SUPPLEMENTAL DATA

Development and validation of prediction models for the subtype diagnosis of patients with primary aldosteronism.

Jacopo Burrello MD[1*], Alessio Burrello MS[2*], Jacopo Pieroni MD[1], Elisa Sconfienza MD[1], Vittorio Forestiero MD[1], Paola Rabbia MD[3], Christian Adolf MD[4], Martin Reincke MD[4], Franco Veglio MD[1], Tracy Ann Williams PhD[1,4], Silvia Monticone PhD[1#], Paolo Mulatero MD[1#].

(1) Division of Internal Medicine and Hypertension, Department of Medical Sciences, University of Torino, Italy. (2) Department of Electrical, Electronic and Information Engineering "Guglielmo Marconi" (DEI), University of Bologna, Italy. (3) Division of Radiology, University of Torino, Italy. (4) Medizinische Klinik und Poliklinik IV, Klinikum der Universität, Ludwig-Maximilians-Universität München, Munich, Germany.

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

Summary

Extended Methods

Diagnosis of Primary Aldosteronism

Primary aldosteronism (PA) was diagnosed in agreement with the Endocrine Society guideline in both the Torino and Munich Hypertension Units [1]. Patients were screened using the aldosterone concentration (AC) to plasma renin activity (PRA) ratio (ARR). If possible, all interfering drugs were withdrawn for at least 4 weeks (6 weeks for diuretics and mineralocorticoid receptor antagonists). Calcium-channel blockers and/or doxazosin were used to control blood pressure when the withdrawal of all antihypertensive drugs was not possible. An ARR greater than 30 ng/dL/ng*mL^{-1}*h^{-1} together with an AC greater than 10 ng/dL were considered as cut-off for a positive screening test. The diagnosis of PA was confirmed by either intravenous saline loading test or a captopril challenge test (Tables 1, S3, and S6), as previously described [2]. Intravenous saline loading test was performed in recumbent position until April 2014, and in seated position from that moment forward [3]; the cut-off chosen for PA confirmation was a post-test AC greater than 5 ng/dL. The cut-off for a positive captopril challenge test was a post-test ARR greater than 30 ng/dL/ng*mL^{-1}*h^{-1}. Patients with a confirmed diagnosis of PA underwent subtype differentiation through computed tomography (CT) scanning and AVS. At the Torino Hypertension Unit, all patients were screened for glucocorticoid-remediable aldosteronism using long PCR technique. Adrenal CT scanning was interpreted as follows: I) nodule: defined as an adrenal mass ≥ 8 mm in diameter; II) unilaterally abnormal: in the presence of a thickening > 4 mm and/or the presence of a nodule (as previously defined) on one side; III) bilaterally abnormal: in the presence of any combination of nodule or thickening > 4 mm on both sides (i.e. nodule on one side + contralateral thickening; bilateral nodules; bilateral thickening); IV) bilaterally normal: absence of any lesion (a thickening up to 4 mm was considered as normal). AVS was performed either with and/or without ACTH infusion (Tables 1, S3, and S6) by the same expert radiologist and was considered successful if the adrenal veins/inferior vena cava cortisol gradients were at least 2 (in Munich) or 3 (in Torino; selectivity index without ACTH) or 5 (selectivity index with ACTH). Lateralization was defined when the aldosterone to cortisol ratio from one adrenal was at least 4 times than the ratio from the other adrenal gland (lateralization index) [4]. For the Torino cohort, AVS was performed under basal conditions in 80 patients, after cosyntropin continuous infusion in 83, and under both unstimulated and after cosyntropin infusion in 52 patients. All patients from the Munich cohort underwent AVS under basal conditions. The diagnosis of unilateral PA was confirmed after pathology evaluation and ICH staining for CYP11B2; none of the adrenalectomized patients displayed absent biochemical success after surgery with the PASO criteria [2].

Diagnostic modelling

Supervised machine learning algorithms were used to evaluate the diagnostic performance of 6 selected variables (AC at screening and after confirmatory testing, lowest potassium, presence/absence of a nodule at CT scanning imaging, nodule diameter, and descriptive CT scanning finding) in predicting the diagnosis of unilateral PA. Machine learning, and in particular linear discriminant analysis (LDA) and random forest (RF) classification algorithms are commonly used in clinical research to formulate predictions about possible outcomes based on a pre-defined set of labeled paired input-output training sample data [5;6].

