Prediction of hyperaldosteronism subtypes when adrenal vein sampling is unilaterally successful

Prediction of hyperaldosteronism subtypes when adrenal vein sampling is unilaterally successful

Jacopo Burrello[1*], Alessio Burrello[2*], Jacopo Pieroni[1], Elisa Sconfienza[1], Vittorio Forestiero[1], Martina Amongero[3], Denis Rossato[4], Franco Veglio[1], Tracy A. Williams[1,5], Silvia Monticone[1#], Paolo Mulatero[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) Department of Mathematical Sciences G. L. Lagrange, Polytechnic University of Torino, Italy. (4) AOU Città della Salute e della Scienza - Service of Radiology, Torino, Italy. (5) 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 di Torino, Via Genova 3, 10126 Torino, Italy. Telephone/Fax number: 0039.011.633.6959 / 0039.011.633.6931. E-mail: paolo.mulatero@unito.it

Manuscript word count: 3,440 excluding tables, references and figure captions.

Abstract word count: 250.

Number of Tables: 3; Number of Figures: 4; 1 Supplementary file.

Short title: Subtype diagnosis in primary aldosteronism

Keywords: aldosterone; primary aldosteronism; contralateral suppression; adrenal venous sampling; machine learning.

ABSTRACT

Objective – Adrenal venous sampling (AVS) is the gold standard to discriminate patients with unilateral primary aldosteronism (UPA) from bilateral disease (BPA). AVS is technically-demanding and in cases of unsuccessful cannulation of adrenal veins, the results may not always be interpreted. The aim of our study was to develop diagnostic models to distinguish UPA from BPA, in cases of unilateral successful AVS and the presence of contralateral suppression of aldosterone secretion.

Design – Retrospective evaluation of 158 patients referred to a tertiary hypertension unit who underwent AVS. We randomly assigned 110 patients to a training cohort and 48 patients to a validation cohort to develop and test the diagnostic models.

Methods – Supervised machine learning algorithms and regression models were used to develop and validate two prediction models and a simple 19-point score system to stratify patients according to their subtype diagnosis.

Results – Aldosterone levels at screening and after confirmatory testing, lowest potassium, ipsilateral and contralateral imaging findings at CT scanning, and contralateral ratio at AVS, were associated with a diagnosis of UPA and were included in the diagnostic models. Machine learning algorithms correctly classified the majority of patients both at training and validation (accuracy 82.9-95.7%). The score system displayed a sensitivity/specificity of 95.2/96.9%, with an AUC of 0.971. A flow-chart integrating our score correctly managed all patients except 3 (98.1% accuracy), avoiding the potential repetition of 77.2% of AVS procedures.

Conclusions – Our score could be integrated in clinical practice and guide surgical decision-making in patients with unilateral successful AVS and contralateral suppression.

ABBREVIATION LIST: A/C, Aldosterone-to-Cortisol ratio; ARR, Aldosterone-to-Renin Ratio; AVS, Adrenal Venous Sampling; BPA, Bilateral Primary Aldosteronism; CLR, Contralateral ratio; CT, Computed Tomography; DDD, Defined Daily Dose; IVC, Inferior Vena Cava; LAV, Left Adrenal Vein; LDA, Linear Discriminant Analysis; LI, Lateralization Index; PA, Primary Aldosteronism; PRA, Plasma Renin Activity; RAV, Right Adrenal Vein; RF, Random Forest; SI, Selectivity Index; UPA, Unilateral Primary Aldosteronism.

INTRODUCTION

Primary aldosteronism (PA) is a common secondary cause of arterial hypertension[1,2], associated with an increased cardiovascular risk compared with patients with essential hypertension[3,4]. This condition may be determined either by unilateral or bilateral hypersecretion of aldosterone by the adrenal glands, justifying the two major subtypes of PA, aldosterone producing adenoma, and bilateral adrenal hyperplasia.

Patients with a diagnosis of unilateral PA (UPA) derive a clinical benefit after surgical adrenalectomy in more than 80% of cases[5-7]. Medical therapy with mineralocorticoid receptor antagonists is recommended for patients with bilateral disease[7]. The correct treatment leads to a significant improvement of long-term outcomes, reverting the cardiovascular risk excess displayed by these patients[8]. For these reasons, an early diagnosis of PA and subtype differentiation is fundamental. Guidelines recommends adrenal venous sampling (AVS) as the gold standard to discriminate UPA from bilateral PA (BPA)[1-3].

However, AVS may not be interpretable when cannulation is unsuccessful for one or both adrenal veins. The cannulation of the left adrenal vein (LAV) is relatively straight forward, because it merges with the inferior phrenic vein to form a common vessel draining into the left renal vein. Conversely, the cannulation of the right adrenal vein (RAV), which is shorter and smaller, directly drains into the inferior vena cava (IVC) at an acute angle and may thus be challenging[9,10]. The correct cannulation

of RAV and LAV is assessed by the selectivity index (SI), which is calculated as the cortisol ratio between adrenal veins and IVC. The ratio between the aldosterone-to-cortisol (A/C) ratio in the dominant adrenal vein and the A/C ratio in the non-dominant vein is defined as lateralization index (LI), and displays high sensitivity and specificity to distinguish UPA from bilateral forms[3,10,11].

Several studies have described the diagnostic performance of the contralateral ratio (CLR, defined as A/C ratio in the non-dominant adrenal vein divided by A/C ratio in the IVC). Contralateral suppression is defined in the presence of an A/C ratio in the non-dominant adrenal vein lower than the A/C ratio in IVC, and a CLR lower than 1 is used in some centers to define lateralization when only one adrenal vein is successfully cannulated[12,13]. Patients with UPA should display a contralateral normal adrenal gland with suppressed aldosterone secretion from the *zona glomerulosa*[14,15]. Nevertheless, the CLR cannot predict UPA by itself, because about one third of patients with BPA also display contralateral suppression[11], and for the occurrence of apparent bilateral aldosterone suppression, defined as adrenal A/C ratio bilaterally lower than that in IVC[16]. Thresholds to interpret CLR and assess lateralization of aldosterone secretion have been not yet defined.

Therefore, the aim of our study was to develop and validate a diagnostic model able to discriminate UPA from BPA, when the AVS is unilaterally successful and the adrenal vein A/C ratio is lower than the A/C ratio in the IVC (so called contralateral suppression).

METHODS

Study cohort and data extraction

We retrospectively evaluated 158 patients who underwent AVS in the Hypertension Unit of Torino (Italy), between 2008 and 2019. All patients gave informed written consent to the study according to Helsinki declaration. The study was approved by the local ethical committee of the hospital "Città della Salute e della Scienza" of Torino. PA was diagnosed according with the Endocrine Society Guideline[1]. Patients were screened using the aldosterone-to-plasma renin activity (PRA) -ratio

(ARR); if possible, all interfering drugs were withdrawn for at least 4 weeks prior to analysis (6 weeks for mineralocorticoid receptor antagonist and diuretics). The cut-off for a positive screening test was an ARR greater than 30 ng/dL/ng*mL-1*h-1 together with an aldosterone concentration greater than 10 ng/dL. Patients with a positive screening test underwent confirmatory/exclusion testing by either intravenous saline loading test or a captopril challenge test[1]. Confirmed patients with PA underwent multi-slice computed tomography (CT) scanning and AVS. An adrenal nodule was reported in the presence of a mass equal or greater than 8 mm. AVS was performed under basal conditions or after continuous cosyntropin (ACTH) infusion (76 vs. 82, respectively). The catheterization of adrenal veins was sequential in both protocols. In case of cosyntropin stimulation, the infusion was performed at 50 ug/h, starting 30 minutes before the procedure. AVS procedures were performed by a single experienced interventional radiologist. The radiologist firstly cannulates the right adrenal vein and subsequently the left adrenal vein (within 15 minutes). ACTH stimulation was necessary for patients who need preparation with steroids before the procedure for a history of contrast allergy (or other allergy; n=13), or when the procedure was not performed in the early morning (n=69)[10]. Adrenal vein cannulation was considered successful in the presence of a SI equal or greater than 3 under basal conditions, and of 5 under cosyntropin infusion. The diagnosis of UPA was established when the LI was at least 4[10].

Patients were included in the study if: 1) the AVS was bilaterally successful; 2) in presence of contralateral suppression, that is when at least one adrenal vein A/C ratio was lower than the A/C ratio in the IVC (CLR < 1). Hypercortisolism was excluded in all patients. Enrolled patients were randomly assigned to a training cohort (N = 110) or to a validation cohort (N = 48), with a ratio 70:30. Contralateral and ipsilateral adrenal glands were defined on the basis of AVS findings. In case of UPA, the contralateral adrenal was the gland on the side opposite to the lateralization. In case of BPA diagnosis, the contralateral adrenal was defined as the gland showing an A/C ratio lower than in the IVC. Ipsilateral and contralateral adrenal imaging at CT scanning were defined accordingly. To build

our model, aldosterone and cortisol measurements from the IVC and from the adrenal vein with A/C ratio lower than in the IVC were used. When UPA is predicted, the ipsilateral adrenal gland should be removed. For all patients, antihypertensive medication was quantified as Defined Daily Dose (DDD), which is the average maintenance dose per day for a drug used for its main indication in adults.

