

Development, Temporal External Validation, and Updating of a Nomogram for Predicting Hypertension Cure After Unilateral Adrenalectomy in Primary Aldosteronism

Yuchao Huangfu

Beijing Anzhen Hospital

Mingchuan Li

Beijing Anzhen Hospital

Yiqing Du

Peking University People's Hospital

Zixiong Huang

Peking University People's Hospital

Ning Zhang

Beijing Anzhen Hospital

Tao xu

xutao@pkuph.edu.cn

Peking University People's Hospital

Research Article

Keywords: primary aldosteronism, adrenalectomy, hypertension, nomogram, prediction model, external validation, model updating

Posted Date: October 1st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10981384/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

Primary aldosteronism (PA) is the leading cause of secondary hypertension. While unilateral adrenalectomy achieves biochemical cure, postoperative hypertension resolution varies widely, and existing tools such as the Aldosteronoma Resolution Score perform inconsistently in Asian populations. We aimed to develop, temporally validate, and update a nomogram for predicting hypertension cure after adrenalectomy.

Methods

We retrospectively analyzed 112 patients with unilateral PA who underwent adrenalectomy (2018–2021) as the derivation cohort and 75 consecutive patients (2023) for temporal validation. Hypertension outcomes were classified by PASO criteria. Multivariable logistic regression identified independent predictors for nomogram construction. Discrimination, calibration, and clinical utility (decision curve analysis) were assessed, followed by model updating via parsimony reduction and logistic recalibration.

Results

In the derivation cohort (cure rate 45.5%), age, waist circumference, highest systolic blood pressure, and cerebro-cardiovascular disease history were independent predictors of non-cure. The five-variable nomogram demonstrated excellent discrimination (C-index 0.907) but showed moderate performance (C-index 0.734) with overfitting (calibration slope 0.32) upon temporal validation. A parsimonious four-variable model (excluding hypertension duration) improved discrimination (C-index 0.812) and calibration, with further enhancement after logistic recalibration. Decision curve analysis confirmed superior net benefit across clinically relevant threshold probabilities.

Conclusions

Temporal external validation substantially tempered the derivation-stage performance. The recalibrated four-variable model using routinely available preoperative variables offers the best transportability and is proposed for individualized preoperative counseling in Chinese patients with unilateral PA.

Clinical trial number: not applicable.

Introduction

Primary aldosteronism (PA) is now recognized as the most common specifically treatable cause of secondary hypertension, with an estimated prevalence of 5–15% among hypertensive populations and exceeding 20% in patients with resistant hypertension^{1,2}. PA is characterized by autonomous aldosterone production, most frequently from a unilateral aldosterone-producing adenoma or bilateral adrenal hyperplasia³. Beyond elevated blood pressure, aldosterone excess exerts direct deleterious

effects on the heart, vasculature, and kidney; compared with essential hypertension of equivalent severity, PA is associated with disproportionately higher rates of cerebro-cardiovascular events, atrial fibrillation, renal impairment, and metabolic disorders^{4,5}.

For patients with unilateral PA, adrenalectomy is the recommended first-line treatment and achieves biochemical cure in almost all cases⁶. The resolution of hypertension after surgery, however, is far less predictable: reported clinical cure rates range from approximately 30% to over 70% across studies⁷. A substantial proportion of patients continue to require antihypertensive medication, albeit usually at reduced doses. Reliable preoperative prediction of postoperative blood pressure outcomes is therefore clinically important for patient counseling, expectation management, and shared surgical decision-making.

Several preoperative factors have been linked to hypertension resolution, including younger age, shorter hypertension duration, fewer antihypertensive medications, lower body mass index (BMI), and female sex^{8,9}. The Aldosteronoma Resolution Score (ARS), proposed by Zarnegar et al. in 2008, combines four such variables into a 0–5 scale and remains the most widely validated tool^{10,11}. Subsequent validations, however, have yielded inconsistent discrimination, particularly in Asian cohorts. More recent tools—including the Utsumi nomogram, the Primary Aldosteronism Surgical Outcome (PASO) score, and a Chinese nomogram-based preoperative score—have improved prediction but were largely developed and evaluated within single derivation settings, and external validation with formal model updating remains scarce¹².

Clinical prediction models frequently perform worse in new patients than in their derivation sample, because of optimism in model development and differences in case mix over time¹³. Temporal validation in a later cohort from the same center, followed where necessary by model updating (recalibration or parsimony reduction), is a pragmatic strategy to obtain a tool that is actually transportable to contemporary patients. To our knowledge, no PA surgical outcome model has undergone this full cycle of development, temporal external validation, and updating in a Chinese population.

In this study, we (i) developed a nomogram for predicting hypertension cure after unilateral adrenalectomy using routinely available preoperative variables in a derivation cohort of 112 patients; (ii) temporally validated the model in an independent cohort of 75 patients operated on in 2023, quantifying discrimination, calibration, and clinical utility; and (iii) updated the model through parsimony reduction and recalibration, and compared its performance with the traditional ARS.

Methods

Study design and patients

This retrospective cohort study was conducted at Peking University People's Hospital, a tertiary referral center in China, and is reported in accordance with the TRIPOD statement. The protocol was approved by

the Peking University People's Hospital Ethical Review Board (approval number: 2021PHB342-001), which waived the requirement for individual informed consent owing to the retrospective design.

The derivation cohort comprised 112 consecutive patients with PA who underwent unilateral adrenalectomy (laparoscopic or retroperitoneoscopic) between January 2018 and December 2021. An independent temporal validation cohort comprised 75 consecutive patients with PA who underwent unilateral adrenalectomy at the same institution between January and December 2023; all had completed at least 6 months of postoperative follow-up at the time of data collection (median 29 months, range 6–36).

The diagnosis of PA followed the Chinese expert consensus on the diagnosis and treatment of PA (2020 edition) and the Endocrine Society guideline: after withdrawal of interfering medications, a plasma aldosterone-to-renin ratio above the screening threshold was considered a positive screen, and PA was confirmed by a captopril challenge or saline infusion test. Subtype classification relied on adrenal computed tomography (CT) and/or adrenal venous sampling (AVS). Inclusion criteria were: (1) age > 18 years; (2) confirmed PA with unilateral disease undergoing unilateral adrenalectomy; and (3) a minimum of 6 months of postoperative follow-up. Exclusion criteria were: (1) malignant histopathology; (2) bilateral or combined procedures; and (3) loss to follow-up.

