

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

Original

Clinical Score and Machine Learning-Based Model to Predict Diagnosis of Primary Aldosteronism in Arterial Hypertension / Buffolo, F., Burrello, J., Burrello, A., Heinrich, D., Adolf, C., Müller, L.M., Chen, R., Forestiero, V., Sconfienza, E., Tetti, M., Veglio, F., Williams, T.A., Mulatero, P., Monticone, S.. - In: HYPERTENSION. - ISSN 0194-911X. - 78:5(2021), pp. 1595-1604. [10.1161/HYPERTENSIONAHA.121.17444]

Availability:

This version is available at: 11583/2978537 since: 2023-05-16T13:05:38Z

Publisher:

LIPPINCOTT WILLIAMS & WILKINS

Published

DOI:10.1161/HYPERTENSIONAHA.121.17444

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

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

3 Fabrizio Buffolo^{1*}, Jacopo Burrello^{1*}, Alessio Burrello^{2*}, Daniel Heinrich³, Christian Adolf³,
4 Lisa Marie Müller³, Rusi Chen¹, Vittorio Forestiero¹, Elisa Sconfienza¹, Martina Tetti¹, Franco
5 Veglio¹, Tracy Ann Williams^{1,3}, Paolo Mulatero^{1†}, Silvia Monticone^{1†}.

6

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

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

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

14 **Corresponding author:** Prof. Paolo Mulatero Division of Internal Medicine and Hypertension
15 Unit, Department of Medical Sciences, University of Torino, Città della Salute e della Scienza,
16 Via Genova 3, 10126 Torino, Italy. Telephone/Fax number: +39.011.633.6959 /
17 +39.011.633.6931. E-mail: paolo.mulatero@unito.it

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

19 Abstract word count: 258

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

21 **Short title:** Prediction models for primary aldosteronism

22 **Acknowledgements:** none

23 **Sources of Funding:** nothing to disclose.

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

1 **Abstract**

2 Primary aldosteronism (PA) is the cause of arterial hypertension in 4-6% of patients, and 30%
3 of patients with PA are affected by unilateral and surgically-curable forms. Current guidelines
4 recommend screening for PA ~50% of patients with hypertension on the basis of individual
5 factors, while some experts suggest screening all patients with hypertension. To define the risk
6 of PA and tailor the diagnostic workup to the individual risk of each patient, we developed a
7 conventional scoring system and supervised machine learning algorithms using a retrospective
8 cohort of 4,059 patients with hypertension. On the basis of 6 widely available parameters, we
9 developed a numerical score and 308 machine learning-based models, selecting the one with
10 the highest diagnostic performance. After validation, we obtained high predictive performance
11 with our score (optimized sensitivity of 90.7% for PA and 92.3% for unilateral primary
12 aldosteronism [UPA]). The machine learning-based model provided the highest performance,
13 with an AUC of 0.834 for PA and 0.905 for diagnosis of UPA, with optimized sensitivity of
14 96.6% for PA and 100.0% for UPA, at validation. The application of the predicting tools
15 allowed the identification of a subgroup of patients with very low risk of PA (0.6 % for both
16 models) and null probability of having UPA. In conclusion, this score and the machine learning
17 algorithm can accurately predict the individual pre-test probability of PA in patients with
18 hypertension and circumvent screening in up to 32.7% of patients using a machine learning
19 based model, without omitting patients with surgically curable unilateral PA.

20

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

1 **Introduction**

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

11 Accumulating evidence indicates that chronic aldosterone excess causes an increased risk of
12 target organ damage, cardio and cerebrovascular events⁷, chronic kidney disease⁸ and overall
13 mortality⁹ in patients affected by PA compared with patients affected by essential hypertension.
14 Treatment with mineralocorticoid receptor antagonists and surgical adrenalectomy allow a
15 significant reduction of the excess risk.⁸⁻¹⁰ Therefore, recognition of the disease is crucial to
16 revert the detrimental effects of aldosterone in a timely manner. Some experts suggest
17 expanding screening for PA to all patients with arterial hypertension^{6,11}, which would however
18 increase the burden on primary care physicians. In this context, a prediction tool to reduce the
19 number of patients for PA screening would be highly desirable.

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

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

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

9

10 **Methods**

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

13

14 *Study Design and PA diagnosis*

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

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

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

17 Left ventricular hypertrophy (LVH) was defined as left ventricular mass index \geq 115 g/m²
18 (male) or \geq 95 g/m² (female).¹⁷ Microalbuminuria was defined as urine albumin concentration
19 of 30 to 300 mg/24 hour and/or by albumin to creatinine ratio between 30 and 300 mg/g.¹⁷

20

21 *Statistical Analysis*

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

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

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

13 *Diagnostic Modeling*

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

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

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

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

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

7

8 **Results**

9 *Characteristics of patients*

10 We included in the study a total of 4,059 patients: 706 (17.4%) with a positive screening test
11 for PA and 3,353 (82.6%) with a negative screening test. Clinical and biochemical
12 characteristics of patients at screening are shown in Table S1. After i.v. saline loading test or
13 captopril challenge test, 431 patients had a confirmed diagnosis of PA (10.6% of the total
14 cohort). A total of 130 patients had UPA (30.2%), 247 bilateral PA (57.3%), while in 54
15 (12.5%) patients subtype diagnosis was not achieved, because the adrenal venous sampling
16 was unsuccessful or patients did not favor AVS.

