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.

Machine Learning Prediction of Cardiovascular Events in Men with Prostate Cancer Receiving Androgen Deprivation Therapy

Wen-Kuan Huang
Linkou Chang Gung Memorial Hospital
Chun-Chi Chen
Linkou Chang Gung Memorial Hospital
Ming-Jer Hsieh
Linkou Chang Gung Memorial Hospital
Chieh-Yu Chang
Linkou Chang Gung Memorial Hospital
Ming-Lung Tsai
Linkou Chang Gung Memorial Hospital
Tien-Hsing Chen
Linkou Chang Gung Memorial Hospital
Po-Jung Su
Linkou Chang Gung Memorial Hospital
See-Tong Pang
Linkou Chang Gung Memorial Hospital
Shao-Wei Chen
Linkou Chang Gung Memorial Hospital
Pao-Hsien Chu
Linkou Chang Gung Memorial Hospital
I-Chang Hsieh
Linkou Chang Gung Memorial Hospital
Dong-Yi Chen
dongyi70@cgmh.org.tw

Linkou Chang Gung Memorial Hospital

Research Article

Keywords: prostate cancer, androgen deprivation therapy, major adverse cardiovascular event, cardio-oncology, machine learning, prediction model

Posted Date: May 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9484197/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

Identifying prostate cancer (PCa) patients at high cardiovascular (CV) risk prior to androgen deprivation therapy (ADT) remains challenging, and no validated tool exists for major adverse cardiovascular event (MACE) risk stratification. To develop and validate a machine learning–based model predicting 10-year MACE risk in PCa patients receiving ADT.

Methods

In this retrospective prognostic study, we analyzed PCa patients treated with ADT from the Taiwan Cancer Registry (2010–2021), with follow-up through 2022. Fifteen sociodemographic, comorbidity, and ADT-related variables were evaluated. The cohort was split into derivation (n = 27,257) and validation (n = 11,681) sets. A random survival forest (RSF) model incorporating eight selected variables was developed to predict MACEs (ischemic stroke, myocardial infarction, and CV death), with non-CV death treated as a competing risk.

Results

Over a mean follow-up of 4.26 years, 5,681 patients (14.6%) experienced MACEs. The final model identified eight predictors: age ≥ 75 years, coronary heart disease, atrial fibrillation, heart failure, prior myocardial infarction, peripheral arterial disease, hypertension, and ADT type (GnRH agonist or orchiectomy). Both the initial 15-variable and final 8-variable RSF models demonstrated satisfactory discrimination, with area under the curve (AUC) of 81.2% (95% CI: 80.5–81.9%) and 78.7% (95% CI: 78.0–79.4%), respectively. A positive correlation was observed between the number of risk factors and the incidence of MACEs.

Conclusions

We developed and validated an 8-variable machine learning model that effectively stratifies MACE risk in PCa patients undergoing ADT, supporting targeted CV risk assessment before treatment initiation.

BACKGROUND

With advances in screening and treatment, the number of prostate cancer (PCa) survivors is increasing, highlighting the importance of survivorship care. [1, 2] Approximately 85% of all PCa survivors in the United States are over 65 years old. [1] The median age of PCa diagnosis in Taiwan is 74, and the 5-year survival rate has increased from 81.4% during 2005–2010 to 87.6% during 2015–2020. [3] As the increasing number of PCa survivors is particularly in the elderly population, cardiovascular (CV) adverse

sequelae associated with androgen-deprivation therapy (ADT) have been one major concern. [4] CV risk monitoring in men receiving ADT, particularly gonadotropin-releasing hormone (GnRH) agonist, has been recommended. [5, 6] Despite these recommendations, intense CV assessment and intervention are not routinely implemented in clinical practice. [7] Financial constraints and feasibility issues may limit cardio-oncology care, particularly for those at high CV risk or with pre-existing CV conditions. Prioritizing high-risk patients for early cardiologist intervention could reduce major adverse CV events (MACE) in prostate cancer (PCa) patients undergoing ADT. Recent guidance recommends using established CV risk calculators for patients receiving ADT. [8] However, a recent study evaluated the predictive performance of established CV risk score in cancer survivors from the UK Biobank showing that the area under the curve (AUC) for predicting CVD (included a composite endpoint of myocardial infarction, stroke, or CV mortality) was 0.62 by SCORE2/SCORE-OP and was 0.63 by QRISK3. [9] Since these traditional risk score were not being developed specifically for the population, superimposed with the not satisfactory performance highlighting the need for population-specific CV risk prediction. While age, previous CV diseases, and CV risk factors can be used to identify the high-risk population, appropriate calibration and integrated analysis of clinical factors have not been established in men with PCa. [10, 11]Notably, CV risk associated with ADT may differ across patient subgroups; older age, pre-existing CV comorbidities, and socioeconomic factors may disproportionately affect certain populations, yet these disparities remain poorly characterized in the context of ADT-related cardiotoxicity. Accordingly, we aimed to develop and validate a convenient risk prediction model based on routinely available information in broad real-world data. We used a nationwide cancer registry linked with an administrative database to predict the first incidence of MACE in men with PCa receiving ADT treatment and presented a predictor-based system for clinical use.

METHODS

Data Sources

The National Health Insurance Research Database (NHIRD) comprises claims data derived from the single-payer National Health Insurance Program, which has been in effect since 1995 and provides coverage to nearly 99% of the entire population in Taiwan. The NHIRD is currently managed by the Health and Welfare Data Science Center (HWDC). The coverage period of database upon our application was between 2010 and 2021. Important CV-related diseases in the NHIRD, such as acute myocardial infarction, [12] ischemic stroke, [13] and diabetes mellitus, [14] have been verified to have high accuracy in terms of quality and registration. [15] Death registration is mandatory in Taiwan and has been validated with high accuracy. [16] Patient data were de-identified before analysis in the NHIRD. This study was approved by the institutional review board (IRB) of Chang Gung Memorial Hospital (IRB number: 202300400B0).

Identification of cohort of PCa under ADT

Patients with new diagnoses of PCa (ICD-9 code 185, ICD-10 code C61) between 1 January 2010 and 31 December 2021 were identified using the Taiwan Cancer Registry database in HWDC. Only patients who received ADT (including bilateral orchiectomy, GnRH agonist, or GnRH antagonist) were included in this study. Treatment of ADT was identified through prescription records. GnRH agonists include leuprorelin (Anatomical Therapeutic Chemical Classification System [ATC] code L02AE02), goserelin (ATC code L02AE03), buserelin (ATC code L02AE01), and triptorelin (ATC code L02AE04). The GnRH antagonist examined in the present study was degarelix (ATC code L02BX02). The index date was defined as the date of orchiectomy or injection of the first dose of GnRH agonist or GnRH antagonist. The end of database was 31 December 2022. Patients who were aged less than 40 years, whose follow up duration less than 3 months, were diagnosed with other cancers and who developed CV events during the 3-month follow up were excluded (Fig. 1). As the primary objective of this study was to evaluate long-term prognostic factors, patients with early events (within the first 3 months) were excluded to focus the analysis on a stable patient population.