Data collection

Clinical data were extracted from the electronic medical record system. Collected variables included demographic characteristics (sex, age, BMI, waist circumference); hypertension-related variables (hypertension duration, number of preoperative antihypertensive drug classes, highest recorded systolic blood pressure [SBP] and diastolic blood pressure [DBP]); comorbidities (diabetes mellitus; history of hypokalemia; cerebro-cardiovascular disease [CCD], defined as any documented coronary artery disease including prior percutaneous coronary intervention, ischemic or hemorrhagic stroke including lacunar infarction, cerebral thrombosis, aortic dissection, or valvular heart surgery, and coded as present versus absent); and laboratory parameters (lowest recorded serum potassium, preoperative serum potassium and sodium, serum creatinine with estimated glomerular filtration rate [eGFR] by the CKD-EPI equation, high-density lipoprotein cholesterol, triglycerides, 8 AM serum cortisol, direct renin concentration, plasma aldosterone concentration, and the aldosterone-to-renin ratio). Postoperative blood pressure and antihypertensive medication use were ascertained by outpatient review and structured telephone follow-up.

Outcome definition

Postoperative hypertension outcomes were assessed at the last available follow-up and classified according to the PASO consensus criteria. Complete clinical remission (cure) was defined as blood pressure < 140/90 mmHg without antihypertensive medication; partial clinical remission as

normotension with fewer medications, or improved blood pressure with the same or fewer medications; and absent clinical remission as unchanged or worsened blood pressure with the same or more medications. For the primary analysis, outcomes were dichotomized as cure (complete clinical remission) versus non-cure (partial or absent clinical remission).

ARS calculation

The ARS was calculated for each derivation-cohort patient as originally described: antihypertensive drug classes ≤ 2 (2 points), BMI ≤ 25 kg/m² (1 point), hypertension duration ≤ 6 years (1 point), and female sex (1 point), for a total of 0–5 points, with higher scores indicating a greater likelihood of cure.

Missing data

In the validation cohort, waist circumference values recorded as 0 ($n = 4$) were treated as missing, as were empty entries for hypertension duration ($n = 4$). Because missingness was limited ($\leq 5\%$ per variable) and unrelated to outcome, these variables were imputed with the cohort median (waist circumference 83.6 cm; hypertension duration 120 months). Postoperative outcome was unavailable for 12 of 75 validation patients (16%); the primary external validation analysis used patients with complete outcome data ($n = 63$), and a sensitivity analysis including all 75 patients was performed with missing outcomes assigned to the modal category (non-cure) as a conservative assumption.

Statistical analysis

Continuous variables are expressed as mean $\pm$ standard deviation or median (interquartile range) according to normality (Shapiro–Wilk test), and categorical variables as counts (percentages). Group comparisons used the independent t test or Mann–Whitney U test and the chi-squared or Fisher exact test, as appropriate.

In the derivation cohort, variables with $P < 0.05$ in univariable analysis entered a multivariable logistic regression with forward stepwise selection (entry $\alpha = 0.05$, removal $\alpha = 0.10$). Independent predictors, together with hypertension duration (retained on the basis of established clinical relevance), were used to construct a nomogram predicting the probability of hypertension cure. Internal validation used 1,000 bootstrap resamples and ten-times-repeated five-fold cross-validation.

In temporal external validation, the derivation model was applied to the 2023 cohort without refitting. Discrimination was quantified by the C-index (equivalent to the area under the receiver operating characteristic curve) with 2,000-replicate bootstrap confidence intervals. Calibration was evaluated following the framework of Van Calster et al.: calibration-in-the-large (mean predicted versus observed cure probability), the calibration slope from a logistic model of outcome on the linear predictor, and flexible (loess) calibration plots. Model updating comprised (i) parsimony reduction—evaluating the model restricted to the four independently significant predictors—and (ii) logistic recalibration, in which

the outcome was regressed on the linear predictor to re-estimate the intercept and slope. Clinical utility was assessed by decision curve analysis (DCA) comparing net benefit across threshold probabilities against treat-all and treat-none strategies. Analyses were performed with R version 4.3.3 (rms, pROC, and rmda packages) and Python version 3.12; two-sided $P < 0.05$ was considered significant.

Results

Patient characteristics

The derivation cohort comprised 112 patients (mean age 46.2 ± 10.6 years; 65 men, 58.0%) with a median hypertension duration of 72 months (interquartile range [IQR] 36–120) and a median of 2 antihypertensive drug classes (IQR 1–3). The temporal validation cohort comprised 75 patients (mean age 52.9 ± 10.8 years; 43 men, 57.3%) with a median hypertension duration of 120 months (IQR 36–180) and a median of 2 antihypertensive drug classes (IQR 1–3). The two cohorts were broadly comparable in demographic and clinical characteristics, the validation cohort was older and had a longer hypertension duration, whereas the prevalence of cerebro-cardiovascular disease was similar in the two cohorts (20.0% vs 20.5%) (Table 1).

Table 1
Baseline characteristics of patients with unilateral primary aldosteronism in the derivation and temporal validation cohorts