17 At screening, patients with PA were slightly older than non-PA, more frequently male and with
18 longer duration of hypertension (Table 1). Patients affected by PA had higher systolic (157±20
19 versus 146±18 mmHg; $P<0.001$) and diastolic BP (95±11 versus 91±11 mmHg; $P<0.001$),
20 higher intensity of antihypertensive treatment (defined by daily defined dose [DDD]) (2.67
21 [1.31-4.33] versus 1.33 [0.33; 2.88]; $P<0.001$) and lower potassium levels (3.6±0.6 versus
22 4.1±0.4 mEq/L; $P<0.001$). Patients with PA had slightly lower BMI and a higher rate of organ
23 damage, defined as left ventricular hypertrophy and/or microalbuminuria and cardiovascular
24 events (13.9% versus 8.9%; $P=0.001$). As expected, PRA was significantly lower and serum
25 aldosterone significantly higher in patients with PA (Table 1).

1 Association of clinical and biochemical characteristics with PA diagnosis has been evaluated
2 by univariate logistic regression (Table S2). A significant association was observed between
3 diagnosis of PA and 10 of 12 patient parameters: age (OR 1.015), male sex (OR 1.368),
4 duration of hypertension (OR 1.003), systolic and diastolic BP (OR 1.025 and 1.033,
5 respectively), anti-hypertensive treatment (OR 1.344), BMI (OR 0.975), lowest potassium
6 levels (OR 0.120), presence of organ damage (OR 2.682) and cardiovascular events (OR
7 1.655). Parameters which were significantly associated with PA diagnosis at univariate
8 analysis were introduced into a multivariate model: male sex, systolic BP, anti-hypertensive
9 treatment, BMI, lowest potassium and organ damage were confirmed as independent predictors
10 of PA (Table S3). These parameters were used as input variables to develop a numerical scoring
11 system and a prediction model based on supervised machine learning.

12

13 *Development and validation of the SToP-PA score*

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

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

1 Figure 1B) and validation (AUC 0.739, 0.796, and 0.882; Figure 1C; Table S6). As expected,
2 the highest performance was reached for the diagnosis of UPA. Performance of SToP-PA score
3 were similar when applied within the validation cohort in patients with and without specific
4 indication for PA screening according to international guidelines (Tables S7-S8-S9)^{1,2}.
5 Additionally, the performances remain similar, in both training and validation cohort, adopting
6 more stringent criteria for screening positivity ($ARR > 40 \text{ ng/dL/ng} \cdot \text{mL}^{-1} \cdot \text{h}^{-1}$ and $AC > 15$
7 ng/dL) and confirmatory test ($AC \text{ post-intravenous saline load test} > 10 \text{ ng/dL}$ or $AC \text{ post-}$
8 $\text{captopril challenge test} > 11 \text{ ng/dL}$) (Tables S10-S11).

9 Diagnostic performance for the detection of screening positivity, PA and UPA diagnosis with
10 different cut-offs are reported in Table S12 for both, training and validation cohorts. The cut-
11 offs with optimized sensitivity were 7.0, 7.5, and 8.0, for prediction of positive screening
12 results, PA, or UPA diagnosis, with respective sensitivity of 90.0%, 93.3% and 100.0% at
13 training and 92.0%, 90.7% and 92.3% at validation. Comparison of performances of SToP-PA
14 score at training and validation revealed a negligible overfitting bias (1.2-2%). Cut-offs of 13.5,
15 14.0 and 15.0 optimized the specificity for the detection of patients with positive screening test,
16 PA, or UPA respectively, achieving a specificity of 92.3% for positive screening, 96.0% for
17 PA and 97.0% for UPA diagnosis at training and 91.8%, 96.4% and 98.1% and validation.

18

19 *Supervised machine learning*

20 Four supervised learning classifiers (LDA, RFr, 1-SVM and rbf-SVM) with or without 3
21 different methods for dataset imbalance correction were applied to the training cohort. A total
22 of 308 models were generated and after tuning of hyperparameters, a RFr composed by 50
23 classification trees and a maximum number of splits equal to 10, with random oversampling
24 correction, displayed the highest accuracy for detection of patients with PA (Tables S13-S14).

