12) and 4.43 (95% CI 3.66–5.38; Fig 4D), respectively. Meanwhile, the risk stratification also presented a significant overall survival (OS) difference, enhancing its clinical utility (Fig. 4E).

**Risk groups based on the risk factor number**

To facilitate rapid assessment of patients’ metastasis probability and address patient inquiries, parameters in the model were converted into risk factors according to Table S1 (available at: <http://links.lww.com/JS9/E13>). Thus, even without a risk score obtained from the model, physicians can make an initial assessment of patients’ condition by calculating the number of risk factors.

Divided into three-risk grouping in training combined validation datasets, respectively, 15.5%, 55.8%, and 28.7% of patients were in the low- (0–2 risk factors), medium- (3–4 risk factors), and high-risk (>4 risk factors) groups. The median MFS was not reached in the low-risk group, 35 months in the medium-risk group, and 18 months in the high-risk group (Fig. 5A and B). In months 12, 34, and 36, the high- and medium-risk

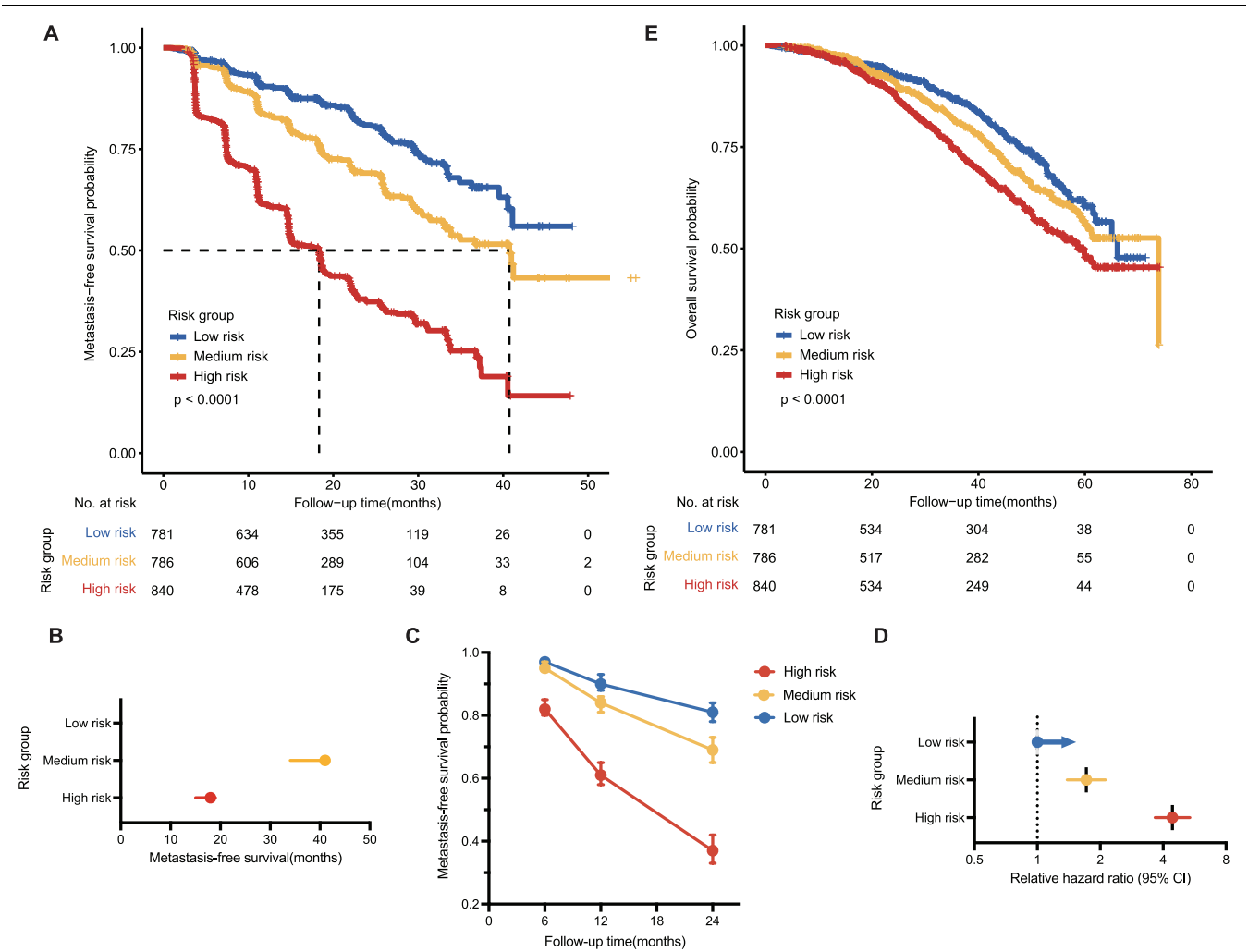