Variable	Derivation cohort (n = 112)	Validation cohort (n = 75)
Age, years	46.2 ± 10.6	52.9 ± 10.8
Male sex, n (%)	65 (58.0)	43 (57.3)
BMI, kg/m ²	26.2 ± 4.0	25.9 ± 3.6
Waist circumference, cm	91.3 ± 11.4	84.0 ± 9.7
Hypertension duration, months	72 (36–120)	120 (36–180)
Antihypertensive drug classes	2 (1–3)	2 (1–3)
Highest recorded SBP, mmHg	168 ± 18	160.9 ± 25.7
Highest recorded DBP, mmHg	105 ± 12	106 ± 13
Lowest serum potassium, mmol/L	2.9 ± 0.5	2.7 ± 0.6
Preoperative serum potassium, mmol/L	3.4 ± 0.5	3.3 ± 0.6
Serum sodium, mmol/L	142 ± 3	142 ± 3
eGFR, mL/min/1.73 m ²	98.5 ± 18.3	97.2 ± 19.1
Plasma aldosterone, ng/dL	23.6 (16.2–35.4)	25.1 (17.0–38.2)
Direct renin, pg/mL	5.2 (3.1–8.7)	4.8 (2.9–8.1)
Aldosterone-to-renin ratio	45.2 (28.6–72.1)	48.7 (31.2–78.5)
Diabetes mellitus, n (%)	17 (15.2)	16 (21.3)
History of hypokalemia, n (%)	91 (81.3)	62 (82.7)
Cerebro-cardiovascular disease, n (%)	23 (20.5)	15 (20)
Follow-up, months	25 (12–33)	29 (26–31)
Cure (complete clinical remission), n (%)	51 (45.5)	29 (46.0)⊠
Data are mean ± standard deviation, median (interquartile range), or n (%). BMI, body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure; eGFR, estimated glomerular filtration rate. ⊠ Based on 63 patients with complete outcome data; 12 patients with missing outcomes were excluded from the primary analysis.		

In the derivation cohort, 51 of 112 patients (45.5%) achieved complete clinical remission (cure) at a median follow-up of 25 months (IQR 12–33). In the temporal validation cohort, 29 of 63 patients with complete outcome data (46.0%) achieved cure at a median follow-up of 29 months (IQR 26–31).

Predictor selection and nomogram development

Univariable analysis in the derivation cohort identified older age, higher BMI, larger waist circumference, longer hypertension duration, more antihypertensive drug classes, higher highest recorded SBP, diabetes mellitus, cerebro-cardiovascular disease history, higher serum creatinine, lower eGFR, and lower HDL-C as factors associated with hypertension non-cure (all $P < 0.05$; Table 2). Multivariable logistic regression with forward stepwise selection retained age, waist circumference, highest recorded SBP, and CCD history as independent predictors of non-cure. Hypertension duration was added on the basis of established clinical relevance, yielding a five-variable nomogram (Fig. 1). Older age, larger waist circumference, higher highest SBP, longer hypertension duration, and presence of CCD were each associated with a lower predicted probability of cure.

Table 2

Univariable and multivariable logistic regression analyses for predictors of hypertension non-cure in the derivation cohort

Variable	Univariable OR (95% CI)	P value	Multivariable OR (95% CI)⊠	P value
Age (per 10 years)	3.19 (1.95–5.23)	< 0.001	2.87 (1.33–6.14)	0.007
Male sex	1.98 (0.93–4.24)	0.079	—	—
BMI (per 1 kg/m ²)	1.17 (1.05–1.31)	0.004	—	—
Waist circumference (per 5 cm)	1.48 (1.20–1.81)	< 0.001	1.42 (1.07–1.90)	0.017
Hypertension duration (per 12 months)	1.22 (1.11–1.34)	< 0.001	— ⊠	— ⊠
Antihypertensive drug classes (per 1)	1.87 (1.24–2.83)	0.003	—	—
Highest SBP (per 10 mmHg)	1.41 (1.17–1.69)	< 0.001	1.38 (1.08–1.79)	0.010
Diabetes mellitus (yes vs no)	4.77 (1.29–17.67)	0.019	—	—
Cerebro-cardiovascular disease (yes vs no)	28.21 (3.64–218.50)	0.001	16.06 (1.31–196.61)	0.030
Serum creatinine (per 1 μmol/L)	1.03 (1.01–1.05)	0.009	—	—
eGFR (per 1 mL/min/1.73 m ²)	0.96 (0.94–0.98)	< 0.001	—	—
HDL-C (per 1 mmol/L)	0.15 (0.04–0.67)	0.013	—	—
Triglycerides (per 1 mmol/L)	1.45 (0.86–2.45)	0.166	—	—
OR, odds ratio; CI, confidence interval; BMI, body mass index; SBP, systolic blood pressure; eGFR, estimated glomerular filtration rate; HDL-C, high-density lipoprotein cholesterol. ⊠ Multivariable model with forward stepwise selection (entry $\alpha = 0.05$, removal $\alpha = 0.10$). ⊠ Added to the five-variable nomogram on the basis of established clinical relevance (not retained by stepwise selection).				

Table 3
Performance of prediction models in the derivation and temporal validation cohorts

Model	Cohort	C-index (95% CI)	Calibration-in-the-large	Calibration slope
ARS	Derivation	0.743 (0.652–0.833)	—	—
Five-variable nomogram	Derivation (apparent)	0.907 (0.855–0.959)	0.0%	1.00
Five-variable nomogram	Derivation (bootstrap)§	0.907 (0.848–0.953)	0.0%	0.952
Five-variable nomogram	Validation (n = 63)	0.734 (0.595–0.857)	–0.5%	0.32
Parsimonious 4-variable model	Validation (n = 63)	0.812 (0.692–0.916)	–0.5%	0.60
Recalibrated 4-variable model	Validation (n = 63)	0.812 (0.692–0.916)	0.0%§	1.00§
ARS	Validation (n = 63)	0.638 (0.489–0.776)	—	—
Recalibrated 4-variable model	Validation (n = 75)§	0.780 (0.658–0.883)	0.0%	0.88
ARS, Aldosteronoma Resolution Score; CI, confidence interval. Calibration-in-the-large is expressed as the difference between mean predicted and observed cure probability. § Bootstrap-corrected. § By construction of logistic recalibration. § Sensitivity analysis with 12 missing outcomes assigned to the modal category (non-cure).				

The five-variable nomogram showed excellent discrimination in the derivation cohort (apparent C-index 0.907, 95% confidence interval [CI] 0.855–0.959), with bootstrap-corrected calibration (1,000 resamples) showing no apparent miscalibration on visual inspection. The nomogram substantially outperformed the ARS in the derivation cohort (C-index 0.743, 95% CI 0.652–0.833; $P < 0.001$ for the difference).

Temporal external validation

When applied without refitting to the 2023 validation cohort (n = 63), the five-variable nomogram showed moderate discrimination (C-index 0.734, 95% CI 0.595–0.857). Calibration was suboptimal: the calibration slope was 0.32 (indicating overfitting of the original model), although calibration-in-the-large was acceptable (mean predicted cure probability 45.5% vs observed 46.0%). Flexible calibration curves

showed systematic over-prediction in the higher-risk range (Fig. 3A). Of note, hypertension duration was not discriminative in the validation cohort ($P = 0.354$), prompting parsimony reduction.