Demographic characteristics, comorbidity, and MACE

We identify comorbidities using ICD-9-CM and ICD-10 diagnostic codes and define comorbidities before and on the index date as baseline characteristics. Comorbidities include hypertension, diabetes mellitus, coronary artery disease, hyperlipidemia, atrial fibrillation, peripheral arterial disease, chronic obstructive pulmonary disease, chronic kidney disease including dialysis, chronic liver disease, heart failure, cerebrovascular disease, myocardial infarction, and prior coronary revascularization. These conditions were identified using validated diagnostic definitions, which required documentation from either at least two outpatient visits or one inpatient hospitalization, as detailed in **Additional file S1**. We also analyzed the use of medications, defined as medications used within 90 days before the index date. The primary outcome was MACE, defined as the first occurrence of hospitalizations for myocardial infarction, ischemic stroke, and CV death. The occurrence of myocardial infarction and ischemic stroke was defined as a principal discharge diagnosis. [12, 13] CV death was defined as death attributable to acute myocardial infarction, sudden cardiac death, heart failure, stroke, CV procedures, CV hemorrhage, or other CV causes, as previously established in the literature. [17] The date, place and causes of death are available via linking to Taiwan Death Registry in HWDC. The ICD codes used to identify the baseline comorbidities and composite outcomes are listed in **Additional file S1**.

As all predictor variables were derived from administrative claims data using binary indicators (presence or absence of ICD-coded diagnoses), and the NHIRD provides complete records for insurance claims and vital status, there were no missing values for the predictor variables or outcome ascertainment in this study.

Random survival forest analysis

We consider variables routinely available to clinicians and patients at the time of receiving ADT. Thus, the candidate predictor variables consisted of 15 variables, including age, ADT type at the index date, comorbidities, history of events requiring hospitalization, and stage of PCa. Due to the large sample size of the current study (n = 38,938) and the incidence of outcome was not rare (> 10%), the patients were randomly divided into two sub-cohorts: a training cohort comprising 70% of the original group, and a validation cohort consisting of the remaining 30%. We applied the traditional regression models and various machine learning-based approaches, including decision tree, linear discriminant analysis, extreme gradient boosting, support vector machine, random forest, k-nearest neighbor classification, Cox proportional hazard model, Fine and Gray subdistribution hazard model, and random survival forest (RSF) accounting for competing risk. Among these, RSF demonstrated the best performance and was selected as the primary model for the study (**Additional file S2**). We then used the RSF model to develop the risk model for predicting MACE at the end of follow-up. [18]

The training cohort was used to construct a random survival forest for competing risks through iterative node splitting, beginning at the root node. At each node, a random subset of candidate variables was selected, and splits were determined using splitting rules adapted for competing risks to maximize the separation in cause-specific outcomes between daughter nodes. Non–cardiovascular death was treated as a competing event rather than as a censoring event. Each tree was grown to full depth, ensuring that each terminal node contained at least one observed event. A total of 500 trees were generated. For each tree, cause-specific cumulative incidence functions (CIF) for MACE were estimated, and the ensemble CIF was obtained by averaging predictions across all trees, thereby providing individualized risk estimates that appropriately accounted for competing mortality. The variance importance (VIMP) was chosen for feature selection (**Additional file S3**). [19] The prediction performance of model was evaluated using the area under the receiver operating characteristic (ROC) curve (AUC), known as the Harrell's C-index. Based on the VIMP ranking from the initial RSF model, we sequentially constructed RSF models incorporating between 1 and 14 features. The results showed a progressive increase in AUC for models including 1 to 8 features, whereas model performance declined when a ninth feature was added (**Additional file S4**). Notably, the AUC of the 8-feature model closely approximated that of the full 15-feature model (78.7% vs. 81.2%). Accordingly, the final model was specified using eight features.

Predictive behavior analysis of RSF variables

To elucidate the relationship between the risk of MACEs and each of the eight selected variables, partial dependence plots were generated. The methodology for creating these plots adhered to protocols established in previous research. [20] The partial dependence analysis revealed a non-linear association between age and MACE risk. Specifically, this observation prompted a clinically relevant dichotomization of the age variable: individuals aged ≥ 75 years were grouped separately from those aged < 75 years (**Additional file S2A**). Based on the partial dependence analysis, both GnRH agonist and bilateral orchiectomy demonstrated elevated predicted CIF, whereas GnRH antagonist was associated with a lower predictive risk. Consequently, ADT types were reclassified into a dichotomized variable for the final

analysis (**Additional file S2G**). Following dichotomization of the above variables, Gray's test was used to formally assess statistical differences in CIFs between the resulting groups.

Statistical analysis

Baseline demographic characteristics and comorbid conditions were compared between the derivation and validation cohorts using appropriate statistical approaches. Categorical variables were evaluated with chi-square tests, whereas continuous variables were examined using independent-sample t tests. Model discrimination was quantified using the AUC. AUCs were calculated and contrasted for both the full 15-feature RSF model and the parsimonious 8-feature RSF model. The finalized 8-feature RSF model was subsequently subjected to external validation using the independent validation cohort. In addition, AUC estimates from the validation cohort without constraints were compared with those obtained from a validation cohort derived from the derivation dataset, in which validation estimates were constrained to match those of the training cohort. To further assess the prognostic relevance of the selected predictors, patients were categorized into five strata according to the number of applicable predictors (0–1, 2, 3, 4, or ≥ 5). CIF estimates for MACE were derived via the Aalen–Johansen estimator, and comparisons among the predefined strata were conducted using univariate Fine–Gray subdistribution hazard models, using the 0–1 predictor group as the reference category. RSF analyses were performed using R version 4.4.2 (Lucent Technologies, Union, NJ) with the packages randomForestSRC, survival, and pROC. All remaining statistical analyses were conducted using SAS version 9.4 (SAS Institute Inc., Cary, NC). Statistical significance was defined as a two-sided *P* value < 0.05. This study was reported in accordance with the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis —Artificial Intelligence (TRIPOD + AI) guideline [37]. A formal study protocol was not prepared. This study was not registered in a public study registry.

RESULTS

Baseline characteristics and outcomes

A total of 38,938 patients were analyzed in the study. Among them, 27,257 patients were in the training cohort and 11,681 patients in the validation cohort (**Additional Figure S1**). The mean age of the study population at baseline was 74.9 ± 8.8 years. Of those, 22,256 (57.2%) had locoregional disease and 16,682 (42.8%) had distant metastasis. Of the ADT type, 35,258 (90.6%) patients received GnRH agonist therapy, 2,270 (5.8%) patients received GnRH antagonist, and 1,410 (3.6%) patients received bilateral orchiectomy. Among the cohort, 23,428 patients (60.2%) presented with hypertension, 10,772 (27.7%) had diabetes mellitus, 11,634 (29.9%) had hyperlipidemia, and 10,359 (26.6%) displayed a history of CV disease (Table 1). During the mean follow-up duration of 4.26 years (standard deviation 2.89 years), 5,681 individuals (14.6% of the cohort) experienced a MACE, as indicated in Table 2. Among these cases, 1,244 patients (3.2%) encountered an ischemic stroke, 594 patients (1.5%) suffered from myocardial infarction, and 4,426 patients (11.4%) had CV death. The all-cause mortality rate at the end of the follow-

up was 48.9%, while the PCa death was 33.0%. In addition, the baseline characteristics of patients with and without MACE in either cohort are provided in **Additional file S5**.