LDA employs linear combinations of variables to maximize the separation between groups by increasing precision estimates by variance reduction. In the model used herein, the algorithm computes a set of coefficients (Supplementary Table 2) for linear combination of each variable to determine the single patient diagnosis. The predicted diagnosis is derived from the following equation: Unilateral PA diagnosis = $LDAcoeff_1$*$Variable_1$ + $LDAcoeff_2$*$Variable_2$ + … + $LDAcoeff_n$*$Variable_n$ > 0.8299.

The canonical plot represents patient distribution after stratification for the linear combination of the 6 selected variables (Figure 1A). The canonical axes of the plot are calculated by the LDA from

weighted linear combination of variables included in the model; each patient is indicated by a point. The crosses indicate the means of (canonical 1; canonical 2) for each group (unilateral *versus* bilateral PA), whereas the ellipses include patients with a linear combination coefficient that falls within the mean +/- SD (canonical 1 +/- SD; canonical 2 +/- SD).

The RF algorithm uses 20 different classification trees with a maximum number of 8 splits for each tree. The predicted diagnosis resulted from the outcome of each classification tree of the forest; if at least 11 of 20 trees of the RF predict unilateral PA, then the patient will be classified as unilateral disease. A representative classification tree is reported in Figure 2A.

Confusion matrix and histogram of 0-1 normalized predictive coefficients are reported for each model (Figure 1B, 1C; Figure 2B, 2C). The machine learning models were internally tested by a 10K-cross validation algorithm. The algorithm randomly divides the cohort into 10 groups; the model is then trained within the first 9 groups, and the remaining group is used for validation. The process is repeated 10 times, with the validation group rotating at each round. The accuracy of internal validation resulted from the mean of the accuracies obtained at each round.

REFERENCES

1. Funder JW, Carey RM, Mantero F, et al. The Management of Primary Aldosteronism: Case Detection, Diagnosis, and Treatment: An ES Clinical Practice Guideline. *J Clin Endocrinol Metab.* 2016;101:1889-916.
2. Williams TA, Lenders JWM, Mulatero P, et al; Primary Aldosteronism Surgery Outcome (PASO) investigators. Outcomes after adrenalectomy for unilateral primary aldosteronism: an international consensus on outcome measures and analysis of remission rates in an international cohort. *Lancet Diabetes Endocrinol.* 2017;5:689-699.
3. Ahmed AH, Cowley D, Wolley M, et al. Seated saline suppression testing for the diagnosis of primary aldosteronism: a preliminary study. *J Clin Endocrinol Metab.* 2014;99:2745-53.
4. Monticone S, Viola A, Rossato D, et al. Adrenal vein sampling in primary aldosteronism: towards a standardised protocol. *Lancet Diabetes Endocrinol.* 2015;3:296-303.
5. Burrello J, Burrello A, Stowasser M, et al. The Primary Aldosteronism Surgical Outcome Score for the Prediction of Clinical Outcomes After Adrenalectomy for Unilateral Primary Aldosteronism. *Ann Surg.* 2019. [Epub ahead of print]
6. Yang Y, Burrello J, Burrello A, et al. Classification of microadenomas in patients with primary aldosteronism by steroid profiling. *J Steroid Biochem Mol Biol.* 2019;189:274-282.
7. Küpers EM, Amar L, Raynaud A, et al. A clinical prediction score to diagnose unilateral primary aldosteronism. *J Clin Endocrinol Metab.* 2012;97:3530-7.
8. Nanba K, Tsuiki M, Nakao K, et al. A subtype prediction score for primary aldosteronism. *J Hum Hypertens.* 2014;28:716-20.
9. Kocjan T, Janez A, Stankovic M, et al. A new clinical prediction criterion accurately determines a subset of patients with bilateral primary aldosteronism before adrenal venous sampling. *Endocr Pract.* 2016;22:587-94.
10. Kobayashi H, Haketa A, Ueno T, et al. Scoring system for the diagnosis of bilateral primary aldosteronism in the outpatient setting before adrenal venous sampling. *Clin Endocrinol (Oxf).* 2017;86:467-472.
11. Kamemura K, Wada N, Ichijo T, et al. Significance of adrenal computed tomography in predicting laterality and indicating adrenal vein sampling in primary aldosteronism. *J Hum Hypertens.* 2017;31:195-199.
12. Kobayashi H, Abe M, Soma M, et al; JPAS Study Group. Development and validation of subtype prediction scores for the workup of primary aldosteronism. *J Hypertens*. 2018;36:2269-2276.
13. Leung HT, Woo YC, Fong CHY, et al. A clinical prediction score using age at diagnosis and saline infusion test parameters can predict aldosterone-producing adenoma from idiopathic adrenal hyperplasia. *J Endocrinol Invest.* 2019. [Epub ahead of print].