Machine learning and statistical analysis

IBM SPSS Statistics 22 (*IBM Corp., Armonk, New York, USA*) and Python 3.5 (library, scikit-learn) were used for statistics. The study followed the TRIPOD statement for transparent reporting of a multivariable prediction model for individual prognosis or diagnosis.

Sample size was calculated on the estimated variability for median value of CLR. With a mean effect size (Cohen's d coefficient) of 1.24, a minimum power (1 - β error probability) of 95%, and a significance level (α error) of 0.05, the estimated total sample size was 44. Sample size calculation was carried out with GPower 3.1.

Kolmogorov–Smirnov test was used to assess variable distribution. Normally and non-normally distributed variables were reported as mean ± standard deviation, or median and interquartile range, and analyzed by T-student, or Mann-Whitney's test, respectively. Categorical variables were reported as absolute number and percentage distribution and analyzed by Chi-square test. Logistic regression analyses were used for univariate and multivariate models; an odds ratio (OR) greater than 1 was associated with an increased likelihood of diagnosis of UPA. Supervised machine learning techniques were used to develop the diagnostic models and the clinical score, as previously described[17]. Briefly, linear discriminant analysis (LDA) and random forest (RF) algorithms were applied to develop two different diagnostic models able to discriminate UPA *vs.* BPA using the best discriminant patients' parameters, previously selected by univariate and multivariate regression analysis. The models were developed after correction of dataset imbalance using a SMOTE algorithm (Synthetic Minority Oversampling Technique). Briefly, SMOTE imputes new patient data (defined as "neighbour"), along

lines that link real patients from the training dataset in the virtual space created by patient parameters, in order to balance the number of UPA and BPA patients at model training. A number of neighbours equal to 3 (K=3), corresponded to the higher accuracy and was used in all analyses.

The LDA computes a set of coefficients for linear combination of each variable to classify patients according with their diagnosis. The predicted diagnosis is derived from the following equation: UPA = $LDAcoeff_1$*$Variable_1$ + $LDAcoeff_2$*$Variable_2$ + … + $LDAcoeff_n$*$Variable_n$ > 9.7425. The RF algorithm is composed by 20 classification trees with a maximum number of 8 splits. The predicted diagnosis resulted from the outcome of each classification tree. If at least 11 of 20 trees of the RF predicted UPA, then the patient is classified accordingly. Machine learning models were tested by a 10-fold validation algorithm: (i) the cohort was randomly divided into 10 groups; (ii) the model was trained within the first 9 groups and validated with the remaining group; (iii) the process was repeated 10 times, rotating the validation group at each round. The accuracy at validation resulted from the mean of the accuracy obtained at each round. A 19-point score based on the same discriminant parameters was finally developed in the training cohort and tested by 10-fold cross validation (with bootstrapping) and in both validation and combined cohorts. Points for each reference category and relative cut-offs were automatically derived to achieve the best accuracy. The analysis of ROC curve was used to assess the area under the curve (AUC) and derive the Youden Index (J = sensitivity + specificity - 1). Accuracy was defined as the ratio between "true positive" plus "true negative" divided by the total number of patients included in each cohort. An online tool was developed to calculate the score and define the diagnosis (available at https://github.com/ABurrello/CLR-score/raw/master/CLR%20Score%20Calculator.xlsm). All data supporting the findings of this study are available from the corresponding author upon reasonable request.

RESULTS

Patient characteristics

One hundred and fifty-eight patients underwent AVS for PA subtyping and were included in the analysis. The mean age at diagnosis of PA was 50 ± 10, mean BP was 164/98, with a duration of hypertension of 67 [23; 130] months. Seventy-six patients underwent AVS under basal conditions, and 82 after ACTH stimulation. No differences were observed between patients who underwent AVS with or without ACTH except for SI higher under ACTH stimulation (P = 0.002) (Supplementary Table S1). For this reason, patients were grouped together for all subsequent analyses, irrespective of the protocol used for AVS.

Clinical and biochemical characteristics of patients after subtype diagnosis are reported in Table 1; 126 patients were diagnosed as UPA, and 32 as BPA. The high prevalence of unilateral forms is expected, because only patients with contralateral suppression at AVS were enrolled in this study. Patients with a diagnosis of UPA were younger (49 ± 10.3 vs. 55 ± 7.1; P = 0.001), frequently females (42.1% *vs.* 21.9%; P = 0.036), with a higher DDD (4 [2.5; 6] *vs.* 2.9 [1.4; 4.3] P = 0.008). The prevalence of microalbuminuria, left ventricular hypertrophy, and prior cardiovascular events was not different between groups.

UPA patients displayed lower potassium levels (3.1 ± 0.6 vs. 3.9 ± 0.4 mEq/L; P < 0.001) and higher aldosterone, both at screening (38.2 [25.9; 49.8] vs. 30.2 [21.1; 41] ng/dL; P = 0.044) and after confirmatory testing (21.9 [13.9; 35.5] vs. 11.4 [7.7; 19.9] ng/dL; P < 0.001). At CT scanning, ipsilateral imaging was normal in 1.6% of patients with UPA, compared to 43.8% of BPA patients; conversely, contralateral imaging was normal in 84.9% of UPA and 53.1% of BPA patients (P < 0.001 for both comparisons). After AVS, contralateral A/C and CLR were both lower in patients with UPA (4.1 [1.5; 8.3] *vs.* 7.9 [3.2; 12] and 0.3 [0.2; 0.5] *vs.* 0.7 [0.5; 0.8], respectively).

Univariate logistic regression analysis confirmed the association with the diagnosis of UPA for all these parameters (Supplementary Table S2). In particular, female sex (OR 2.59), age at diagnosis (OR 0.94), DDD (OR 1.43), lowest potassium (0.07), aldosterone at screening (OR 1.01) and after confirmatory testing (1.01), a normal ipsilateral and contralateral imaging (OR 0.02 and OR 4.97,

197 respectively), contralateral A/C (OR 0.91), and CLR (OR 0.01). Six of these 10 variables were
198 selected considering their discriminative performance and expert knowledge[18], and were introduced
199 in the multivariate model (Table 2), which confirmed their independent association with UPA
200 diagnosis (female sex, age at diagnosis, DDD, and contralateral A/C were no longer associated with
201 UPA when introduced in the multivariate analysis; data not shown).
202
203 ***Linear discriminant analysis and random forest models***
204 Patients included in the analysis were randomized into two subgroups: the training cohort (N = 110)
205 and the validation cohort (N = 48). No differences were found for all the evaluated variables
206 (Supplementary Table S3).
207 Supervised machine learning algorithms were used to develop two diagnostic models: an LDA model
208 and a RF classification algorithm. The models were developed in the training cohort and subsequently
209 tested by both internal validation (10-fold cross validation applied to the same training cohort) and
210 external validation (on the randomly selected validation cohort). Supplementary Table S4 reports the
211 diagnostic performance at training, internal and external validation, and on the combined cohort. The
212 linear combination of the 6 above-selected variables is represented in the canonical plot (Figure 1A);
213 each patient is represented by a point, and the clear separation according to their subtype diagnosis
214 indicates the high diagnostic performance of the LDA model (accuracy 92.3%,88.9%, 84.6%,
215 respectively at training, internal, and external validation). In the combined cohort, 112 of 126 patients
216 with UPA and 30 of 32 with BPA, were correctly classified, resulting in a sensitivity and specificity
217 of 88.8% and 93.4%, respectively (Figure 1B). Internal and external validation confirmed the high
218 predictive performance, with a minimum overfitting effect (3.4-7.7%).
219 The RF classification algorithm (the first tree of the RF is reported in Figure 2A) correctly
220 discriminated 145 of 158 patients (accuracy 91.7%) from the combined cohort, with a sensitivity and
221 specificity of 92.1% and 90.3%, respectively (Figure 2B). Also in this case, the accuracy at external

and internal validation by 10-fold cross validation was high (accuracy 82.9% and 89.6%, respectively; overfitting effect 6.0-12.7%; see Supplementary Table S4). Ipsilateral imaging and CLR were the strongest predictors in the LDA model, whereas lowest potassium levels and aldosterone post-confirmatory test in the RF (Figures 1C and 2C). Estimate predictor importance is reported in Supplementary Table S5 for both LDA and RF models.