Table 1
The baseline characteristics of patients in the derivation and validation cohorts

Variable	Total (*n* = 38,938)	Derivation (*n* = 27,257)	Validation (*n* = 11,681)	*P* value
Urbanization level of the residence				0.933
Low	4,947 (12.7)	3,469 (12.7)	1,478 (12.7)	
Moderate	11,833 (30.4)	8,266 (30.3)	3,567 (30.5)	
High	11,430 (29.4)	8,024 (29.4)	3,406 (29.2)	
Very High	10,728 (27.6)	7,498 (27.5)	3,230 (27.7)	
Region				0.613
North	17,705 (45.5)	12,411 (45.5)	5,294 (45.3)	
West	9,375 (24.1)	6,520 (23.9)	2,855 (24.4)	
South	10,819 (27.8)	7,586 (27.8)	3,233 (27.7)	
East	1,039 (2.7)	740 (2.7)	299 (2.6)	
Age, year	74.9 ± 8.8	74.94 ± 8.83	74.89 ± 8.88	0.631
ADT type at the index date				0.074
GnRH agonist	35,258 (90.6)	24,741 (90.8)	10,517 (90.0)	
GnRH antagonist (Degarelix)	2,270 (5.8)	1,549 (5.7)	721 (6.2)	
Bilateral orchiectomy	1,410 (3.6)	967 (3.6)	443 (3.8)	
Comorbidities				
Hypertension	23,428 (60.2)	16,418 (60.2)	7,010 (60.0)	0.682
Diabetes mellitus	10,772 (27.7)	7,516 (27.6)	3,256 (27.9)	0.545
Coronary heart disease	4,771 (12.3)	3,331 (12.2)	1,440 (12.3)	0.768
Hyperlipidemia	11,634 (29.9)	8,119 (29.8)	3,515 (30.1)	0.547

Abbreviation: ADT, androgen deprivation therapy; ACEI, angiotensin-converting enzyme inhibitor; ARB, angiotensin receptor blocker; ARNI, angiotensin receptor-neprilysin inhibitor; Cox-2, cyclooxygenase-2; DOACs, direct oral anticoagulants; GnRH, gonadotropin-releasing hormone; NSAIDs, non-steroidal anti-inflammatory drugs; PCa, prostate cancer; STD, standardized difference.

* Anyone of coronary heart disease, peripheral arterial disease, myocardial infarction, and stroke;

Data are presented as frequency (percentage), mean ± standard deviation, or median [25th percentile, 75th percentile].

Variable	Total (*n* = 38,938)	Derivation (*n* = 27,257)	Validation (*n* = 11,681)	*P* value
Atrial fibrillation	1,616 (4.2)	1,119 (4.1)	497 (4.3)	0.498
Peripheral arterial disease	1,533 (3.9)	1,080 (4.0)	453 (3.9)	0.695
Chronic obstructive pulmonary disease	4,461 (11.5)	3,115 (11.4)	1,346 (11.5)	0.788
Chronic kidney disease	8,539 (21.9)	5,977 (21.9)	2,562 (21.9)	0.992
Cardiovascular disease*	10,359 (26.6)	7,218 (26.5)	3,141 (26.9)	0.403
History of event				
Myocardial infarction	1,427 (3.7)	975 (3.6)	452 (3.9)	0.159
Coronary revascularization	2,894 (7.4)	2,008 (7.4)	886 (7.6)	0.452
Heart failure hospitalization	1,770 (4.6)	1,217 (4.5)	553 (4.7)	0.242
Stroke	4,197 (10.8)	2,906 (10.7)	1,291 (11.1)	0.255
Duration between PCa diagnosis and ADT, month	1.63 [0.77, 5.37]	1.63 [0.77, 5.37]	1.63 [0.77, 5.37]	0.796
Disease extent				0.883
Locoregional	22,256 (57.2)	15,586 (57.2)	6,670 (57.1)	
Distant metastasis	16,682 (42.8)	11,671 (42.8)	5,011 (42.9)	
Medication				
NSAIDs	23,367 (60.0)	16,380 (60.1)	6,987 (59.8)	0.606
Cox-2 inhibitor	6,440 (16.5)	4,510 (16.6)	1,930 (16.5)	0.954
Antiplatelet agents (aspirin/ clopidogrel/ ticagrelor)	10,931 (28.1)	7,586 (27.8)	3,345 (28.6)	0.105
Anticoagulant agents (Warfarin/ DOACs)	1,532 (3.9)	1,073 (3.9)	459 (3.9)	0.974
Oral hypoglycemic agents	8,898 (22.9)	6,184 (22.7)	2,714 (23.2)	0.239

Abbreviation: ADT, androgen deprivation therapy; ACEI, angiotensin-converting enzyme inhibitor; ARB, angiotensin receptor blocker; ARNI, angiotensin receptor-neprilysin inhibitor; Cox-2, cyclooxygenase-2; DOACs, direct oral anticoagulants; GnRH, gonadotropin-releasing hormone; NSAIDs, non-steroidal anti-inflammatory drugs; PCa, prostate cancer; STD, standardized difference.

* Anyone of coronary heart disease, peripheral arterial disease, myocardial infarction, and stroke;

Data are presented as frequency (percentage), mean ± standard deviation, or median [25th percentile, 75th percentile].

Variable	Total (*n* = 38,938)	Derivation (*n* = 27,257)	Validation (*n* = 11,681)	*P* value
Insulin	1,124 (2.9)	783 (2.9)	341 (2.9)	0.801
ACEI / ARB / ARNI	11,849 (30.4)	8,257 (30.3)	3,592 (30.8)	0.368
Beta-blocker	9,338 (24.0)	6,607 (24.2)	2,731 (23.4)	0.069
Dihydropyridine calcium channel blockers	11,357 (29.2)	7,898 (29.0)	3,459 (29.6)	0.206
MRAs (spironolactone/ eplerenone)	1,157 (3.0)	826 (3.0)	331 (2.8)	0.295
Loop diuretics	4,914 (12.6)	3,464 (12.7)	1,450 (12.4)	0.421
Thiazide	2,097 (5.4)	1,458 (5.4)	639 (5.5)	0.627
Statin	9,372 (24.1)	6,550 (24.0)	2,822 (24.2)	0.786
Fibrate	987 (2.5)	665 (2.4)	322 (2.8)	0.068
Anti-androgen medication				
Flutamide	2,408 (6.2)	1,699 (6.2)	709 (6.1)	0.539
Bicalutamide	20,731 (53.2)	14,541 (53.4)	6,190 (53.0)	0.519
Cyproterone	8,332 (21.4)	5,794 (21.3)	2,538 (21.7)	0.299
Novel anti-androgen (enzalutamide/ abiraterone)	811 (2.1)	570 (2.1)	241 (2.1)	0.859
Follow up year	4.26 ± 2.89	4.28 ± 2.90	4.22 ± 2.88	0.078

Abbreviation: ADT, androgen deprivation therapy; ACEI, angiotensin-converting enzyme inhibitor; ARB, angiotensin receptor blocker; ARNI, angiotensin receptor-neprilysin inhibitor; Cox-2, cyclooxygenase-2; DOACs, direct oral anticoagulants; GnRH, gonadotropin-releasing hormone; NSAIDs, non-steroidal anti-inflammatory drugs; PCa, prostate cancer; STD, standardized difference.

