Clinical Score and Machine Learning-Based Model to Predict Diagnosis of Primary Aldosteronism in Arterial Hypertension

Clinical score and machine learning-based model to predict diagnosis of primary aldosteronism in arterial hypertension

Fabrizio Buffolo[1*], Jacopo Burrello[1*], Alessio Burrello[2*], Daniel Heinrich[3], Christian Adolf[3], Lisa Marie Müller[3], Rusi Chen[1], Vittorio Forestiero[1], Elisa Sconfienza[1], Martina Tetti[1], Franco Veglio[1], Tracy Ann Williams[1,3], Paolo Mulatero[1†], Silvia Monticone[1†].

(1) Division of Internal Medicine and Hypertension Unit, Department of Medical Sciences, University of Torino, Via Genova 3, 10126, Torino, Italy. (2) Department of Electrical, Electronic and Information Engineering "Guglielmo Marconi" (DEI), University of Bologna, Italy. (3) Medizinische Klinik und Poliklinik IV, Klinikum der Universität, Ludwig-Maximilians-Universität München, Munich, Germany.

* Contributed equally and should be considered as joint first authors.

† Contributed equally and should be considered as joint last authors.

Corresponding author: Prof. Paolo Mulatero Division of Internal Medicine and Hypertension Unit, Department of Medical Sciences, University of Torino, Città della Salute e della Scienza, Via Genova 3, 10126 Torino, Italy. Telephone/Fax number: +39.011.633.6959 / +39.011.633.6931. E-mail: paolo.mulatero@unito.it

Manuscript word count: 3,570 excluding references, tables and figure captions.

Abstract word count: 258

Number of Tables: 1; Number of Figures: 3; 1 Supplemental file.

Short title: Prediction models for primary aldosteronism

Acknowledgements: none

Sources of Funding: nothing to disclose.

Disclosure: P.M. received fees for educational speech from DIASORIN.

Abstract

Primary aldosteronism (PA) is the cause of arterial hypertension in 4-6% of patients, and 30% of patients with PA are affected by unilateral and surgically-curable forms. Current guidelines recommend screening for PA ~50% of patients with hypertension on the basis of individual factors, while some experts suggest screening all patients with hypertension. To define the risk of PA and tailor the diagnostic workup to the individual risk of each patient, we developed a conventional scoring system and supervised machine learning algorithms using a retrospective cohort of 4,059 patients with hypertension. On the basis of 6 widely available parameters, we developed a numerical score and 308 machine learning-based models, selecting the one with the highest diagnostic performance. After validation, we obtained high predictive performance with our score (optimized sensitivity of 90.7% for PA and 92.3% for unilateral primary aldosteronism [UPA]). The machine learning-based model provided the highest performance, with an AUC of 0.834 for PA and 0.905 for diagnosis of UPA, with optimized sensitivity of 96.6% for PA and 100.0% for UPA, at validation. The application of the predicting tools allowed the identification of a subgroup of patients with very low risk of PA (0.6 % for both models) and null probability of having UPA. In conclusion, this score and the machine learning algorithm can accurately predict the individual pre-test probability of PA in patients with hypertension and circumvent screening in up to 32.7% of patients using a machine learning based model, without omitting patients with surgically curable unilateral PA.

Key words: primary aldosteronism, machine learning, clinical score, aldosterone producing adenoma, arterial hypertension

Introduction

Primary aldosteronism (PA) is the leading cause of endocrine hypertension with a prevalence of 4-6% in the general hypertensive population,[1,2] and an estimate of 70 million of affected patients worldwide. Endocrine Society (ES) and European Society of Hypertension (ESH) guidelines recommend screening for PA in patients at moderate-high risk of the disease, encompassing 50-60% of patients with arterial hypertension.[1,2] Nevertheless, PA is often overlooked, with less than 3% and 8% of patients actually screened in the US and Europe, respectively.[3–5] Several reasons underlie the current underdiagnosis, including low awareness and underestimation of the true prevalence of the disease, reluctance to medication withdrawal and complex interpretation of diagnostic tests.[4,6]

Accumulating evidence indicates that chronic aldosterone excess causes an increased risk of target organ damage, cardio and cerebrovascular events[7], chronic kidney disease [8] and overall mortality[9] in patients affected by PA compared with patients affected by essential hypertension. Treatment with mineralocorticoid receptor antagonists and surgical adrenalectomy allow a significant reduction of the excess risk.[8–10] Therefore, recognition of the disease is crucial to revert the detrimental effects of aldosterone in a timely manner. Some experts suggest expanding screening for PA to all patients with arterial hypertension[6,11], which would however increase the burden on primary care physicians. In this context, a prediction tool to reduce the number of patients for PA screening would be highly desirable.

Conventional scoring systems are based on non-adaptive linear and logistic regression algorithms, derived from a limited number of clinical predictors. Machine learning relies on adaptable, non-linear algorithms derived from large dataset of multidimensional variables, allowing more accurate prediction by computation of multiple simultaneous features. In PA, supervised machine learning have been successfully applied for prediction of subtype diagnosis and clinical outcomes post-adrenalectomy.[12,13] Only two studies have previously investigated

the impact of clinical scores for prediction of PA in patients with arterial hypertension, with limited diagnostic performance, relatively small sample size, and/or absence of a validation cohort.[14,15] No study has previously applied machine learning algorithm for prediction of PA in a large cohort of patients with arterial hypertension.

In this study, we build and validate prediction models based on supervised learning algorithms and a conventional scoring system (Score To Predict Primary Aldosteronism, SToP-PA score), in a large internal cohort of more than 4,000 patients with arterial hypertension screened for PA and in an external cohort of 584 patients with arterial hypertension.

Methods

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Study Design and PA diagnosis

Between 2008 and 2020, 4,059 patients completed the diagnostic workup for PA at the Division of Internal Medicine – Hypertension Unit of the University of Torino and were included in the analysis. Eligible patients were randomly assigned to a training cohort (n=3,045) or to a validation cohort (n=1,014). An independent external cohort of 584 patients recruited from the Munich Klinikum der Universität was used for external validation. The study complied with the Declaration of Helsinki and was approved by local ethical committees.

