

External Validation of Clinical Prediction Models in Unilateral Primary Aldosteronism

Davis Sam,^{1,○} Gregory A. Kline,² Benny So,^{3,○} Gregory L. Hundemer,⁴ Janice L. Pasieka,⁵ Adrian Harvey,⁵ Alex Chin,^{6,7} Stefan J. Przybojewski,³ Cori E. Caughlin,³ and Alexander A. Leung^{2,8,○}

BACKGROUND

Targeted treatment of primary aldosteronism (PA) is informed by adrenal vein sampling (AVS), which remains limited to specialized centers. Clinical prediction models have been developed to help select patients who would most likely benefit from AVS. Our aim was to assess the performance of these models for PA subtyping.

METHODS

This external validation study evaluated consecutive patients referred for PA who underwent AVS at a tertiary care referral center in Alberta, Canada during 2006–2018. In alignment with the original study designs and intended uses of the clinical prediction models, the primary outcome was the presence of lateralization on AVS. Model discrimination was evaluated using the C-statistic. Model calibration was assessed by comparing the observed vs. predicted probability of lateralization in the external validation cohort.

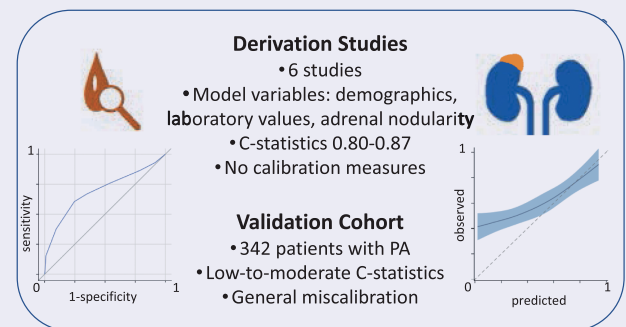
RESULTS

The validation cohort included 342 PA patients who underwent AVS (mean age, 52.1 years [SD, 11.5]; 201 [58.8%] male; 186 [54.4%] with lateralization). Six published models were assessed. All models demonstrated low-to-moderate discrimination in the validation set (C-statistics; range, 0.60–0.72), representing a marked decrease compared with the derivation sets (range, 0.80–0.87). Comparison of observed and predicted probabilities of unilateral PA revealed significant miscalibration. Calibration-in-the-large for every model was >0 (range, 0.35–1.67), signifying systematic underprediction of lateralizing disease. Calibration slopes were consistently <1 (range, 0.35–0.87), indicating poor performance at the extremes of risk.

CONCLUSIONS

Overall, clinical prediction models did not accurately predict AVS lateralization in this large cohort. These models cannot be reliably used to inform the decision to pursue AVS for most patients.

GRAPHICAL ABSTRACT



Conclusion:
Prediction models do not accurately predict AVS lateralization

Keywords: adrenal vein sampling; blood pressure; decision rules; hypertension; prediction models; primary aldosteronism.

<https://doi.org/10.1093/ajh/hpab195>

Correspondence: Alexander A. Leung (aacleung@ucalgary.ca).

Initially submitted November 15, 2021; date of first revision December 3, 2021; accepted for publication December 22, 2021; online publication December 27, 2021.

¹Department of Medicine, University of British Columbia, Vancouver, British Columbia, Canada; ²Department of Medicine, University of Calgary, Calgary, Alberta, Canada; ³Department of Radiology, University of Calgary, Calgary, Alberta, Canada; ⁴Department of Medicine, University of Ottawa, Ottawa, Ontario, Canada; ⁵Department of Surgery, University of Calgary, Calgary, Alberta, Canada; ⁶Alberta Precision Laboratories, Alberta Health Services, Calgary, Alberta, Canada; ⁷Department of Pathology and Laboratory Medicine, University of Calgary, Calgary, Alberta, Canada; ⁸Department of Community Health Sciences, University of Calgary, Calgary, Alberta, Canada.

© The Author(s) 2021. Published by Oxford University Press on behalf of American Journal of Hypertension, Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Primary aldosteronism (PA) is a common cause of hypertension, affecting at least 10% of all people with high blood pressure.¹ Distinguishing between unilateral vs. bilateral disease helps to guide treatment. Specifically, adrenalectomy is preferred for patients with unilateral disease, whereas mineralocorticoid receptor antagonist therapy is recommended for patients with bilateral disease.²⁻⁵ Subtyping is one of the most challenging diagnostic steps, owing to the unreliability of diagnostic imaging and limited availability of adrenal vein sampling (AVS),^{2,6-10} creating a large care gap.¹¹

Consequently, multiple clinical prediction models have been proposed to simplify the workup by helping to prioritize patients most likely to benefit from AVS.¹²⁻¹⁹ These frequently incorporate demographic, laboratory, and/or imaging characteristics to stratify patients based on the probability of having lateralizing disease. However, their performance in clinical practice remains unclear, as previous evaluations of these models have generally been limited by nonstandardized reporting and the absence of key measures of discrimination or calibration.^{14,15,18-22}

As performance is commonly overestimated at the development stage, external validation of a prediction model is requisite for determining its generalizability before it can be incorporated into clinical practice.²³ Addressing this, we sought to externally validate the performance of multiple published clinical prediction models for subtyping in a large and diverse cohort of patients with PA, using AVS as a reference standard, exactly as the models intended. Specifically, we evaluated discrimination and calibration to determine whether each model provided discriminate and reliable AVS lateralization predictions in an independent dataset.

METHODS

This study was reported according to the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement²⁴ and was approved by the University of Calgary Conjoint Health Research Ethics Board.

Validation study population

Data were prospectively collected. We identified all consecutive patients who underwent AVS in the workup of PA from January 2006 to May 2018 using the regional AVS database in Calgary, AB, Canada. This database has been used in numerous studies evaluating PA workup and contains detailed diagnostic information on all AVS performed in Southern Alberta.^{9,25,26} The catchment includes Alberta, the interior of British Columbia, and Saskatchewan, representing an ethnically and sociodemographically diverse population in western Canada.