* Anyone of coronary heart disease, peripheral arterial disease, myocardial infarction, and stroke;

Data are presented as frequency (percentage), mean ± standard deviation, or median [25th percentile, 75th percentile].

Table 2
The number of events at the end of follow-up in the derivation and validation cohorts

Outcome	Total (*n* = 38,938)	Derivation (*n* = 27,257)	Validation (*n* = 11,681)	*P* value
Cardiovascular event				
Ischemic stroke	1,244 (3.2)	892 (3.3)	352 (3.0)	0.183
Myocardial infarction	594 (1.5)	416 (1.5)	178 (1.5)	0.986
Cardiovascular death	4,426 (11.4)	3,070 (11.3)	1,356 (11.6)	0.325
MACE*	5,681 (14.6)	3,982 (14.6)	1,699 (14.5)	0.870
Other outcome				
All-cause death	19,058 (48.9)	13,349 (49.0)	5,709 (48.9)	0.856
PCa related death	12,829 (33.0)	9,000 (33.0)	3,829 (32.8)	0.645
HF hospitalization	646 (1.7)	451 (1.7)	195 (1.7)	0.917

Abbreviation: HF, Heart failure hospitalization; MACE, major adverse cardiac events; PCa, prostate cancer; STD, standardized difference.

* Anyone of ischemic stroke, myocardial infarction, and cardiovascular death.

Data are presented as frequency (percentage).

Predictors and RSF model for risk of MACE

The initial RSF model (including fifteen variables) exhibited a satisfactory AUC value of 81.2% (95% confidence interval [CI]: 80.5 to 81.9%) (Fig. 2A). The VIMP analysis, detailed in **Additional file S3**, identified age as the most influential predictor (VIMP: 18.19%), followed by coronary heart disease (VIMP: 10.59%) and atrial fibrillation (VIMP: 7.04%). The performance metrics of RSF models with varying predictor numbers are presented in **Additional file S4**. According to the VIMP ranking, the RSF model with top eight variables showed an AUC value of 78.8% (95% CI: 78.0 to 79.4%), which had small difference from the initial RSF model. However, the RSF model with top nine variables showed more loss of AUC value (78.6%) compared to the eight variables model. Finally, eight variables, including age, coronary heart disease, atrial fibrillation, heart failure, myocardial infarction, peripheral arterial disease, ADT type, and hypertension, were selected based on variable importance in the final RSF model (Fig. 1). In the validation sample, the unrestricted and restricted estimates of the final RSF model exhibited AUCs of 81.5% and 80.1%, respectively (Fig. 2B). This little difference in the two AUCs implied a good external generalizability of the model. Partial dependency plots for the eight selected predictors are illustrated in **Additional file S2**. The CIFs under Aalen–Johansen estimator of MACE according to the eight selected features are shown in the supplements (**Additional file S2**).

Prognostic significance of identified risk factors

High-risk factors for MACEs were identified as age ≥ 75 years, coronary heart disease, atrial fibrillation, heart failure, myocardial infarction, peripheral arterial disease, ADT type of GnRH agonist or bilateral orchiectomy, and hypertension. These thresholds were validated using Aalen–Johansen–based CIF analyses with Gray's test in the training cohort (**Additional file S2**). Analysis of the CIF of MACEs revealed a positive correlation between event rates and the number of predictors present, consistent across both cohorts (Fig. 3). In the training cohort, stratification by risk factor count yielded the following 10-year MACE CIFs: 8.7% for patients with 0–1 risk factors, 16.6% for those with 2 risk factors, 28.5% for those with 3 risk factors, 37.2% for those with 4 risk factors, and 46.7% for individuals with ≥ 5 risk factors (*P* trend < 0.001; Fig. 3A). The validation cohort demonstrated similar trends, with 10-year MACE CIFs of 9.7%, 18.4%, 28.6%, 39.6% and 44.9% for patients with 0–1, 2, 3, 4, and ≥ 5 risk factors, respectively (*P* trend < 0.001; Fig. 3B). This replication of findings in the validation cohort substantiates the potential robustness of the risk stratification model.

DISCUSSION

We developed and validated a multivariate prediction model for developing MACE in PCa patients undergoing ADT. We identified eight clinical variables, including age, several CV comorbidities, and ADT type, that were the most determinant predictors in the model. The prediction model exhibited robust performance, with high calibration and discrimination, as indicated by a Harrell's C statistic of 78.7%. The predictor variables used in this model can be easily obtained from routine clinical information. The predictor system can be implemented into clinical practice using automated approaches within electronic medical records, enabling uro-oncologists or cardiologists to assess an individual patient's risk of MACE in routine clinical care.

A number of CV factors, including older age and hypertension, were in this prediction model, which was in line with the proposed mechanisms of ADT-induced MACE. Patients with pre-existing atherosclerotic CV diseases are known as the high-risk population. [21] Correspondingly, a history coronary heart disease or myocardial infarction were included in this prediction model. Of note, heart failure is a crucial predictor for MACE in this model. A previous study showed that lower testosterone levels predict all-cause mortality risk in patients with HFrEF and HFpEF. [22, 23] An observational study from the PCBaSe Sweden database reported ADT associated with increased heart failure risk. [24] A multicenter study comprising 421 patients with PCa demonstrated that ADT with GnRH agonist therapy, GnRH antagonist therapy, and bilateral orchiectomy were associated with an increased risk of cardiac dysfunction. [25] Androgen deprivation has also been suggested to be associated with a decrease in left ventricular longitudinal, circumferential, and radial strain measures, as assessed by speckle-tracking echocardiography. [26] Mechanistically, testosterone activates L-type calcium channels, which lead to an increase in the intracellular levels of calcium and thus regulate myocardial contractility. [27] Testosterone suppression by ADT may have an adverse effect and further worsen CV dysfunction.

Overall, most variables in this prediction model are either CV risks factors or CV diseases, reassuring the importance of comprehensive CV assessment before ADT commences.

Of note, ADT type was selected as a predictor of MACEs in the RSF model. Previous studies have suggested a varied association with CV risks among the ADT types. In our analysis, ADT with GnRH agonists or bilateral orchiectomy was identified as risk factors for MACEs. In a nationwide cohort study including 14,715 patients with PCa, orchiectomy was associated with rates of ischemic CV events comparable to those observed with GnRH agonist therapy. [28] In contrast, GnRH antagonists were not identified as a risk factor in our risk prediction model. This observation is concordant with prior studies suggesting a lower risk of ischemic CV events with GnRH antagonists compared with GnRH agonists. [29, 30] In a pooled analysis of six phase III prospective trials involving 2328 patients, GnRH antagonists use had a significantly lower incidence of cardiac events compared to those receiving GnRH agonists (Hazard ratio = 0.44, 95% CI, 0.26–0.74). [21] Supportively, a recent randomized Phase III trial revealed that treatment with relugolix, a GnRH antagonist, led to a significant reduction in adverse CV events when compared to leuprolide, a GnRH agonist, in patients with advanced PCa. [31] By utilizing the machine learning approach, the prediction model comprehensively capture the CV risk associated with all clinically relevant ADT options. This approach ensures that the model is applicable across the spectrum of ADT choices in contemporary clinical practice.