PA diagnosis was performed according to ES and ESH recommendations.[1,2,16] Interfering drugs were withdrawn before assessment of serum aldosterone and plasma renin activity (PRA) or direct renin concentration (DRC) for screening test. When the complete withdrawal of anti-hypertensive drugs was not feasible, screening test was performed with medications that have only small or minimal impact on ARR (including α1-antagonists, moxonidine, dihydropyridine

and non-dihydropyridine calcium channel blockers). A threshold of aldosterone-to-renin ratio (ARR) of 30 ng/dL/ng*mL^{-1}*h^{-1} was considered for positive screening test, together with aldosterone levels greater than 10 ng/dL, for the internal cohort. For the external cohort, a threshold of ARR of 2.0 ng/dL/mU/L, using DRC, was used for positive screening test, together with aldosterone greater than 10 ng/dL. Patients with a positive screening test underwent confirmatory test by intravenous saline load test or captopril challenge test. Patients with aldosterone greater than 5 ng/dL, after intravenous saline load test, or ARR greater than 30 ng/dL/ng*mL^{-1}*h^{-1} (2.0 ng/dL/mU/L, using DRC) after captopril challenge test were considered as affected by PA. Subtype diagnosis was performed by adrenal computed tomography and adrenal venous sampling according to ES and ESH recommendations.[2,16] Successful cannulation of adrenal glands was defined by selectivity index≥3 for unstimulated and ≥5 for cosyntropin stimulated AVS. Unilateral PA (UPA) was defined in case of lateralization index≥4 or ≥3 with contralateral ratio <1, (defined as aldosterone/cortisol of non-dominant adrenal vein divided for aldosterone/cortisol of peripheral vein). For the external cohort, a selectivity index≥2 was used to define successful cannulation and lateralization index≥4 do define UPA.

Left ventricular hypertrophy (LVH) was defined as left ventricular mass index ≥ 115 g/m^2 (male) or ≥ 95 g/m^2 (female).[17] Microalbuminuria was defined as urine albumin concentration of 30 to 300 mg/24 hour and/or by albumin to creatinine ratio between 30 and 300 mg/g.[17]

Statistical Analysis

IBM SPSS Statistics version 26.0 (IBM Corp, Armonk, New York) and GraphPad Prism 8.0 (GraphPad, La Jolla, CA) were used for statistics. Variables were treated as parametric or non-parametric according to their distribution. Continuous variables with a normal distribution were expressed as mean ± standard deviation. Non-normally distributed variables were expressed as

median [interquartile range]. Categorical variables were expressed as absolute number and percentage. Significance was defined by Student *t* test and ANOVA 1-way with Bonferroni post hoc tests for parametric variables, and Mann-Whitney U test and Kruskall-Wallis for non-parametric variables, respectively. $\chi 2$ was used for comparison of categorical variables. Univariate and multivariate logistic regression were used to assess the association of clinical / biochemical variables with PA; an odds ratio (OR) greater than 1 indicates an increased likelihood of PA, and an OR less than 1 a decreased likelihood. Receiver operator characteristics (ROC) curves were analyzed to assess the discrimination performance of the proposed models; area under the curve (AUC) was reporter together with 95% confidence interval. A *P*-value of less than 0.05 was considered significant.

Pearson correlation was performed to correlate aldosterone-to-renin ratio with SToP-PA score and machine learning coefficients.

Diagnostic Modeling

Python 3.5 (library, scikit-learn) was used to generate and test the prediction models. An overview of the development strategy is provided in Figure S1. Univariate and multivariate logistic regression were used to select independent predictors of PA.

Points and cut-offs for the numerical score (SToP-PA score) were automatically assigned by a computational algorithm in order to achieve the highest accuracy in the training cohort. The score was then tested in the validation cohort.

Four machine learning (ML) classifiers (linear discriminant analysis [LDA], random forest regressor [RFr], support vector machine with linear [l-SVM], or gaussian kernel [rbf-SVM]) were applied to the training cohort, with or without 3 techniques of data imbalance correction (random oversampling methods, synthetic minority over-sampling technique [SMOTE], SMOTE and nearest neighbors), and with grid-search of hyperparameters, resulting in 308 different prediction models. After tuning of hyperparameters, a random forest regressor

algorithm with random oversampling correction for dataset imbalance was selected as the most accurate for PA prediction and was tested in the validation cohorts. A detailed description of the applied supervised learning methods is provided in the extended methods section of the Data Supplement.

Free-downloadable tools for both the SToP-PA score and the RFr-model are available at the following link: https://github.com/CentroIpertenUnito/SToP-PA.

Results

Characteristics of patients

We included in the study a total of 4,059 patients: 706 (17.4%) with a positive screening test for PA and 3,353 (82.6%) with a negative screening test. Clinical and biochemical characteristics of patients at screening are shown in Table S1. After i.v. saline loading test or captopril challenge test, 431 patients had a confirmed diagnosis of PA (10.6% of the total cohort). A total of 130 patients had UPA (30.2%), 247 bilateral PA (57.3%), while in 54 (12.5%) patients subtype diagnosis was not achieved, because the adrenal venous sampling was unsuccessful or patients did not favor AVS.

At screening, patients with PA were slightly older than non-PA, more frequently male and with longer duration of hypertension (Table 1). Patients affected by PA had higher systolic (157±20 versus 146±18 mmHg; *P*=<0.001) and diastolic BP (95±11 versus 91±11 mmHg; *P*=<0.001), higher intensity of antihypertensive treatment (defined by daily defined dose [DDD]) (2.67 [1.31-4.33] versus 1.33 [0.33; 2.88]; *P*=<0.001) and lower potassium levels (3.6±0.6 versus 4.1±0.4 mEq/L; *P*=<0.001). Patients with PA had slightly lower BMI and a higher rate of organ damage, defined as left ventricular hypertrophy and/or microalbuminuria and cardiovascular events (13.9% versus 8.9%; *P*=0.001). As expected, PRA was significantly lower and serum aldosterone significantly higher in patients with PA (Table 1).

Association of clinical and biochemical characteristics with PA diagnosis has been evaluated by univariate logistic regression (Table S2). A significant association was observed between diagnosis of PA and 10 of 12 patient parameters: age (OR 1.015), male sex (OR 1.368), duration of hypertension (OR 1.003), systolic and diastolic BP (OR 1.025 and 1.033, respectively), anti-hypertensive treatment (OR 1.344), BMI (OR 0.975), lowest potassium levels (OR 0.120), presence of organ damage (OR 2.682) and cardiovascular events (OR 1.655). Parameters which were significantly associated with PA diagnosis at univariate analysis were introduced into a multivariate model: male sex, systolic BP, anti-hypertensive treatment, BMI, lowest potassium and organ damage were confirmed as independent predictors of PA (Table S3). These parameters were used as input variables to develop a numerical scoring system and a prediction model based on supervised machine learning.