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

5 At ROC curves analysis, RFr-model showed a reliable performance, with an AUC of 0.796 for
6 screening results, 0.871 for PA, and 0.938 for UPA diagnosis, with similar performance at
7 validation and very low overfitting bias (1.0%, 1.8% and 1.7% respectively) (Figure 2C-D).

8 Performance of RFr-model were similar when applied within the validation cohort in patients
9 with and without specific indication for PA screening according to international guidelines
10 (Tables S7-S8-S9)^{1,2}, and adopting more stringent criteria for screening positivity (ARR>40
11 ng/dL/ng*mL⁻¹*h⁻¹ and AC>15 ng/dL) and confirmatory test (AC post-intravenous saline load
12 test>10 ng/dL or AC post-captopril challenge test>11 ng/dL), for both training and validation
13 cohort (Tables S15-S16).

14 Diagnostic performances and confusion matrix for the RFr-model are reported in Table S17.

15 The RFr allowed the calculation of a classification coefficient for each patient. The ML
16 prediction model displayed optimized sensitivity with a coefficient cut-off of 0.23 for positive
17 screening, 0.24 for PA, and 0.35 for UPA detection, with sensitivity of 97.4% for screening
18 positivity, 99.7% for PA and 100.0% for UPA diagnosis at training and 95.5%, 96.6% and
19 100.0% at validation. The cut-offs for optimized specificity were 0.70 for positive screening
20 (94.8% specificity at training and 94.6% at validation), 0.76 for PA (95.6% specificity at
21 training and 96.2% at validation) and 0.81 for UPA detection (97.5% specificity at training and
22 96.9% at validation).

23

24 *Pre-test probability stratification*

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

6 Considering risk distribution, patients were stratified in 4 subgroups with an increasing pre-
7 test probability of PA. Then, we integrated SToP-PA score and the RFr prediction model in a
8 flowchart for the management of patients with arterial hypertension (Figure 3). Applying the
9 flowchart to the overall cohort, patients with SToP-PA score < 5.0 (469 patients) had a pre-test
10 probability of 0.6% and 0.0% of PA or UPA, respectively (Figure 3A). Applying the ML
11 model, patients with a RFr coefficient < 0.30 (1,329 patients) had a probability of 0.6% and
12 0.0% of PA or UPA, respectively (Figure 3B). Considering the very low prevalence of PA and
13 the absence of unilateral surgically curable form, we propose to avoid screening test in this
14 subgroup of patients. On the counterpart, we suggest to screen all the patients with SToP-PA
15 score ≥ 5.0 or RFr coefficient ≥ 0.30 . Among these patients, the probability of having PA
16 increases progressively in the groups with intermediate risk: 4.2% with SToP-PA score of 5-
17 10.5 and 21.1% with SToP-PA score 11.0-15.5, with a prevalence of UPA of 0.3% and 6.5%
18 respectively. Similarly, RFr coefficient progressively stratifies the risk of PA in patients with
19 intermediate probability: 6.9% of PA and 0.9% of UPA with RFr coefficient between 0.30 and
20 0.59, and 20.5% of PA and 4.5% of UPA between 0.60 and 0.79. Finally, patients with SToP-
21 PA score ≥ 16.0 or RFr coefficient ≥ 0.80 display a very high probability of having PA (52.7%
22 for both SToP-PA score and RFr-model) and UPA (32.8% with SToP-PA score and 26.1%
23 with RFr-model). Clinicians should consider direct referral to hypertension centers for these
24 patients.

25 *External Validation*

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

12

13 **Discussion**

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

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

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

12 PA prevalence increases with the severity of hypertension²⁰: ES guideline recommends
13 screening for PA in patients with SBP \geq 150 mmHg and DBP \geq 100 mmHg. Similarly, the recent
14 ESH consensus recommend screening of patients with grade 2 and 3 of hypertension (SBP \geq 160
15 mmHg and DBP \geq 100 mmHg).¹ However, while definition of hypertension grade presents no
16 difficulty in never-treated patients with newly diagnosed hypertension, it is less straightforward
17 in patients under hypertensive treatment. The importance of anti-hypertensive therapy is only
18 partially considered by current guidelines, who recommend screening for patients with resistant
19 hypertension (3 or more anti-hypertensive drugs at full-dose, including a diuretic), but do not
20 consider lower number of anti-hypertensive drugs or intermediate dosage. SToP-PA score and
21 RFr-model take into account both blood pressure levels and intensity of anti-hypertensive
22 treatment (defined by DDD), allowing a more accurate risk stratification also in patients under
23 anti-hypertensive treatment. Moreover, in our study DDD was calculated with medical therapy
24 at the first visit, that included potentially interfering drugs, subsequently stopped or substituted
25 by non-interfering ones for screening test. Considering the reluctance of general practitioners

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

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

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

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

1 Kietsiriroje et al. proposed a second score based on age, BMI, diabetes, antihypertensive
2 treatments, serum sodium and potassium, built on 420 patients with arterial hypertension and
3 displaying an AUC 0.87.¹⁵ The major limitation of the study was the absence of a validation
4 cohort, hampering the generalizability of the model. None of the studies performed an external
5 validation.