Prediction score and management of PA patients

The 6 variables used in the LDA and RF models were used to develop the CLR score, a 19-points scoring system to discriminate patients with a diagnosis of UPA *vs.* BPA. The CLR score was developed in the training cohort and then tested by 10-fold cross validation, and in the validation and combined cohorts. Figure 3A show how each parameter was categorized and points assigned. The analysis of the ROC curve confirmed the reliable performance of CLR score, with an AUC of 0.971 (95% CI 0.947-0.994; Figure 3B). In the training cohort, a CLR score greater than 11 displayed the higher accuracy (96.4%), with the correct classification of 84 of 88 UPA patients (sensitivity 95.5%), whereas a score equal or lower than 11 classified all the BPA patients (specificity 100%). By 10-fold cross validation, we confirmed a reliable accuracy (96.3%). The diagnostic performance of the CLR score at external validation remained very high, with an accuracy, sensitivity, and specificity of 93.7%, 94.7% and 90.0%, respectively, higher than the machine learning algorithms (confusion matrix for training, validation, and combined cohorts are reported in Supplementary Tables S4 and S6). A cut-off of greater than 8 optimized sensitivity, correctly classified 86 of 88, and 37 of 38 patients with UPA, respectively at training and at validation, with a sensitivity of 97.8% and 97.4%. A cut-off lower than 13 maximized specificity, and correctly identified all patients with a diagnosis of BPA, both at training and at validation (specificity 100%).

Using 11 as cut-off, patients which underwent AVS under basal conditions were correctly discriminated in 94.7% of cases, whereas after stimulation with ACTH in 96.3% of cases.

Figure 4A shows the stratification of patients with a diagnosis of UPA or BPA for the CLR score. The scoring system was directly correlated with the proportion of patients with a diagnosis of UPA. In addition, all patients with a score greater than 13 had UPA (N = 88), whereas all patients with a score equal or lower than 3 had BPA (N = 6; Supplementary Table S7).

Overall, 102 of the 126 patients with UPA underwent adrenalectomy. According to the PASO criteria[5], at follow-up of 6-12 months after surgery, patients with UPA displayed complete clinical and biochemical success in 50% and 96.1% of cases, respectively (Supplementary Table S8). All the 6 patients with UPA misclassified as BPA displayed partial or absent clinical outcome (4 with a complete and 2 with a partial biochemical outcome). Forty three percent (41 of 96) of correctly predicted patients with UPA displayed a partial clinical outcome and 4% absent clinical outcome; biochemical outcome was complete in 94 of 96 correctly classified patients and partial in the remaining 2 patients. No statistically significant differences were observed in the clinical and biochemical outcomes for patients who underwent unstimulated *vs.* ACTH stimulated AVS (Supplementary Table 8).

The CLR score was finally integrated into a flow chart for the management of patients with PA with a unilateral successful AVS and contralateral suppression (CLR < 1; Figure 4B and 4C). Patients with a score equal to or greater than 13 were classified as "probable UPA" with indication to surgical intervention. Notably, all patients with UPA identified in this way (N = 104) correctly underwent unilateral adrenalectomy. Patients with a score equal to or lower than 8 were classified as "probable BPA" and treated with MRA (N = 18), resulting in 15 patients with BPA correctly managed and 3 UPA patients, who missed the possibility to be operated. All the other patients (N = 36), with a score ranging between 8.5 and 12.5 should repeat AVS. The application of the prediction score in our clinical context would result in the correct management of 155 of 158 patients (accuracy 98.1%), allowing a subtype diagnosis with the repetition of only 22.8% of AVS procedures (sensitivity 97.6% and specificity 100.0%).

DISCUSSION

We propose two prediction models based on supervised machine learning algorithms, and an easily applicable scoring system, to discriminate UPA from BPA. An online free-downloadable tool is provided and allows the application of a flow-chart for the management of patients undergoing unilateral successful AVS in the presence of contralateral suppression.

Although the rate of bilaterally successful cannulation during AVS may be higher than 90% when performed by an experienced radiologist[19], several centers do not have a comparable reliability, with a successful rate ranging between 8.0 and 30.5%, especially where a low number of AVS are performed per year[20-22]. Given its course and anatomical variability[23], the main reason for unsuccessful AVS is the failed cannulation of RAV because of difficulties in identification of the vessel. Therefore, over the last few years, some studies have tried to develop indices to predict subtype diagnosis from incomplete AVS data[24-28].

Kline et al. demonstrated a high performance of the CLR in the identification of UPA; sensitivity and specificity were 90% and 94%, respectively, using a CLR < 1.4 as cut-off[24]. However, the authors did not validate their approach in a second cohort of patients and observed a high accuracy only in patients who underwent cosyntropin-stimulated AVS, whereas in unstimulated procedures the accuracy was lower. Many patients with bilateral disease may be misclassified using a CLR cut-off of 1.4 that includes subjects without contralateral suppression of aldosterone secretion. On the other hand, Lin et al. proposed very low cut-offs (CLR < 0.07 for left and 0.08 for right lateralisation) in a cohort of 160 patients who underwent AVS in basal conditions. Specificity was 100%, but sensitivity was unacceptable for clinical application, ranging between 27.5 and 40%, thus missing the majority of patients with a diagnosis of UPA[28]. In a subsequent study, Durivage et al. applied multinomial regression modelling in patients with unsuccessful right adrenal vein cannulation. The authors correctly identified 75.1% and 64.1% of UPA patients in cosyntropin-stimulated and unstimulated

AVS, respectively[26]. In another study performed in 62 AVS procedures[25], a CLR equal or lower than 0.5 displayed a specificity and positive predictive value of 100% but missed about half of patients with UPA (sensitivity 58%). Strajina et al. tested the same cut-off in a cohort of 150 patients; the specificity was 100%, but 20% of patients with UPA were classified as BPA and missed the possibility to be cured by adrenalectomy[27].

In our cohort, 79.7% of patients were diagnosed as UPA, thus being representative of the prevalence of unilateral disease in patients with contralateral suppression[24-28]. With our machine learning-based models and with the CLR score, we obtained a higher diagnostic performance compared to previous methods. We achieved a 97.6% sensitivity and 100% specificity with the proposed flow-chart for patient management. Combining the cut-offs which optimized sensitivity and specificity in our flow-chart, all patients predicted as UPA by the CLR score were correctly diagnosed (positive predictive value 100%), whereas only 3 patients with a diagnosis of UPA were misclassified as BPA and would have missed the possibility of surgical treatment but would have a long-term pharmacological treatment. The cut-off with the maximal specificity of the CLR score displayed a sensitivity higher than 80% both at training and validation, outperforming previous scores. The accuracy was similarly very high, independent of the AVS protocol (with or without cosyntropin stimulation).

In our study, 20.3% of patients had BPA and displayed a CLR lower than 1, thus confirming a high prevalence of contralateral suppression in patients with bilateral disease[11]. Using a CLR equal or lower than 0.5[25,27], 8 of 32 patients with BPA (25%) still received a diagnosis of UPA and would be operated inappropriately. Indeed, a large multicenter study demonstrated that contralateral suppression was not associated with a better clinical and biochemical outcome after surgery[13], whereas the studies evaluating the diagnostic performance of CLR did not assess post-surgical outcomes[24,25,27,28]. With our flow-chart, only 3 UPA patients were classified as BPA would have missed the possibility to be surgically treated. When assessed with the PASO criteria[5], these 3 patients

displayed partial clinical success, without normalization of aldosterone levels in 2 cases (partial biochemical success).

Contralateral suppression alone is not considered a reliable parameter to recommend adrenalectomy[10]. Similarly, CT scanning and/or clinical and biochemical parameters do not allow to predict subtype diagnosis in patients with PA[7,29].

Combining CLR with CT scanning findings and potassium and aldosterone levels, we considerably improved the diagnostic performance described for the CLR alone[24-28]. To evaluate the net benefit of our clinical score, compared to CLR alone, we assessed the respective diagnostic performance by ROC curve analysis. The AUC for CLR was 0.836 (CI 95% 0.759-0.913), compared with 0.971 for the clinical score (+16.1%). A CLR < 0.45 displayed an accuracy of 84.8% in the discrimination of patients with UPA from those with BPA. In the combined cohort, the score, LDA and RF models reached an accuracy of 95.5%, 90.0%, and 91.7% with a net benefit ranging between 6.1% and 12.6%. Our score could be of interest for centers with a low rate of bilateral adrenal vein cannulation, since it correctly classifies 98.1% of patients that display contralateral suppression and avoid the repetition of AVS in 77.2% of the cases. This approach would allow to recommend adrenalectomy (100% positive predictive value and specificity), without the repetition of the AVS in case of unsuccessful right adrenal vein cannulation.

There are some limitations to this study. This was a retrospective analysis of patients referred to a single center: we validated our scoring system in a separate randomly selected population, but the algorithm should be validated in independent prospective cohorts from other centers, to assess his generalizability and exclude selection bias. The strengths of our study include the high accuracy of the proposed prediction models, integrating machine learning algorithms and a scoring-system, allowing the correct management of patients with unilateral successful AVS and contralateral

suppression. Moreover, we developed an online tool to calculate the CLR score automatically, allowing the easy application of our flow-chart in clinical practice.

In conclusion, AVS remains a technically demanding and invasive procedure but still essential to distinguish UPA from BPA. Our findings suggest that integrating clinical and biochemical parameters, CT scanning imaging, and CLR, partial AVS data could be used to define the subtype diagnosis, facilitating surgical decision-making in case of missed cannulation of one adrenal vein.