Supplementary Table 1. Patient Characteristics of Study Cohort

Variable (ref. UPA)	OR (CI 95%)	*P*-value
Sex (ref. female)	2.41 (1.30 – 4.47)	**0.005**
Age at diagnosis (years)	0.98 (0.96 – 1.01)	0.282
Duration of HTN (months)	1.01 (1.01 – 1.01)	**0.037**
Systolic BP (mmHg)	1.01 (0.99 – 1.02)	0.610
Diastolic BP (mmHg)	0.99 (0.97 – 1.02)	0.872
Antihypertensive medication (DDD)	1.18 (1.01 – 1.38)	**0.037**
eGFR (mL/min)	1.01 (0.99 – 1.02)	0.273
Lowest Potassium (mEq/L)	0.10 (0.05 – 0.21)	**< 0.001**
PRA at screening (ng/mL/h)	1.19 (0.41 – 3.41)	0.750
Aldosterone at screening (ng/dL)	1.04 (1.02 – 1.07)	**< 0.001**
PRA post-confirmatory test (ng/mL/h)	1.73 [0.29 – 10.19]	0.547
Aldosterone post-confirmatory test (ng/dL)	1.09 (1.05 – 1.12)	**< 0.001**
Microalbuminuria (ref. yes)	1.10 (0.52 – 2.32)	0.800
LVH at Echo (ref. yes)	1.06 (0.57 – 2.00)	0.831
CV events (ref. yes)	0.65 (0.28 - 1.52)	0.322
Nodule at CT scanning (ref. presence)	8.33 (4.35 – 16.67)	**< 0.001**
Largest nodule at CT scanning (diameter, mm)	1.12 (1.07 – 1.16)	**< 0.001**
CT scanning findings (ref. unilateral abnormality)	9.91 (3.50 – 28.05)	**< 0.001**

Univariate logistic regression analysis was performed to assess the odds ratio (OR) and the 95% confidence interval (CI) for each variable. An OR greater than 1 indicates an increased likelihood of unilateral primary aldosteronism (UPA), and an OR less than 1 a decreased likelihood. The DDD is the assumed average maintenance dose per day for a drug used for its main indication in adults. AVS, Adrenal Venous Sampling; HTN, Hypertension; BP, Blood Pressure; DDD, Defined Daily Dose; eGFR, estimated Glomerular Filtration Rate; PRA, Plasma Renin Activity; LVH, Left Ventricular Hypertrophy; Echo, Echocardiography; CV, Cardiovascular; CT, Computed Tomography. Normally and non-normally distributed variables were reported as mean ± standard deviation or median [interquartile range], respectively. Categorical variables were reported as absolute number (n) and proportion (%).

Supplementary Table 2. LDA coefficients for subtype diagnosis

Variable (ref. UPA)	LDA Coefficients	Normalized Coefficients
Aldosterone at screening (ng/dL)	-0.027	0.387
Lowest Potassium (mEq/L)	1.746	1.000
Aldosterone post-confirmatory test (ng/dL)	-0.0295	0.373
Nodule at CT scanning (ref. presence)	1.954	0.788
Largest nodule at CT scanning (diameter, mm)	-0.00091	0.004
CT scanning findings (ref. unilateral abnormality)	0.672	0.403

LDA model coefficients and normalized coefficients (absolute values between 0 and 1). LDA coefficients can be used in combination with each single variable to predict subtype diagnosis (unilateral *versus* bilateral PA). Each variable is multiplied by its corresponding LDA coefficient and the adjusted coefficients are summed to derive value X according the following equation: Diagnosis of unilateral PA = $LDAcoeff_1$*$Variable_1$ + $LDAcoeff_2$*$Variable_2$ + ... + $LDAcoeff_n$*$Variable_n$ > 0.8299. If the value of X is more than the given cut-off (0.8299) then unilateral PA is diagnosed. UPA, Unilateral Primary Aldosteronism; CT, Computed Tomography.