Development and validation of the SToP-PA score

All the patients included in the study were randomly assigned to the training (n=3,045) or to the validation cohort (n=1,014). No differences were found between the two groups (Table S4). The 6 independent predictors of PA at multivariate analysis (male sex, systolic BP, anti-hypertensive treatment, BMI, lowest potassium, and organ damage) were used for the development of a numerical scoring system (SToP-PA score, 0-21.5 points) to discriminate between patients with and without PA. Cut-offs and points for the 6 variables were automatically determined in patients from the training cohort by a computational algorithm designed to obtain the highest diagnostic performance (Table S5 and Figure 1A). The SToP-PA score was then applied to detect patients with positive screening test, PA, or UPA in the validation cohort.

The analysis of ROC curves showed high performances both at training (AUC 0.734, 0.822, and 0.903 for the detection of patients with positive screening test, PA, or UPA, respectively;

Figure 1B) and validation (AUC 0.739, 0.796, and 0.882; Figure 1C; Table S6). As expected, the highest performance was reached for the diagnosis of UPA. Performance of SToP-PA score were similar when applied within the validation cohort in patients with and without specific indication for PA screening according to international guidelines (Tables S7-S8-S9)[1,2]. Additionally, the performances remain similar, in both training and validation cohort, adopting more stringent criteria for screening positivity (ARR>40 ng/dL/ng*mL^{-1}*h^{-1} and AC>15 ng/dL) and confirmatory test (AC post-intravenous saline load test>10 ng/dL or AC post-captopril challenge test>11 ng/dL) (Tables S10-S11).

Diagnostic performance for the detection of screening positivity, PA and UPA diagnosis with different cut-offs are reported in Table S12 for both, training and validation cohorts. The cut-offs with optimized sensitivity were 7.0, 7.5, and 8.0, for prediction of positive screening results, PA, or UPA diagnosis, with respective sensitivity of 90.0%, 93.3% and 100.0% at training and 92.0%, 90.7% and 92.3% at validation. Comparison of performances of SToP-PA score at training and validation revealed a negligible overfitting bias (1.2-2%). Cut-offs of 13.5, 14.0 and 15.0 optimized the specificity for the detection of patients with positive screening test, PA, or UPA respectively, achieving a specificity of 92.3% for positive screening, 96.0% for PA and 97.0% for UPA diagnosis at training and 91.8%, 96.4% and 98.1% and validation.

Supervised machine learning

Four supervised learning classifiers (LDA, RFr, l-SVM and rbf-SVM) with or without 3 different methods for dataset imbalance correction were applied to the training cohort. A total of 308 models were generated and after tuning of hyperparameters, a RFr composed by 50 classification trees and a maximum number of splits equal to 10, with random oversampling correction, displayed the highest accuracy for detection of patients with PA (Tables S13-S14).

The ML model was based on the same six predictors of PA used to develop the SToP-PA score. The strongest predictor was lowest potassium, followed by anti-hypertensive treatment, and systolic BP (Figure 2A). A representative classification tree from the RFr-model is shown in Figure 2B.

At ROC curves analysis, RFr-model showed a reliable performance, with an AUC of 0.796 for screening results, 0.871 for PA, and 0.938 for UPA diagnosis, with similar performance at validation and very low overfitting bias (1.0%, 1.8% and 1.7% respectively) (Figure 2C-D). Performance of RFr-model were similar when applied within the validation cohort in patients with and without specific indication for PA screening according to international guidelines (Tables S7-S8-S9)[1,2], and adopting more stringent criteria for screening positivity (ARR>40 ng/dL/ng*mL^{-1}*h^{-1} and AC>15 ng/dL) and confirmatory test (AC post-intravenous saline load test>10 ng/dL or AC post-captopril challenge test>11 ng/dL), for both training and validation cohort (Tables S15-S16).

Diagnostic performances and confusion matrix for the RFr-model are reported in Table S17. The RFr allowed the calculation of a classification coefficient for each patient. The ML prediction model displayed optimized sensitivity with a coefficient cut-off of 0.23 for positive screening, 0.24 for PA, and 0.35 for UPA detection, with sensitivity of 97.4% for screening positivity, 99.7% for PA and 100.0% for UPA diagnosis at training and 95.5%, 96.6% and 100.0% at validation. The cut-offs for optimized specificity were 0.70 for positive screening (94.8% specificity at training and 94.6% at validation), 0.76 for PA (95.6% specificity at training and 96.2% at validation) and 0.81 for UPA detection (97.5% specificity at training and 96.9% at validation).

Pre-test probability stratification

The pre-test probability of having a positive screening test, PA, or UPA diagnosis progressively raised with increasing SToP-PA score, as showed in Figure 1D and Table S18. Similarly, increasing RFr coefficients reflected a higher pre-test probability of having a positive screening test, PA, or UPA diagnosis (Figure 2E and Table S19). A correlation between aldosterone-to-renin ratio and both SToP-PA score and RFr coefficients was present (R 0.336, P<0.001).

Considering risk distribution, patients were stratified in 4 subgroups with an increasing pre-test probability of PA. Then, we integrated SToP-PA score and the RFr prediction model in a flowchart for the management of patients with arterial hypertension (Figure 3). Applying the flowchart to the overall cohort, patients with SToP-PA score<5.0 (469 patients) had a pre-test probability of 0.6% and 0.0% of PA or UPA, respectively (Figure 3A). Applying the ML model, patients with a RFr coefficient <0.30 (1,329 patients) had a probability of 0.6% and 0.0% of PA or UPA, respectively (Figure 3B). Considering the very low prevalence of PA and the absence of unilateral surgically curable form, we propose to avoid screening test in this subgroup of patients. On the counterpart, we suggest to screen all the patients with SToP-PA score ≥ 5.0 or RFr coefficient ≥ 0.30. Among these patients, the probability of having PA increases progressively in the groups with intermediate risk: 4.2% with SToP-PA score of 5-10.5 and 21.1% with SToP-PA score 11.0-15.5, with a prevalence of UPA of 0.3% and 6.5% respectively. Similarly, RFr coefficient progressively stratifies the risk of PA in patients with intermediate probability: 6.9% of PA and 0.9% of UPA with RFr coefficient between 0.30 and 0.59, and 20.5% of PA and 4.5% of UPA between 0.60 and 0.79. Finally, patients with SToP-PA score ≥16.0 or RFr coefficient ≥0.80 display a very high probability of having PA (52.7% for both SToP-PA score and RFr-model) and UPA (32.8% with SToP-PA score and 26.1% with RFr-model). Clinicians should consider direct referral to hypertension centers for these patients.