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

9 Recent studies progressively expanded the spectrum of PA towards milder forms, highlighting
10 the continuum from earliest form of renin independent aldosteronism towards overt PA.²¹⁻²³

11 On the basis of these findings, some authors proposed to widen the cohort of subjects eligible
12 for PA screening to all patients with hypertension.^{6,11} This choice would have two
13 consequences: on one side no patients with PA would be missed and could benefit of target
14 medical or surgical therapy, according to subtype diagnosis; on the other side, general
15 practitioners should face an increasing number of patients undergoing PA screening. The
16 medical cost of a screening test is not high *per se*, but its application in a large scale would
17 undoubtedly amplify its economic impact for the management of the patients with a positive
18 screening test in term of confirmatory/exclusion tests and subtype diagnosis, including CT
19 scanning and adrenal vein sampling. The application of the SToP-PA score and RFr-model
20 allows the identification of a subgroup of patients with very low pre-test probability of having
21 PA, that can avoid PA screening, significantly reducing the burden on general practitioners and
22 hypertension centers. The application of SToP-PA score would reduce by 11.6% the number
23 patients undergoing PA screening, with strikingly higher reduction (32.7%) with RFr-model.
24 At the same time, SToP-PA score and RFr-model identify subgroups of patient with
25 intermediate-high and very high risk of having PA and UPA. Taken together, these two

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

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

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

17

18 **Perspective**

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

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

1 **References**

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

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

- 12 3. Ruhle BC, White MG, Alsafran S, Kaplan EL, Angelos P, Grogan RH. Keeping primary
13 aldosteronism in mind: Deficiencies in screening at-risk hypertensives. *Surgery*.
14 2019;165(1):221-227. doi:10.1016/j.surg.2018.05.085

- 15 4. Mulatero P, Monticone S, Burrello J, Veglio F, Williams TA, Funder J. Guidelines for
16 primary aldosteronism: uptake by primary care physicians in Europe. *J Hypertens*.
17 2016;34(11):2253-2257. doi:10.1097/HJH.0000000000001088

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

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

- 1 7. Monticone S, D'Ascenzo F, Moretti C, Williams TA, Veglio F, Gaita F, Mulatero P.
2 Cardiovascular events and target organ damage in primary aldosteronism compared with
3 essential hypertension: a systematic review and meta-analysis. *Lancet Diabetes*
4 *Endocrinol.* 2018;6(1):41-50. doi:10.1016/S2213-8587(17)30319-4
- 5 8. Monticone S, Sconfienza E, D'Ascenzo F, Buffolo F, Satoh F, Sechi LA, Veglio F,
6 Mulatero P. Renal damage in primary aldosteronism: a systematic review and meta-
7 analysis. *J Hypertens.* 2020;38(1):3-12. doi:10.1097/HJH.0000000000002216
- 8 9. Hundemer GL, Curhan GC, Yozamp N, Wang M, Vaidya A. Cardiometabolic outcomes
9 and mortality in medically treated primary aldosteronism: a retrospective cohort study.
10 *Lancet Diabetes Endocrinol.* 2018;6(1):51-59. doi:10.1016/S2213-8587(17)30367-4
- 11 10. Hundemer GL, Curhan GC, Yozamp N, Wang M, Vaidya A. Incidence of Atrial
12 Fibrillation and Mineralocorticoid Receptor Activity in Patients With Medically and
13 Surgically Treated Primary Aldosteronism. *JAMA Cardiol.* 2018;3(8):768-774.
14 doi:10.1001/jamacardio.2018.2003
- 15 11. Funder JW. Primary aldosteronism: a consensus statement. *J Hypertens.*
16 2020;38(10):1937-1939. doi:10.1097/HJH.0000000000002568
- 17 12. Burrello J, Burrello A, Stowasser M, Nishikawa T, Quinkler M, Prejbisz A, Lenders
18 JWM, Satoh F, Mulatero P, Reincke M, Williams TA. The Primary Aldosteronism
19 Surgical Outcome Score for the Prediction of Clinical Outcomes After Adrenalectomy
20 for Unilateral Primary Aldosteronism. *Ann Surg.* 2020;272(6):1125-1132.
21 doi:10.1097/SLA.0000000000003200
- 22 13. Burrello J, Burrello A, Pieroni J, Sconfienza E, Forestiero V, Rabbia P, Adolf C,
23 Reincke M, Veglio F, Williams TA, Monticone S, Mulatero P. Development and