Patients were included if they had a diagnosis of PA (based on an elevated aldosterone–renin ratio [ARR] >550 pmol/l per ng/ml/h or >60 pmol/l per mIU/l and high-probability features of PA [resistant hypertension, adrenal mass, spontaneous hypokalemia, and/or diuretic-induced hypokalemia]) and successful AVS, as defined by cannulation of both adrenal veins.^{25,27} Prior to AVS, potassium-sparing diuretics were stopped for at least 4 weeks and other interfering

medications were discontinued for 2 weeks; alpha blockers, calcium channel blockers, and/or hydralazine were instead used for blood pressure control.²

Outcome

The study outcome was the presence (or absence) of unilateral PA as defined by AVS based on a standardized protocol and interpretation criteria. Lateralization with AVS was selected to facilitate comparisons with other studies and to align with the original intention of these clinical prediction models. Post-treatment outcomes, such as biochemical or clinical response to adrenalectomy,²⁸ were not considered because these were not relevant for the intended use of the AVS prediction models being studied; rather, post-treatment outcomes would only be useful for assessing whether models designed to predict postoperative success are effective at prioritizing patients for surgical referral (which is entirely different than and outside the scope of the present study). We have previously published that our AVS interpretation criteria and subsequent surgical decisions are associated with excellent outcomes.²⁶

A selectivity index of $\geq 2:1$ at baseline or $\geq 3:1$ following cosyntropin stimulation (250 μ g cosyntropin bolus followed by 6.25 μ g infusion over 15 minutes) indicated successful cannulation during AVS.²⁶ The presence of unilateral aldosterone hypersecretion was based on a lateralization index of $>3:1$, comparing the dominant to the nondominant adrenal gland, using unstimulated aldosterone–cortisol ratios.^{9,26}

Clinical prediction models

We defined a “clinical prediction model” as a model that combined 2 or more independent variables to obtain an estimate of the probability of having unilateral or bilateral subtypes for PA in accordance with our institutional AVS interpretation practice, as above. A sensitivity analysis was also performed where we applied the same AVS selectivity index and lateralization index thresholds used in the original studies for each model to define successful AVS and lateralization, respectively, and then reassessed model performance.

We searched MEDLINE from inception to November 2020 using medical subject headings (“hyperaldosteronism” and “clinical decision rules”) and author-supplied keywords (“prediction score,” “prediction model,” “subtype prediction”). Reference lists of included articles were also hand-searched to identify other relevant studies. Studies were considered for inclusion if they reported model derivation using original data. Those that incorporated variables that were not routinely collected at our institution (e.g., ARR following captopril challenge) were excluded, as these models could not be externally validated with our data. We identified 6 models for inclusion.^{12-14,16-18} We additionally assessed a prediction rule adopted by several societies suggesting that the combination of age under 35 years, hypokalemia, plasma aldosterone concentration (PAC) >30 ng/dl, and a unilateral adrenal adenoma >1.0 cm is specific for unilateral disease.^{2,29}

Predictor variables

Based on the clinical prediction models identified, information on potential predictors of lateralization was retrieved from clinical records. These included age (at the time of AVS), sex, and baseline measures of serum potassium, renal function, PAC, and ARR. The lowest serum potassium on record was used to define hypokalemia. PAC was measured by either solid-phase radioimmunoassay (Siemens Coat-A-Count Aldosterone assay; Siemens Healthcare Diagnostics, Tarrytown, NY) or chemiluminescent immunoassay (Diasorin Liaison XL platform; Diasorin, Mississauga, ON). Both assays performed similarly across the range of measurements.²⁷ Renin was variably reported as plasma renin activity (PRA), using the GammaCoat PRA I¹²⁵ assay (Diasorin, Stillwater, MN), or as direct renin concentration (DRC), using the Diasorin Liaison XL platform. To facilitate comparisons, DRC was converted to PRA (DRC of 14.1 mIU/l = PRA of 1 ng/ml/h).²⁷ The highest ARR prior to AVS (in the absence of confounding medications) was used. Finally, computed tomography and magnetic resonance imaging reports were reviewed to determine the presence of adrenal nodules. Where appropriate, we restricted the analysis to patients without visible adrenal nodularity >1.0 cm to better match the population where the clinical prediction models were originally derived.^{16,17}

Statistical analysis

We evaluated model performance in predicting unilateral PA.^{23,30} For consistency, reverse coding of the outcome was used for models designed to predict bilateral disease.^{16,18} For each model, we restricted our analysis to individuals who had data for all the predictors of interest for complete case analysis. Discrimination was examined by calculating the C-statistic (where a value of 1.0 represents perfect discrimination and 0.5 reflects no discriminative ability). We considered a C-statistic <0.60 to indicate poor discrimination, 0.60–0.75 to represent moderate discrimination, and >0.75 to be acceptable discrimination.³¹ Calibration was assessed using locally estimated scatterplot smoothing plots, comparing observed and predicted probabilities of lateralization (where perfect calibration is represented as a 45° straight line with a slope of one and an intercept of zero).^{23,32} Whenever possible, the expected probability of having the unilateral subtype was calculated using the published intercept and regression coefficients from the original model,¹³ where the predicted probability was equal to $1/(1 + e^{-\text{linear predictor}})$. Otherwise, the predicted probability was based on published frequencies of the outcome in the derivation dataset for each stratum of risk,^{12,17,18} or was estimated from the reported sensitivities and specificities given for at least 2 strata.¹⁶ Calibration assessment was not possible for 1 model because none of the above were available.¹⁴ The calibration slope was calculated by fitting a logistic regression model using the linear predictor from the original model as an independent variable then estimating the corresponding regression coefficient.³³ The calibration intercept (“calibration-in-the-large”) was determined by fixing the calibration slope at one and estimating the corresponding model intercept.³³

Additionally, we recalibrated the widely cited Küpers model by adjusting the model intercept to our local dataset.^{13,34} The model intercept reflects the baseline risk of the outcome, which may be different between populations.³⁵ By updating the intercept, the model may perform better with improved calibration, provided that the original model was developed appropriately.^{34,36} Recalibration was possible here because it was the only full model (with regression coefficients and intercept) reported. The regression coefficients were fixed at their original values while a new intercept was estimated as a free parameter.³⁶ After updating the intercept, calibration was reassessed. *P* value <0.05 was used to indicate statistical significance. Statistical analyses were performed using SAS version 9.4 (SAS Institute, Cary, NC).

RESULTS

Study participants

In total, 354 patients underwent AVS for the investigation of PA during the study period. After excluding 12 patients where cannulation of both adrenal veins could not be confirmed, the validation cohort included 342 people (reflecting a 96.6% AVS success rate).

Overall, the mean age was 52.1 (SD, 11.5) years and 201 (58.8%) participants were male (Table 1). All patients had cross-sectional adrenal imaging; 318 (93%) received computed tomography and 24 (7%) magnetic resonance imaging. A discrete unilateral adrenal nodule measuring >1.0 cm was detected in 45% of patients. Hypokalemia (serum potassium <3.5 mmol/l) was present in 68.8% (218 of 317 known values). The median PAC was 522 (interquartile range, 457) pmol/l and median ARR was 2,930 (interquartile range, 3,632) pmol/l per ng/ml/h (i.e., 5.3 times the upper limit of normal). There were 186 (54.4%) people with lateralization on AVS.