External Validation

An independent external cohort of 584 patients was included. Clinical characteristics of the external cohort are reported in Table S4. A total of 208 patients (35.6%) had positive screening test, 129 (22.1%) had confirmed diagnosis of PA and 72 (12.3%) of UPA. ROC curves showed high performance for the detection of PA and UPA, with AUCs not significantly different from the internal validation (Table S6), but lower for the detection of patients with screening positivity. Diagnostic performances and confusion matrix for SToP-PA score and RFr-model in the external cohort are reported in Table S20. Patient stratification (Figure S2) showed a progressive increase of pre-test probability of PA and UPA along with higher SToP-PA score and RFr coefficient; 4.5% of patients with SToP-PA score<5 had PA and 5.5% patients with RFr coefficient<0.3 had PA, but none of them had a diagnosis of UPA (see also Extended Results in Data Supplement).

Discussion

In this study, we developed a conventional scoring system and an advanced computational model based on supervised learning, using six widely available parameters, to predict likelihood of PA and of unilateral and thus surgically-curable PA in a large cohort of 4,059 patients with arterial hypertension. We also proposed a flow-chart for the management of these patients according to their pre-test probability of a positive screening result, a diagnosis of PA, or a diagnosis of UPA. We suggest to screen only patients with a SToP-PA score≥5.0 or a RFr coefficient≥0.30. This approach would avoid screening for PA in 11.6% of patients affected by arterial hypertension after stratification for the SToP-PA score and 32.7% of patients with the RFr-model, without missing any patients with UPA. Moreover, the SToP-PA score and the RFr model were integrated in two user friendly free-downloadable tools, which allow their practical application in clinical practice. By the use of these tools, general practitioners can assess the probability of having PA and/or UPA case-by-case and tailor the diagnostic workup for each

individual patient. On the other side, in the time of patient-centered medicine, the results of SToP-PA score and RFr-model can provide the patient with a definite probability of having a surgically-curable form of hypertension, reinforcing clinical recommendation in case PA screening is suggested.

Serum potassium had the highest relevance in RFr-model. Hypokalemia has been acknowledged as the most typical feature of PA from its first description, together with high blood pressure levels.[18] However, with the widespread use of the ARR and the increase detection of mild forms of PA, several studies have demonstrated that hypokalemia is present at diagnosis in a minority (9-37%) of affected patients.[1] Nonetheless, hypokalemia remains a key biochemical feature of patients with PA and probability of PA progressively increase with lower serum potassium levels.[19]

PA prevalence increases with the severity of hypertension[20]: ES guideline recommends screening for PA in patients with SBP≥150 mmHg and DBP≥100 mmHg. Similarly, the recent ESH consensus recommend screening of patients with grade 2 and 3 of hypertension (SBP≥160 mmHg and DBP≥100 mmHg).[1] However, while definition of hypertension grade presents no difficulty in never-treated patients with newly diagnosed hypertension, it is less straightforward in patients under hypertensive treatment. The importance of anti-hypertensive therapy is only partially considered by current guidelines, who recommend screening for patients with resistant hypertension (3 or more anti-hypertensive drugs at full-dose, including a diuretic), but do not consider lower number of anti-hypertensive drugs or intermediate dosage. SToP-PA score and RFr-model take into account both blood pressure levels and intensity of anti-hypertensive treatment (defined by DDD), allowing a more accurate risk stratification also in patients under anti-hypertensive treatment. Moreover, in our study DDD was calculated with medical therapy at the first visit, that included potentially interfering drugs, subsequently stopped or substituted by non-interfering ones for screening test. Considering the reluctance of general practitioners

to modify medical therapy for PA screening test, this aspect significantly simplifies the applicability of both SToP-PA score and RFr-model.

Several studies and two large meta-analysis demonstrated that PA is associated with higher rate of hypertension-mediated organ damage, including left ventricular hypertrophy and microalbuminuria.[7,8] Therefore, it is not surprising that the presence of organ damage was independently associated with PA diagnosis at multivariate analysis and therefore included in both models.

The likelihood of having positive screening test, PA diagnosis and UPA progressively increase with the increase of SToP-PA score and RFr coefficient (Figure 1D and Figure 2E), in both internal and external cohort. The continuum of renin-independent aldosterone production, from mildest form (low renin hypertension) to overt forms (PA and UPA), is reflected by the gradual shift of the probability curves towards the upper part of the histograms (corresponding to the highest scores of SToP-PA and RFr coefficients). The diagnostic consequence is that both tools (SToP-PA and RFr-model) show the highest discriminatory capacity in patient with UPA diagnosis, followed by patients with PA and then patients with low renin hypertension, especially for patients with intermediate risk. This concept is easily visualized by the gradual shift of ROC curves with highest diagnostic performance in patient with PA, especially unilateral forms (Figure 1B-1C-2C-2D).

Two previous studies developed a numerical score to predict PA.[14,15] Yamashita and colleagues first designed a score based on sex, serum potassium and urinary pH (PFK score).[14] Limitations of the PFK score are the small sample size (130 patients and 24 affected by PA) and the development in untreated patients, that limits its applicability to subjects with newly-diagnosed hypertension. Performances of the score were relatively low at validation, and patients with the lowest score had 11% of probability of having PA. Therefore, based on the results of this score, it was not possible to identify a subgroup of patients that could avoid screening for PA.

Kietsiriroje et al. proposed a second score based on age, BMI, diabetes, antihypertensive treatments, serum sodium and potassium, built on 420 patients with arterial hypertension and displaying an AUC 0.87.[15] The major limitation of the study was the absence of a validation cohort, hampering the generalizability of the model. None of the studies performed an external validation.

In our models, we adopted 6 parameters that are widely available in the clinical practice, including primary care setting: male sex, systolic BP, anti-hypertensive treatment, BMI, lowest potassium, and organ damage.

Recent studies progressively expanded the spectrum of PA towards milder forms, highlighting the continuum from earliest form of renin independent aldosteronism towards overt PA.[21–23] On the basis of these findings, some authors proposed to widen the cohort of subjects eligible for PA screening to all patients with hypertension.[6,11] This choice would have two consequences: on one side no patients with PA would be missed and could benefit of target medical or surgical therapy, according to subtype diagnosis; on the other side, general practitioners should face an increasing number of patients undergoing PA screening. The medical cost of a screening test is not high *per se*, but its application in a large scale would undoubtedly amplify its economic impact for the management of the patients with a positive screening test in term of confirmatory/exclusion tests and subtype diagnosis, including CT scanning and adrenal vein sampling. The application of the SToP-PA score and RFr-model allows the identification of a subgroup of patients with very low pre-test probability of having PA, that can avoid PA screening, significantly reducing the burden on general practitioners and hypertension centers. The application of SToP-PA score would reduce by 11.6% the number patients undergoing PA screening, with strikingly higher reduction (32.7%) with RFr-model. At the same time, SToP-PA score and RFr-model identify subgroups of patient with intermediate-high and very high risk of having PA and UPA. Taken together, these two

subgroups account for 34.2% of patients by SToP-PA score and 22.2% by RFr-model. In low- and -medium income countries, where health-related resources are limited, the application of RFr-model and SToP-PA score may help clinicians and health-systems to direct diagnostic efforts in patients with high or intermediate-high probability of PA.