**Figure 3.** (A–B) Calibration curves of the training and testing cohort, respectively, in 6, 12, and 24 months. (C–H) Decision curve analysis of the training and testing cohort in 6, 12, and 24 months.

groups showed higher nonmetastasis probabilities than the low-risk groups (Fig. 5C). Compared with the low-risk group, the metastasis HRs for the medium- and high-risk groups were 1.98 (95% CI 1.50–2.61) and 4.17 (95% CI 3.16–5.52; Fig 5D), respectively. By contrast, metastasis risk stratification based on

risk count presented accuracy similar to that of risk number. In order to conform to new therapy guidelines approved by the FDA, we excluded placebo patients, and NHT patients still exhibited prognostic discrepancy according to division by the risk number (Fig. 5E).





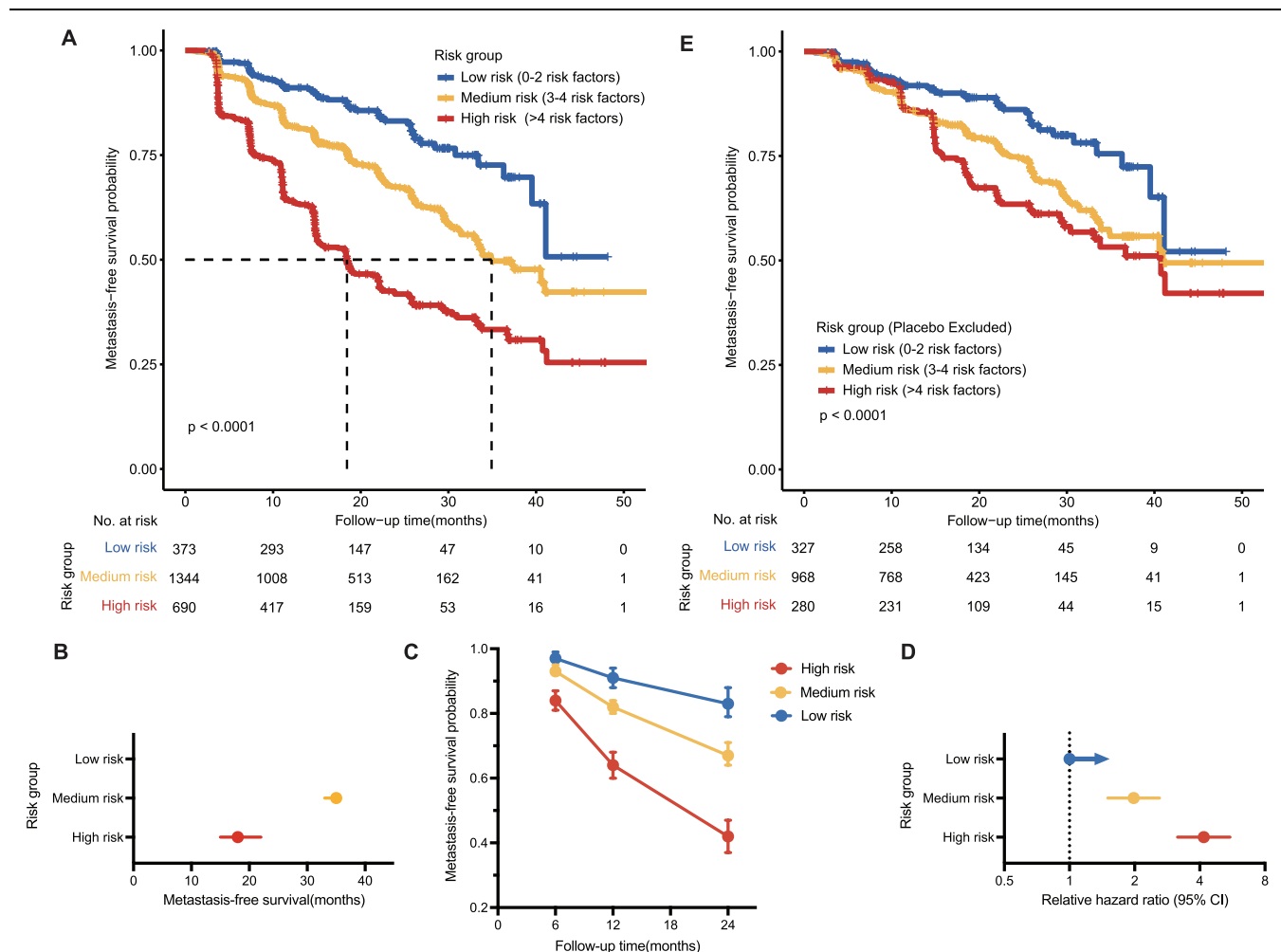
**Figure 4.** Risk groups based on the risk score. (A) Kaplan-Meier survival curves for MFS of the three-risk groups in the training combined validation sets. (B-D) Mean MFS times, nonmetastasis probability, and relative hazard ratio with 95% CI for the three-risk groups. (E) Kaplan-Meier survival curves for OS of the three-risk groups in the training combined validation sets.