Additional subgroups were defined to facilitate prediction rule comparisons, one that excluded patients with bilateral adrenal lesions (>1.0 cm)¹⁷ and another which only included patients with normal cross-sectional imaging (lesions absent or <1.0 cm),¹⁶ comprising 328 and 174 patients, respectively. The baseline characteristics of these subgroups were broadly similar to those of the overall validation cohort (Table 1). Rates of lateralization were also similar at 54.8% and 40.2%, respectively.

Prediction models

Six clinical prediction models were included, based out of Japan,^{16–18} France,¹³ Italy,¹² and the United Kingdom¹⁴ (Supplementary Table S1 online). Most of the original studies were small with typically <50 people with lateralizing disease.^{12–14,16} Rates of lateralization varied, ranging from 10.5%¹⁶ to 65.3%.¹⁴ The reported demographic characteristics in the derivation cohorts were generally similar to those in the validation cohort. Variables incorporated into the models included sex, serum potassium, renal function, PAC and/or ARR levels, and presence of an adrenal nodule (Supplementary Figure S1 online). In the derivation sets, discrimination was frequently reported to be excellent with

Table 1. Characteristics of the validation cohort

Characteristics*	Total AVS group			Bilateral lesions excluded group			Normal imaging group		
	Stratified by AVS result			Stratified by AVS result			Stratified by AVS result		
	All (n = 342)	Unilateral (n = 186)	Bilateral (n = 156)	All (n = 328)	Unilateral (n = 180)	Bilateral (n = 148)	All (n = 174)	Unilateral (n = 70)	Bilateral (n = 104)
Age, mean (SD), years [n = 342; n = 328; n = 174]	52.1 (11.5)	52.3 (11.3)	52.0 (11.8)	51.8 (11.5)	52.0 (11.3)	51.7 (11.8)	52.1 (11.9)	54.3 (10.3)	50.6 (12.7)
Age <35 years, no. (%)	29 (8.5)	13 (7.0)	16 (10.3)	29 (8.8)	13 (7.2)	16 (10.8)	17 (9.8)	2 (2.9)	15 (14.4)
Male sex, no. (%) [n = 342; n = 328; n = 174]	201 (58.8%)	119 (64.0%)	82 (52.6%)	189 (57.6%)	114 (63.3%)	75 (50.7%)	109 (62.6%)	52 (74.3%)	57 (54.8%)
Unilateral adrenal nodule >0.8 cm, no. (%) [n = 342; n = 328; n = 174]	173 (50.6%)	122 (65.6%)	51 (32.7%)	173 (52.7%)	122 (67.8%)	51 (34.5%)	23 (13.2%)	14 (20.0%)	9 (8.7%)
Unilateral adrenal nodule >1 cm, no. (%) [n = 342; n = 328; n = 174]	154 (45.0%)	110 (59.1%)	44 (28.2%)	154 (47.0%)	110 (61.1%)	44 (29.7%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Serum potassium, mean (SD), mmol/l [n = 317; n = 304; n = 159]	3.2 (0.6)	3.0 (0.6)	3.4 (0.6)	3.2 (0.6)	3.0 (0.6)	3.4 (0.6)	3.3 (0.5)	3.2 (0.5)	3.3 (0.5)
Serum potassium >3.9, no. (%) of known	33 (10.4%)	12 (6.7%)	21 (15.1%)	33 (10.9%)	12 (7.0%)	21 (15.9%)	19 (11.9%)	7 (10.4%)	12 (13.0%)
Serum potassium 3.5–3.9, no. (%) of known	66 (20.8%)	29 (16.3%)	37 (26.6%)	61 (20.1%)	26 (15.1%)	35 (26.5%)	47 (29.6%)	18 (26.9%)	29 (31.5%)
Serum potassium <3.5, no. (%) of known	218 (68.8%)	137 (77.0%)	81 (58.3%)	210 (69.1%)	134 (77.9%)	76 (57.6%)	93 (58.5%)	42 (62.7%)	51 (55.4%)
PAC, median (IQR), pmol/l [n = 276; n = 264; n = 139]	522 (457)	619 (555)	431 (329)	517 (457)	618 (562)	428 (321)	448 (346)	521 (435)	439 (297)
PAC <583 pmol/l (<210 pg/ml), no. (%) of known	156 (56.5%)	70 (46.1%)	86 (69.4%)	151 (57.2%)	68 (46.3%)	83 (70.9%)	95 (68.3%)	35 (61.4%)	60 (73.2%)
PAC >694 pmol/l (>25 ng/dl), no. (%) of known	91 (33.0%)	64 (42.1%)	27 (21.8%)	87 (33.0%)	62 (42.2%)	25 (21.4%)	31 (22.3%)	18 (31.6%)	13 (15.9%)
PAC >831 pmol/l (>30 ng/dl), no. (%) of known	62 (22.5)	46 (30.3)	16 (12.9)	60 (22.7)	45 (30.6)	15 (12.8)	17 (12.2)	10 (17.5)	7 (8.5)
ARR, median (IQR), pmol/l per ng/ml/h [n = 289; n = 276; n = 148]	2,930 (3,632)	3,473 (4,436)	2,300 (2,864)	2,906 (3,671)	3,457 (4,581)	2,300 (2,829)	2,617 (2,932)	2,822 (3,275)	2,380 (2,781)
ARR <1,526 pmol/l per ng/ml/h (<550 pg/ml per ng/ml/h), no. (%) of known	70 (24.2%)	37 (23.4%)	33 (25.2%)	68 (24.6%)	37 (24.2%)	31 (25.2%)	40 (27.0%)	16 (26.7%)	24 (27.3%)
ARR <1,720 pmol/l per ng/ml/h (<620 pg/ml per ng/ml/h), no. (%) of known	77 (26.6%)	38 (24.1%)	39 (29.8%)	75 (27.2%)	38 (24.8%)	37 (30.1%)	45 (30.4%)	17 (28.3%)	28 (31.8%)
eGFR, mean (SD), ml/min/1.73 m ² [n = 264; n = 252; n = 130]	83.7 (20.9)	84.0 (19.8)	83.1 (22.4)	83.7 (20.5)	84.4 (19.5)	82.7 (22.0)	83.2 (21.1)	80.2 (19.9)	85.7 (21.9)
eGFR >100, no. (%) of known	55 (20.8%)	34 (21.7%)	21 (19.6%)	52 (20.6%)	33 (21.9%)	19 (18.9%)	24 (18.5%)	7 (12.1%)	17 (23.6%)
eGFR 80–100, no. (%) of known	97 (36.7%)	60 (38.2%)	37 (34.6%)	93 (36.9%)	57 (37.7%)	36 (35.5%)	51 (39.2%)	23 (39.7%)	28 (38.9%)
eGFR <80, no. (%) of known	112 (42.4%)	63 (40.1%)	49 (45.8%)	107 (42.5%)	61 (40.4%)	46 (45.5%)	55 (42.3%)	28 (48.3%)	27 (37.5%)
AVS lateralization index, median (IQR) [n = 342; n = 328; n = 174]	3.6 (11.3)	11.8 (18.7)	1.5 (0.8)	3.6 (11.4)	11.8 (18.9)	1.5 (0.7)	2.1 (4.2)	8.4 (14.2)	1.5 (0.7)