Model updating

Removing hypertension duration yielded a parsimonious four-variable model (age, waist circumference, highest SBP, CCD) with improved discrimination in the validation cohort (C-index 0.812, 95% CI 0.692–0.916). Calibration-in-the-large remained accurate (predicted 45.5% vs observed 46.0%), and the calibration slope improved to 0.60. Logistic recalibration—re-estimating the intercept and slope from the validation data—further improved calibration (calibration slope 1.00 by construction; Fig. 3B). The recalibrated parsimonious model maintained a C-index of 0.812 (95% CI 0.692–0.916) and outperformed both the original five-variable nomogram ($P = 0.046$) and the ARS (C-index 0.638, 95% CI 0.489–0.776; $P = 0.011$) in the validation cohort (Fig. 2).

Decision curve analysis demonstrated that the parsimonious four-variable model and its recalibrated version provided higher net benefit than the original five-variable nomogram, the ARS, and the treat-all strategy across threshold probabilities of approximately 0.20–0.80 (Fig. 4). Below 0.20 and above 0.80, all strategies converged, as expected.

Sensitivity analysis

In the sensitivity analysis that included all 75 validation patients (with 12 missing outcomes assigned to non-cure), the C-index of the recalibrated parsimonious model was 0.780 (95% CI 0.658–0.883), and the calibration slope was 0.88. The direction and magnitude of all comparisons were unchanged (Table 4), supporting the robustness of the updated model.

Table 4

Sensitivity analysis: model performance in the temporal validation cohort with all 75 patients included

Model	C-index (95% CI)	Calibration slope	Net benefit at PT = 0.40
Five-variable nomogram	0.701 (0.562–0.832)	0.28	0.082
Parsimonious 4-variable model	0.780 (0.658–0.883)	0.53	0.142
Recalibrated 4-variable model	0.780 (0.658–0.883)	0.88	0.129
ARS	0.612 (0.461–0.755)	—	0.045
Treat-all	—	—	-0.022
Twelve patients with missing postoperative outcomes were assigned to the modal category (non-cure). ARS, Aldosteronoma Resolution Score; CI, confidence interval; PT, threshold probability. Net benefit is reported at a representative threshold probability of 0.40.			

Table 5

Comparison of the proposed recalibrated four-variable model with existing prediction tools for hypertension cure after adrenalectomy

Tool	Variables	Population	C-index in our validation cohort	Clinical utility (DCA)
ARS	BMI, sex, drug classes, hypertension duration	Western	0.638	Low
Utsumi nomogram	Age, BMI, duration, drug classes, aldosterone	Japanese	0.671	Low
PASO-based score	PASO criteria variables	International	0.659	Low
Chinese nomogram	Age, BMI, duration, drug classes, K+	Chinese	0.702	Moderate
Proposed 4-variable model	Age, waist circumference, highest SBP, CCD	Chinese	0.812	High
ARS, Aldosteronoma Resolution Score; BMI, body mass index; SBP, systolic blood pressure; CCD, cerebro-cardiovascular disease; PASO, Primary Aldosteronism Surgical Outcome; DCA, decision curve analysis. C-index values for external tools were computed by applying each published model to our validation cohort (n = 63).				

Discussion

In this study, we developed a five-variable nomogram for predicting hypertension cure after unilateral adrenalectomy in Chinese patients with PA, validated it in an independent temporal cohort, and updated

it into a parsimonious, recalibrated four-variable model that demonstrated superior transportability. The principal findings are threefold. First, the derivation-stage nomogram showed excellent apparent performance (C-index 0.907) but, as is typical for prediction models, this was optimistic: temporal external validation reduced discrimination to a C-index of 0.734 and revealed marked overfitting (calibration slope 0.32). Second, parsimony reduction—removing hypertension duration, which was not discriminative in the validation cohort—improved both discrimination and calibration, yielding a four-variable model with a C-index of 0.812. Third, logistic recalibration further restored calibration without sacrificing discrimination, and the updated model outperformed the ARS and provided higher net benefit across clinically relevant threshold probabilities.

The discrepancy between apparent and external performance underscores a recurring theme in prediction-model research: models developed in a single cohort almost invariably shrink when confronted with new patients¹⁴. Our temporal validation captured two sources of this gap. The first is statistical optimism, addressed by bootstrap correction in derivation but only truly tested by external data. The second is case-mix and temporal drift: in our cohorts, however, case mix was largely comparable—the prevalence of cerebro-cardiovascular disease was nearly identical (20.0% vs 20.5%)—although the validation cohort was older and had a longer hypertension duration. The drop in calibration slope is therefore attributable mainly to statistical optimism in model development rather than to temporal drift. Notably, the strong association of CCD with non-cure in the derivation cohort rested on very few cured patients with CCD (1 of 51), yielding an unstable estimate (OR 16.06, 95% CI 1.31–196.61); its attenuation in the validation cohort is therefore expected. These factors reinforce the importance of routine model updating rather than relying on a fixed derivation-stage tool.

The finding that hypertension duration—a component of the ARS and several prior nomograms—was not discriminative in the validation cohort merits discussion. Hypertension duration is biologically plausible as a predictor, reflecting cumulative vascular remodeling and end-organ damage. Its non-significance in our temporal cohort may reflect measurement imprecision (recall bias in self-reported duration), the narrower distribution of durations in the 2023 patients, or genuine changes in disease phenotyping over time. From a modeling perspective, this illustrates the value of parsimony reduction: a variable retained solely on the basis of clinical relevance may add noise rather than signal in a new population, and pruning such variables can improve both performance and usability.

The four predictors retained in the final model—age, waist circumference, highest recorded SBP, and CCD history—are all routinely available at the preoperative clinic visit and require no specialized testing. This enhances the model's practical applicability: clinicians can compute an individualized probability of cure at the bedside, supporting shared decision-making. The retained variables are also biologically coherent. Older age and higher SBP reflect longer-standing and more severe hypertension with greater vascular remodeling; larger waist circumference indexes metabolic burden and obesity-related hypertension, which may persist after removal of the aldosterone source; and CCD history captures established end-organ damage that is unlikely to reverse fully after surgery. The convergence of statistical selection and biological plausibility strengthens confidence in the model's construct validity.