### Discussion

In this study, we developed a prognostic model and risk stratification for metastasis based on two latest independent Phase III clinical trials' cohorts, which complements the blank of the multivariate prognostic model for MFS and OS in nmCRPC patients. 10 algorithms were utilized to construct 101 kinds of machine learning models for MFS, and LASSO + Stepwise Cox model presented several advantages over other models such as a better predictive ability and simplicity. The performance of the selected prognostic model was comprehensively evaluated from three perspectives: accuracy (AUC in internal and external validation), calibration (calibration curve), and clinical utility (DCA). By risk score and variants acquired by the LASSO + Stepwise Cox model, we built two risk stratification strategies to address clinical requirements for disease stage assessment and individual therapy scheme design in nmCRPC patients.

Prior studies have focused on exploring single risk factors for MFS<sup>[28-32]</sup>, but a multivariate prognostic model is lacking to evaluate and stratify individuals' metastatic risks. The chosen

model identified eight prognostic predictors of MFS: treatment, Gleason score, previous therapy (both surgery and radiotherapy, or neither), Race, PSADT, HGB, and IgPSA. It included the widely accepted factors with prognostic value in prostate cancer patients such as NHT application<sup>[14,15]</sup>, Gleason score<sup>[33]</sup>, PSADT<sup>[5]</sup>, and PSA<sup>[29,34]</sup>. In addition, we have identified several new prognostic factors for MFS: previous therapy (surgery, radiotherapy, or both), race, and HGB levels. We found that patients who had previously undergone local surgery and radiotherapy had a higher risk of distant metastasis. This may be attributed to the selective pressure exerted on tumor cells by local treatment, leading to a higher malignancy and increased metastatic potential in cells that recurred after both surgery and radiotherapy. Moreover, the analysis revealed that being of White race became a significant risk factor in this model (HR 1.27 with 95% CI 1.04–1.55). For ethnicity impact on prognosis, the earlier finding showed white men had a better prognosis than black men<sup>[35]</sup>. However, recent studies have observed the prognostic discrepancy lies in the differing access to medical settings. Specifically, after adjusting for confounding factors,



**Figure 5.** Risk groups based on the risk factors. (A) Kaplan–Meier survival curves for the three-risk groups in the training combined validation sets. (B–D) Mean MFS times, nonmetastasis probability, and relative hazard ratio with 95% CI for the three-risk groups. (E) Kaplan–Meier survival curves for MFS of the three risk groups excluded placebo patients in the training combined validation sets.

American black patients were found to have a reduced risk of metastatic disease or death compared to white patients<sup>[36]</sup>. Our previous study proposed that Asians with de novo metastatic prostate cancer have a better prognosis than White<sup>[37]</sup>. However, direct evidence supporting similar conclusions in nmCRPC is lacking. Thirdly, HGB levels were identified as a risk factor across various stages of prostate cancer<sup>[38,39]</sup>, indicating a relationship with the response and tolerance capacity to prior therapy. It is reasonable to deduce that nmCRPC patients with decreasing HGB levels have a lower response to treatment, resulting in high metastatic probability.