Abbreviations: ARR, aldosterone–renin ratio; AVS, adrenal vein sampling; eGFR, estimated glomerular filtration rate; IQR, interquartile range; PAC, plasma aldosterone concentration. Number of patients with missing data (no.; % of total AVS group) included: serum potassium (25; 7.3%), PAC (66; 19.3%), ARR (53; 15.5%), and eGFR (78; 22.8%). Number of patients with missing data (no.; % of bilateral lesions excluded group) included: serum potassium (24; 7.3%), PAC (64; 19.5%), ARR (52; 15.9%), and eGFR (76; 23.2%). Number of patients with missing data (no.; % of normal imaging group) included: serum potassium (15; 8.6%), PAC (35; 20.1%), ARR (26; 14.9%), and eGFR (44; 25.3%).

*n shown for total AVS cohort, bilateral lesions excluded cohort, and normal imaging cohort, respectively.

C-statistics of 0.80–0.87.^{13,16,18} None of the studies provided measures of calibration.

Model performances

When externally validated, the rates of lateralization generally increased with each point assigned for the models that predicted the unilateral subtype^{12–14,17} (and the reverse for those designed to predict bilateral disease)^{16,18} (Table 2). In many cases, however, differences between strata were small and the range in the reported frequencies of lateralization was narrow. Correspondingly, the models had low-to-modest discriminative ability overall (Figure 1), with C-statistics between 0.60 (95% confidence intervals [CIs], 0.51, 0.69) and 0.72 (95% CI, 0.66, 0.78).

For the 5 models where calibration assessment was possible, plots were used to examine the observed vs. predicted probabilities of the unilateral subtype (Figure 2). Three models were unable to accommodate a large spectrum of probabilities.^{12,16,17} Specifically, in the Kamemura model, patients were universally classified as having a high probability of bilateral disease. In contrast, most patients were considered to have at least a 50% chance of lateralization using the Mulatero model. Calibration intercepts (calibration-in-the-large) were uniformly >0 (from 0.35 [95% CI, 0.07, 0.64] to 1.67 [95% CI, 1.32, 2.01]; i.e., showing systematic bias with underprediction of the unilateral subtype) and the corresponding slopes were all <1 (from 0.35 [95% CI, 0.21, 0.50] to 0.87 [95% CI, 0.18, 1.70]; i.e., indicating that lateralizing disease was generally underestimated and bilateral disease was overestimated at the extremes of risk). After recalibration, the Küpers model demonstrated some improvement, but substantial bias remained with significant variability at the extremes of the predicted probabilities (Supplementary Figure S2 online).

The guideline-based prediction rule was assessed separately, as it was designed to confirm unilateral disease in high-probability cases (rather than broadly differentiating between unilateral vs. bilateral subtypes).^{2,29} The presence of all 4 predictors from the rule was highly specific for lateralization but lacked sensitivity (specificity, 100.0% [95% CI, 97.0%, 100.0%]; sensitivity, 2.6% [95% CI, 0.7%, 6.6%]).

Sensitivity analysis

When the models were reexamined using the AVS interpretation criteria originally described in the derivation cohorts, there was incremental improvement in performance across all models, although our overall findings remained similar (Supplementary Figures S3 and S4 online). The largest improvements were seen in the Kobayashi model, which had a C-statistic of 0.80 (95% CI, 0.74, 0.85), calibration intercept crossing zero (−0.21 [95% CI, −0.50, 0.08]) and slope crossing one (0.89 [95% CI, 0.67, 1.13]). Calibration measures also improved in the Mulatero and Umakoshi models, but these continued to be limited by a narrow range of predictions and modest discrimination.

DISCUSSION

In this study, we attempted external validation of 6 clinical prediction models for detecting lateralization on AVS in a large and diverse cohort of patients with PA,^{12–14,16–18} and consistently found that these had low-to-modest discrimination and calibration. The range of predictions from these models was generally narrow and the frequency of lateralization was commonly underestimated. Given the importance of accurate subtyping for informing treatment decisions (i.e., both to minimize missed opportunities for intervention in patients with unilateral PA and to avoid unnecessary surgery in those with bilateral disease), our findings suggest that clinical assessment alone cannot be reliably used to decide upon whether AVS should be offered. Indeed, a robust prediction tool with good discrimination that remains well-calibrated across diverse settings is needed, but presently lacking.

Previous evaluations of these models have been limited in scope.^{23,30,32} Generally, these have confirmed modest discrimination, but none have examined calibration. The Küpers model has been the subject of greatest study with external validations (with C-statistic, if available) performed in German,²⁰ British,¹⁴ Chinese (C-statistic, 0.64),¹⁵ and Japanese (C-statistic, 0.80)¹⁸ cohorts. A recent study evaluated the Küpers, Kamemura, and Kobayashi models in an Italian population¹⁹; while no C-statistics or measures of calibration were provided, the investigators reported only modest model accuracy at distinguishing between unilateral vs. bilateral subtypes (accuracy range, 72.7%–74.1%). Consistent with the common rule that prognostic models generalize best to populations that are similar to the development population, a limited validation of the Kamemura model in an independent Japanese cohort—one that shared many similar characteristics with the derivation cohort—reported fair discrimination (C-statistic, 0.78).¹⁸ Finally, the guideline-based prediction rule has only been evaluated in small groups, all showing limited sensitivity, but high specificity for unilateral disease.^{21,22} In contrast to previous validation studies, ours was the first to assess multiple dimensions of model performance, and demonstrated global miscalibration with systematic underprediction of lateralization. Importantly, good discrimination, even if present, is insufficient to guide diagnostic decisions, as it only captures relative ranks between subjects. Calibration, however, is a crucial property of a prognostic model because it reflects absolute risk for individuals.²³