Supplementary Table 3. Patient Characteristics of Study Cohort

Variable	Combined Cohort (N = 215)	Training Cohort (N = 150)	Validation Cohort (N = 65)	*P*-value
Diagnosis of UPA	133 (61.9)	93 (62.0)	40 (61.5)	0.949
Female sex, n (%)	75 (34.9)	57 (38.0)	18 (27.7)	0.145
Age at diagnosis (years)	49 ± 9.5	49 ± 9.5	50 ± 9.7	0.383
Duration of HTN (months)	68 [27; 128]	67 [27; 128]	73 [27; 160]	0.843
Systolic BP (mmHg)	164 ± 23.3	164 ± 23.8	164 ± 22.3	0.936
Diastolic BP (mmHg)	99 ± 13.4	100 ± 13.0	98 ± 14.5	0.557
Antihypertensive medication (DDD)	3.3 [2.0; 5.0]	3.4 [2.1; 5.1]	3.0 [1.5; 4.6]	0.311
eGFR (mL/min)	96 [81; 106]	94 [82; 106]	98 [79; 108]	0.639
Lowest Potassium (mEq/L)	3.4 ± 0.7	3.4 ± 0.7	3.4 ± 0.5	0.576
PRA at screening (ng/mL/h)	0.25 [0.18; 0.40]	0.20 [0.20; 0.40]	0.30 [0.13; 0.40]	0.721
Aldosterone at screening (ng/dL)	33.4 [23.5; 45.6]	33.3 [23.3; 46.2]	34.5 [23.6; 44.0]	0.954
Confirmatory testing Saline infusion test, n (%) Captopril Challenge test, n (%)	 165 (76.7) 50 (23.3)	 117 (78.0) 33 (22.0)	 48 (73.8) 17 (26.2)	0.508
PRA post-confirmatory test (ng/mL/h)	0.15 [0.10; 0.20]	0.15 [0.10; 0.20]	0.15 [0.10; 0.21]	0.174
Aldosterone post-confirmatory test (ng/dL)	16.4 [10.5; 27.2]	16.1 [9.8; 27.4]	17.4 [11.1; 27.3]	0.461
Microalbuminuria, n (%)	66 (30.7)	40 (26.5)	26 (40.5)	0.101
LVH at Echo, n (%)	129 (60.1)	88 (58.7)	41 (63.5)	0.558
CV events, n (%)	32 (14.9)	23 (15.3)	9 (13.7)	0.787
Presence of nodule at CT scanning, n (%)	148 (68.8)	103 (68.9)	45 (69.2)	0.935
Largest nodule at CT scanning (diameter, mm)	13 [10; 20]	13 [10; 18]	13 [10; 20]	0.689
CT scanning findings Bilaterally Normal Bilaterally Abnormal Unilateral Abnormality	 25 (11.6) 37 (17.2) 153 (71.2)	 17 (11.3) 22 (14.7) 111 (74.0)	 8 (12.3) 15 (23.1) 42 (64.6)	0.292
AVS protocol Basal, n (%) ACTH continuous infusion, n (%) Both (Basal + ACTH), n (%)	 80 (37.2) 83 (38.6) 52 (24.2)	 52 (34.7) 64 (42.7) 34 (22.7)	 28 (43.1) 19 (29.2) 18 (27.7)	0.178
Lateralization Index at AVS	6.0 [2.2; 14.8]	6.2 [2.1; 16.3]	5.8 [2.3; 10.7]	0.444
Clinical outcome: Complete, n (%) [only for UPA] Partial, n (%) Absent, n (%)	72 (54.1) 55 (41.4) 6 (4.5)	54 (58.1) 35 (37.6) 4 (4.3)	18 (45.0) 20 (50.0) 2 (5.0)	0.378
Biochemical outcome: Complete, n (%) [only for UPA] Partial, n (%) Absent, n (%)	131 (98.5) 2 (1.5) 0 (0.0)	91 (97.8) 2 (2.2) 0 (0.0)	40 (100.0) 0 (0.0) 0 (0.0)	0.350

Clinical characteristics of patients included in the analysis: patients from the combined cohort (N = 215) were randomly assigned to training cohort (N = 150), or validation cohort (N = 65). The DDD is the assumed average maintenance dose per day for a drug used for its main indication in adults. UPA, Unilateral Primary Aldosteronism; AVS, Adrenal Venous Sampling; HTN, Hypertension; BP, Blood Pressure; DDD, Defined Daily Dose; eGFR, estimated Glomerular Filtration Rate; PRA, Plasma Renin Activity; LVH, Left Ventricular Hypertrophy; Echo, Echocardiography; CV, Cardiovascular; CT, Computed Tomography. Normally and non-normally distributed variables were reported as mean ± standard deviation or median [interquartile range], respectively. Categorical variables were reported as absolute number (n) and proportion (%).