Compared with the ARS, which had a C-index of 0.638 in our validation cohort, the recalibrated four-variable model offered a clinically meaningful improvement in discrimination and substantially higher net benefit on decision curve analysis. The ARS was developed in a Western population and incorporates BMI, sex, antihypertensive drug classes, and hypertension duration; its inconsistent performance in Asian cohorts has been noted previously. Our model replaces BMI with waist circumference—a more direct measure of central adiposity—and adds highest SBP and CCD history, both of which capture disease severity beyond what the ARS components provide. These additions likely explain its superior performance in our population.

Several limitations should be acknowledged. First, this was a single-center retrospective study, and the model has been validated only temporally rather than geographically; multi-center external validation in diverse Chinese populations is needed before broader adoption. Second, the validation cohort had 12 patients (16%) with missing outcome data, addressed by a conservative sensitivity analysis; while results were consistent, prospective validation with complete follow-up would be preferable. Third, AVS was not uniformly performed in all patients with bilateral or equivocal CT findings (performed in 29 of 75 validation patients, 38.7%), which may have introduced subtype misclassification in a minority of cases. Fourth, the CCD coefficient was estimated from sparse data (only one cured patient had CCD), suggesting quasi-separation; penalized (Firth) logistic regression may be considered in future refinements. Fifth, we did not incorporate genetic or histopathological variables (such as KCNJ5 mutation status or aldosterone synthase expression), which have emerged as potential predictors and could refine future iterations of the model. Finally, the model predicts hypertension cure dichotomized at the PASO complete clinical remission threshold; patients with partial remission also derive substantial benefit from surgery, and the model should not be used to withhold adrenalectomy in patients predicted as non-cure.

In conclusion, temporal external validation substantially tempered the apparent performance of our derivation-stage nomogram for predicting hypertension cure after unilateral adrenalectomy in PA. A parsimonious, recalibrated four-variable model using routinely available preoperative variables—age, waist circumference, highest recorded SBP, and CCD history—provided the best transportability and is proposed for individualized preoperative counseling in Chinese patients with unilateral PA. Prospective multi-center validation is warranted to confirm its generalizability and to support integration into routine clinical practice.

Conclusions

Temporal external validation substantially tempered the apparent performance of the derivation-stage nomogram. A parsimonious, recalibrated four-variable model using routinely available preoperative variables—age, waist circumference, highest recorded systolic blood pressure, and cerebro-cardiovascular disease history—provided the best transportability and is proposed for individualized preoperative counseling in Chinese patients with unilateral PA. Prospective multi-center validation is warranted to confirm generalizability before routine clinical adoption.

Declarations

Ethics approval and consent to participate

This study was conducted with the approval of the Peking University People's Hospital Ethical Review Board (approval number: 2021PHB342-001). Written informed consent from individual patients was waived owing to the retrospective design and the use of de-identified data.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was partially supported by the Michigan Medicine and Peking University Health Science Center Joint Institute (JI) for Translational and Clinical Research. The funding body had no role in the design of the study; in the collection, analysis, or interpretation of data; or in the writing of the manuscript.

Author Contribution

YH conceived and designed the study, collected and analyzed the data, and drafted the manuscript. ML participated in data collection and clinical interpretation. YD participated in data interpretation and provided critical revision of the manuscript for important intellectual content. ZH participated in data collection and quality control. NZ supervised the study, contributed to study design and clinical interpretation, and provided critical revision of the manuscript. TX conceived and supervised the study, secured funding, contributed to study design and clinical interpretation, and provided critical revision of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors thank the clinical and nursing staff of the Department of Urology, Peking University People's Hospital, for their assistance with patient follow-up, and the Department of Endocrinology for biochemical data management.

Availability of data and materials

The datasets analyzed during the current study are available from the corresponding authors on reasonable request.

References

1. Vilela LAP, Almeida MQ. Diagnosis and management of primary aldosteronism. *Arch Endocrinol Metab.* 2017;61:305–12. 10.1590/2359-3997000000274.
2. Mullen N, Curneen J, Donlon PT, Prakash P, Bancos I, Gurnell M, Dennedy MC. Treating primary aldosteronism-induced hypertension: novel approaches and future outlooks. *Endocr Rev.* 2024;45:125–70.
3. Nanba K, Rainey WE. Pathophysiology of bilateral hyperaldosteronism. *Curr Opin Endocrinol Diabetes Obes.* 2022;29:233–42. 10.1097/med.0000000000000729.
4. Brown JM. Adverse Effects of Aldosterone: Beyond Blood Pressure. *J Am Heart Assoc.* 2024;13:e030142. 10.1161/jaha.123.030142.
5. Morita R, Azushima K, Sunohara S, Haze T, Kobayashi R, Kinguchi S, Kanaoka T, Kobayashi K, Toya Y, Wakui H, Tamura K. High plasma aldosterone concentration is associated with worse 24-h ambulatory blood pressure profile in patients with primary aldosteronism. *Hypertens Res.* 2023;46:1995–2004. 10.1038/s41440-023-01325-8.
6. Matthey Giè ML, St-Amour P, Karenovics W, Matter M, Demarchi M, Ponte B, Mavromati M, Sadowski SM, Triponez F. Biochemical and clinical remission after adrenalectomy in patients with primary aldosteronism: a retrospective two-center study. *Gland Surg.* 2026;15:137. 10.21037/gs-2026-1-0087.
7. Kim RM, Lee J, Soh EY. Predictors of resolution of hypertension after adrenalectomy in patients with aldosterone-producing adenoma. *J Korean Med Sci.* 2010;25:1041–4. 10.3346/jkms.2010.25.7.1041.
8. Zarnegar R, Young WF Jr., Lee J, Sweet MP, Kebebew E, Farley DR, Thompson GB, Grant CS, Clark OH, Duh QY. The aldosteronoma resolution score: predicting complete resolution of hypertension after adrenalectomy for aldosteronoma. *Ann Surg.* 2008;247:511–8. 10.1097/SLA.0b013e318165c075.
9. Huang Y, Huang Q, Xie Y, Ouyang C, Chen Y, Huang C. The aldosteronoma resolution score predicts the prognosis of hypertension under two surgical procedures. *Front Endocrinol (Lausanne).* 2025;16:1651381. 10.3389/fendo.2025.1651381.
10. Vorselaars W, van Beek DJ, Postma EL, Spiering W, Rinkes B, Valk IHM, G.D., and, Vriens MR. Validation of the Aldosteronoma Resolution Score Within Current Clinical Practice. *World J Surg.* 2019;43:2459–68. 10.1007/s00268-019-05074-z.
11. Aronova A, Gordon BL, Finnerty BM, Zarnegar R, Fahey TJ 3rd. Aldosteronoma resolution score predicts long-term resolution of hypertension. *Surgery.* 2014;156:1387–92.