Finally, it should be noted that in the group with very low probability of PA, the chance of having low renin hypertension is relatively low: 5.2% with the SToP-PA score and 3.2% with RFr-model. Therefore, in the internal cohort, our prediction tools miss a very limited number of patients even considering the mildest forms of renin-independent aldosteronism.

Finally, we validated our models within an independent external cohort of 584 patients, with performance that were not significantly different from the internal validation for the detection of patients with PA and UPA, providing optimized sensitivity greater than 90% for both models. The external cohort was characterized by patients with a more severe hypertensive phenotype and significantly higher prevalence of screening positivity, PA and UPA diagnosis. The high performances in such different cohorts widen the applicability of both models, suggesting that SToP-PA score and RFr-model can be reliably applied in different settings, from primary care to tertiary referral centers.

Perspective

In this study we developed and validated a clinical score and a machine learning-based tool to predict diagnosis of PA in patients with arterial hypertension in a large cohort of more than 4-thousands patients and in an external cohort of more than 5-hundred patients. The application of these computational models would allow a tailored management of patients with arterial hypertension, based on their pre-test probability of having PA, avoiding unnecessary screening test in patients with negligible probability of PA and without missing any patients with potential curable forms. Reducing up to one third the number of patients undergoing screening

test for PA, the use of our prediction tools allows a more accurate allocation of health-related resources.

References

1. Mulatero P, Monticone S, Deinum J, Amar L, Prejbisz A, Zennaro M-C, Beuschlein F, Rossi GP, Nishikawa T, Morganti A, Seccia TM, Lin Y-H, Fallo F, Widimsky J. Genetics, prevalence, screening and confirmation of primary aldosteronism: a position statement and consensus of the Working Group on Endocrine Hypertension of The European Society of Hypertension. *J Hypertens*. 2020;38(10):1919-1928. doi:10.1097/HJH.0000000000002510

2. Funder JW, Carey RM, Mantero F, Murad MH, Reincke M, Shibata H, Stowasser M, Young WF. The Management of Primary Aldosteronism: Case Detection, Diagnosis, and Treatment: An Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab*. 2016;101(5):1889-1916. doi:10.1210/jc.2015-4061

3. Ruhle BC, White MG, Alsafran S, Kaplan EL, Angelos P, Grogan RH. Keeping primary aldosteronism in mind: Deficiencies in screening at-risk hypertensives. *Surgery*. 2019;165(1):221-227. doi:10.1016/j.surg.2018.05.085

4. Mulatero P, Monticone S, Burrello J, Veglio F, Williams TA, Funder J. Guidelines for primary aldosteronism: uptake by primary care physicians in Europe. *J Hypertens*. 2016;34(11):2253-2257. doi:10.1097/HJH.0000000000001088

5. Cohen JB, Cohen DL, Herman DS, Leppert JT, Byrd JB, Bhalla V. Testing for Primary Aldosteronism and Mineralocorticoid Receptor Antagonist Use Among U.S. Veterans : A Retrospective Cohort Study. *Ann Intern Med*. Published online December 29, 2020. doi:10.7326/M20-4873

6. Vaidya A, Carey RM. Evolution of the Primary Aldosteronism Syndrome: Updating the Approach. *J Clin Endocrinol Metab*. 2020;105(12). doi:10.1210/clinem/dgaa606

7. Monticone S, D'Ascenzo F, Moretti C, Williams TA, Veglio F, Gaita F, Mulatero P. Cardiovascular events and target organ damage in primary aldosteronism compared with essential hypertension: a systematic review and meta-analysis. *Lancet Diabetes Endocrinol*. 2018;6(1):41-50. doi:10.1016/S2213-8587(17)30319-4

8. Monticone S, Sconfienza E, D'Ascenzo F, Buffolo F, Satoh F, Sechi LA, Veglio F, Mulatero P. Renal damage in primary aldosteronism: a systematic review and meta-analysis. *J Hypertens*. 2020;38(1):3-12. doi:10.1097/HJH.0000000000002216

9. 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(1):51-59. doi:10.1016/S2213-8587(17)30367-4

10. 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(8):768-774. doi:10.1001/jamacardio.2018.2003

11. Funder JW. Primary aldosteronism: a consensus statement. *J Hypertens*. 2020;38(10):1937-1939. doi:10.1097/HJH.0000000000002568

12. Burrello J, Burrello A, Stowasser M, Nishikawa T, Quinkler M, Prejbisz A, Lenders JWM, Satoh F, Mulatero P, Reincke M, Williams TA. The Primary Aldosteronism Surgical Outcome Score for the Prediction of Clinical Outcomes After Adrenalectomy for Unilateral Primary Aldosteronism. *Ann Surg*. 2020;272(6):1125-1132. doi:10.1097/SLA.0000000000003200

13. 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(10). doi:10.1210/clinem/dgaa379

14. Yamashita T, Shimizu S, Koyama M, Ohno K, Mita T, Tobisawa T, Takada A, Togashi N, Ohnuma Y, Hasegawa T, Tsuchida A, Endo T, Ando T, Yoshida H, Fukuma S, Fukuhara S, Moniwa N, Miura T. Screening of primary aldosteronism by clinical features and daily laboratory tests: combination of urine pH, sex, and serum K. *J Hypertens*. 2018;36(2):326-334. doi:10.1097/HJH.0000000000001511

15. Kietsiriroje N, Wonghirundecha R, Suntornlohanakul O, Murray RD. Construction of a predictive scoring system as a guide to screening and confirmation of the diagnosis of primary aldosteronism. *Clin Endocrinol*. 2020;92(3):196-205. doi:10.1111/cen.14142

16. 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(10):1929-1936. doi:10.1097/HJH.0000000000002520

17. Williams B, Mancia G, Spiering W, Agabiti Rosei E, Azizi M, Burnier M, Clement DL, Coca A, de Simone G, Dominiczak A, Kahan T, Mahfoud F, Redon J, Ruilope L, Zanchetti A, Kerins M, Kjeldsen SE, Kreutz R, Laurent S, Lip GYH, McManus R, Narkiewicz K, Ruschitzka F, Schmieder RE, Shlyakhto E, Tsioufis C, Aboyans V, Desormais I, ESC Scientific Document Group. 2018 ESC/ESH Guidelines for the management of arterial hypertension. *Eur Heart J*. 2018;39(33):3021-3104. doi:10.1093/eurheartj/ehy339

18. Conn JW. Presidential address. I. Painting background. II. Primary aldosteronism, a new clinical syndrome. *J Lab Clin Med*. 1955;45(1):3-17.