Supplementary Table 4. Score development and validation

	AVS Score Accuracy	Predicted Diagnosis		Performance	
Real Diagnosis (Cut-off > 8)	**Training cohort** (N = 150)	UPA	BPA	Accuracy (%)	74.7
	UPA	91	2	Sensitivity (%)	97.8
	BPA	36	21	Specificity (%)	36.8
	Validation cohort (N = 65)	UPA	BPA	Accuracy (%)	69.2
	UPA	38	2	Sensitivity (%)	95.0
	BPA	18	7	Specificity (%)	28.0
	Combined cohort (N = 215)	UPA	BPA	Accuracy (%)	73.0
	UPA	129	4	Sensitivity (%)	97.0
	BPA	54	28	Specificity (%)	34.1
Real Diagnosis (Cut-off > 16)	**Training cohort** (N = 150)	UPA	BPA	Accuracy (%)	66.7
	UPA	44	49	Sensitivity (%)	47.3
	BPA	1	56	Specificity (%)	98.2
	Validation cohort (N = 65)	UPA	BPA	Accuracy (%)	58.5
	UPA	15	25	Sensitivity (%)	37.5
	BPA	2	23	Specificity (%)	92.0
	Combined cohort (N = 215)	UPA	BPA	Accuracy (%)	64.2
	UPA	59	74	Sensitivity (%)	44.4
	BPA	3	79	Specificity (%)	96.3

The table shows the real and predicted subtype diagnosis, accuracy sensitivity, specificity for the training cohort (N = 150), the validation cohort (N = 65), and the combined cohort (N = 215). A cut-off of greater than 8 identifies patients with a diagnosis of unilateral primary aldosteronism (UPA) with an optimized sensitivity; cut-off of greater than 16 identifies patients with a diagnosis of UPA with an optimized specificity. BPA, Bilateral Primary Aldosteronism.

Supplementary Table 5. Diagnostic performance of proposed score-systems

Score Performance	N	Cut-off	Sens (%)	Spec (%)	PPV (%)	NPV (%)	Acc (%)
Küpers Score [7]	87	≥5	53.1	100.0	100.0	62.3	73.6
Validation*	215		61.6	76.8	81.2	55.3	67.4
Nanba Score [8]	71	≥5	75.0	94.9	92.3	82.2	85.9
Validation*	50		90.3	57.9	77.8	78.6	78.0
Kocjan Score [9]	67	<3	100.0	28.2	50.0	100.0	58.2
Validation*	165		96.1	28.6	68.5	81.8	70.3
Kobayashi Score (2017) [10]	73	<5	94.6	66.7	74.5	92.3	80.8
Validation*	50		93.5	42.1	72.5	80.0	74.0
Kamemura Score [11]	228	<2	75.0	74.0	25.4	96.2	74.1
Validation*	103		84.5	36.8	85.5	35.0	75.7
Kobayashi Score (2018) [12]	1278	<8	87.0	66.7	52.2	92.5	72.7
Validation*	215		96.2	20.7	66.3	77.3	67.4
Leung Score (2019) [13]	80	>2	84.2	88.1	94.6	86.0	86.3
Validation*	165		70.6	61.9	75.0	56.5	67.3
LDA Model	215	N.A.	86.5	73.2	83.9	76.9	81.4
Validation	N.A.		85.7	68.3	81.4	74.7	79.1
RF Model	215	N.A.	99.2	82.9	90.4	98.5	93.0
Validation	N.A.		94.0	75.6	86.2	88.6	87.0
SPACE Score - Best Accuracy	150	>12	93.5	82.5	89.7	88.7	89.3
Internal Validation	65		87.5	72.0	83.3	78.3	81.5
Combined	215		91.7	79.3	87.8	85.5	87.0
External Validation	118		87.7	70.5	73.5	86.0	78.8
SPACE Score (Sensitivity Optimization)	150	>8	97.8	36.8	71.7	91,3	74.7
Internal Validation	65		95.0	28.0	67.9	77.8	69.2
Combined	215		97.0	34.1	70.5	87.5	73.0
External Validation	118		93.0	63.9	70.7	90.7	78.0
SPACE Score (Specificity Optimization)	150	>16	47.3	98.2	97.8	53.3	66.7
Internal Validation	65		37.5	92.0	88.2	47.9	58.5
Combined	215		44.4	96.3	95.2	51.6	64.2
External Validation	118		57.9	96.7	94.3	71.1	78.0

Sensitivity (Sens), specificity (Spec), positive/negative predictive value (PPV/ NPV), and accuracy (Acc) for proposed scores (each indicator is derived considering unilateral PA as referral diagnosis). *Each score was validated on Torino combined cohort. Nanba and Kobayashi (2016) scores were validated on patients with PA confirmed by captopril challenge test (N = 50); Kocjan and Leung scores on patients with PA confirmed by saline infusion test (N = 165); Kamemura score on patients with a unilateral nodule with diameter ≥ 10 mm at imaging. LDA, Linear Discriminant Analysis; RF, Random Forest. The external validation of the score was performed on the Munich Cohort.