- 1 Validation of Prediction Models for Subtype Diagnosis of Patients With Primary
2 Aldosteronism. *J Clin Endocrinol Metab.* 2020;105(10). doi:10.1210/clinem/dgaa379
- 3 14. Yamashita T, Shimizu S, Koyama M, Ohno K, Mita T, Tobisawa T, Takada A, Togashi
4 N, Ohnuma Y, Hasegawa T, Tsuchida A, Endo T, Ando T, Yoshida H, Fukuma S,
5 Fukuhara S, Moniwa N, Miura T. Screening of primary aldosteronism by clinical
6 features and daily laboratory tests: combination of urine pH, sex, and serum K. *J*
7 *Hypertens.* 2018;36(2):326-334. doi:10.1097/HJH.0000000000001511
- 8 15. Kietsiroje N, Wonghirundecha R, Suntornlohanakul O, Murray RD. Construction of a
9 predictive scoring system as a guide to screening and confirmation of the diagnosis of
10 primary aldosteronism. *Clin Endocrinol.* 2020;92(3):196-205. doi:10.1111/cen.14142
- 11 16. Mulatero P, Sechi LA, Williams TA, Lenders JWM, Reincke M, Satoh F, Januszewicz
12 A, Naruse M, Doumas M, Veglio F, Wu VC, Widimsky J. Subtype diagnosis, treatment,
13 complications and outcomes of primary aldosteronism and future direction of research:
14 a position statement and consensus of the Working Group on Endocrine Hypertension of
15 the European Society of Hypertension. *J Hypertens.* 2020;38(10):1929-1936.
16 doi:10.1097/HJH.0000000000002520
- 17 17. Williams B, Mancia G, Spiering W, Agabiti Rosei E, Azizi M, Burnier M, Clement DL,
18 Coca A, de Simone G, Dominiczak A, Kahan T, Mahfoud F, Redon J, Ruilope L,
19 Zanchetti A, Kerins M, Kjeldsen SE, Kreutz R, Laurent S, Lip GYH, McManus R,
20 Narkiewicz K, Ruschitzka F, Schmieder RE, Shlyakhto E, Tsioufis C, Aboyans V,
21 Desormais I, ESC Scientific Document Group. 2018 ESC/ESH Guidelines for the
22 management of arterial hypertension. *Eur Heart J.* 2018;39(33):3021-3104.
23 doi:10.1093/eurheartj/ehy339

- 1 18. Conn JW. Presidential address. I. Painting background. II. Primary aldosteronism, a new
2 clinical syndrome. *J Lab Clin Med.* 1955;45(1):3-17.
- 3 19. Burrello J, Monticone S, Losano I, Cavaglià G, Buffolo F, Tetti M, Covella M, Rabbia
4 F, Veglio F, Pasini B, Williams TA, Mulatero P. Prevalence of Hypokalemia and
5 Primary Aldosteronism in 5100 Patients Referred to a Tertiary Hypertension Unit.
6 *Hypertension.* 2020;75(4):1025-1033. doi:10.1161/HYPERTENSIONAHA.119.14063
- 7 20. Mosso L, Carvajal C, González A, Barraza A, Avila F, Montero J, Huete A, Gederlini
8 A, Fardella CE. Primary aldosteronism and hypertensive disease. *Hypertension.*
9 2003;42(2):161-165. doi:10.1161/01.HYP.0000079505.25750.11
- 10 21. Brown JM, Robinson-Cohen C, Luque-Fernandez MA, Allison MA, Baudrand R, Ix JH,
11 Kestenbaum B, de Boer IH, Vaidya A. The Spectrum of Subclinical Primary
12 Aldosteronism and Incident Hypertension: A Cohort Study. *Ann Intern Med.*
13 2017;167(9):630-641. doi:10.7326/M17-0882
- 14 22. Baudrand R, Guarda FJ, Fardella C, Hundemer G, Brown J, Williams G, Vaidya A.
15 Continuum of Renin-Independent Aldosteronism in Normotension. *Hypertension.*
16 2017;69(5):950-956. doi:10.1161/HYPERTENSIONAHA.116.08952
- 17 23. Brown JM, Siddiqui M, Calhoun DA, Carey RM, Hopkins PN, Williams GH, Vaidya A.
18 The Unrecognized Prevalence of Primary Aldosteronism: A Cross-sectional Study. *Ann*
19 *Intern Med.* 2020;173(1):10-20. doi:10.7326/M20-0065