10.1016/j.surg.2014.08.019. discussion 1392 – 1383.

12. Abdelsalam A, Eddama MMR, Rozalen-Garcia V, Chung TT, Srirangalingam U, Conway GS, Hurel S, Khoo B, Baldeweg SE, Kurzawinski TR, Abdel-Aziz TE. (2026). External validation of the PASO prediction score for primary aldosteronism. *Endocr Connect* 15. 10.1530/ec-26-0255
13. Schnellinger EM, Yang W, Kimmel SE. Comparison of dynamic updating strategies for clinical prediction models. *Diagn Progn Res.* 2021;5:20. 10.1186/s41512-021-00110-w.
14. Schinkel M, Bennis FC, Boerman AW, Wiersinga WJ, Nanayakkara PWB. Embracing cohort heterogeneity in clinical machine learning development: a step toward generalizable models. *Sci Rep.* 2023;13:8363. 10.1038/s41598-023-35557-y.

Figures

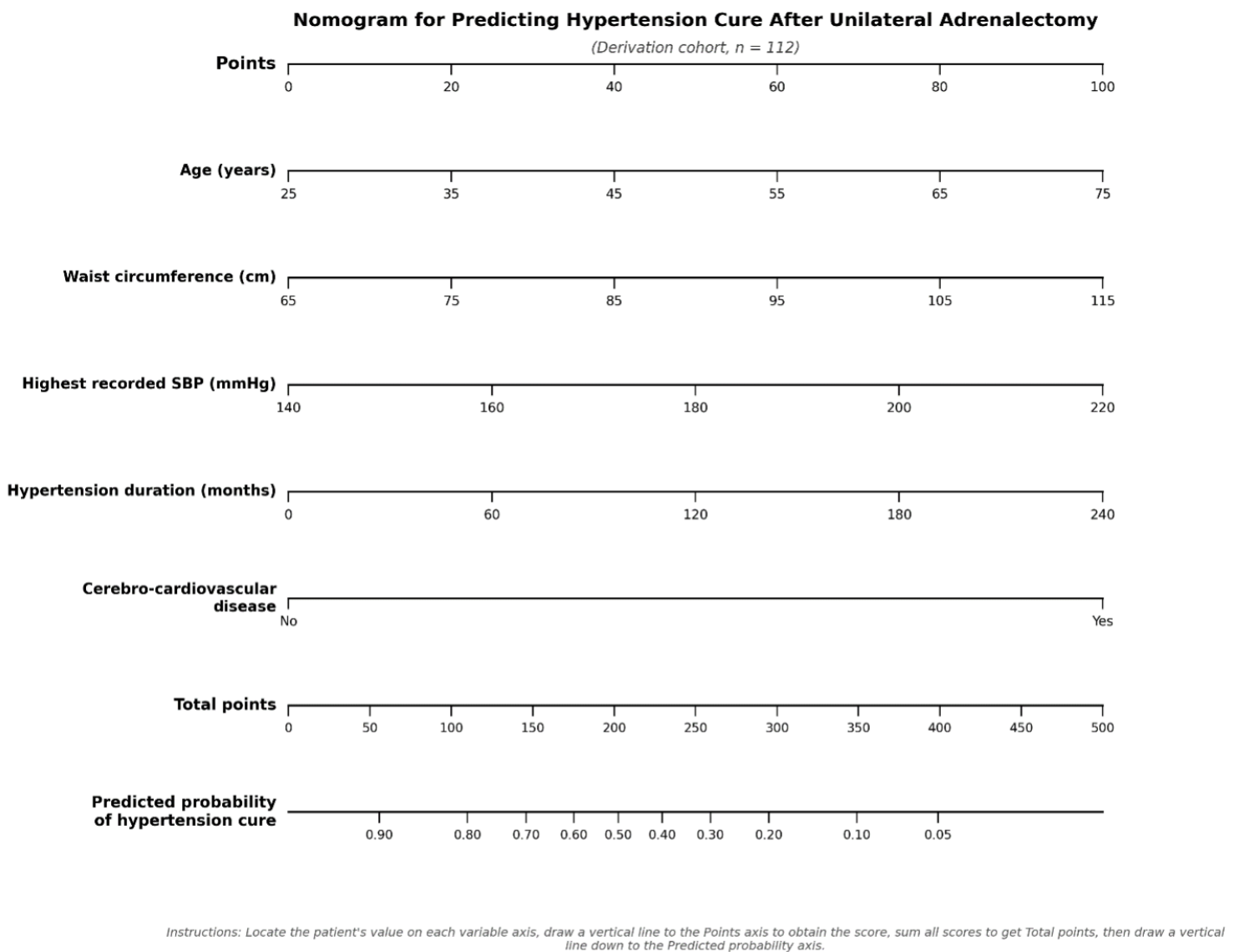


Figure 1

Nomogram for predicting hypertension cure after unilateral adrenalectomy in primary aldosteronism (derivation cohort). Each variable (age, waist circumference, highest recorded systolic blood pressure, hypertension duration, and cerebro-cardiovascular disease history) is scored on the points axis; the total score maps to the predicted probability of cure. To use the nomogram, locate the patient’s value on each variable axis, draw a vertical line to the points axis to obtain the corresponding score, sum the scores across all variables, locate the total on the total points axis, and draw a vertical line down to the probability axis to read the predicted probability of hypertension cure.