There were multiple possible reasons why the studied models did not perform well externally, including methodological factors (e.g., overfitting of the original models),³⁷ clinical heterogeneity (e.g., diversity in participant characteristics, laboratory assays, and interpretation criteria),³⁰ and incomplete data (e.g., potential omission of important predictors that were either unknown or unmeasured). First, the inclusion of many variables in a model with comparatively few outcomes may lead to overly optimistic measures of performance that only become apparent when the same model is evaluated in an independent dataset. Addressing clinical heterogeneity, we partially accounted for differences in baseline risk using recalibration techniques and also accounted for differences in AVS interpretation criteria,^{6,13,36} but found

Table 2. Frequencies of the unilateral subtype according to stratum for each prediction model and the guideline-based prediction rule

Model (no.)	Points	No. in category	No. with unilateral subtype (%)	No. with bilateral subtype (%)
Mulatero [<i>n</i> = 271] ^a	Negative	215	104 (48.4)	111 (51.6)
	Positive	56	48 (85.7)	8 (14.3)
Küpers [<i>n</i> = 263]	0	20	11 (55.0)	9 (45.0)
	1	23	10 (43.5)	13 (56.5)
	2	44	17 (38.6)	27 (61.4)
	3	34	14 (41.2)	20 (58.8)
	4	21	10 (47.6)	11 (52.4)
	5	52	36 (69.2)	16 (30.8)
	6	44	35 (79.6)	9 (20.4)
	7	25	24 (96.0)	1 (4.0)
Sze [<i>n</i> = 263] ^b	0	9	4 (44.4)	5 (55.6)
	1	21	10 (47.6)	11 (52.4)
	2	50	21 (42.0)	29 (58.0)
	3	38	17 (44.7)	21 (55.3)
	4	19	6 (31.6)	13 (68.4)
	5	54	39 (72.2)	15 (27.8)
	6	47	36 (76.6)	11 (23.4)
	7	25	24 (96.0)	1 (4.0)
Kamemura [<i>n</i> = 141] ^c	0	52	26 (50.0)	26 (50.0)
	1	60	26 (43.3)	34 (56.7)
	2	25	7 (28.0)	18 (72.0)
	3	4	0 (0)	4 (100)
Umakoshi [<i>n</i> = 304] ^d	Low	66	25 (37.9)	41 (62.1)
	Low-moderate	28	13 (46.4)	15 (53.6)
	High-moderate	93	42 (45.2)	51 (54.8)
	High	117	92 (78.6)	25 (21.4)
Kobayashi [<i>n</i> = 269] ^e	0	37	26 (83.9)	5 (16.1)
	1	27	24 (88.9)	3 (11.1)
	2	21	12 (70.6)	5 (29.4)
	3	39	27 (61.4)	17 (38.6)
	4	17	9 (52.9)	8 (47.1)
	5	22	12 (50.0)	12 (50.0)
	6	20	8 (38.1)	13 (61.9)
	7	23	10 (45.4)	12 (54.6)
	8	24	11 (42.3)	15 (57.7)
	9	21	7 (33.3)	14 (66.7)
	10	10	3 (27.3)	8 (72.7)
	11	7	2 (28.6)	5 (71.4)
	12	1	0 (0)	1 (100)
Endocrine Society and European Society of Hypertension [<i>n</i> = 271] ^e	Negative	267	148 (55.4)	119 (44.6)
	Positive	4	4 (100)	0 (0)

Data are no. (%).

^aBased on the combined presence of serum potassium <3.0 mmol/l and plasma aldosterone concentration >25 ng/dl (>694 pmol/l).

^bA modification of the clinical prediction model derived by Küpers *et al.*

^cThe models by Kamemura *et al.* and Kobayashi *et al.* were developed to predict bilateral disease (i.e., higher scores correlating with a lower frequency of the unilateral subtype).

^dAfter excluding patients with bilateral adrenal lesions seen on computed tomography (CT), patients were categorized into 4 categories in ascending frequencies of having a unilateral subtype (normal CT with normokalemia [serum potassium \geq 3.5 mmol/l]; unilateral adenoma with normokalemia; normal CT with hypokalemia; and unilateral adenoma with hypokalemia).

^eBased on the combined presence of age <35 years, serum potassium <3.5 mmol/l, plasma aldosterone concentration >30 ng/dl (>831 pmol/l), and a unilateral adrenal adenoma >1.0 cm with a contralateral normal adrenal gland.

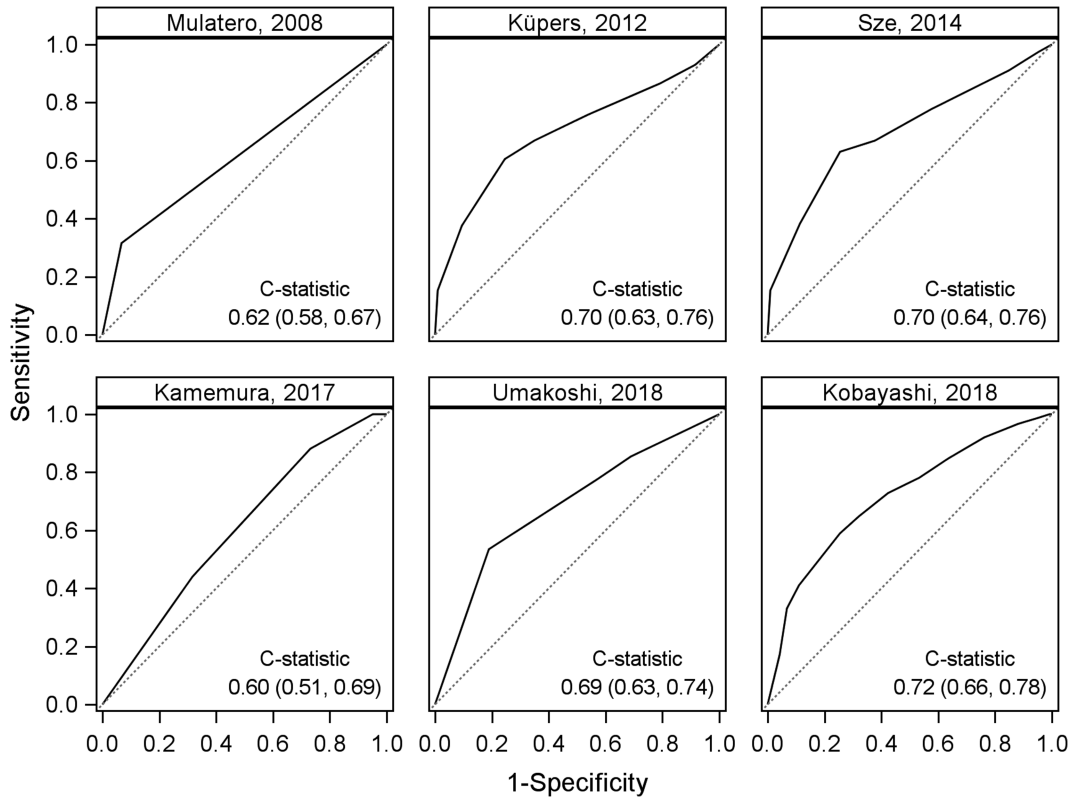


Figure 1. Receiver operating characteristic curves of the clinical prediction models. Sensitivity and specificity were estimated for each category of risk of unilateral primary aldosteronism within each model in the validation cohort, using adrenal vein sampling as the reference. C-Statistics with 95% confidence intervals are shown.