20

1 **Novelty and Significance**

2 *What Is New?*

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

6 *What Is Relevant?*

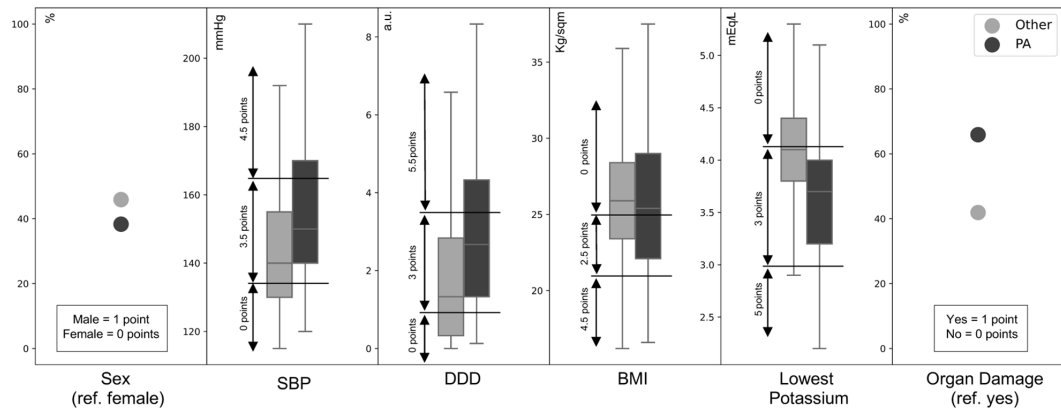
7 Both prediction models showed high sensitivity for the detection of patients with PA, with
8 remarkably high diagnostic performances with the selected machine learning-based tool.

9 *Summary*

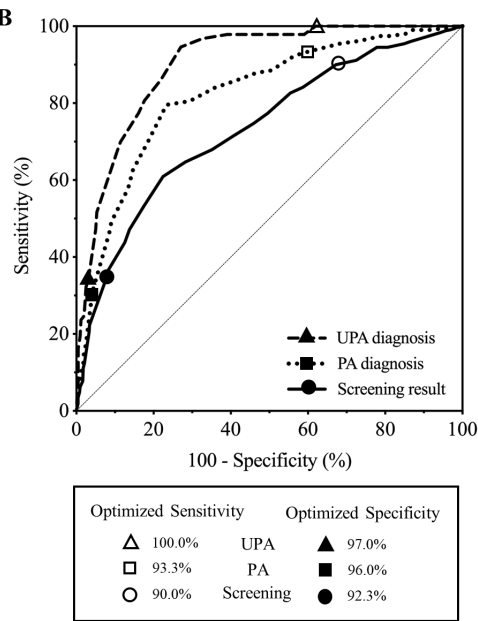
10 The clinical application of PA prediction models allows the identification of a subgroup of
11 patients with very low probability of PA, avoiding unnecessary screening in up to one third of
12 patients with arterial hypertension.