Our model focused on pretreatment survival prediction and did not incorporate post-treatment prostate-specific antigen (PSA) kinetics or radiological features, which have been shown to have a significant correlation with prognosis<sup>[28,40,41]</sup>. Utilizing post-treatment data resulted in the exclusion of patients who experienced metastasis or dropped out of the clinical trial within 3 and 6 months, causing available data loss and population bias toward the NHT group. In contrast, our pretreatment model provides more comprehensive coverage for nmCRPC patients. By considering therapy tolerance in conjunction with the

primary cancer condition, our model significantly facilitates clinical decision-making and enhances its practical application. The estimation and stratification of prognosis can assist clinicians in determining the appropriate aggressiveness of therapy. This enables individualized treatment approaches to be implemented in the context of clinical trials, with more aggressive strategies and combination therapies for high-risk patients while offering alternative medical schemes with fewer side effects and improved quality of life for low-risk patients. This can also help explore combination treatment strategies for high-risk individuals in future research, such as the ongoing clinical trial NCT05849298. Additionally, the value of local therapies in controlling the progression of high-risk nmCRPC should also be considered.

Our research has some limitations. First, the exclusion of data from recent trials is inevitable due to missing value, leading to a reduction in the available dataset and may have introduced a potential selection bias. Secondly, although the variables included have clinical rationale and statistical significance in univariate analysis, outcomes for individuals in clinical practice may exist differently from the populations from clinical trials.



Both the SPARTAN and ARAMIS trials predominantly enrolled patients with PSA doubling time <10 months, which limits the generalizability of our model to lower-risk populations, such as those with PSA doubling time >10 months or Gleason 3 + 3 disease. Therefore, external validation should be conducted in independent cohorts with diverse settings, particularly in real-world contexts, to assess the broader applicability of the model.

Additionally, it is essential to note that with the advancement of novel molecular imaging techniques, nearly half of the nmCRPC patients can now be identified with PSMA-positive metastases<sup>[42]</sup>. However, this does not diminish the utility of our model. First, patients with PSMA-positive metastases and conventional imaging-negative nmCRPC may exhibit biological differences compared to conventional imaging-detected mCRPC. According to Suter *et al.*<sup>[43]</sup>, patients with oligoM CSPC detected by advanced molecular imaging had a lower incidence of TP53 mutation (17% vs. 28%) than those detected by conventional imaging. Therefore, the direct incorporation of PSMA for guiding treatment decisions in nmCRPC is controversial, as current disease classification, treatment, and outcome assessment are based on conventional imaging. For instance, it remains unclear whether the improved outcomes observed in PSMA-defined oligoM CRPC (conventional imaging-negative nmCRPC) are due to a lower tumor burden or indolent biology. Furthermore, it would be inappropriate to classify this subset of PSMA-PET positive but conventional imaging-negative patients directly as mCRPC. Despite having lesions detected by novel molecular imaging, their prognosis is significantly better than that of mCRPC patients defined by conventional imaging according to current guidelines. For instance, in ARAMIS, the median overall survival (OS) for nmCRPC patients treated with darolutamide was 40.4 months, and in SPARTAN, the median MFS for nmCRPC patients treated with apalutamide was 40.5 months. In contrast, the median OS in existing prognostic models for mCRPC patients is only 32.7 months<sup>[10]</sup>. Therefore, the subgroup of CRPC patients who are negative on conventional imaging but positive on PSMA-PET represents a distinct patient population with unique prognostic characteristics.

Second, PSMA PET, as a novel imaging modality, lacks high-level evidence supporting its role in the prognostic stratification of nmCRPC patients. Only retrospective studies, such as the one conducted by Manuel *et al.*<sup>[42]</sup>, have been performed, demonstrating that only nmCRPC patients with more than five metastatic lesions detected by PSMA PET imaging showed a statistically significant long-term survival disadvantage. Furthermore, due to the cohort being constructed in the pre-NHA era (2013–2018), the treatment approaches in the cohort were diverse. Some patients received local or regional therapies, while others underwent stereotactic body radiation therapy for metastatic lesions. Additionally, most patients did not receive NHA treatment when transitioning to the CRPC stage. Considering these confounding treatment factors, the prognostic value of PSMA PET in nmCRPC patients remains uncertain. Future prospective studies are necessary to evaluate the prognostic advantage or alteration conferred by PSMA PET detection. Therefore, there is still a long way to go before new modalities can reclassify nmCRPC patients, and during this period, our model can demonstrate its clinical utility.