Supplementary Table 6. Patient Characteristics: Torino *vs.* Munich Cohort

Variable	Torino Cohort (N = 215)	Munich Cohort (N = 118)	*P*-value
Diagnosis of UPA	133 (61.9)	57 (48.3)	**0.017**
Female sex, n (%)	75 (34.9)	46 (39.0)	0.457
Age at diagnosis (years)	49 ± 9.5	51 ± 10.8	0.274
Duration of HTN (months)	68 [27; 128]	99 [26; 213]	0.137
Systolic BP (mmHg)	164 ± 23.3	153 ± 21.2	**< 0.001**
Diastolic BP (mmHg)	99 ± 13.4	94 ± 12.5	**0.001**
Antihypertensive medication (DDD)	3.3 [2.0; 5.0]	2.5 [1.0; 4.0]	**0.003**
eGFR (mL/min)	96 [81; 106]	89 [74; 103]	0.095
Lowest Potassium (mEq/L)	3.4 ± 0.7	3.1 ± 0.5	**< 0.001**
PRA at screening (ng/mL/h)	0.25 [0.18; 0.40]	0.29 [0.17; 0.69]	**0.036**
Aldosterone at screening (ng/dL)	33.4 [23.5; 45.6]	17.9 [11.3; 27.8]	**< 0.001**
Confirmatory testing Saline infusion test, n (%) Captopril Challenge test, n (%)	 165 (76.7) 50 (23.3)	 116 (98.3) 2 (1.7)	**< 0.001**
PRA post-confirmatory test (ng/mL/h)	0.15 [0.10; 0.20]	0.21 [0.16; 0.44]	**< 0.001**
Aldosterone post-confirmatory test (ng/dL)	16.4 [10.5; 27.2]	11.2 [7.3; 20.4]	**< 0.001**
Microalbuminuria, n (%)	66 (30.7)	41 (34.7)	0.491
LVH at Echo, n (%)	129 (60.1)	59 (50.0)	0.086
CV events, n (%)	32 (14.9)	17 (14.4)	0.915
Presence of nodule at CT scanning, n (%)	148 (68.8)	72 (61.0)	0.149
Largest nodule at CT scanning (diameter, mm)	13 [10; 20]	14 [9; 17]	0.229
CT scanning findings Bilaterally Normal Bilaterally Abnormal Unilateral Abnormality	 25 (11.6) 37 (17.2) 153 (71.2)	 46 (39.0) 18 (15.3) 54 (45.8)	**< 0.001**
AVS protocol Basal, n (%) ACTH continuous infusion, n (%) Both (Basal + ACTH), n (%)	 80 (37.2) 83 (38.6) 52 (24.2)	 118 (100.0) 0 (0.0) 0 (0.0)	**< 0.001**
Lateralization Index at AVS	6.0 [2.2; 14.8]	3.3 [1.8; 19.5]	0.246
Clinical outcome: Complete, n (%) [only for UPA] Partial, n (%) Absent, n (%)	72 (54.1) 55 (41.4) 6 (4.5)	19 (33.3) 26 (45.6) 12 (21.1)	**< 0.001**
Biochemical outcome: Complete, n (%) [only for UPA] Partial, n (%) Absent, n (%)	131 (98.5) 2 (1.5) 0 (0.0)	57 (100.0) 0 (0.0) 0 (0.0)	0.082

Clinical characteristics of patients included in the analysis: patients from the Torino cohort (N = 215) were compared to patients from the Munich cohort (N = 118). The DDD is the assumed average maintenance dose per day for a drug used for its main indication in adults. UPA, Unilateral Primary Aldosteronism; AVS, Adrenal Venous Sampling; HTN, Hypertension; BP, Blood Pressure; DDD, Defined Daily Dose; eGFR, estimated Glomerular Filtration Rate; PRA, Plasma Renin Activity; LVH, Left Ventricular Hypertrophy; Echo, Echocardiography; CV, Cardiovascular; CT, Computed Tomography. Normally and non-normally distributed variables were reported as mean ± standard deviation or median [interquartile range], respectively. Categorical variables were reported as absolute number (n) and proportion (%).