CV risk stratification prior to potentially cardiotoxic cancer treatment is recommended by the European Society of Medical Oncology (ESMO) and the American Society of Clinical Oncology (ASCO) society. [32, 33] The Heart Failure Association of the European Society of Cardiology, in collaboration with the International Cardio-Oncology Society, provided specific CV risk assessment proformas tailored to different classes of cardiotoxic cancer therapies such as anthracyclines, HER2 inhibitors, tyrosine kinase inhibitors, and ADTs. [8] They suggested using established CV disease risk calculators for CV risk assessment specific to ADT in PCa to calculate a patient's 10-year risk of atherosclerotic events. Patients with known pre-existing CV disease or a 10-year CVD risk score ≥ 20% were categorized as high risk, whereas a score of 10 ~ 19% was classified as medium risk and a score of < 10% was considered low risk. In the derivation cohort, patients with ≤ 1 risk factor exhibited a 10-year cumulative MACE incidence of 8.7%, corresponding to a low-risk group. Those with two risk factors had a 10-year MACE incidence of 16.6% and were classified as an intermediate-risk group, whereas patients with ≥ 3 risk factors experienced substantially higher event rates (≥ 28.5% at 10 years), consistent with a high-risk phenotype. The cumulative incidence and subdistribution hazard ratios increased progressively across risk groups, indicating good risk discrimination by this simplified model. While the 2022 European Society of Cardiology (ESC) Guidelines on cardio-oncology recommend using SCORE2 or SCORE2-OP to stratify CV risk in ADT recipients without pre-existing CV disease, these tools were developed for general populations rather than cancer patients. [8] Since these risk calculators were not developed from cohorts of PCa treated with ADT, this reflects the unmet need for CV risk prediction in this population. Furthermore, SCORE2 and SCORE2-OP are exclusively applicable to patients without pre-existing CVD, offering no risk stratification guidance for those with established CV disease. In contrast, our study is the first to identify predictors and develop a validated multivariate predictive model specifically for prostate

cancer patients undergoing ADT, regardless of prior CV status. The MACE risk prediction of the current study would be more applicable in real-world practice for patients with PCa under ADT.

We used machine learning techniques, specifically RSF, to construct the risk model. This model uses a binary tree decision methodology, allowing for unbiased feature selection. [34] The RSF model is superior for risk modeling based on uncovering nonlinear variable effects and interactions than traditional Cox regression models. [35] Through partial dependence plots in RSF analysis, we identified nonlinear effects of variables (i.e., age) on the MACE outcome, which could further provide clinically relevant thresholds. Moreover, rigorous internal and pseudo-external validation procedures confirmed the reliability of the RSF model across a diverse patient population. Strengths of this study include that the model was developed in a large, longitudinal, population-representative cohort. The Taiwan National Health Insurance Research Database demonstrates high accuracy and completeness in capturing clinical diagnoses and prescribed medications, effectively mitigating selection and recall bias and making it an ideal database for developing a robust prediction model. [36] This is the first study to identify predictors and develop a multivariate prediction model for MACE in prostate patients undergoing ADT. The predictor variables in this risk model comprise patient demographics and clinical information that are easily accessible, providing a simple and reliable approach that can be employed to assess MACE risk when initiating ADT, thus facilitating informed decision-making and guiding subsequent CV management.

This study also has several limitations. First, candidate variables do not include all potential risk factors for MACE, such as smoking status, blood pressure, serum glucose, lipid profile, level of FSH, luteinizing hormone, and testosterone, which was lacking in the NHIRD. The incorporation of blood pressure indices would likely enhance the predictive accuracy of our model. Second, residual or unmeasured confounding may exist, such as familial history or lifestyle factors. Third, during the study period, novel anti-androgens (abiraterone and enzalutamide) were exclusively reimbursed for metastatic castration-resistant prostate cancer (mCRPC) patients in our healthcare system, with metastatic hormone-sensitive prostate cancer (mHSPC) reimbursement approval not occurring until July 2020. Consequently, their utilization was extremely limited in our study population (n = 811, 2.1%), precluding their inclusion as variables in our prediction model. Lastly, the risk model was developed and validated in the cohort from Taiwan, and there was a lack of examination of the generalizability across ethnicity and race. Furthermore, the fairness of the model across different racial, ethnic, and sociodemographic groups has not been evaluated, and caution is warranted when applying the model to populations outside of Taiwan. For clinical implementation, the eight predictor variables can be ascertained from electronic health records or administrative claims databases. In settings where certain predictor information is unavailable, clinicians should exercise caution and supplement with clinical judgment. The model is designed for use by oncologists, urologists, or cardiologists and does not require specialized technical expertise beyond access to the patient's medical history.

CONCLUSIONS

We developed and internally validated a prediction model for MACE in patients with PCa receiving ADT, derived from machine learning-based analysis of a nationwide cohort. This RSF model demonstrated that eight clinical variables, including age ≥ 75 years, coronary heart disease, atrial fibrillation, heart failure, myocardial infarction, peripheral arterial occlusive disease, ADT type of GnRH agonist or bilateral orchiectomy, and hypertension. These predictors from this model enable personalized CV risk assessment when considering ADT treatment.

Declarations

Ethics approval and consent to participate

Ethical approval was granted by the institutional review board of Chang Gung Memorial Hospital (202300400B0), which waived the need for informed consent from participants. The study was performed in accordance with the Declaration of Helsinki.

Clinical trial number

Not applicable.

Consent for publication

Not applicable.

Patient and public involvement

Patients or the public were not involved in the design, conduct, reporting, or dissemination plans of this research.

Availability of data and materials

The primary data that support the findings of this study are derived from the Taiwan NHIRD. The NHIRD contains encrypted personal information to protect patient privacy and is not publicly available due to legal restrictions imposed by the government of Taiwan. The secondary data analyzed in this study can be provided by the corresponding author upon reasonable request. The R code used for the random survival forest analysis (packages: randomForestSRC, survival, pROC) is available from the corresponding author upon reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by Chang Gung Memorial Hospital (grant number CFRPG3N0051), and by Ministry of Health and Welfare (grant numbers MOHW111-TDU-B-221-114002, MOHW112-TDU-B-222-

124002).

Authors' contributions

Wen-Kuan Huang and Dong-Yi Chen: grant acquisition, conceptualization, methodology, data curation, project administration, investigation, model validation, visualization, and initial draft of manuscript.

Ming-Jer Hsieh, Chun-Chi Chen, Chieh-Yu Chang, and Ming-Lung Tsai: conceptualization, methodology, investigation, visualization, and model validation

Po-Jung Su, Shao-Wei Chen and Tien-Hsing Chen: data curation, software, and statistical analysis.

Pao-Hsien Chu, See-Tong Pang, and I-Chang Hsieh: conceptualization, methodology, supervision, and visualization.

Acknowledgments

The authors thank the statistical assistance and wish to acknowledge the support of the Maintenance Project of the Center for Big Data Analytics and Statistics at Chang Gung Memorial Hospital for study design and monitor, data analysis and interpretation.