Declarations

Conflict(s) of Interest/Disclosure(s): the authors declare no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Source(s) of Funding: This research did not receive any specific grant from any funding agency in the public, commercial or not-for-profit sector.

REFERENCES

1. Funder JW, Carey RM, Mantero F, Murad MH, Reincke M, Shibata H, Stowasser M, Young WF Jr. The Management of Primary Aldosteronism: Case Detection, Diagnosis, and Treatment: An ES Clinical Practice Guideline. *Journal of Clinical Endocrinology and Metabolism* 2016 **101** 1889-1916.
2. Mulatero P, Monticone S, Deinum J, Amar L, Prejbisz A, Zennaro MC, Beuschlein F, Rossi GP, Nishikawa T, Morganti A, et al. 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 [Epub ahead of print].
3. Mulatero P, Sechi LA, Williams TA, Lenders JWM, Reincke M, Satoh F, Januszewicz A, Naruse M, Doumas M, Veglio F, et al. 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 [Epub ahead of print].
4. 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. *The Lancet Diabetes & Endocrinology* 2018 **6** 41-50.
5. Williams TA, Lenders JWM, Mulatero P, Burrello J, Rottenkolber M, Adolf C, Satoh F, Amar L, Quinkler M, Deinum J, et al. Outcomes after adrenalectomy for unilateral primary aldosteronism: an international consensus on outcome measures and analysis of remission rates in an international cohort. *The Lancet Diabetes & Endocrinology* 2017 **5** 689-699.
6. Williams TA, Burrello J, Sechi LA, Fardella CE, Matrozova J, Adolf C, Baudrand R, Bernardi S, Beuschlein F, Catena C, et al. Computed Tomography and Adrenal Venous Sampling in the Diagnosis of Unilateral Primary Aldosteronism. *Hypertension* 2018 **72** 641-649.

7. Takeda M, Yamamoto K, Akasaka H, Rakugi H, Naruse M, Takeda Y, Kurihara I, Itoh H, Umakoshi H, Tsuiki M, et al., JPAS Study Group. Clinical characteristics and postoperative outcomes of primary aldosteronism in the elderly. *Journal of Clinical Endocrinology and Metabolism* 2018 **103** 3620-3629.
8. Hundemer GL, Curhan GC, Yozamp N, Wang M, Vaidya A. Cardiometabolic outcomes and mortality in medically treated primary aldosteronism: a retrospective cohort study. *The Lancet Diabetes & Endocrinology* 2018 **6** 51-59.
9. Daunt N. Adrenal vein sampling: how to make it quick, easy, and successful. *Radiographics* 2005 **25** S143-S158.
10. Monticone S, Viola A, Rossato D, Veglio F, Reincke M, Gomez-Sanchez C, Mulatero P. Adrenal vein sampling in primary aldosteronism: towards a standardised protocol. *The Lancet Diabetes & Endocrinology* 2015 **3** 296-303.
11. Young WF, Stanson AW, Thompson GB, Grant CS, Farley DR, van Heerden JA. Role for adrenal venous sampling in primary aldosteronism. *Surgery* 2004 **136** 1227-1235.
12. Wolley MJ, Gordon RD, Ahmed AH, Stowasser M. Does contralateral suppression at adrenal venous sampling predict outcome following unilateral adrenalectomy for primary aldosteronism? A retrospective study. *Journal of Clinical Endocrinology and Metabolism* 2015 **100** 1477-1484.
13. Monticone S, Satoh F, Viola A, Fischer E, Vonend O, Bernini G, Lucatello B, Quinkler M, Ronconi V, Morimoto R, et al. Aldosterone suppression on contralateral adrenal during adrenal vein sampling does not predict blood pressure response after adrenalectomy. *Journal of Clinical Endocrinology and Metabolism* 2014 **99** 4158-4166.
14. Espiner EA, Ross DG, Yandle TG, Richards AM, Hunt PJ. Predicting surgically remedial primary aldosteronism: role of adrenal scanning, posture testing, and adrenal vein sampling. *Journal of Clinical Endocrinology and Metabolism* 2003 **88** 3637-3644.

15. Auchus RJ, Wians FH Jr, Anderson ME, Dolmatch BL, Trimmer CK, Josephs SC, Chan D, Toomay S, Nwariaku FE. What we still do not know about adrenal vein sampling for primary aldosteronism. *Hormone and Metabolic Research* 2010 **42** 411-415.
16. Shibayama Y, Wada N, Naruse M, Kurihara I, Ito H, Yoneda T, Takeda Y, Umakoshi H, Tsuiki M, Ichijo T, et al. The Occurrence of Apparent Bilateral Aldosterone Suppression in Adrenal Vein Sampling for Primary Aldosteronism. *Journal of the Endocrine Society* 2018 **2** 398-407.
17. Burrello J, Burrello A, Stowasser M, Nishikawa T, Quinkler M, Prejbisz A, Lenders JWM, Satoh F, Mulatero P, Reincke M, et al. The Primary Aldosteronism Surgical Outcome Score for the Prediction of Clinical Outcomes After Adrenalectomy for Unilateral Primary Aldosteronism. *Annals of Surgery* 2019 [Epub ahead of print].
18. Roe KD, Jawa V, Zhang X, Chute CG, Epstein JA, Matelsky J, Shpitser I, Taylor CO. Feature engineering with clinical expert knowledge: A case study assessment of machine learning model complexity and performance. *PLoS One* 2020 **15** e0231300.
19. Young WF, Stanson AW. What are the keys to successful adrenal venous sampling (AVS) in patients with primary aldosteronism? *Clinical Endocrinology* 2009 **70** 14-17.
20. Vonend O, Ockenfels N, Gao X, Allolio B, Lang K, Mai K, Quack I, Saleh A, Degenhart C, Seufert J, et al. Adrenal venous sampling: evaluation of the German Conn's registry. *Hypertension* 2011 **57** 990-995.
21. Elliott P, Holmes DT. Adrenal vein sampling: substantial need for technical improvement at regional referral centres. *Clinical Biochemistry* 2013 **46** 1399-1404.
22. Harsha A, Trerotola SO. Technical aspects of adrenal vein sampling. *Journal of Vascular and Interventional Radiology* 2015 **26** 239.
23. Matsuura T, Takase K, Ota H, Yamada T, Sato A, Satoh F, Takahashi S. Radiologic anatomy of the right adrenal vein: preliminary experience with MDCT. *AJR American journal of roentgenology* 2008 **191** 402-408.

24. Kline GA, Chin A, So B, Harvey A, Pasieka JL. Defining contralateral adrenal suppression in primary aldosteronism: implications for diagnosis and outcome. *Clinical Endocrinology* 2015 **83** 20-27.

25. Pasternak JD, Epelboym I, Seiser N, Wingo M, Herman M, Cowan V, Gosnell JE, Shen WT, Kerlan RK Jr, Lee JA, et al. Diagnostic utility of data from adrenal venous sampling for primary aldosteronism despite failed cannulation of the right adrenal vein. *Surgery* 2016 **159** 267-273.

26. Durivage C, Blanchette R, Soulez G, Chagnon M, Gilbert P, Giroux MF, Bourdeau I, Oliva VL, Lacroix A, Therasse E. Adrenal venous sampling in primary aldosteronism: multinomial regression modeling to detect aldosterone secretion lateralization when right adrenal sampling is missing. *Journal of Hypertension* 2017 **35** 362-368.

27. Strajina V, Al-Hilli Z, Andrews JC, Bancos I, Thompson GB, Farley DR, Lyden ML, Dy BM, Young WF, McKenzie TJ. Primary aldosteronism: making sense of partial data sets from failed adrenal venous sampling-suppression of adrenal aldosterone production can be used in clinical decision making. *Surgery* 2018 **163** 801-806.

28. Lin L, Zhou L, Guo Y, Liu Z, Chen T, Liu Z, Wang K, Li J, Zhu Y, Ren Y. Can incomplete adrenal venous sampling data be used in predicting the subtype of primary aldosteronism? *Annales Endocrinology* 2019 **80** 301-307.

29. K Kempers MJ, Lenders JW, van Outheusden L, van der Wilt GJ, Schultze Kool LJ, Hermus AR, Deinum J. Systematic review: diagnostic procedures to differentiate unilateral from bilateral adrenal abnormality in primary aldosteronism. *Annals of Internal Medicine* 2009 **151** 329-337.

FIGURE LEGENDS

Legend to Figure 1 – **Diagnostic Modelling: Linear Discriminant Analysis.** The LDA model included the 6 variables with the highest prediction power for subtype diagnosis in the training cohort (N = 110). Contralateral adrenal gland corresponds to the adrenal with the A/C ratio lower than the A/C ratio in the inferior vena cava, while ipsilateral gland corresponds to the opposite side. When unilateral PA (UPA) is predicted, the ipsilateral adrenal gland should be removed. **Panel A**, canonical plot representing diagnostic performance of LDA; each patient is indicated by a point and subtype diagnosis are reported by colour (UPA, black; BPA, bilateral PA, grey). The axes (canonical component 1 and 2) are calculated by weighted linear combination of the 6 variables included in the model to maximize the separation between groups. The crosses indicate the means of (canonical 1; canonical 2) for patients with UPA or BPA, the ellipse included patients with a linear combination coefficient that falls within the mean ± SD. **Panel B**, confusion matrix reporting real and predicted diagnosis, accuracy, sensitivity, specificity in the combined cohort, and 10-fold cross validation. **Panel C**, histogram representing normalized LDA coefficients for each variable included in the model (see also Supplementary Table S5).