Supplementary Table 7. Distribution of PA patients according to the score

Score points	Total (N)	UPA		BPA	
		(N)	(%)	(N)	(%)
0.0-2.0	7	0	0,0	7	100,0
2.1-4.0	10	2	20,0	8	80,0
4.1-6.0	3	1	33,3	2	66,7
6.1-8.0	12	1	8,3	11	91,7
8.1-10.0	21	4	19,0	17	81,0
10.1-12.0	23	3	13,0	20	87,0
12.1-14.0	42	30	71,4	12	28,6
14.1-16.0	35	33	94,3	2	5,7
16.1-18.0	31	28	90,3	3	9,7
18.1-20.0	31	31	100,0	0	0,0
Total	215	133	N.A.	82	N.A.

The number (N) and the proportion (%) of patients stratified for subtype diagnosis (unilateral PA *versus* bilateral PA) is shown according to the AVS score in the combined cohort (N = 215). N.A., Not Applicable.

Supplementary Table 8. Correlation of the SPACE score with lateralization index

Pearson's R test	**Regression Line**	**R coefficient**	***P*-value**
Torino Training cohort (n=150)	Y = -6.7 + 1.4*X	0.455	**< 0.001**
Torino Validation cohort (n=65)	Y = -7.0 + 1.4*X	0.361	**< 0.001**
Munich Validation cohort (n=118)	Y = -8.8 + 2.8*X	0.307	**< 0.001**

LI stratified according to SPACE score points	**All patients**	**Torino Cohort**	**Munich Cohort**
SPACE Score </= 8	1.9 [1.5; 2.8]	1.8 [1.3; 2.7]	1.9 [1.6; 2.8]
SPACE Score 8.5 - 16	4.5 [2.2; 11.5]	4.8 [2.3; 10.2]	3.2 [2.1; 23.0]
SPACE Score > 16	17.0 [9.2; 31.0]	15.6 [9.8; 30.6]	18.3 [6.7; 42.7]

Correlation between SPACE score and lateralization index (LI) was evaluated by Pearson's R test. The table reports regression line, R coefficient, and *P*-value for Torino and Munich cohorts. LI median and interquartile range are also shown after stratification for SPACE score points.

Supplementary Table 9. Diagnostic modelling: sub-analysis according to patient outcome

Score accuracy	All patients	Torino Cohort	Munich Cohort
UPA predicted as UPA (true prediction)			
Complete clinical outcome, n (%)	88 (51.2)	70 (57.4)	18 (36.0)
Partial clinical outcome, n (%)	70 (40.7)	48 (39.3)	22 (44.0)
Absent clinical outcome, n (%)	14 (8.1)	4 (3.3)	10 (20.0)
Complete biochemical outcome, n (%)	171 (99.4)	121 (99.2)	50 (100.0)
Partial biochemical outcome, n (%)	1 (0.6)	1 (0.8)	0 (0.0)
Absent biochemical outcome, n (%)	0 (0.0)	0 (0.0)	0 (0.0)
UPA predicted as BPA (false prediction)			
Complete clinical outcome, n (%)	3 (16.7)	2 (18.2)	1 (14.3)
Partial clinical outcome, n (%)	11 (61.1)	7 (63.6)	4 (57.1)
Absent clinical outcome, n (%)	4 (22.2)	2 (18.2)	2 (28.6)
Complete biochemical outcome, n (%)	17 (94.4)	10 (90.9)	7 (100.0)
Partial biochemical outcome, n (%)	1 (5.6)	1 (9.1)	0 (0.0)
Absent biochemical outcome, n (%)	0 (0.0)	0 (0.0)	0 (0.0)

Clinical and biochemical outcomes according to the PASO criteria were reported for patients with unilateral primary aldosteronism (UPA) after a follow-up of 6-12 months. Patients were stratified according to the SPACE score predicted diagnosis (true *vs.* false predictions).

Supplementary Table 10. Diagnostic modelling: sub-analysis according to confirmatory test

Confirmatory testing	Score accuracy
Saline Infusion Test (patients correctly predicted, %)	236 of 281 (84.0%)
Captopril Challenge Test (patients correctly predicted, %)	44 of 52 (84.6%)

SPACE score accuracy in patients which underwent saline infusion test *vs.* captopril challenge test. For this analysis, patients from Torino and Munich cohorts (N = 333) were grouped together.