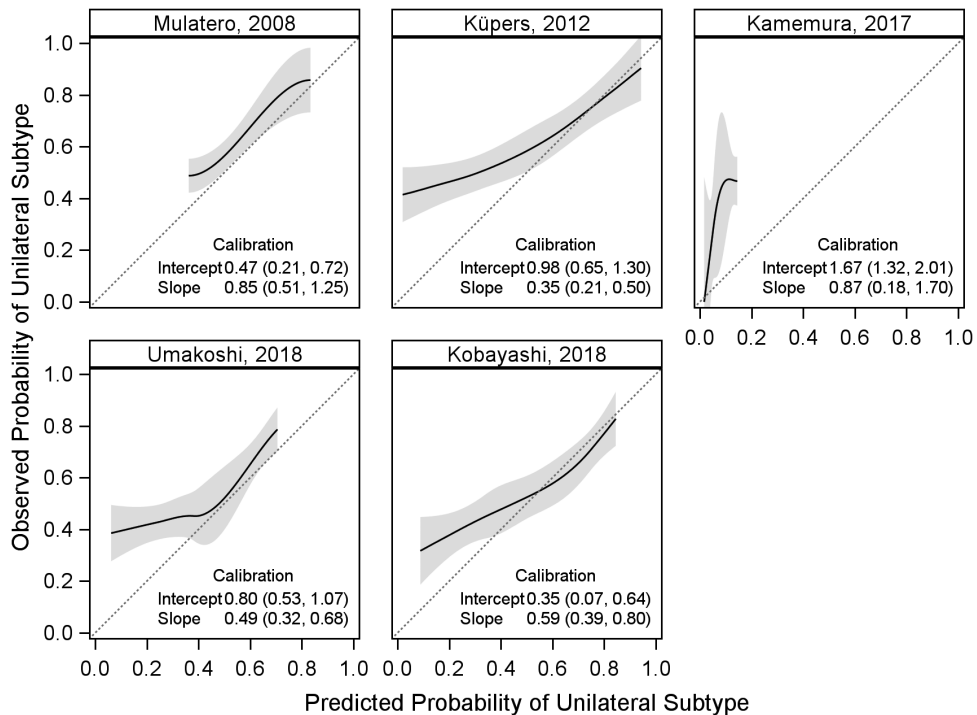


Figure 2. Calibration plots of the clinical prediction models. The curves compare the observed (from adrenal vein sampling) and predicted (from each model) probabilities of unilateral primary aldosteronism in the validation cohort. Calibration intercept and slope are shown with 95% confidence intervals.

minimal improvement in overall performance. Admittedly, wide variability exists between practices for patients with PA, and it is impossible to find a single “representative” population.⁶ Indeed, many PA patients are never identified or referred for subtyping.^{1,11} However, these facts simply reinforce the difficulty in generalizing clinical prediction models for PA subtyping. The third factor was likely the most significant limitation of every model. It is not surprising that the major determinants of lateralization are more complex than can be captured with current data. For instance, imaging abnormalities and hypokalemia were the strongest predictors of lateralization, and were included in most models, but these still proved to be unreliable.^{2,13,14,16–18,29} Specifically, an adrenal nodule may frequently be discordant with AVS,⁹ and the absence of a visible nodule does not rule-out unilateral disease.³⁸ Previous studies supporting the use of computed tomography alone were limited by small sample sizes and inconsistent verification with AVS.^{10,39} Moreover, hypokalemia is frequently detected in patients with both unilateral and bilateral subtypes.⁴⁰ As such, the common variables purported to be associated with lateralization are not sufficiently sensitive to guide most clinical decisions. While there may be a limited role for the highly specific guideline-based prediction rule for identifying a selected group of patients who are likely to have unilateral disease, it cannot be used to decide upon AVS for the majority of PA patients.^{2,29}

This study had multiple strengths (i.e., at our center, AVS has been verified to be a reliable reference standard with a high degree of success and a low rate of false positive lateralization; baseline characteristics, including mean age and the proportion male, were generally similar to those in the derivation studies, thus allowing for fair comparisons; there were a large number of patients who had lateralization providing sufficient power to evaluate each model; and we were able to assess performance using a range of AVS interpretation criteria).^{6,26,35} There were some weaknesses. First, there were occasional missing data for the predictors of interest in the validation cohort. However, we performed complete case analysis for each model, only including people where all variables were known. Second, we were unable to evaluate models that used variables which were not routinely collected at our center.^{18,19} Even so, we were able to assess the commonly referenced models in this study.

In conclusion, the clinical prediction models evaluated here lack generalizability and cannot reliably predict lateralization. The presence of young age, hypokalemia, high aldosterone, and a unilateral adrenal nodule, even in combination, are not sufficiently accurate to distinguish between unilateral and bilateral disease in most people. Correspondingly, the absence of these factors cannot be used to withhold AVS, as this is likely to miss many lateralizing cases. Therefore, whenever available, AVS should still be offered to guide diagnostic subtyping decisions for most patients with PA.

SUPPLEMENTARY MATERIAL

Supplementary data are available at *American Journal of Hypertension* online.

FUNDING

This work was supported by the Canadian Institutes of Health Research Project Grant (159533). Dr Leung is supported by the Heart and Stroke Foundation of Canada's National New Investigator Award.

DISCLOSURE

The authors declared no conflict of interest.

DATA AVAILABILITY

The data underlying this article cannot be shared publicly for the privacy of individuals that participated in the study. The data may be shared on reasonable request to the corresponding author.