Legend to Figure 2 – **Diagnostic Modelling: Random Forest**. The RF algorithm included the 6 variables with the highest classification power for subtype diagnosis in the training cohort (N = 110). Contralateral adrenal gland corresponds to the adrenal with the A/C ratio lower than the A/C ratio in the inferior vena cava, while ipsilateral gland corresponds to the opposite side. When unilateral PA (UPA) is predicted, the ipsilateral adrenal gland should be removed. **Panel A**, the first classification tree of the forest is shown for the prediction of UPA *vs.* BPA (bilateral PA). **Panel B**, confusion matrix reporting real and predicted diagnosis, accuracy, sensitivity, specificity in the combined cohort, and 10-fold cross validation. **Panel C**, histogram representing normalized predictive coefficients for each variable included in the model (see also Supplementary Table S5).

Legend to Figure 3 – **Score development.** Univariate/multivariate regression analyses and coefficients from the LDA and RF models were used to assign points to each variable according to stratification level. The score was developed in the training cohort (N = 110) and tested on the validation cohort (N = 48). Contralateral adrenal gland corresponds to the adrenal with the A/C ratio lower than the A/C ratio in the inferior vena cava, while ipsilateral gland corresponds to the opposite side. When unilateral PA (UPA) is predicted, the ipsilateral adrenal gland should be removed. **Panel A**, table showing included variables and scoring-point system. **Panel B**, receiver operating characteristics (ROC) curve to assess AUC (area under the curve) and the best cut-off for the score in the combined cohort (N = 158).

Legend to Figure 4 – **Score performance and management of PA patients.** Flow chart for the management of patients with PA, unilateral successful AVS, and CLR < 1. Contralateral adrenal gland corresponds to the adrenal with the A/C ratio lower than the A/C ratio in the inferior vena cava, while ipsilateral gland corresponds to the opposite side. When unilateral PA (UPA) is predicted, the ipsilateral adrenal gland should be removed. **Panel A**, histogram showing the proportion of patients (y-axis, %) for each subtype diagnosis (UPA, black; BPA, bilateral PA, grey), stratified by score points (x-axis) on the combined cohort. The total number of patients (N) for each AVS score level and their proportion (%) are reported in Table S5. **Panel B**, PA patient management using our score; the number of patients is indicated in bold; cut-offs and misclassified patients are indicated in grey. **Panel C**, confusion matrix representing real and predicted subtype diagnosis, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). CLR, Contralateral Ratio; AVS, Adrenal Venous Sampling; MRA, Mineral Receptor Antagonist.

TABLE LEGENDS

Legend to Table 1 – Characteristics of patients included in the analysis stratified for diagnosis: unilateral PA (UPA; N=126) vs. bilateral PA (BPA; N=32). Antihypertensive medication is expressed as Defined Daily Dose (DDD). DDD is the average maintenance dose per day for a drug used for its main indication in adults. HTN, Hypertension; BP, Blood Pressure; PRA, Plasma Renin Activity; LVH, Left Ventricular Hypertrophy; Echo, Echocardiography; CV, Cardiovascular; CT, Computed Tomography; SI, Selectivity Index; A/C, Aldosterone to Cortisol ratio. 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 (%).

Legend to Table 2 – Logistic regression analysis was performed to assess the odds ratio (OR), β estimate and the 95% confidence interval (CI) for each variable. Univariate and multivariate analysis are shown as indicated. An OR greater than 1 indicates an increased likelihood of unilateral PA (UPA), and an OR less than 1 a decreased likelihood (i.e. an OR of 1.01 for aldosterone means an increase of 1% of the likelihood of UPA for each 1 ng/dL increase of aldosterone concentration). Aldosterone at screening, lowest potassium, aldosterone post-confirmatory test, and contralateral ratio were treated as continuous variables; ipsilateral and contralateral imaging (defined as normal in absence of nodules, or with thickening < 4 mm) were treated as categorical variables.

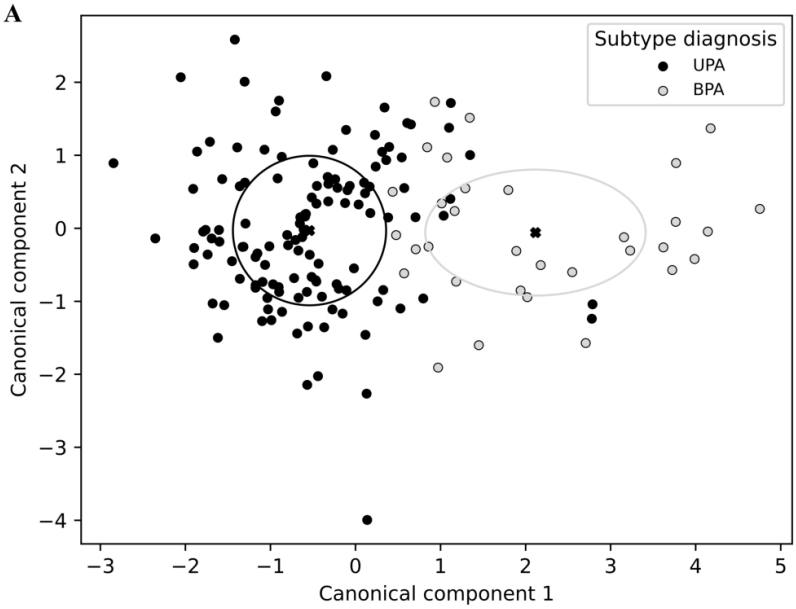
Table 1. Patient Characteristics of Study Cohort

Variable	UPA (N = 126)	BPA (N = 32)	*P*-value
Female sex, n (%)	53 (42.1)	7 (21.9)	**0.036**
Age at diagnosis (years)	49 ± 10.3	55 ± 7.1	**0.001**
Duration of HTN (months)	66 [22; 156]	75 [44; 124]	0.721
Systolic BP (mmHg)	164 ± 25.7	164 ± 20.3	0.934
Diastolic BP (mmHg)	97 ± 14.3	98 ± 12.7	0.824
Antihypertensive medication (DDD)	4.0 [2.5; 6.0]	2.9 [1.4; 4.3]	**0.008**
PRA at screening (ng/mL/h)	0.3 [0.20; 0.40]	0.20 [0.10; 0.22]	**0.014**
Aldosterone at screening (ng/dL)	38.2 [25.9; 49.8]	30.2 [21.1; 41.0]	**0.044**
Lowest potassium (mEq/L)	3.1 ± 0.6	3.9 ± 0.4	**< 0.001**
Aldosterone post-confirmatory test (ng/dL)	21.9 [13.9; 35.5]	11.4 [7.7; 19.9]	**< 0.001**
Microalbuminuria, n (%)	37 (29.6)	6 (19.0)	0.333
LVH at Echo, n (%)	70 (55.2)	24 (76.0)	0.058
CV events, n (%)	17 (13.6)	6 (18.5)	0.519
Largest nodule at CT scanning (diameter, mm)	13 [10; 18]	11 [9; 20]	0.433
Normal ipsilateral imaging, n (%)	2 (1.6)	14 (43.8)	**<0.001**
Normal contralateral imaging, n (%)	107 (84.9)	17 (53.1)	**<0.001**
Peripheral A/C ratio	15.9 [3.7; 26.2]	13.3 [7.2; 18.1]	0.145
Contralateral A/C ratio	4.1 [1.5; 8.3]	7.9 [3.2; 12.0]	**0.012**
Contralateral ratio	0.3 [0.2; 0.5]	0.7 [0.5; 0.8]	**< 0.001**

Legend to Table 1 – Characteristics of patients included in the analysis stratified for diagnosis: unilateral PA (UPA; N=126) vs. bilateral PA (BPA; N=32). Antihypertensive medication is expressed as Defined Daily Dose (DDD). DDD is the average maintenance dose per day for a drug used for its main indication in adults. HTN, Hypertension; BP, Blood Pressure; PRA, Plasma Renin Activity; LVH, Left Ventricular Hypertrophy; Echo, Echocardiography; CV, Cardiovascular; CT, Computed Tomography; SI, Selectivity Index; A/C, Aldosterone to Cortisol ratio. 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 (%).