Third, our model utilizes data from robust phase 3 trials, incorporating comprehensive and reliable demographic parameters, clinical indicators, and pathological parameters, with

less selection bias and a large number of patients, thereby providing strength in model performance. The external validation results prompt a quick translation into clinical practice. In contrast, no current phase 3 nmCRPC trial incorporates PSMA PET, significantly impacting the evidence's strength. Lastly, the user-friendliness of our model provides convenience and cost-saving in clinical usage, trial planning, and patient consultation, especially considering that PSMA PET is not reimbursed for nmCRPC in many countries.

In conclusion, we have successfully constructed and validated a prognostic model of eight variables for patients with nmCRPC. These prognostic factors can calculate a risk score, serving as an eligibility criterion and aiding in stratified randomization for clinical trials. Additionally, the risk score can help identify subsets of patients with significantly different outcomes regarding metastasis-free survival, allowing for swift identification of high-risk individuals who may require more aggressive treatment approaches.

## Ethical approval

The data for this study were derived from two phase III clinical trials, SPARTAN (NCT01946204) and ARAMIS (NCT02200614). Both ARAMIS and SPARTAN studies were conducted in accordance with the Declaration of Helsinki, the International Conference on Harmonisation for Good Clinical Practice, and local regulations. The study protocols were approved by the authorities and the ethics committees of the respective institutions, and signed informed consent was obtained from all patients. Informed consent specifically for the construction and validation of the nmCRPC model was not required due to the de-identified nature of the individual patient data from ARAMIS and SPARTAN. The patient-level data from the SPARTAN clinical trial were obtained through the Yale University Open Data Access Project (#2022-4942), while the ARAMIS data were provided by Bayer and made available through Vivli, Inc. (00008577).

## Consent

None.

## Sources of funding

This study was supported by Oriental Scholar Professorship (TP2022051), Shanghai Municipal Commission of Education, China; Program of Shanghai Academic/Technology Research Leader (23XD1420600); National Nature Science Foundation of China (82473381, 82172621, 81972375); Shanghai Medical Innovation Research Special Project (21Y11904300); Shanghai Shenkang Research Physician Innovation and Transformation Ability Training Project (SHDC2022CRD035); General Program of Beijing Xisike Clinical Oncology Research Foundation (Y-MSDZD2021-0230, Y-2019AZMS0012); Chinese Anti-Cancer Association – Hengrui PARP Inhibitor Cancer Research Foundation; Fudan University (FDUROP No.23108); Medical Engineering Fund of Fudan University; Clinical Scientist Program of Fudan University; and the Medical Science Data Center in Shanghai Medical College of Fudan University.

## Author contributions

X.N., D.Y., and Y.Z.: integrity of entire study; Z.W., X.N., X.L., and J.S.: statistical analysis; X.N. and Z.W.: writing – original draft. All authors: study concepts/study design or data acquisition or data analysis/interpretation; manuscript drafting or manuscript revision for important intellectual content; approval of final version of submitted manuscript; agrees to ensure any questions related to the work are appropriately resolved; writing – review & editing.

## Conflicts of interest disclosure

The authors declare no potential conflicts of interest.

## Research registration unique identifying number (UIN)

SPARTAN (NCT01946204); ARAMIS (NCT02200614).

## Guarantor

Yao Zhu.

## Provenance and peer review

Not commissioned, externally peer-reviewed.

## Data availability statement

The data that support the findings of this study are available from YODA (<https://yoda.yale.edu/applications>) and Vivli (<https://search.vivli.org/>), but restrictions apply to the availability of these data, which were used under license for the current study. Data are however available from the authors upon reasonable request and with permission of YODA and Vivli.