REFERENCES

1. Brown JM, Siddiqui M, Calhoun DA, Carey RM, Hopkins PN, Williams GH, Vaidya A. The unrecognized prevalence of primary aldosteronism: a cross-sectional study. *Ann Intern Med* 2020; 173:10–20.
2. Funder JW, Carey RM, Mantero F, Murad MH, Reincke M, Shibata H, Stowasser M, Young WF Jr. The management of primary aldosteronism: case detection, diagnosis, and treatment: an Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab* 2016; 101:1889–1916.
3. Hundemer GL, Curhan GC, Yozamp N, Wang M, Vaidya A. Incidence of atrial fibrillation and mineralocorticoid receptor activity in patients with medically and surgically treated primary aldosteronism. *JAMA Cardiol* 2018; 3:768–774.
4. Hundemer GL, Curhan GC, Yozamp N, Wang M, Vaidya A. Cardiometabolic outcomes and mortality in medically treated primary aldosteronism: a retrospective cohort study. *Lancet Diabetes Endocrinol* 2018; 6:51–59.
5. Vorselaars WMCM, Nell S, Postma EL, Zarnegar R, Drake FT, Duh QY, Talutis SD, McAneny DB, McManus C, Lee JA, Grant SB, Grogan RH, Romero Arenas MA, Perrier ND, Peiper BJ, Mongelli MN, Castellino T, Mitmaker EJ, Parente DN, Pasternak JD, Engelsman AF, Sywak M, D'Amato G, Raffaelli M, Schuermans V, Bouvy ND, Eker HH, Bonjer HJ, Vaarzon Morel NM, Nieveen van Dijkum EJM, Vrieling OM, Kruijff S, Spiering W, Borel Rinkes IHM, Valk GD, Vriens MR; International CONNsortium study group. Clinical outcomes after unilateral adrenalectomy for primary aldosteronism. *JAMA Surg* 2019; 154:e185842.
6. Kempers MJ, Lenders JW, van Outhousden L, van der Wilt GJ, Schultze Kool LJ, Hermus AR, Deinum J. Systematic review: diagnostic procedures to differentiate unilateral from bilateral adrenal abnormality in primary aldosteronism. *Ann Intern Med* 2009; 151:329–337.
7. Lim V, Guo Q, Grant CS, Thompson GB, Richards ML, Farley DR, Young WF Jr. Accuracy of adrenal imaging and adrenal venous sampling in predicting surgical cure of primary aldosteronism. *J Clin Endocrinol Metab* 2014; 99:2712–2719.
8. Nanba AT, Nanba K, Byrd JB, Shields JJ, Giordano TJ, Miller BS, Rainey WE, Auchus RJ, Turcu AF. Discordance between imaging and immunohistochemistry in unilateral primary aldosteronism. *Clin Endocrinol (Oxf)* 2017; 87:665–672.
9. Sam D, Kline GA, So B, Leung AA. Discordance between imaging and adrenal vein sampling in primary aldosteronism irrespective of interpretation criteria. *J Clin Endocrinol Metab* 2019; 104:1900–1906.
10. Thiesmeyer JW, Ullmann TM, Stamatou AT, Limberg J, Stefanova D, Beninato T, Finnerty BM, Vignaud T, Leclerc J, Fahey TJ III, Brunaud L, Mirallie E, Zarnegar R. Association of adrenal venous sampling with