- 1 **Legends to figures**
- 2 **Figure 1. Development of the SToP-PA score**

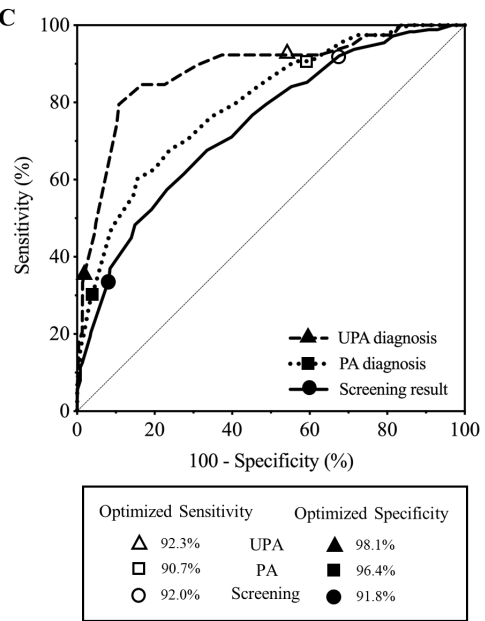
A



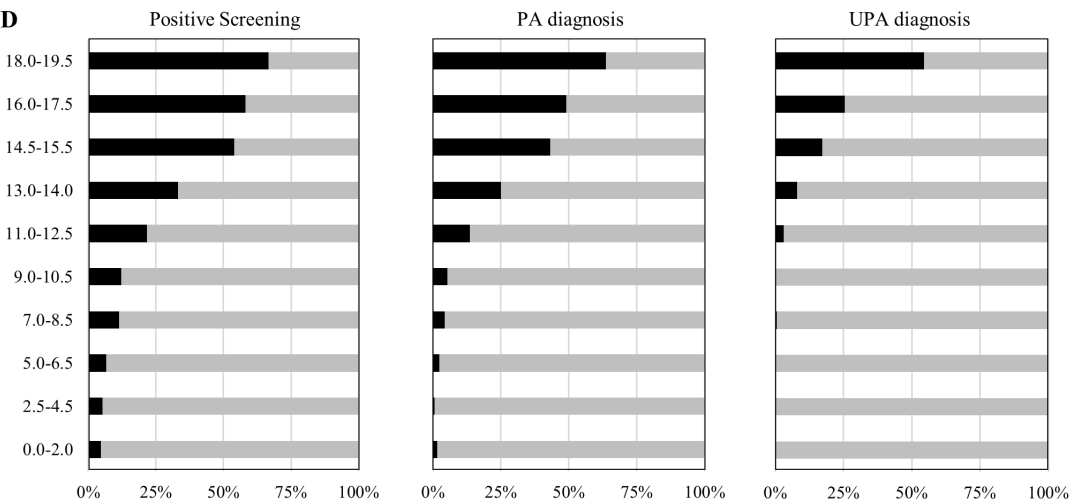
B



C



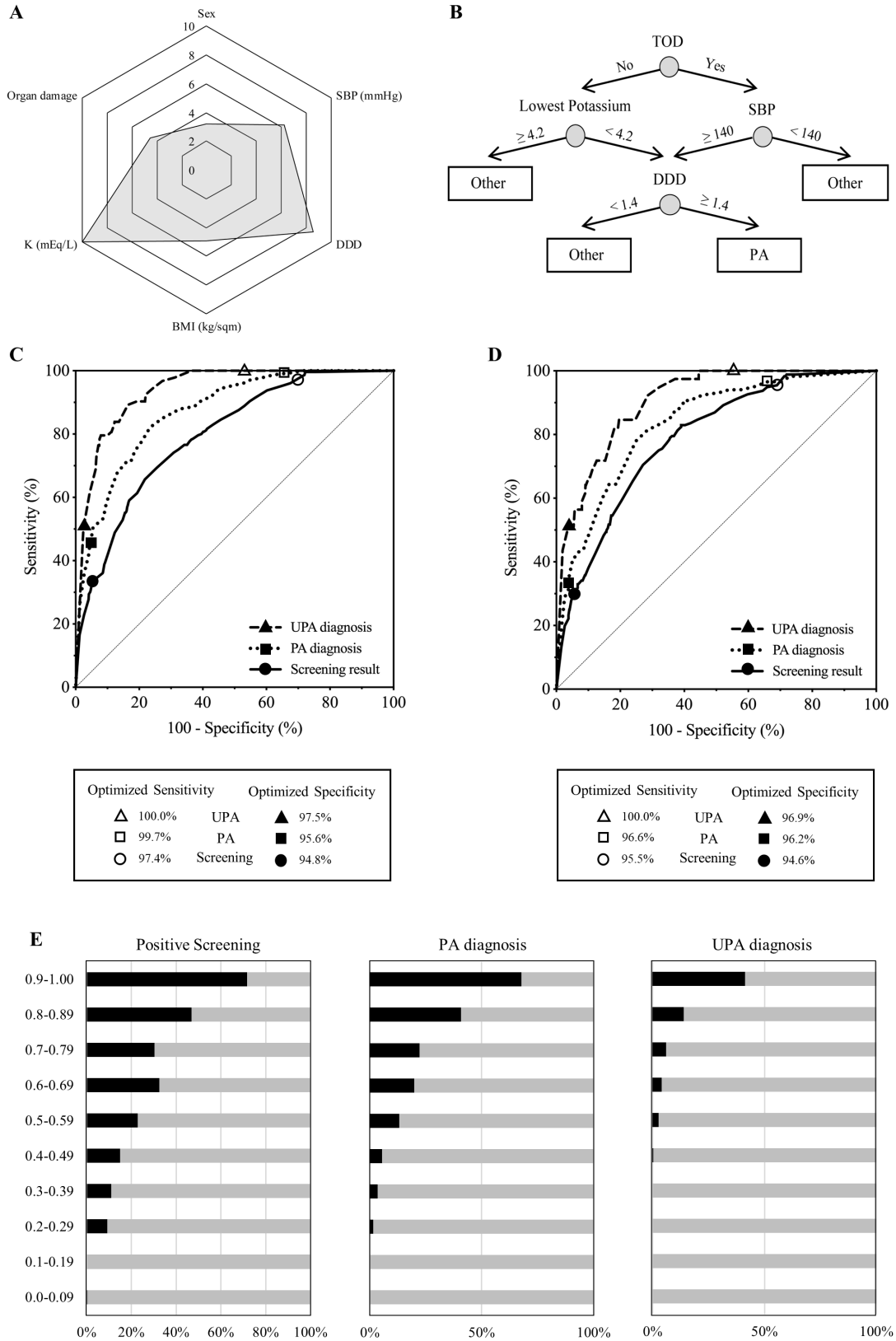
D



1 The PA prediction score was built in the training cohort (n=3,045) and tested in the validation
2 cohort (n=1,045). (A) Cut-offs and score points for each variable after categorization (Primary
3 Aldosteronism, PA, black; other patients, grey); the boxes indicate median and interquartile
4 range. Sex (Female = 0 points; Male = 1 point); Systolic blood pressure (SBP; <135 mmHg =
5 0 points; 135-164 mmHg = 3.5 points; \geq 165 mmHg = 4.5 points); Defined Daily Dose
6 (DDD; <0.9 = 0 points; 0.9-3.59 = 3 points; \geq 3.6 = 5.5 points); BMI (<21 = 4.5 points; 21-
7 24.9 = 2.5 points; \geq 25 = 0 points); Lowest Potassium (<3 = 5 points; 3-4 = 3 points; \geq 4.1
8 = 0 points); Organ damage (No = 0 points; Yes = 1 point). ROC (Receiver Operating
9 Characteristics) curve were used to assess the area under the curve (AUC) and the best cut-off
10 for the SToP-PA score in the training (B) and validation cohort (C). ROC curve analysis was
11 performed to detect patients with a positive screening result, a diagnosis of PA, or a diagnosis
12 of unilateral PA (UPA). AUC for the training cohort were 0.903 (0.879-0.927) for UPA, 0.822