References

1. Miller KD, Nogueira L, Devasia T, Mariotto AB, Yabroff KR, Jemal A, et al. Cancer treatment and survivorship statistics, 2022. CA Cancer J Clin. 2022;72:409–36. https://doi.org/10.3322/caac.21731.
2. Donnelly DW, Gavin A, Downing A, Hounsome L, Kearney T, McNair E, et al. Regional Variations in Quality of Survival Among Men with Prostate Cancer Across the United Kingdom. Eur Urol. 2019;76:228–37. https://doi.org/10.1016/j.eururo.2019.04.018.
3. Health Promotion Administration. Taiwan cancer registry online system. In: Ministry of Health and Welfare. https://cris.hpa.gov.tw/pagepub/Home.aspx
4. Muniyan S, Xi L, Datta K, Das A, Teply BA, Batra SK, et al. Cardiovascular risks and toxicity - The Achilles heel of androgen deprivation therapy in prostate cancer patients. Biochim Biophys Acta Rev Cancer. 2020;1874:188383. https://doi.org/10.1016/j.bbcan.2020.188383.
5. Levine GN, D'Amico AV, Berger P, Clark PE, Eckel RH, Keating NL, et al. Androgen-deprivation therapy in prostate cancer and cardiovascular risk: a science advisory from the American Heart Association, American Cancer Society, and American Urological Association: endorsed by the American Society for Radiation Oncology. Circulation. 2010;121:833–40. https://doi.org/10.1161/CIRCULATIONAHA.109.192695.
6. Skolarus TA, Wolf AM, Erb NL, Brooks DD, Rivers BM, Underwood W 3, et al. American Cancer Society prostate cancer survivorship care guidelines. CA Cancer J Clin. 2014;64:225–49. https://doi.org/10.3322/caac.21234.

7. Atalar K, Rajan P. Understanding and managing the cardiovascular risks of androgen deprivation therapy for prostate cancer. Nat Rev Urol. 2023;20:3–4. https://doi.org/10.1038/s41585-022-00671-5.
8. Lyon AR, Dent S, Stanway S, Earl H, Brezden-Masley C, Cohen-Solal A, et al. Baseline cardiovascular risk assessment in cancer patients scheduled to receive cardiotoxic cancer therapies: a position statement and new risk assessment tools from the Cardio-Oncology Study Group of the Heart Failure Association of the European Society of Cardiology in collaboration with the International Cardio-Oncology Society. Eur J Heart Fail. 2020;22:1945–60. https://doi.org/10.1002/ejhf.1920.
9. McCracken C, Condurache D-G, Szabo L, Elghazaly H, Walter FM, Mead AJ, et al. Predictive performance of cardiovascular risk scores in cancer survivors from the UK Biobank. Cardio Oncol. 2024;6:575–88. https://doi.org/10.1016/j.jaccao.2024.05.015.
10. Narayan V, Ross AE, Parikh RB, Nohria A, Morgans AK. How to Treat Prostate Cancer With Androgen Deprivation and Minimize Cardiovascular Risk: A Therapeutic Tightrope. JACC CardioOncol. 2021;3:737–41. https://doi.org/10.1016/j.jaccao.2021.09.014.
11. Hu JR, Duncan MS, Morgans AK, Brown JD, Meijers WC, Freiberg MS, et al. Cardiovascular Effects of Androgen Deprivation Therapy in Prostate Cancer: Contemporary Meta-Analyses. Arterioscler Thromb Vasc Biol. 2020;40:e55–64. https://doi.org/10.1161/ATVBAHA.119.313046.
12. Cheng C-L, Lee C-H, Chen P-S, Li Y-H, Lin S-J, Yang Y-HK. Validation of acute myocardial infarction cases in the national health insurance research database in Taiwan. J Epidemiol. 2014;24:500–7. https://doi.org/10.2188/jea.JE20140076.
13. Cheng CL, Kao YH, Lin SJ, Lee CH, Lai ML. Validation of the National Health Insurance Research Database with ischemic stroke cases in Taiwan. Pharmacoepidemiol Drug Saf. 2011;20:236–42. https://doi.org/10.1002/pds.2087.
14. Wu CS, Lai MS, Gau SS, Wang SC, Tsai HJ. Concordance between patient self-reports and claims data on clinical diagnoses, medication use, and health system utilization in Taiwan. PLoS ONE. 2014;9:e112257. https://doi.org/10.1371/journal.pone.0112257.
15. Huang YT, Wei T, Huang YL, Wu YP, Chan KA. Validation of diagnosis codes in healthcare databases in Taiwan, a literature review. Pharmacoepidemiol Drug Saf. 2023. https://doi.org/10.1002/pds.5608.
16. Lu TH, Lee MC, Chou MC. Accuracy of cause-of-death coding in Taiwan: types of miscoding and effects on mortality statistics. Int J Epidemiol. 2000;29:336–43. https://doi.org/10.1093/ije/29.2.336.
17. Hicks KA, Mahaffey KW, Mehran R, Nissen SE, Wiviott SD, Dunn B, et al. 2017 Cardiovascular and Stroke Endpoint Definitions for Clinical Trials. Circulation. 2018;137:961–72. https://doi.org/10.1161/circulationaha.117.033502.
18. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. Ann Appl Stat. 2008;2:841–60. https://doi.org/10.1214/08-AOAS169.

19. Ishwaran H, Kogalur UB, Chen X, Minn AJ. Random survival forests for high-dimensional data. Stat Anal Data Min. 2011;4:115–32. https://doi.org/10.1002/sam.10103.
20. Chang CY, Chen CC, Tsai ML, Hsieh MJ, Chen TH, Chen SW, et al. Predicting Mortality and Hospitalization in Heart Failure With Preserved Ejection Fraction by Using Machine Learning. JACC Asia. 2024;4:956–68. https://doi.org/10.1016/j.jacasi.2024.09.003.
21. Albertsen PC, Klotz L, Tombal B, Grady J, Olesen TK, Nilsson J. Cardiovascular morbidity associated with gonadotropin releasing hormone agonists and an antagonist. Eur Urol. 2014;65:565–73. https://doi.org/10.1016/j.eururo.2013.10.032.
22. Yoshihisa A, Suzuki S, Sato Y, Kanno Y, Abe S, Miyata M, et al. Relation of testosterone levels to mortality in men with heart failure. Am J Cardiol. 2018;121:1321–7. https://doi.org/10.1016/j.amjcard.2018.01.052.
23. Güder G, Frantz S, Bauersachs J, Allolio B, Ertl G, Angermann CE, et al. Low circulating androgens and mortality risk in heart failure. Heart. 2010;96:504–9. https://doi.org/10.1136/hrt.2009.181065.
24. Van Hemelrijck M, Garmo H, Holmberg L, Ingelsson E, Bratt O, Bill-Axelson A, et al. Absolute and relative risk of cardiovascular disease in men with prostate cancer: results from the Population-Based PCBaSe Sweden. J Clin Oncol. 2010;28:3448–56. https://doi.org/10.1200/JCO.2010.29.1567.
25. Chen DY, Lee CH, Tsai ML, Hsieh MJ, Chuang CK, Pang ST, et al. Cancer Therapy-Related Cardiac Dysfunction in Patients With Prostate Cancer Undergoing Androgen Deprivation Therapy. J Am Heart Assoc. 2023;12:e030447. https://doi.org/10.1161/jaha.123.030447.
26. Kanar BG, Ozben B, Sunbul M, Şener E, Ozkan O, Tınay I, et al. Androgen-deprivation therapy impairs left ventricle functions in prostate cancer patients. Int Urol Nephrol. 2019;51:1107–12. https://doi.org/10.1007/s11255-019-02184-4.
27. Curl CL, Delbridge LMD, Canny BJ, Wendt IR. Testosterone modulates cardiomyocyte Ca(2+) handling and contractile function. Physiol Res. 2009;58:293–7. https://doi.org/10.33549/physiolres.931460.
28. Chen DY, See LC, Liu JR, Chuang CK, Pang ST, Hsieh IC, et al. Risk of Cardiovascular Ischemic Events After Surgical Castration and Gonadotropin-Releasing Hormone Agonist Therapy for Prostate Cancer: A Nationwide Cohort Study. J Clin Oncol. 2017;35:3697–705. https://doi.org/10.1200/JCO.2016.71.4204.
29. Chen DY, Su PJ, See LC, Liu JR, Chuang CK, Pang ST, et al. Gonadotropin-releasing hormone antagonist associated with lower cardiovascular risk compared with gonadotropin-releasing hormone agonist in prostate cancer: A nationwide cohort and in vitro study. Prostate. 2021;81:902–12. https://doi.org/10.1002/pros.24187.
30. Cicione A, Nacchia A, Guercio A, Gravina C, Franco A, Grimaldi MC, et al. Cardiovascular adverse events-related to GnRH agonists and GnRH antagonists: analysis of real-life data from Eudra-Vigilance and Food and Drug Administration databases entries. Prostate Cancer Prostatic Dis. 2023. https://doi.org/10.1038/s41391-022-00640-4.