19. 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(4):1025-1033. doi:10.1161/HYPERTENSIONAHA.119.14063

20. Mosso L, Carvajal C, González A, Barraza A, Avila F, Montero J, Huete A, Gederlini A, Fardella CE. Primary aldosteronism and hypertensive disease. *Hypertension*. 2003;42(2):161-165. doi:10.1161/01.HYP.0000079505.25750.11

21. Brown JM, Robinson-Cohen C, Luque-Fernandez MA, Allison MA, Baudrand R, Ix JH, Kestenbaum B, de Boer IH, Vaidya A. The Spectrum of Subclinical Primary Aldosteronism and Incident Hypertension: A Cohort Study. *Ann Intern Med*. 2017;167(9):630-641. doi:10.7326/M17-0882

22. Baudrand R, Guarda FJ, Fardella C, Hundemer G, Brown J, Williams G, Vaidya A. Continuum of Renin-Independent Aldosteronism in Normotension. *Hypertension*. 2017;69(5):950-956. doi:10.1161/HYPERTENSIONAHA.116.08952

23. 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(1):10-20. doi:10.7326/M20-0065

Novelty and Significance

What Is New?

We developed and validated a conventional scoring system and, for the first time, machine learning-based models to predict primary aldosteronism (PA) in a large cohort of more than 4-thousands patients with arterial hypertension.

What Is Relevant?

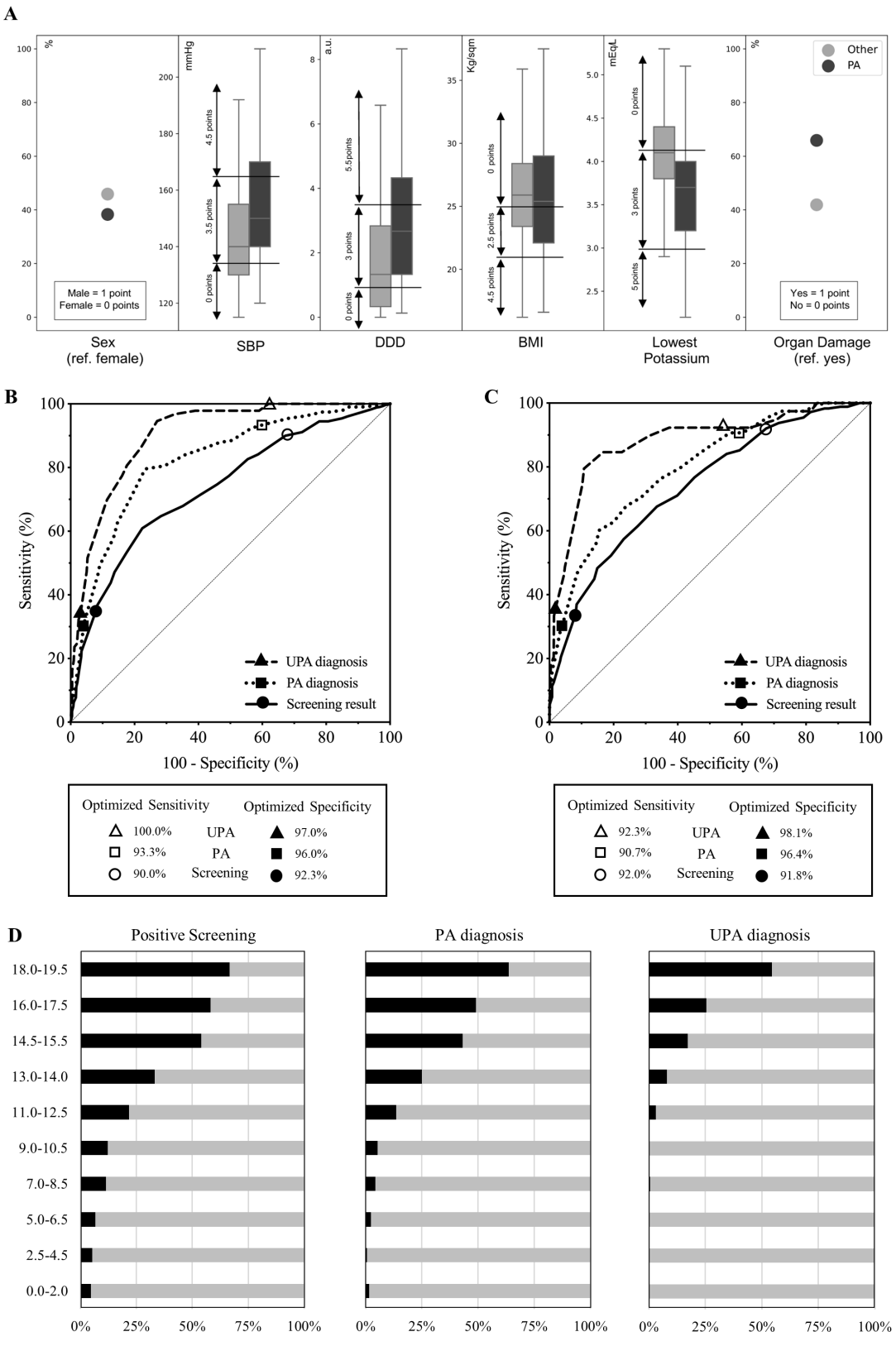
Both prediction models showed high sensitivity for the detection of patients with PA, with remarkably high diagnostic performances with the selected machine learning-based tool.

Summary

The clinical application of PA prediction models allows the identification of a subgroup of patients with very low probability of PA, avoiding unnecessary screening in up to one third of patients with arterial hypertension.