## References

- [1] Saad F, Bögemann M, Suzuki K, Shore N. Treatment of nonmetastatic castration-resistant prostate cancer: focus on second-generation androgen receptor inhibitors. *Prostate Cancer Prostatic Dis* 2021;24:323–34.
- [2] Scher HI, Morris MJ, Stadler WM, *et al.* Trial design and objectives for castration-resistant prostate cancer: updated recommendations from the prostate cancer clinical trials working group 3. *J Clin Oncol* 2016;34:1402–18.
- [3] Halabi S, Kelly WK, Ma H, *et al.* Meta-analysis evaluating the impact of site of metastasis on overall survival in men with castration-resistant prostate cancer. *J Clin Oncol* 2016;34:1652–59.
- [4] Dai B, Wang H, Shi B, *et al.* CACA guidelines for holistic integrative management of prostate cancer. *Holist Integr Oncol* 2024;3:47.
- [5] Bögemann M, Shore ND, Smith MR, *et al.* Efficacy and safety of darolutamide in patients with nonmetastatic castration-resistant prostate cancer stratified by prostate-specific antigen doubling time: planned subgroup analysis of the phase 3 ARAMIS trial. *Eur Urol* 2023;83:212–21.
- [6] Saad F, Sternberg CN, Efstathiou E, *et al.* Prostate-specific antigen progression in enzalutamide-treated men with nonmetastatic castration-resistant prostate cancer: any rise in prostate-specific antigen may require closer monitoring. *Eur Urol* 2020;78:847–53.
- [7] Smith MR, Saad F, Oudard S, *et al.* Denosumab and bone metastasis-free survival in men with nonmetastatic castration-resistant prostate cancer: exploratory analyses by baseline prostate-specific antigen doubling time. *J Clin Oncol* 2013;31:3800–06.
- [8] Caffo O, Maines F, Kinspergher S, Vecchia A, Messina C. To treat or not to treat: is it acceptable to avoid active therapies in advanced prostate cancer today? *Expert Rev Anticancer Ther* 2021;21:389–400.
- [9] Halabi S, Lin C-Y, Kelly WK, *et al.* Updated prognostic model for predicting overall survival in first-line chemotherapy for patients with metastatic castration-resistant prostate cancer. *J Clin Oncol* 2014;32:671–77.
- [10] Armstrong AJ, Lin P, Higano CS, *et al.* Development and validation of a prognostic model for overall survival in chemotherapy-naïve men with metastatic castration-resistant prostate cancer. *Ann Oncol* 2018;29:2200–07.
- [11] Halabi S, Yang Q, Roy A, *et al.* External validation of a prognostic model of overall survival in men with chemotherapy-naïve metastatic castration-resistant prostate cancer. *J Clin Oncol* 2023;41:2736–46.
- [12] Hussain M, Fizazi K, Saad F, *et al.* Enzalutamide in men with nonmetastatic, castration-resistant prostate cancer. *N Engl J Med* 2018;378:2465–74.
- [13] Smith MR, Saad F, Chowdhury S, *et al.* Apalutamide treatment and metastasis-free survival in prostate cancer. *N Engl J Med* 2018;378:1408–18.
- [14] Saad F, Cella D, Basch E, *et al.* Effect of apalutamide on health-related quality of life in patients with non-metastatic castration-resistant prostate cancer: an analysis of the SPARTAN randomised, placebo-controlled, phase 3 trial. *Lancet Oncol* 2018;19:1404–16.
- [15] Fizazi K, Shore N, Tammela TL, *et al.* Darolutamide in nonmetastatic, castration-resistant prostate cancer. *N Engl J Med* 2019;380:1235–46.
- [16] Grimes DA, Hubacher D, Nanda K, *et al.* The Good Clinical Practice guideline: a bronze standard for clinical research. *Lancet* 2005;366:172–74.
- [17] Dixon JR. The international conference on harmonization good clinical practice guideline. *Quality Assurance* 1999;6:65–74.
- [18] World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects. *Nurs Ethics* 2002;9:105–09.
- [19] Sauerbrei W, Taube SE, McShane LM, Cavenagh MM, Altman DG. Reporting recommendations for tumor marker prognostic studies (REMARK): an abridged explanation and elaboration. *J Natl Cancer Inst* 2018;110:803–11.
- [20] Liu Z, Liu L, Weng S, *et al.* Machine learning-based integration develops an immune-derived lncRNA signature for improving outcomes in colorectal cancer. *Nat Commun* 2022;13:816.
- [21] Harrell FE. In *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. Springer International Publishing; 2015:475–519.
- [22] Ranstam J, Cook JA. LASSO regression. *J Br Surg* 2018;105:1348–1348.
- [23] Tibshirani R. The lasso method for variable selection in the cox model. *Stat Med* 1997;16:385–95.
- [24] Binder H, Allignol A, Schumacher M, Beyersmann J. Boosting for high-dimensional time-to-event data with competing risks. *Bioinformatics* 2009;25:890–96.
- [25] Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 587–604. Springer, New York, 2009.
- [26] Whitney CA, Howard LE, Freedland SJ, *et al.* Impact of age, comorbidity, and PSA doubling time on long-term competing risks for mortality among men with non-metastatic castration-resistant prostate cancer. *Prostate Cancer Prostatic Dis* 2019;22:252–60.
- [27] Collins GS, Dhiman P, Ma J, *et al.* Evaluation of clinical prediction models (part 1): from development to external validation. *BMJ* 2024;384:e074819.
- [28] Saad F, Small EJ, Feng FY, *et al.* Deep prostate-specific antigen response following addition of apalutamide to ongoing androgen deprivation therapy and long-term clinical benefit in SPARTAN. *Eur Urol* 2022;81:184–92.
- [29] De Giorgi U, Hussain M, Shore N, *et al.* Consistent survival benefit of enzalutamide plus androgen deprivation therapy in men with nonmetastatic castration-resistant prostate cancer: PROSPER subgroup analysis by age and region. *Eur J Cancer* 2021;159:237–46.
- [30] Perez-Ruixo C, Ackaert O, Ouellet D, *et al.* Efficacy and safety exposure-response relationships of apalutamide in patients with nonmetastatic castration-resistant prostate cancer. *Clin Cancer Res* 2020;26:4460–67.