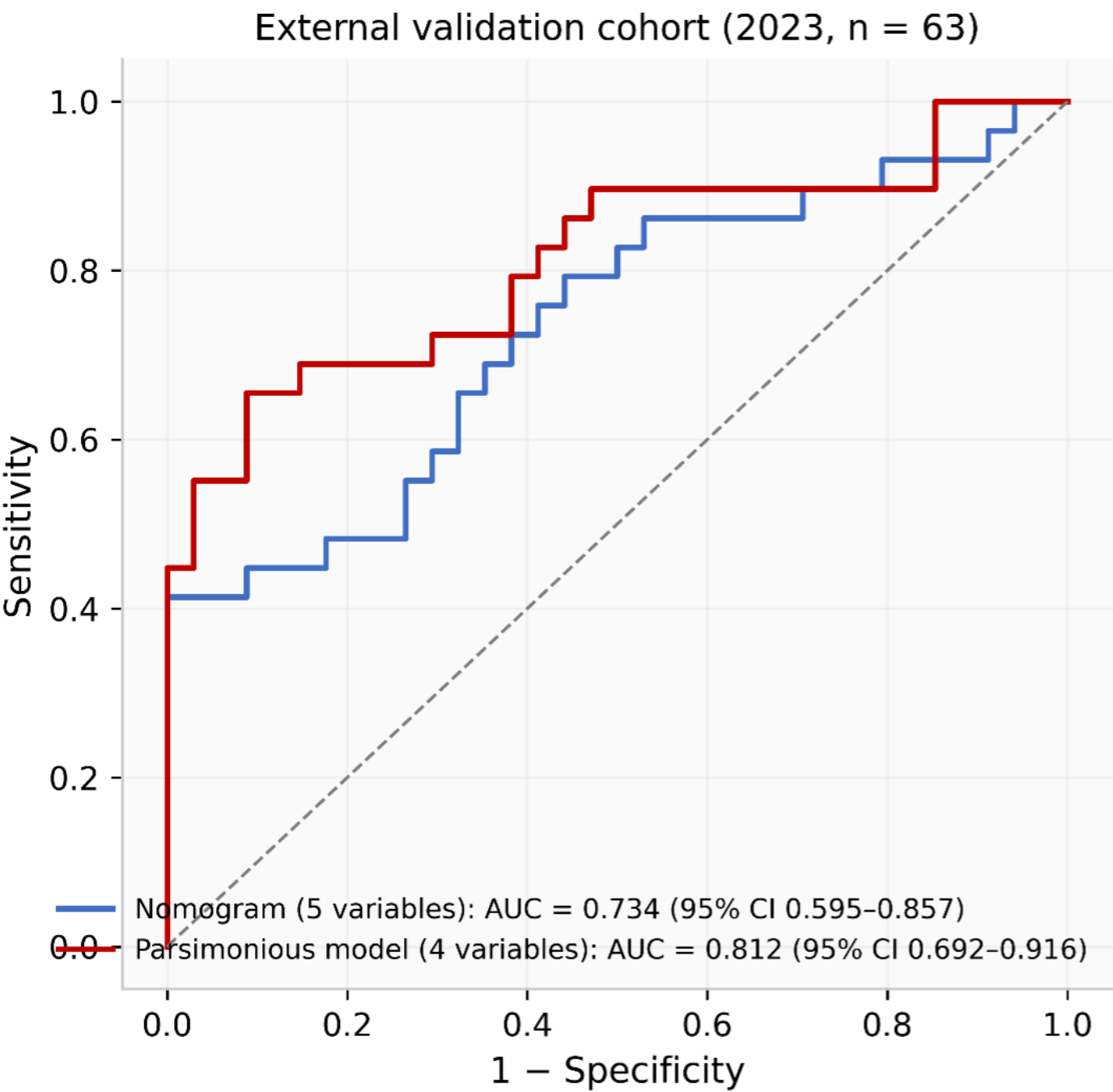


Figure 2

Receiver operating characteristic curves in the temporal validation cohort (2023; n = 63). The parsimonious four-variable model (red; C-index 0.812, 95% CI 0.692–0.916) discriminated better than the original five-variable nomogram (blue; C-index 0.734, 95% CI 0.595–0.857) and the Aldosteronoma Resolution Score (green; C-index 0.638, 95% CI 0.489–0.776). The diagonal dashed line represents a non-informative reference (C-index 0.5).