Supplementary Table 11. Previously proposed score-systems

Küpers score [7]		
Cut-off >/= 5 [= UPA]	3	Adrenal nodule >/= 10 mm and normal contralateral gland
	2	Potassium levels < 3.5 mEq/L
	0	eGFR (MDRD) < 80 mL/min/1.73sqm
	1	eGFR (MDRD) 80-99 mL/min/1.73sqm
	2	eGFR (MDRD) >/= 100 mL/min/1.73sqm
Namba score [8]		
Cut-off >/= 5 [= UPA]	2	Potassium levels < 3.5 mEq/L
	3	Aldosterone at screening >/= 16.5 ng/dL
	3	ARR post captopril challenge test >/= 100 [ng/dL]/[ng/mL/h]
Kocjan score [9]		
Cut-off >/= 3 [= BPA]	1	Potassium Levels >/= 3.5 mEq/L
	1	Aldosterone post saline infusion test < 18 ng/dL
	1	Bilateral normal or abnormal CT scanning
Kamemura score [11]		
Cut-off >/= 2 [= BPA]	1	Female sex
	1	Potassium levels >/= 3.8 mEq/L
	1	ARR at screening </= 55 [ng/dL]/[ng/mL/h]
Kobayashi score [10]		
Cut-off >/= 5 [= BPA]	2	Lowest Potassium >/= 3.5 mEq/L
	2	ARR post captopril challenge test < 49 [ng/dL]/[ng/mL/h]
	3	No adrenal nodules at CT scanning
Kobayashi score [12]		
Cut-off >/= 8 [= BPA]	4	Potassium levels > 3.9 mEq/L
	3	Potassium levels 3.5-3.9 mEq/L
	3	No adrenal nodules at CT scanning
	2	Aldosterone at screening < 21 ng/dL
	2	ARR at screening < 62 [ng/dL]/[ng/mL/h]
	1	Female sex
Leung score [13]		
Cut-off >/= 2 [= UPA]	1	Age < 50 years
	1	PRA pre saline infusion test </= 0.26 ng/mL/h
	1	Aldosterone post saline infusion test >/= 15.3 ng/dL

UPA, Unilateral Primary Aldosteronism; BPA, Bilateral Primary Aldosteronism; eGFR, estimated Glomerular Filtration Rate; ARR, Aldosterone-to-Renin Ratio; CT, Computed Tomography; PRA, Plasma Renin Activity.

Supplementary Figure 1. Diagnostic modelling: Torino *vs.* Munich Cohort

Performance on TURIN Cohort (Internal Validation)

SCORE (validation cohort)		PREDICTED	
		UPA	BPA
REAL	UPA	35	5
	BPA	7	18
Accuracy		81.5 %	
Sensitivity		87.5 %	
Specificity		72.0 %	
Macro Average Accuracy		79.8 %	

LDA (10k-Cross-validation)		PREDICTED	
		UPA	BPA
REAL	UPA	114	19
	BPA	26	56
Accuracy		79.1 %	
Sensitivity		85.7 %	
Specificity		68.3 %	
Macro Average Accuracy		77.0 %	

FLOW CHART (validation cohort)		PREDICTED	
		UPA	BPA
REAL	UPA	38	2
	BPA	2	23
Accuracy		93.8 %	
Sensitivity		95.0 %	
Specificity		92.0 %	
Macro Average Accuracy		93.5 %	

RF (10k-Cross-validation)		PREDICTED	
		UPA	BPA
	UPA	125	8
	BPA	20	62
Accuracy		87.0 %	
Sensitivity		94.0 %	
Specificity		75.6 %	
Macro Average Accuracy		84.8 %	

Performance on MUNICH Cohort (External Validation)

SCORE		PREDICTED	
		UPA	BPA
REAL	UPA	50	7
	BPA	18	43
Accuracy		78.8 %	
Sensitivity		87.7 %	
Specificity		70.5 %	
Macro Average Accuracy		79.1 %	

LDA		PREDICTED	
		UPA	BPA
REAL	UPA	51	6
	BPA	19	42
Accuracy		78.8 %	
Sensitivity		89.5 %	
Specificity		68.9 %	
Macro Average Accuracy		79.2 %	

FLOW CHART		PREDICTED	
		UPA	BPA
REAL	UPA	53	4
	BPA	2	59
Accuracy		94.9 %	
Sensitivity		93.0 %	
Specificity		96.7 %	
Macro Average Accuracy		94.9 %	

RF		PREDICTED	
		UPA	BPA
REAL	UPA	55	2
	BPA	21	40
Accuracy		80.5 %	
Sensitivity		96.5 %	
Specificity		65.6 %	
Macro Average Accuracy		81.0 %	

The confusion matrix reports accuracy, sensitivity, specificity, and macro average accuracy (mean of sensitivity and specificity) for the SPACE score, the flow chart for patient management, and the linear discriminant analysis (LDA) and random forest (RF) models. Data are shown for Torino cohort, as internal validation, *vs*. Munich cohort, as external validation.