Table 2. Selected discriminant variables for a diagnosis of unilateral PA

Variable (ref. BPA)	Univariate analysis			Multivariate Analysis		
	OR (CI 95%)	β	*P*	OR (CI 95%)	β	*P*
Aldosterone at screening (ng/dL)	1.01 (1.01 – 1.01)	0.003	**0.046**	1.01 (1.01-1.03)	0.001	**0.032**
Lowest potassium (mEq/L)	0.07 (0.03 – 0.20)	-2.639	**< 0.001**	0.04 (0.01 – 0.26)	-3.139	**0.001**
Aldosterone post-confirmatory test (ng/dL)	1.01 (1.01 – 1.02)	0.006	**0.003**	1.01 (1.01 – 1.02)	0.004	**0.038**
Normal ipsilateral imaging	0.02 (0.01 – 0.10)	-3.876	**< 0.001**	0.02 (0.01 – 0.19)	-4.120	**0.001**
Normal contralateral imaging	4.97 (2.13 – 11.61)	1.603	**< 0.001**	5.35 (1.02 – 33.11)	1.678	**0.041**
Contralateral ratio	0.01 (0.01 – 0.03)	-5.710	**< 0.001**	0.01 (0.01 – 0.08)	-6.532	**0.001**



B

Diagnosis		Predicted	
		UPA	BPA
Real	UPA	112	14
	BPA	2	30
		Accuracy (%)	90.0
		Sensitivity (%)	88.8
		Specificity (%)	94.7
		10-fold cross validation (%)	88.9

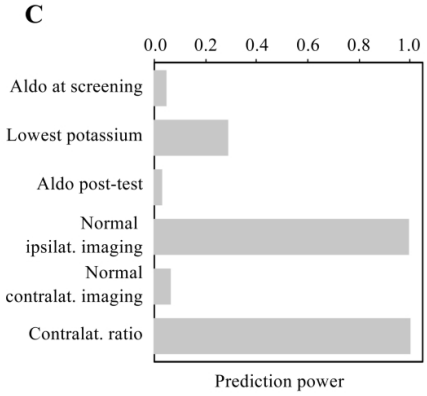
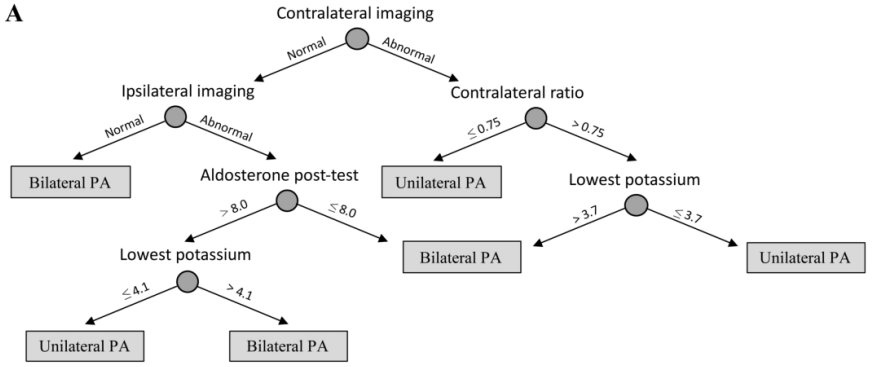


Figure 1. Diagnostic Modelling: Linear Discriminant Analysis.

126x141mm (600 x 600 DPI)



B

Diagnosis		Predicted	
		UPA	BPA
Real	UPA	116	10
	BPA	3	29
	Accuracy (%)	91.7	
	Sensitivity (%)	92.1	
	Specificity (%)	90.3	
	10-fold cross validation (%)	92.2	

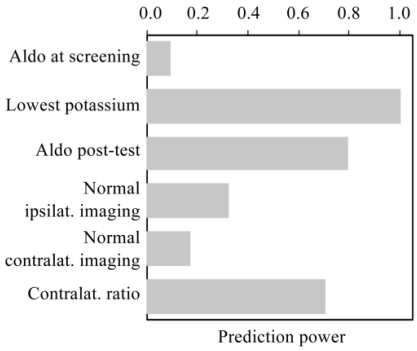


Figure 2. Diagnostic Modelling: Random Forest

127x110mm (600 x 600 DPI)

A

Variable	Label	Points
Aldosterone at screening (ng/dL)	</= 25.0	0
	25.1 – 54.9	0.5
	>/= 55.0	1
Lowest Potassium (mEq/L)	</= 3.1	5
	3.2 - 3.9	2
	>/= 4.0	0
Aldosterone post-confirmatory test (ng/dL)	</= 12.0	0
	12.1 – 19.9	0.5
	>/= 20.0	1
Ipsilateral Imaging	Normal	0
	Abnormal	5
Contralateral Imaging	Normal	4
	Abnormal	0
Contralateral Ratio	</= 0.30	3
	0.31-0.69	1.5
	>/= 0.70	0

B

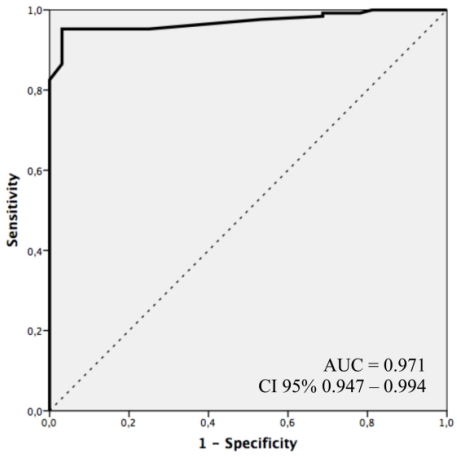
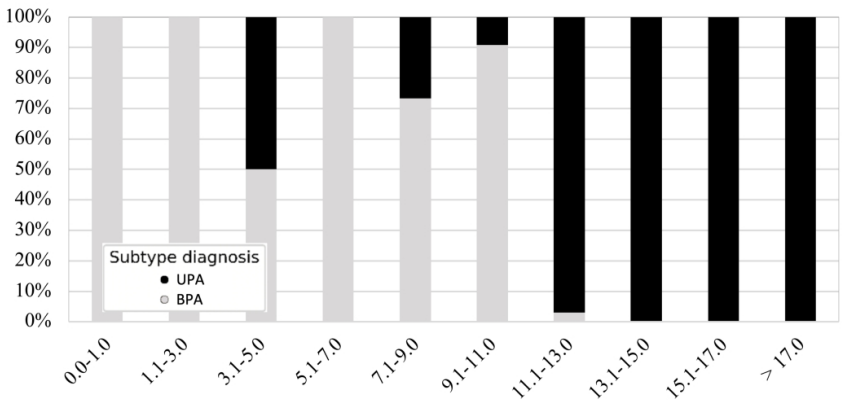


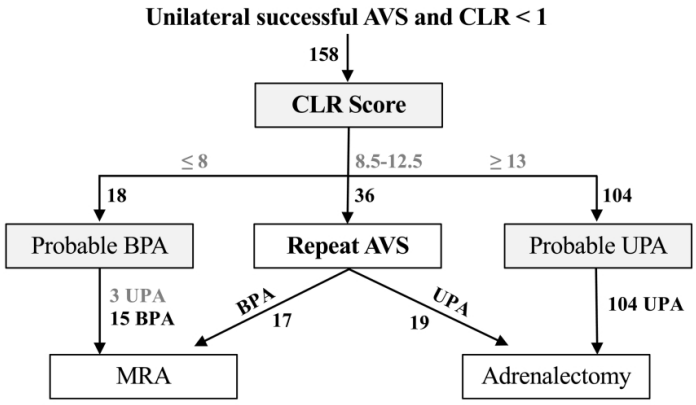
Figure 3. Score development

139x75mm (600 x 600 DPI)

A



B



C

Diagnosis		Predicted			
		UPA	BPA	Sensitivity (%)	97.6
Real	UPA	123	3	Specificity (%)	100.0
	BPA	0	32	PPV (%)	100.0
				NPV (%)	91.4

Figure 4. Score performance and management of PA patients

131x155mm (600 x 600 DPI)

Supplementary Table 1. Patient Characteristics of Study Cohort

Variable	Basal (N = 76)	ACTH (N = 82)	*P*-value
Diagnosis of UPA, n (%)	57 (75.0)	69 (84.1)	0.153
Female sex, n (%)	27 (35.5)	33 (40.2)	0.542
Age at diagnosis (years)	51 ± 9.3	49 ± 10.5	0.156
Duration of HTN (months)	69 [24; 176]	66 [22; 125]	0.342
Systolic BP (mmHg)	163 ± 25.2	164 ± 23.8	0.821
Diastolic BP (mmHg)	98 ± 13.0	97 ± 15.0	0.756
Antihypertensive medication (DDD)	4.0 [2.4; 5.7]	3.7 [2.2;5.0]	0.671
PRA at screening (ng/mL/h)	0.20 [0.10; 0.30]	0.20 [0.15; 0.40]	0.656
Aldosterone at screening (ng/dL)	38.9 [25.8; 50.5]	34.8 [24.3; 45.3]	0.198
Lowest potassium (mEq/L)	3.3 ± 0.7	3.1 ± 0.7	0.085
Aldosterone post-confirmatory test (ng/dL)	19.6 [11.9; 32.7]	19.6 [10.8; 32.0]	0.738
Microalbuminuria, n (%)	23 (30.4)	21 (25.0)	0.541
LVH at Echo, n (%)	49 (65.1)	44 (53.7)	0.188
CV events, n (%)	10 (12.7)	13 (16.4)	0.549
Largest nodule at CT scanning (diameter, mm)	12 [10; 16]	14 [10; 20]	0.189
Normal ipsilateral imaging, n (%)	8 (10.5)	8 (9.8)	0.873
Normal contralateral imaging, n (%)	62 (81.6)	62 (75.6)	0.362
Contralateral SI	14.6 [7.5; 33.3]	23.0 [13.9; 48.6]	**0.002**
Peripheral A/C ratio	17.8 [3.4; 28.7]	14.4 [4.1; 19.7]	0.348
Contralateral A/C ratio	4.4 [1.2; 9.9]	4.9 [1.6; 8.6]	0.867
Contralateral ratio	0.3 [0.2; 0.6]	0.4 [0.3; 0.6]	0.274