31. Shore ND, Saad F, Cookson MS, George DJ, Saltzstein DR, Tutrone R, et al. Oral Relugolix for Androgen-Deprivation Therapy in Advanced Prostate Cancer. N Engl J Med. 2020;382:2187–96. https://doi.org/10.1056/NEJMoa2004325.
32. Armenian SH, Lacchetti C, Barac A, Carver J, Constine LS, Denduluri N, et al. Prevention and Monitoring of Cardiac Dysfunction in Survivors of Adult Cancers: American Society of Clinical Oncology Clinical Practice Guideline. J Clin Oncol. 2017;35:893–911. https://doi.org/10.1200/jco.2016.70.5400.
33. Curigliano G, Lenihan D, Fradley M, Ganatra S, Barac A, Blaes A, et al. Management of cardiac disease in cancer patients throughout oncological treatment: ESMO consensus recommendations. Ann Oncol. 2020;31:171–90. https://doi.org/10.1016/j.annonc.2019.10.023.
34. Chang C-Y, Chen C-C, Tsai M-L, Hsieh M-J, Chen T-H, Chen S-W, et al. Predicting Mortality and Hospitalization in Heart Failure With Preserved Ejection Fraction by Using Machine Learning. JACC Asia. 2024;4:956–68. https://doi.org/10.1016/j.jacasi.2024.09.003.
35. Dietrich S, Floegel A, Troll M, Kühn T, Rathmann W, Peters A, et al. Random Survival Forest in practice: a method for modelling complex metabolomics data in time to event analysis. Int J Epidemiol. 2016;45:1406–20. https://doi.org/10.1093/ije/dyw145.
36. Lin LY, Warren-Gash C, Smeeth L, Chen PC. Data resource profile: the National Health Insurance Research Database (NHIRD). Epidemiol Health. 2018;40:e2018062. https://doi.org/10.4178/epih.e2018062.
37. Collins GS, Moons KGM, Dhiman P, Riley RD, Altman DG, van Smeden M, et al. TRIPOD + AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378. https://doi.org/10.1136/bmj-2023-078378.

Figures

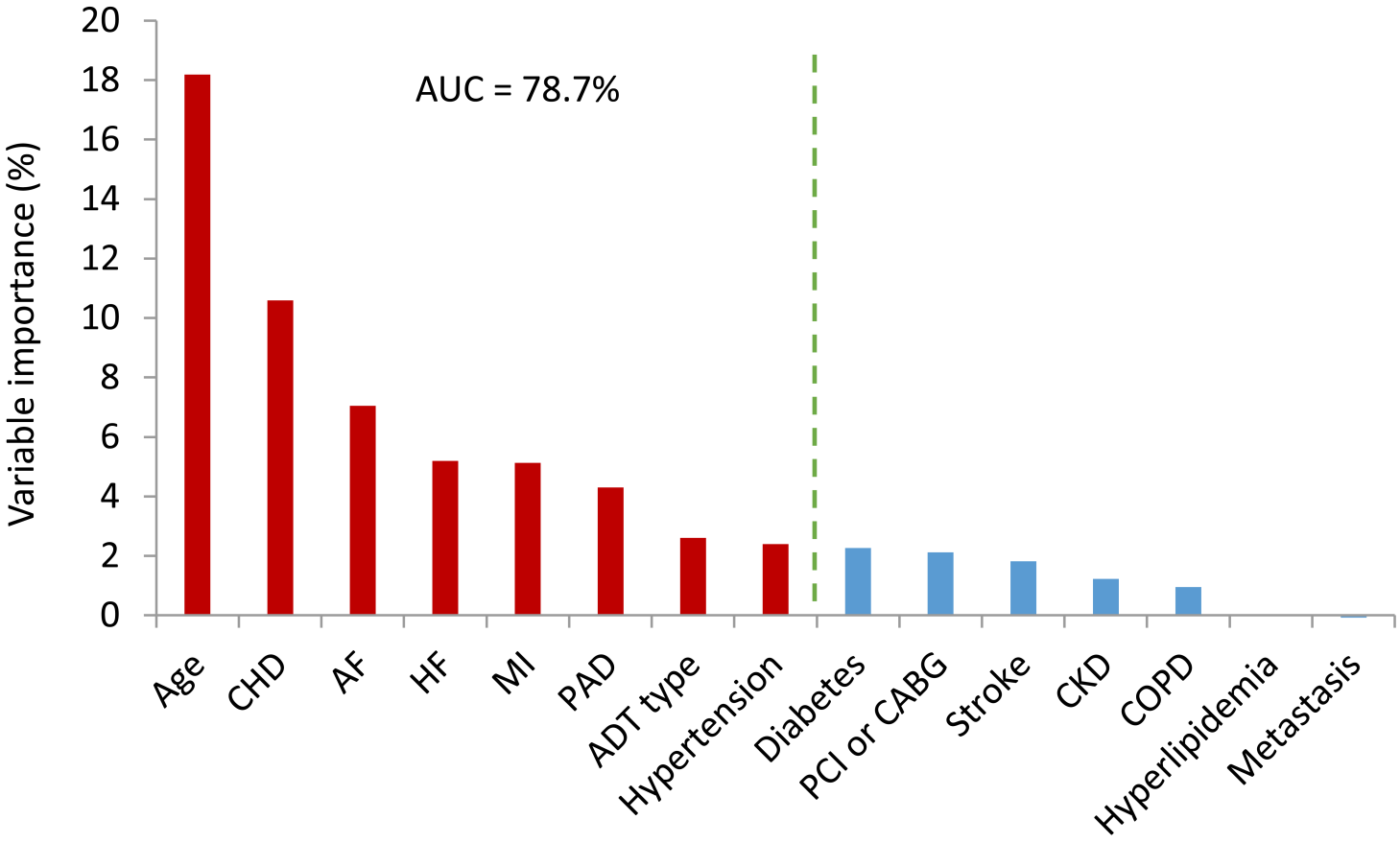


Figure 1

The importance of variables. The variable importance distribution of the 15 variables by the initial random survival forest model was presented. The top 8 variables were selected for the final random survival forest model.