- outcomes in primary aldosteronism for unilateral adenomas. *JAMA Surg* 2021; 156:165–171.
11. Liu YY, King J, Kline GA, Padwal RS, Pasięka JL, Chen G, So B, Harvey A, Chin A, Leung AA. Outcomes of a specialized clinic on rates of investigation and treatment of primary aldosteronism. *JAMA Surg* 2021; 156:541–549.
 12. Mulatero P, Bertello C, Rossato D, Mengozzi G, Milan A, Garrone C, Giraudo G, Passarino G, Garabello D, Verhovez A, Rabbia F, Veglio F. Roles of clinical criteria, computed tomography scan, and adrenal vein sampling in differential diagnosis of primary aldosteronism subtypes. *J Clin Endocrinol Metab* 2008; 93:1366–1371.
 13. Küpers EM, Amar L, Raynaud A, Plouin PF, Steichen O. A clinical prediction score to diagnose unilateral primary aldosteronism. *J Clin Endocrinol Metab* 2012; 97:3530–3537.
 14. Sze WC, Soh LM, Lau JH, Reznik R, Sahdev A, Matson M, Riddoch F, Carpenter R, Berney D, Grossman AB, Chew SL, Akker SA, Druce MR, Waterhouse M, Monson JB, Drake WM. Diagnosing unilateral primary aldosteronism—comparison of a clinical prediction score, computed tomography and adrenal venous sampling. *Clin Endocrinol (Oxf)* 2014; 81:25–30.
 15. Zhang Y, Niu W, Zheng F, Zhang H, Zhou W, Shen Z, Xu J, Tang X, Zhang J, Gao PJ, Wang JG, Zhu L. Identifying unilateral disease in Chinese patients with primary aldosteronism by using a modified prediction score. *J Hypertens* 2017; 35:2486–2492.
 16. Kamemura K, Wada N, Ichijo T, Matsuda Y, Fujii Y, Kai T, Fukuoka T, Sakamoto R, Ogo A, Suzuki T, Umakoshi H, Tsuiki M, Naruse M. Significance of adrenal computed tomography in predicting laterality and indicating adrenal vein sampling in primary aldosteronism. *J Hum Hypertens* 2017; 31:195–199.
 17. Umakoshi H, Tsuiki M, Takeda Y, Kurihara I, Itoh H, Katabami T, Ichijo T, Wada N, Yoshimoto T, Ogawa Y, Kawashima J, Sone M, Inagaki N, Takahashi K, Watanabe M, Matsuda Y, Kobayashi H, Shibata H, Kamemura K, Otsuki M, Fujii Y, Yamamoto K, Ogo A, Yanase T, Suzuki T, Naruse M; JPAS Study Group. Significance of computed tomography and serum potassium in predicting subtype diagnosis of primary aldosteronism. *J Clin Endocrinol Metab* 2018; 103:900–908.
 18. Kobayashi H, Abe M, Soma M, Takeda Y, Kurihara I, Itoh H, Umakoshi H, Tsuiki M, Katabami T, Ichijo T, Wada N, Yoshimoto T, Ogawa Y, Kawashima J, Sone M, Inagaki N, Takahashi K, Watanabe M, Matsuda Y, Shibata H, Kamemura K, Yanase T, Otsuki M, Fujii Y, Yamamoto K, Ogo A, Nanba K, Tanabe A, Suzuki T, Naruse M; JPAS Study Group. Development and validation of subtype prediction scores for the workup of primary aldosteronism. *J Hypertens* 2018; 36:2269–2276.
 19. Burrello J, Burrello A, Pieroni J, Sconfienza E, Forestiero V, Rabbia P, Adolf C, Reincke M, Veglio F, Williams TA, Monticone S, Mulatero P. Development and validation of prediction models for subtype diagnosis of patients with primary aldosteronism. *J Clin Endocrinol Metab* 2020; 105:e3706–e3717.
 20. Riestler A, Fischer E, Degenhart C, Reiser MF, Bidlingmaier M, Beuschlein F, Reincke M, Quinkler M. Age below 40 or a recently proposed clinical prediction score cannot bypass adrenal venous sampling in primary aldosteronism. *J Clin Endocrinol Metab* 2014; 99:E1035–E1039.
 21. Umakoshi H, Ogasawara T, Takeda Y, Kurihara I, Itoh H, Katabami T, Ichijo T, Wada N, Shibayama Y, Yoshimoto T, Ogawa Y, Kawashima J, Sone M, Inagaki N, Takahashi K, Watanabe M, Matsuda Y, Kobayashi H, Shibata H, Kamemura K, Otsuki M, Fujii Y, Yamamoto K, Ogo A, Yanase T, Okamura S, Miyamoto S, Suzuki T, Tsuiki M, Naruse M. Accuracy of adrenal computed tomography in predicting the unilateral subtype in young patients with hypokalaemia and elevation of aldosterone in primary aldosteronism. *Clin Endocrinol (Oxf)* 2018; 88:645–651.
 22. Williams TA, Burrello J, Sechi LA, Fardella CE, Matrozoza J, Adolf C, Baudrand R, Bernardi S, Beuschlein F, Catena C, Doumas M, Fallo F, Giacchetti G, Heinrich DA, Saint-Hilary G, Jansen PM, Januszewicz A, Kocjan T, Nishikawa T, Quinkler M, Satoh F, Umakoshi H, Widimský J Jr, Hahner S, Douma S, Stowasser M, Mulatero P, Reincke M. Computed tomography and adrenal venous sampling in the diagnosis of unilateral primary aldosteronism. *Hypertension* 2018; 72:641–649.
 23. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J* 2014; 35:1925–1931.
 24. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* 2015; 162:55–63.
 25. Kline GA, Pasięka JL, Harvey A, So B, Dias VC. High-probability features of primary aldosteronism may obviate the need for confirmatory testing without increasing false-positive diagnoses. *J Clin Hypertens (Greenwich)* 2014; 16:488–496.
 26. Kline G, Leung A, So B, Chin A, Harvey A, Pasięka JL. Application of strict criteria in adrenal venous sampling increases the proportion of missed patients with unilateral disease who benefit from surgery for primary aldosteronism. *J Hypertens* 2018; 36:1407–1413.
 27. Leung AA, Orton DJ, Chin A, Sadrzadeh H, Kline GA. Novel approach to establishing an aldosterone:renin ratio cutoff for primary aldosteronism. *Hypertension* 2017; 69:450–456.
 28. Williams TA, Lenders JWM, Mulatero P, Burrello J, Rottenkolber M, Adolf C, Satoh F, Amar L, Quinkler M, Deinum J, Beuschlein F, Kitamoto KK, Pham U, Morimoto R, Umakoshi H, Prejbisz A, Kocjan T, Naruse M, Stowasser M, Nishikawa T, Young WF Jr, Gomez-Sanchez CE, Funder JW, Reincke M; Primary Aldosteronism Surgery Outcome (PASO) investigators. Outcomes after adrenalectomy for unilateral primary aldosteronism: an international consensus on outcome measures and analysis of remission rates in an international cohort. *Lancet Diabetes Endocrinol* 2017; 5:689–699.
 29. Mulatero P, Sechi LA, Williams TA, Lenders JWM, Reincke M, Satoh F, Januszewicz A, Naruse M, Doumas M, Veglio F, Wu VC, Widimsky J. Subtype diagnosis, treatment, complications and outcomes of primary aldosteronism and future direction of research: a position statement and consensus of the Working Group on Endocrine Hypertension of the European Society of Hypertension. *J Hypertens* 2020; 38:1929–1936.
 30. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009; 338:b605.
 31. Alba AC, Agoritsas T, Walsh M, Hanna S, Iorio A, Devereaux PJ, McGinn T, Guyatt G. Discrimination and calibration of clinical prediction models: users' guides to the medical literature. *JAMA* 2017; 318:1377–1384.
 32. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010; 21:128–138.
 33. Crowson CS, Atkinson EJ, Therneau TM. Assessing calibration of prognostic risk scores. *Stat Methods Med Res* 2016; 25:1692–1706.
 34. Janssen KJ, Moons KG, Kalkman CJ, Grobbee DE, Vergouwe Y. Updating methods improved the performance of a clinical prediction model in new patients. *J Clin Epidemiol* 2008; 61:76–86.
 35. Riley RD, Ensor J, Snell KI, Debray TP, Altman DG, Moons KG, Collins GS. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. *BMJ* 2016; 353:i3140.
 36. Janssen KJ, Vergouwe Y, Kalkman CJ, Grobbee DE, Moons KG. A simple method to adjust clinical prediction models to local circumstances. *Can J Anaesth* 2009; 56:194–201.
 37. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996; 15:361–387.
 38. Sam D, Kline GA, So B, Przybojewski SJ, Leung AA. Unilateral disease is common in patients with primary aldosteronism without adrenal nodules. *Can J Cardiol* 2021; 37:269–275.
 39. Dekkers T, Prejbisz A, Kool LJS, Groenewoud HJMM, Velema M, Spiering W, Kołodziejczyk-Kruk S, Arntz M, Kądziela J, Langenhuijsen JF, Kerstens MN, van den Meiracker AH, van den Born BJ, Sweep FCGJ, Hermus ARMM, Januszewicz A, Ligthart-Naber AF, Makai P, van der Wilt GJ, Lenders JWM, Deinum J; SPARTACUS Investigators. Adrenal vein sampling versus CT scan to determine treatment in primary aldosteronism: an outcome-based randomised diagnostic trial. *Lancet Diabetes Endocrinol* 2016; 4:739–746.
 40. Burrello J, Monticone S, Losano I, Cavaglià G, Buffolo F, Tetti M, Covella M, Rabbia F, Veglio F, Pasini B, Williams TA, Mulatero P. Prevalence of hypokalemia and primary aldosteronism in 5100 patients referred to a tertiary hypertension unit. *Hypertension* 2020; 75:1025–1033.