Clinical characteristics of patients included in the analysis: patients were stratified according to the protocol of adrenal venous sampling (basal *vs.* ACTH). Antihypertensive medication is expressed as Defined Daily Dose (DDD). DDD is the average maintenance dose per day for a drug used for its main indication in adults. UPA, Unilateral Primary Aldosteronism; HTN, Hypertension; BP, Blood Pressure; PRA, Plasma Renin Activity; LVH, Left Ventricular Hypertrophy; Echo, Echocardiography; CV, Cardiovascular; CT, Computed Tomography; SI, Selectivity Index; A/C, Aldosterone to Cortisol ratio. 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. Patient Characteristics of Study Cohort

Variable (ref. BPA)	OR (CI 95%)	β	*P*-value
Female sex, n (%)	2.59 (1.04 – 6.44)	0.943	**0.040**
Age at diagnosis (years)	0.94 (0.90 – 0.98)	-0.065	**0.006**
Duration of HTN (months)	1.01 (0.99 – 1.01)	0.002	0.236
Systolic BP (mmHg)	1.00 (0.98 – 1.02)	-0.001	0.933
Diastolic BP (mmHg)	1.00 (0.96 – 1.03)	-0.004	0.822
Antihypertensive medication (DDD)	1.43 (1.09 – 1.87)	0.357	**0.009**
PRA at screening (ng/mL/h)	4.40 (0.56 – 34.43)	1.480	0.159
Aldosterone at screening (ng/dL)	1.01 (1.01 – 1.01)	0.003	**0.046**
Lowest potassium (mEq/L)	0.07 (0.03 – 0.20)	-2.639	**< 0.001**
Aldosterone post-confirmatory test (ng/dL)	1.01 (1.01 – 1.02)	0.006	**0.003**
Microalbuminuria, n (%)	1.79 (0.54 – 5.88)	0.582	0.337
LVH at Echo, n (%)	0.39 (0.14 – 1.05)	-0.942	0.063
CV events, n (%)	0.69 (0.23 – 2.13)	-0.368	0.521
Largest nodule at CT scanning (diameter, mm)	0.99 (0.92 – 1.06)	-0.015	0.683
Normal ipsilateral imaging, n (%)	0.02 (0.01 – 0.10)	-3.876	**< 0.001**
Normal contralateral imaging, n (%)	4.97 (2.13 – 11.61)	1.603	**< 0.001**
Peripheral A/C ratio	1.03 (1.00 – 1.06)	0.028	0.071
Contralateral A/C ratio	0.91 (0.85 – 0.97)	-0.093	**0.007**
Contralateral ratio	0.01 (0.01 – 0.03)	-5.710	**< 0.001**

Univariate logistic regression analysis was performed to assess the odds ratio (OR), β estimate 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 (i.e. an OR of 1.01 for aldosterone means an increase of 1% of the likelihood of UPA for each 1 ng/dL increase of aldosterone concentration). Antihypertensive medication is expressed as Defined Daily Dose (DDD). DDD is the average maintenance dose per day for a drug used for its main indication in adults. HTN, Hypertension; BP, Blood Pressure; PRA, Plasma Renin Activity; LVH, Left Ventricular Hypertrophy; Echo, Echocardiography; CV, Cardiovascular; CT, Computed Tomography; SI, Selectivity Index; A/C, Aldosterone to Cortisol ratio. Ipsilateral and contralateral imaging were defined as "normal" in absence of nodules, or with thickening < 4 mm.

Supplementary Table 3. Patient Characteristics of Study Cohort

Variable	Combined Cohort (N = 158)	Training Cohort (N = 110)	Validation Cohort (N = 48)	*P*-value
ACTH protocol, n (%)	82 (51.9)	53 (48.2)	29 (60.4)	0.157
Diagnosis of UPA, n (%)	126 (79.7)	88 (80.0)	38 (79.2)	0.905
Female sex, n (%)	60 (38.0)	40 (36.4)	20 (41.7)	0.528
Age at diagnosis (years)	50 ± 10.0	50 ± 9.3	50 ± 11.5	0.949
Duration of HTN (months)	67 [23; 130]	66 [23; 128]	68 [24; 161]	0.825
Systolic BP (mmHg)	164 ± 24.5	163 ± 24.5	166 ± 24.6	0.534
Diastolic BP (mmHg)	98 ± 13.9	98 ± 13.8	97 ± 14.3	0.925
Antihypertensive medication (DDD)	3.8 [2.2; 5.0]	3.7 [2.4; 5.0]	4.0 [2.2; 5.5]	0.986
PRA at screening (ng/mL/h)	0.20 [0.12; 0.40]	0.29 [0.15; 0.40]	0.20 [0.10; 0.30]	0.166
Aldosterone at screening (ng/dL)	36.0 [25.5; 47.8]	35.9 [25.6; 47.3]	37.9 [25.0; 49.8]	0.867
Lowest potassium (mEq/L)	3.2 ± 0.7	3.2 ± 0.7	3.2 ± 0.7	0.658
Aldosterone post-confirmatory test (ng/dL)	19.6 [11.3; 32.0]	20.1 [13.1; 32.3]	16.4 [9.5; 33.0]	0.172
Microalbuminuria, n (%)	43 (27.5)	32 (29.1)	11 (21.9)	0.394
LVH at Echo, n (%)	94 (59.2)	65 (59.1)	29 (59.5)	0.963
CV events, n (%)	23 (14.6)	16 (14.4)	7 (15.0)	0.934
Largest nodule at CT scanning (diameter, mm)	13 [10; 19]	14 [10; 20]	12 [10; 18]	0.219
Normal ipsilateral imaging, n (%)	16 (10.1)	10 (9.1)	6 (12.5)	0.514
Normal contralateral imaging, n (%)	124 (78.5]	84 (76.4)	40 (83.3)	0.327
Contralateral SI	18.8 [10.2; 41.5]	18.8 [11.3; 47.5]	20.2 [7.9; 30.3]	0.463
Peripheral A/C ratio	17.8 [3.4; 28.7]	15.9 [3.7; 23.1]	14.3 [5.4; 26.0]	0.950
Contralateral A/C ratio	4.4 [1.2; 9.9]	4.4 [1.4; 9.1]	5.5 [1.8; 9.3]	0.456
Contralateral ratio	0.3 [0.2; 0.6]	0.3 [0.2; 0.6]	0.4 [0.2; 0.7]	0.393

Characteristics of patients included in the analysis: patients from the combined cohort (N = 158) were randomly assigned to training (N = 110), or validation cohort (N = 48). Antihypertensive medication is expressed as Defined Daily Dose (DDD). DDD is the average maintenance dose per day for a drug used for its main indication in adults. UPA, Unilateral Primary Aldosteronism; HTN, Hypertension; BP, Blood Pressure; PRA, Plasma Renin Activity; LVH, Left Ventricular Hypertrophy; Echo, Echocardiography; CV, Cardiovascular; CT, Computed Tomography; SI, Selectivity Index; A/C, Aldosterone to Cortisol ratio. 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. Diagnostic performance