13 (0.798-0.847) for PA diagnosis and 0.734 (0.710-0.759) for screening positivity. At validation
14 AUC were 0.882 (0.819-0.944) for UPA, 0.796 (0.754-0.839) for PA diagnosis and 0.739
15 (0.699-0.780) for screening positivity. Maximized sensitivity and specificity for UPA, PA
16 diagnosis and positive screening test are reported below respective ROC curves. (D)
17 Histograms showing the proportion of patients (x-axis, %) stratified according to their
18 diagnosis (positive screening result, PA, or UPA, black vs. other patients, grey) in the overall
19 cohort (n=4,059); the y-axis reports the assigned points for the prediction score. A user-friendly
20 downloadable tool to apply the StoP-PA score is available at:
21 <https://github.com/CentroIpertenUnito/StoP-PA/raw/main/Score%20Calculator.xlsm>

22
23
24
25

1 **Figure 2. Development of the PA screening ML model**



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

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

18

19

20

21

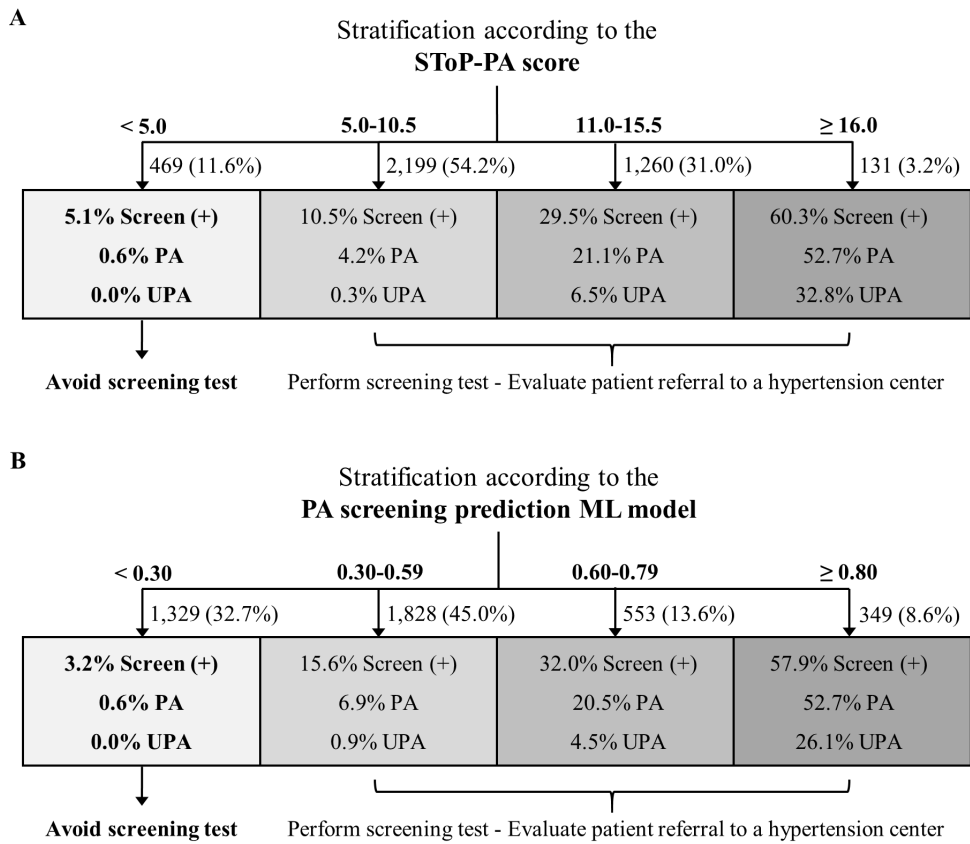
22

23

24

25

1 **Figure 3. Patient management**



2

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

1 **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

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