ADT, androgen deprivation therapy; AF, atrial fibrillation; CABG, coronary artery bypass graft; CHD, coronary heart disease; CKD, chronic kidney disease; COPD, chronic obstructive pulmonary disease; HF, heart failure; MI, myocardial infarction; PCI, percutaneous coronary intervention.

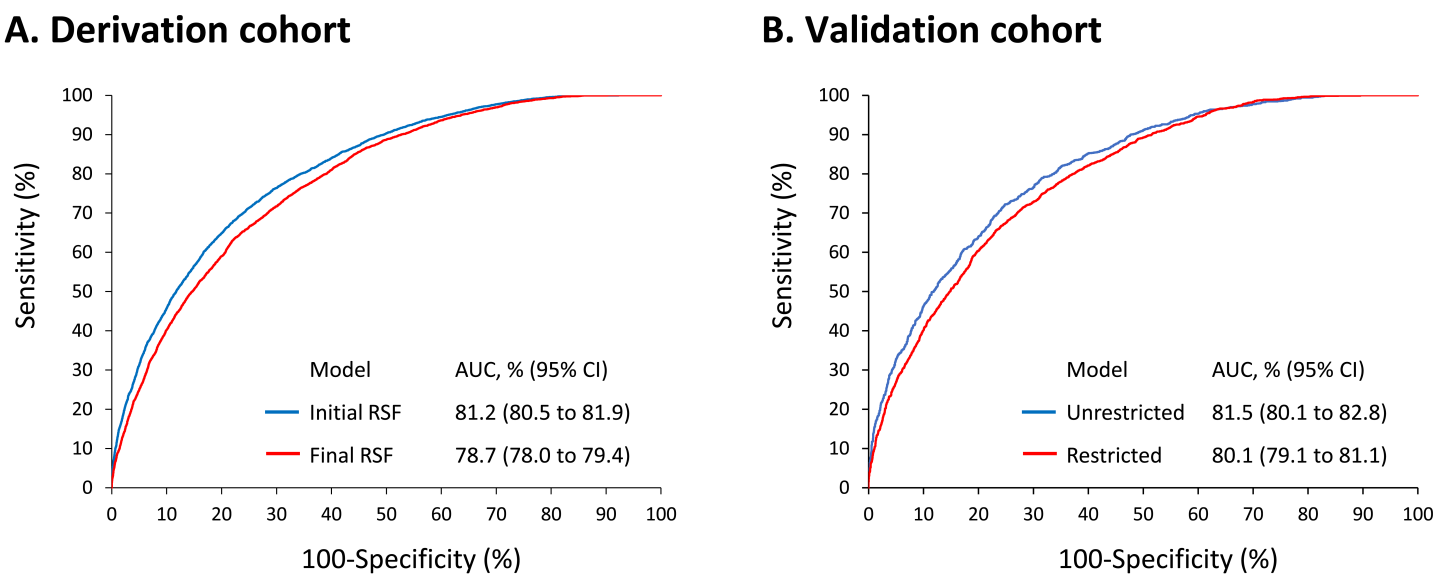


Figure 2

The discrimination performance of random survival forest models. (A) This includes all 15 variables (initial RSF) and 8 selected variables (final RSF) in the derivation sample. **(B)**The unrestricted and restricted estimates under the final RSF model in the validation sample. AUC, area under curve; CI, confidence interval; RSF, random survival forest.

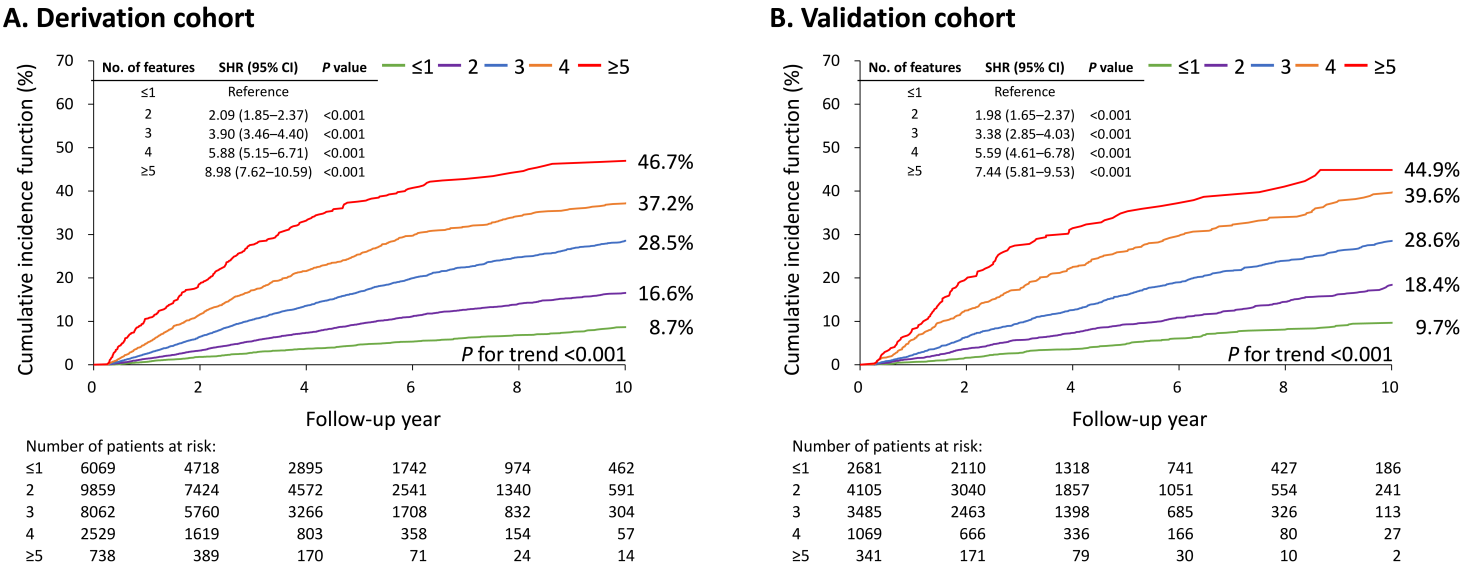


Figure 3

Risk stratification based on predictor variables. Cumulative incidence function via the Aalen–Johansen estimatorof major adverse cardiovascular events (non-cardiovascular death was considered a competing risk) stratified by different number of risk factors identified in random survival forest model: (A) derivation and (B) validation cohorts. CI, confidence interval; SHR, subdistribution hazard ratio.

Supplementary Files

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

- SuppdocumentADTprediction20260208DYC.docx