Legends to figures

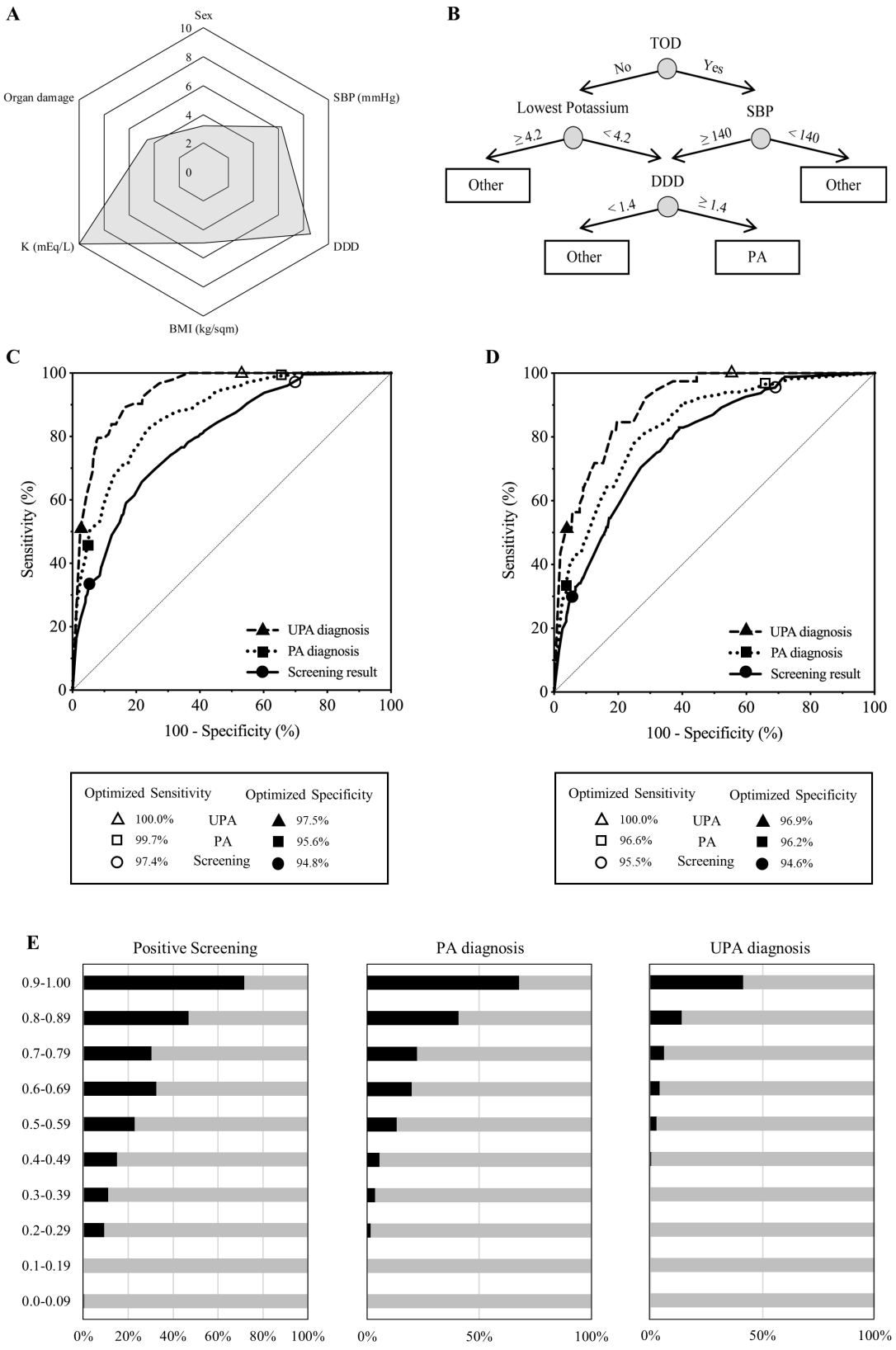
Figure 1. Development of the SToP-PA score


A
%
mmHg
a.u.
Kg/sqm
mEq/L
%
Other
PA
Male = 1 point
Female = 0 points
4.5 points
3.5 points
0 points
5.5points
3 points
0 points
0 points
2.5 points
4.5 points
0 points
3 points
5 points
Yes = 1 point
No = 0 points
Sex
(ref. female)
SBP
DDD
BMI
Lowest
Potassium
Organ Damage
(ref. yes)
B
Sensitivity (%)
100 - Specificity (%)
UPA diagnosis
PA diagnosis
Screening result
Optimized Sensitivity
Optimized Specificity
100.0% UPA 97.0%
93.3% PA 96.0%
90.0% Screening 92.3%
C
Sensitivity (%)
100 - Specificity (%)
UPA diagnosis
PA diagnosis
Screening result
Optimized Sensitivity
Optimized Specificity
92.3% UPA 98.1%
90.7% PA 96.4%
92.0% Screening 91.8%
D
Positive Screening
PA diagnosis
UPA diagnosis
18.0-19.5
16.0-17.5
14.5-15.5
13.0-14.0
11.0-12.5
9.0-10.5
7.0-8.5
5.0-6.5
2.5-4.5
0.0-2.0
0%
25%
50%
75%
100%

1

The PA prediction score was built in the training cohort (n=3,045) and tested in the validation cohort (n=1,045). (**A**) Cut-offs and score points for each variable after categorization (Primary Aldosteronism, PA, black; other patients, grey); the boxes indicate median and interquartile range. Sex (Female = 0 points; Male = 1 point); Systolic blood pressure (SBP; <135 mmHg = 0 points; 135-164 mmHg = 3.5 points; >/=165 mmHg = 4.5 points); Defined Daily Dose (DDD; <0.9 = 0 points; 0.9-3.59 = 3 points; >/=3.6 = 5.5 points); BMI (<21 = 4.5 points; 21-24.9 = 2.5 points; >/=25 = 0 points); Lowest Potassium (<3 = 5 points; 3-4 = 3 points; >/=4.1 = 0 points); Organ damage (No = 0 points; Yes = 1 point). ROC (Receiver Operating Characteristics) curve were used to assess the area under the curve (AUC) and the best cut-off for the SToP-PA score in the training (**B**) and validation cohort (**C**). ROC curve analysis was performed to detect patients with a positive screening result, a diagnosis of PA, or a diagnosis of unilateral PA (UPA). AUC for the training cohort were 0.903 (0.879-0.927) for UPA, 0.822 (0.798-0.847) for PA diagnosis and 0.734 (0.710-0.759) for screening positivity. At validation AUC were 0.882 (0.819-0.944) for UPA, 0.796 (0.754-0.839) for PA diagnosis and 0.739 (0.699-0.780) for screening positivity. Maximized sensitivity and specificity for UPA, PA diagnosis and positive screening test are reported below respective ROC curves. (**D**) Histograms showing the proportion of patients (x-axis, %) stratified according to their diagnosis (positive screening result, PA, or UPA, black vs. other patients, grey) in the overall cohort (n=4,059); the y-axis reports the assigned points for the prediction score. A user-friendly downloadable tool to apply the StoP-PA score is available at: https://github.com/CentroIpertenUnito/StoP-PA/raw/main/Score%20Calculator.xlsm