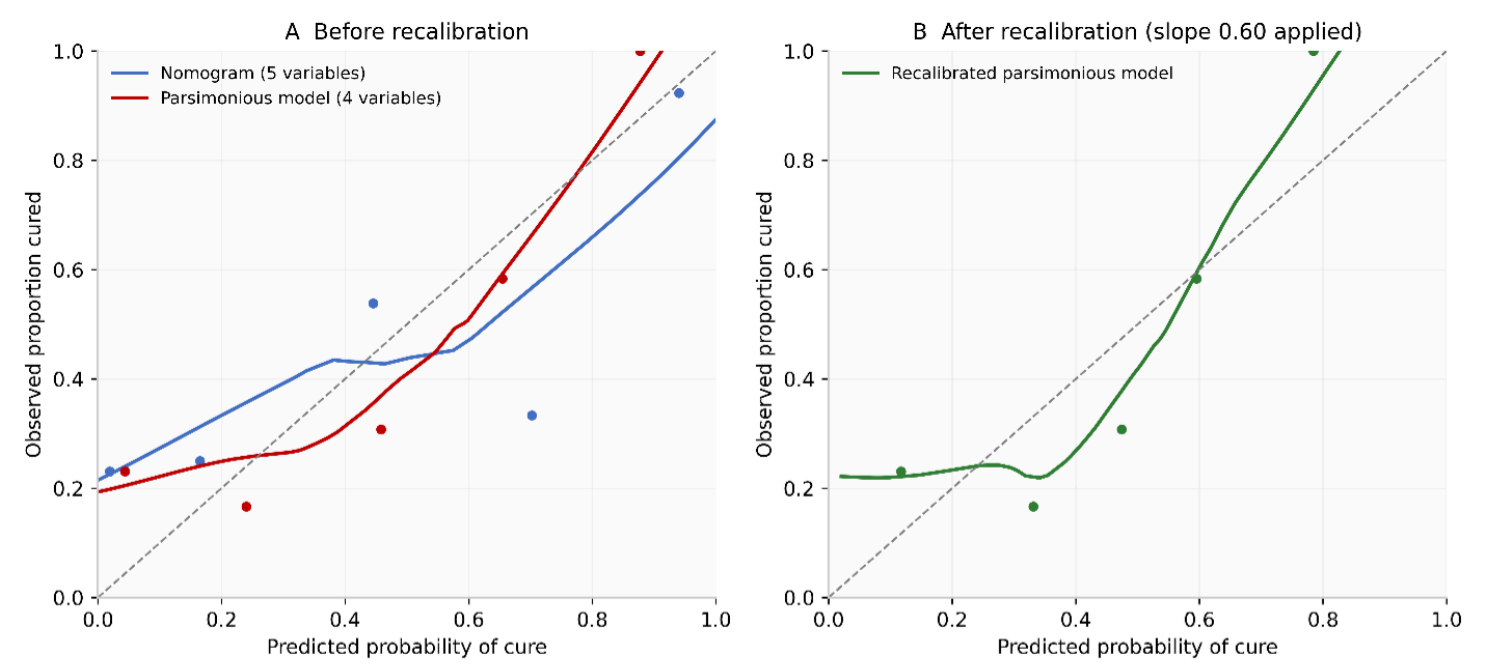


Figure 3

Calibration of prediction models in the temporal validation cohort. (A) Before recalibration: flexible (loess) calibration curves and quintile-based observed proportions for the five-variable nomogram (blue) and the parsimonious four-variable model (red). Both models show some departure from the ideal diagonal, with the five-variable nomogram exhibiting greater over-prediction in the higher-risk range. (B) The recalibrated parsimonious model (green) shows improved alignment with the ideal diagonal. The dashed line indicates perfect calibration. Circles represent observed cure proportions within quintiles of predicted probability; shaded bands represent 95% confidence intervals from 1,000 bootstrap resamples.

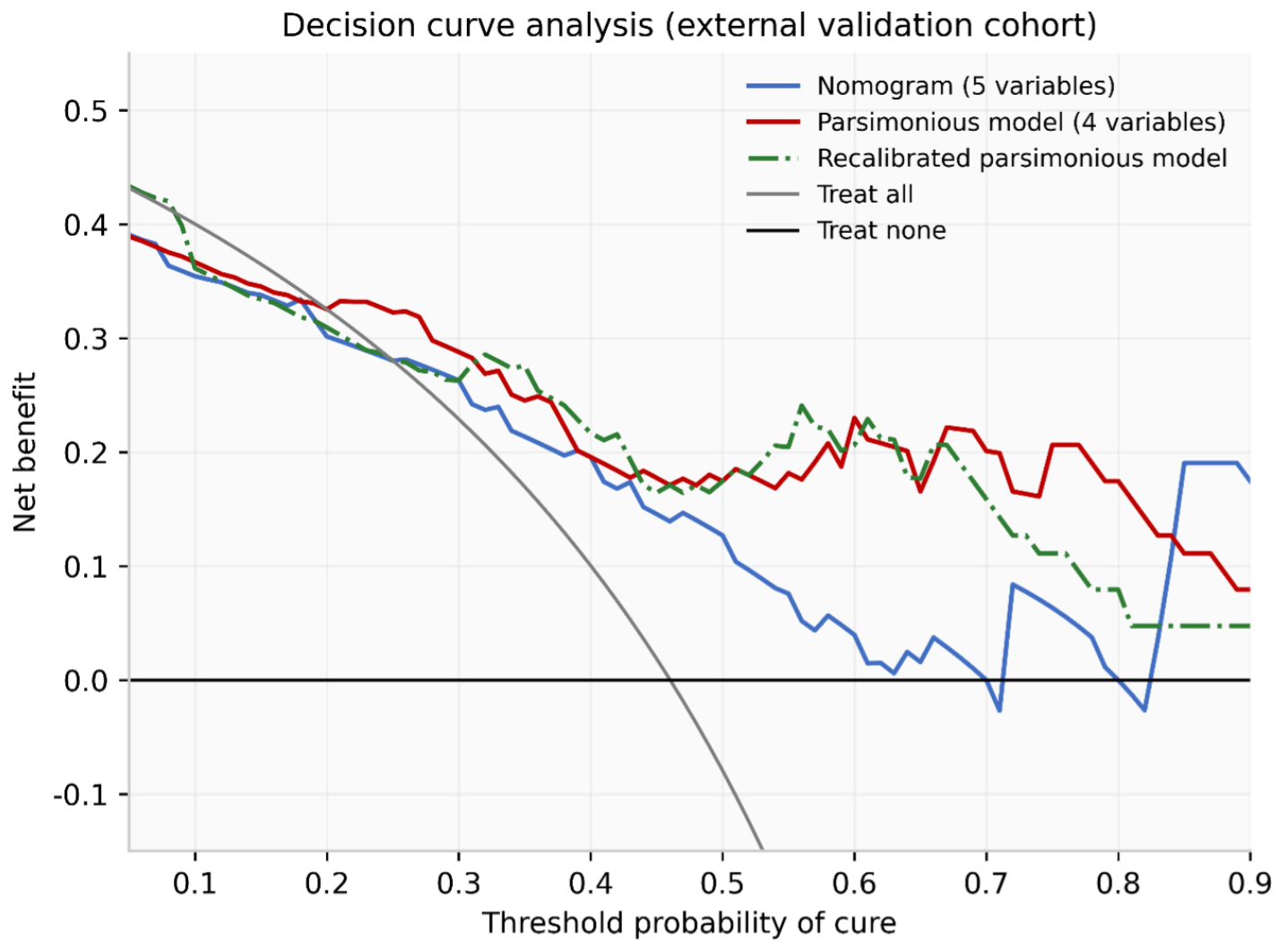


Figure 4

Decision curve analysis in the temporal validation cohort. Net benefit is plotted against threshold probability of cure for the five-variable nomogram (blue), the parsimonious four-variable model (red), the recalibrated parsimonious model (green), and the treat-all (grey) and treat-none (black) strategies. The recalibrated parsimonious model provides the highest net benefit across threshold probabilities of approximately 0.20–0.80. Below 0.20 and above 0.80, all prediction-based strategies converge with the default strategies, as expected.