Score Performance	N	Cut-off	Sens (%)	Spec (%)	PPV (%)	NPV (%)	Acc (%)
LDA Model (Training)	110	N.A.	90.6 (88.2-93.0)	99.4 (96.4-100.0)	99.8 (99.0-100.0)	72.6 (67.4-77.8)	92.3 (90.1-94.5)
Internal Validation	110		88.9 (85.6-92.2)	88.9 (77.7-100.0)	97.0 (94.1-99.9)	66.8 (59.5-74.1)	88.9 (85.3-92.5)
External Validation	48		84.7 (82.7-86.6)	84.2 (73.1-95.3)	95.4 (92.2-98.6)	59.1 (55.1-63.1)	84.6 (82.1-87.1)
Combined cohort	158		88.8 (86.8-90.8)	94.7 (91.1-98.3)	98.5 (97.5-99.5)	68.3 (64.7-71.9)	90.0 (88.5-91.6)
RF Model (Training)	110	N.A.	95.8 (94.1-97.5)	94.9 (91.7-98.1)	98.7 (97.9-99.5)	85.1 (80.0-90.2)	95.6 (94.2-97.0)
Internal Validation	110		92.2 (89.1-95.3)	79.5 (66.5-92.6)	94.7 (91.6-97.8)	71.9 (63.6-80.2)	89.6 (86.1-93.1)
External Validation	48		83.6 (79.0-88.2)	80.0 (77.8-82.2)	94.1 (93.8-94.4)	56.4 (49.2-63.6)	82.9 (79.3-86.5)
Combined cohort	158		92.1 (90.2-94.0)	90.3 (88.1-92.5)	97.4 (96.8-98.0)	74.5 (69.8-79.2)	91.7 (90.1-93.3)
CLR Score - Best Accuracy	110	> 11	95.5 (88.9-98.2)	100.0 (85.1-100.0)	100.0 (96.0-100.0)	84.7 (65.7-93.3)	96.4 (88.1-98.6)
Internal Validation	110		95.3 (83.9-100.0)	100.0 (97.5-100.0)	100.0 (98.4-100.0)	88.3 (59.9-95.4)	96.3 (87.7-100.0)
External Validation	48		94.7 (82.7-99.1)	90.0 (59.6-99.5)	97.3 (88.6-99.9)	81.7 (47.6-96.7)	93.7 (77.9-99.2)
Combined Cohort	158		95.2 (90.0-97.8)	96.9 (84.3-99.8)	99.2 (95.8-99.9)	83.7 (68.2-92.0)	95.5 (88.8-98.2)
CLR Score (Sensitivity Optimization)	110	> 8	97.8 (92.1-99.6)	45.5 (26.9-65.3)	87.8 (83.4-92.0)	83.8 (46.0-97.6)	87.3 (79.1-92.7)
Internal Validation	110		96.9 (87.6-100.0)	51.7 (28.0-75.4)	88.8 (72.0-100.0)	75.0 (47.0-96.9)	87.3 (71.0-94.7)
External Validation	48		97.4 (86.5-99.9)	50.0 (23.7-76.3)	88.1 (81.2-94.1)	83.5 (31.6-99.5)	87.5 (73.4-95.0)
Combined Cohort	158		97.6 (93.2-99.4)	46.9 (30.9-63.6)	87.9 (84.2-91.5)	83.2 (53.6-96.4)	87.3 (80.6-92.1)
CLR Score (Specificity Optimization)	110	≥ 13	81.8 (72.5-88.5)	100.0 (85.1-100.0)	100.0 (95.1-100.0)	57.9 (43.6-68.5)	85.4 (75.0-90.8)
Internal Validation	110		79.2 (51.4-93.2)	100.0 (97.8-100.0)	100.0 (97.3-100.0)	54.5 (24.8-83.1)	82.7 (59.5-95.8)
External Validation	48		84.2 (69.6-92.6)	100.0 (72.2-100.0)	100.0 (90.5-100.0)	62.5 (38.5-78.1)	87.5 (70.1-94.1)
Combined Cohort	158		82.5 (75.0-88.2)	100.0 (89.3-100.0)	100.0 (96.5-100.0)	59.2 (47.6-68.3)	86.0 (77.9-90.6)

Sensitivity (Sens), specificity (Spec), positive/negative predictive value (PPV/ NPV), and accuracy (Acc) for proposed models (each indicator is derived considering unilateral PA as referral diagnosis and reported together with its 95% binomial confidence interval; confidence intervals are derived by normal approximation method and capped to 100%). Diagnostic indices are provided at training, at internal validation by 10-fold cross validation, at external validation, and in the combined cohort. A cut-off of greater than 11 corresponded to the higher model accuracy, as assessed by ROC curve analysis. The cut-offs of > 8 and ≥ 13 increased sensitivity or specificity, respectively, providing an overall model accuracy of at least 85% at validation. LDA, Linear Discriminant Analysis; RF, Random Forest.

Supplementary Table 5. Estimate predictor importance for LDA and RF models

Variable	LDA	RF
Aldosterone at screening (ng/dL)	0.047626	0.093263
Lowest potassium (mEq/L)	-0.287831	1.000000
Aldosterone post-confirmatory test (ng/dL)	0.033038	0.797178
Normal ipsilateral imaging, n (%)	-0.994998	0.327111
Normal contralateral imaging, n (%)	0.064665	0.172839
Contralateral ratio	-1.000000	0.707113

Normalized LDA and RF importance coefficients computed at the training of the models. A highest absolute value corresponds to the best predictor in each model.

Supplementary Table 6. Score development and validation

AVS Score Accuracy			Predicted Diagnosis		Performance	
Real Diagnosis (Cut-off > 11)	**Training cohort** (N = 110)		UPA	BPA	Accuracy (%)	96.4
		UPA	84	4	Sensitivity (%)	95.5
		BPA	0	22	Specificity (%)	100.0
	Validation cohort (N = 48)		UPA	BPA	Accuracy (%)	93.7
		UPA	36	2	Sensitivity (%)	94.7
		BPA	1	9	Specificity (%)	90.0
	Combined cohort (N = 158)		UPA	BPA	Accuracy (%)	95.5
		UPA	120	6	Sensitivity (%)	95.2
		BPA	1	31	Specificity (%)	96.9
Real Diagnosis (Cut-off > 8)	**Training cohort** (N = 110)		UPA	BPA	Accuracy (%)	87.3
		UPA	86	2	Sensitivity (%)	97.8
		BPA	12	10	Specificity (%)	45.5
	Validation cohort (N = 48)		UPA	BPA	Accuracy (%)	87.5
		UPA	37	1	Sensitivity (%)	97.4
		BPA	5	5	Specificity (%)	50.0
	Combined cohort (N = 158)		UPA	BPA	Accuracy (%)	87.3
		UPA	123	3	Sensitivity (%)	97.6
		BPA	17	15	Specificity (%)	46.9
Real Diagnosis (Cut-off >/= 13)	**Training cohort** (N = 110)		UPA	BPA	Accuracy (%)	85.4
		UPA	72	16	Sensitivity (%)	81.8
		BPA	0	22	Specificity (%)	100.0
	Validation cohort (N = 48)		UPA	BPA	Accuracy (%)	87.5
		UPA	32	6	Sensitivity (%)	84.2
		BPA	0	10	Specificity (%)	100.0
	Combined cohort (N = 158)		UPA	BPA	Accuracy (%)	86.0
		UPA	104	22	Sensitivity (%)	82.5
		BPA	0	32	Specificity (%)	100.0

The table shows the real and predicted subtype diagnosis, accuracy sensitivity, specificity for the training cohort (N = 110), the validation cohort (N = 48), and the combined cohort (N = 158). A cut-off of greater than 11 identifies patients with unilateral primary aldosteronism (UPA) with the higher accuracy. A cut-off of greater than 8 identifies patients with a diagnosis of UPA with an optimized sensitivity; cut-off of greater than or equal to 13 identifies patients with a diagnosis of UPA with an optimized specificity. BPA, Bilateral Primary Aldosteronism.

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

Score points	Total (N)	UPA		BPA	
		(N)	(%)	(N)	(%)
0.0-1.0	1	0	0.0	1	100.0
1.1-3.0	5	0	0.0	5	100.0
3.1-5.0	2	1	50.0	1	50.0
5.1-7.0	3	0	0.0	3	100.0
7.1-9.0	15	4	26.7	11	73.3
9.1-11.0	11	1	9.1	10	90.9
11.1-13.0	33	32	97.0	1	3.0
13.1-15.0	21	21	100.0	0	0.0
15.1-17.0	34	34	100.0	0	0.0
> 17.0	33	33	100.0	0	0.0
Total	158	126	N.A.	32	N.A.

The number (N) and the proportion (%) of patients stratified for subtype diagnosis (unilateral PA *versus* bilateral PA) is shown according to the score in the combined cohort (N = 158).

Supplementary Table 8. Sub-analysis according to patient outcome

Subgroup of patients	Clinical Outcome	Biochemical Outcome
UPA predicted as UPA (true prediction)		
Complete, n (%)	51 (53.1)	94 (97.9)
Partial, n (%)	41 (42.7)	2 (2.1)
Absent, n (%)	4 (4.2)	0 (0.0)
UPA predicted as BPA (false prediction)		
Complete, n (%)	0 (0.0)	4 (66.7)
Partial, n (%)	5 (83.3)	2 (33.3)
Absent, n (%)	1 (16.7)	0 (0.0)
Unstimulated AVS		
Complete, n (%)	21 (42.9)	48 (98.0)
Partial, n (%)	27 (55.1)	1 (2.0)
Absent, n (%)	1 (2.0)	0 (0.0)
ACTH-stimulated AVS		
Complete, n (%)	30 (56.7)	50 (94.3)
Partial, n (%)	19 (35.8)	3 (5.7)
Absent, n (%)	4 (7.5)	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 CLR score predicted diagnosis (true *vs.* false predictions) or AVS protocol (ACTH-stimulated *vs.* unstimulated procedures).