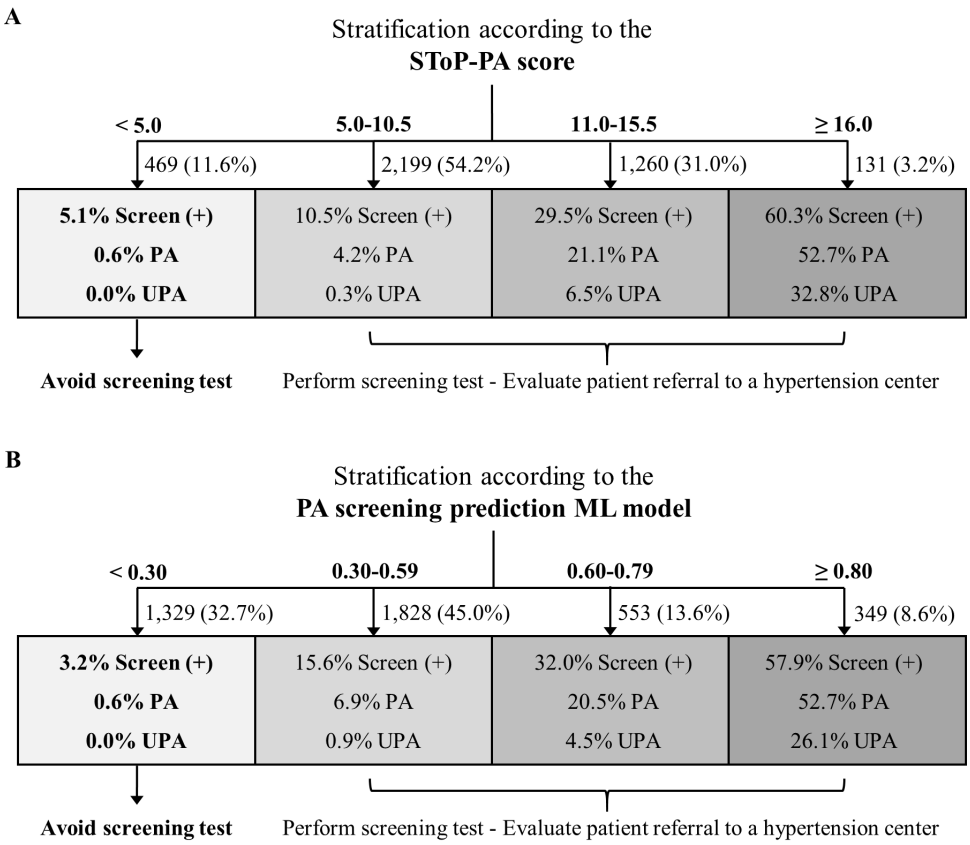
Figure 2. Development of the PA screening ML model



The PA prediction ML model was built in the training cohort (n=3,045) and tested in the validation cohort (n=1,045). (**A**) Radar charts reporting the 6 normalized predictors associated to the diagnosis of primary aldosteronism (PA). (**B**) Representative classification tree from the random forest model. ROC (Receiver Operating Characteristics) curve were used to assess the area under the curve (AUC) and the best cut-off for the PA screening ML model in the training (**C**) and validation cohort (**D**). ROC curve analysis was performed to detect patients with a positive screening result, a diagnosis of PA, or a diagnosis of unilateral PA (UPA). AUC for the training cohort were 0.938 (0.921-0.955) for UPA, 0.871 (0.853-0.890) for PA diagnosis and 0.796 (0.777-0.816) for screening positivity. At validation AUC were 0.905 (0.868-0.942) for UPA, 0.834 (0.797-0.871) for PA diagnosis and 0.786 (0.752-0.821) for screening positivity. Maximized sensitivity and specificity for UPA, PA diagnosis and positive screening test are reported below respective ROC curves. (E) Histograms showing the proportion of patients (x-axis, %) stratified according to their diagnosis (positive screening result, PA, or UPA, black vs. other patients, grey) in the overall cohort (n=4,059); the y-axis reports the ML model coefficients.

A user-friendly downloadable tool to apply the RFr PA prediction model is available at: https://github.com/CentroIpertenUnito/SToP-PA/raw/main/Prediction%20Tool.zip

Figure 3. Patient management



The panels show the suggested management of patient (n=4,059) according to their probability to have a positive screening test, a diagnosis of primary aldosteronism (PA), or a diagnosis of unilateral PA, after application of the SToP-PA score (**A**), or the PA screening ML model (**B**). Number and percentage of patients are indicated at each level of stratification.

Table 1. Characteristics of patients with primary aldosteronism

Variable	PA (n=431)	Other (n=3,628)	*P*-value
Age at screening (years)	50 ± 10.2	48 ± 12.9	**< 0.001**
Female sex, n (%)	165 (38.3)	1,666 (45.9)	**0.003**
Duration of HTN (months)	68 [22; 134]	41 [14; 101]	**< 0.001**
Systolic BP (mmHg)	157 ± 20.5	146 ± 18.5	**< 0.001**
Diastolic BP (mmHg)	95 ± 11.2	91 ± 10.8	**< 0.001**
Antihypertensive medication (DDD)	2.67 [1.31; 4.33]	1.33 [0.33; 2.88]	**< 0.001**
BMI (Kg/sqm)	25.7 ± 4.48	26.2 ± 4.32	**0.044**
Lowest Potassium (mEq/L)	3.6 ± 0.64	4.1 ± 0.42	**< 0.001**
Creatinine (mg/dL)	0.89 ± 0.214	0.91 ± 0.220	0.147
Diabetes, n (%)	31 (7.2)	260 (7.2)	0.984
Organ damage, n (%)	284 (65.9)	1,519 (41.9)	**< 0.001**
CV events, n (%)	60 (13.9)	323 (8.9)	**0.001**
PRA at screening (ng/mL/h)	0.20 [0.10; 0.40]	1.80 [0.80; 4.10]	**< 0.001**
Aldosterone at screening (ng/dL)	28.5 [21.1; 38.6]	16.9 [10.1; 25.7]	**< 0.001**

The table shows characteristics of patients with primary aldosteronism (PA; n=431) compared to the others (n=3,628). Variables are reported as mean ± standard deviation, median [interquartile range], or absolute number (percentage, %), as appropriated. Differences were considered significant when $p<0.05$ and reported in bold. HTN, Hypertension; BP, Blood Pressure; DDD, Defined Daily Dose (average maintenance dose per day for a drug used for its main indication in adults); CV, Cardiovascular; PRA, Plasma Renin Activity.