- [31] Howard LE, Moreira DM, De Hoedt A, *et al.* Thresholds for PSA doubling time in men with non-metastatic castration-resistant prostate cancer. *BJU Int* 2017;120:E80–e86.
- [32] Feng FY, Thomas S, Saad F, *et al.* Association of molecular subtypes with differential outcome to apalutamide treatment in nonmetastatic castration-resistant prostate cancer. *JAMA Oncol* 2021;7:1005–14.
- [33] Egevad L, Granfors T, Karlberg L, Bergh A, Stattin P. Prognostic value of the Gleason score in prostate cancer. *BJU Int* 2002;89: 538–42.
- [34] Smith MR, Kabbavar F, Saad F, *et al.* Natural history of rising serum prostate-specific antigen in men with castrate nonmetastatic prostate cancer. *J Clin Oncol* 2005;23:2918–25.
- [35] Evans S, Metcalfe C, Ibrahim F, Persad R, Ben-Shlomo Y. Investigating black-white differences in prostate cancer prognosis: a systematic review and meta-analysis. *Int J Cancer* 2008;123:430–35.
- [36] Rasmussen KM, Patil V, Li C, *et al.* Survival outcomes by race and ethnicity in veterans with nonmetastatic castration-resistant prostate cancer. *JAMA Netw Open* 2023;6:e2337272.
- [37] Ni X, Michael L, Ma W, *et al.* Impact of Asian race on prognosis in de novo metastatic prostate cancer. *J Natl Compr Canc Netw* 2023;21: 733–741.e733.
- [38] Beer TM, Tangen CM, Bland LB, *et al.* The prognostic value of hemoglobin change after initiating androgen-deprivation therapy for newly diagnosed metastatic prostate cancer: a multivariate analysis of Southwest oncology group study 8894. *Cancer* 2006;107:489–96.
- [39] Ebbinge M, Berglund A, Varenhorst E, Hedlund PO, Sandblom G. Clinical and prognostic significance of changes in haemoglobin concentration during 1 year of androgen-deprivation therapy for hormone-naïve bone-metastatic prostate cancer. *BJU Int* 2018;122:583–91.
- [40] Small EJ. Post hoc analysis of rapid and deep prostate-specific antigen decline and patient-reported health-related quality of life in SPARTAN and TITAN patients with advanced prostate cancer. *Eur Urol Oncol* 2024;7:844–52.
- [41] Sonpavde G, Pond GR, Berry WR, *et al.* The association between radiographic response and overall survival in men with metastatic castration-resistant prostate cancer receiving chemotherapy. *Cancer* 2011;117:3963–71.
- [42] Weber M, Fendler WP, Ravi Kumar AS, *et al.* Prostate-specific membrane antigen positron emission tomography-detected disease extent and overall survival of patients with high-risk nonmetastatic castration-resistant prostate cancer: an international multicenter retrospective study. *Eur Urol* 2024;85:511–16.
- [43] Suter P, Song Y, Van der Eecken K, *et al.* Clinical and genomic differences between advanced molecular imaging-detected and conventional imaging-detected metachronous oligometastatic castration-sensitive prostate cancer. *Eur Urol* 2023;84:531–35.