Exclusion tests in Unilateral Primary Aldosteronism (ExcluPA) Study

Subtitle: A systematic review and meta-analysis

Author names: Rui Zhu, MD, PhD[1,2]*; Tungalagtamir Shagjaa, MD[1,3]*; Giacomo Rossitto, MD, PhD[1,4]; Brasilina Caroccia, PhD[1]; Teresa Maria Seccia, MD, PhD[1]; Dario Gregori, MA, PhD[5]; Gian Paolo Rossi, MD, FACC, FAHA[1]

Affiliation:
[1] Internal & Emergency Medicine Unit, Department of Medicine – DIMED, University of Padua, Padua, Italy;
[2] Department of Endocrinology, Sichuan Academy of Medical Sciences & Sichuan Provincial People's Hospital, Chengdu, China;
[3]Department of Neurology, Mongolian National University of Medical Sciences, Ulaanbaatar, Mongolia;
[4] Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow UK;
[5] Unit of Biostatistics, Epidemiology and Public Health, Department of Cardiac, Thoracic and Vascular Sciences and Public Health, University of Padua, Padua, Italy.

ORCID number: 0000-0002-7963-0931 (G. P. Rossi)

*These authors have equally contributed.

Correspondence and reprint request to:
Prof. Gian Paolo Rossi, MD, FACC, FAHA
Emergency Internal Medicine Unit & Specialized Center for blood pressure disorders
Department of Medicine -DIMED
University of Padua
via Giustiniani, 2, 35122 Padova, Italy.
E-mail: gianpaolo.rossi@unipd.it

Disclaimers: none
Disclosure of relationship and activities: none

Total word: 3452
Number of Figures and tables: figures 4, tables 2
Funding source: This study supported by the following research Grants: FORICA (The Foundation for advanced Research In Hypertension and CArdiovascular diseases) to G.P. Rossi, the China Scholarship Council (CSC) for PhD to RZ, and The International PhD Program in Arterial Hypertension and Vascular Biology University of Padua, and University of Padua DOR2045593/20 to TMS.

Key points

Question: Are exclusion tests necessary to diagnose unilateral primary aldosteronism?

Findings: A meta-analysis of 31 separate datasets comprising 4,242 patients showed that the exclusion tests furnished no diagnostic gain over the aldosterone-to-renin ratio.

Meaning: The systematic use in clinical practice of exclusion tests for primary aldosteronism is not supported by available evidences.

Abstract

Importance: Determining the diagnostic accuracy of "exclusion" tests for primary aldosteronism (PA) compared to the aldosterone-to-renin ratio (ARR) is fundamental to avoid invasive subtyping in false-positive patients at screening.

Objective: To assess the accuracy of exclusion tests for PA using the diagnosis of unilateral PA as reference.

Data Sources: PubMed, EMBASE, Web of Science and Cochrane Library databases.

Study Selection: Studies that met tight quality criteria published from January 1st, 1970 to December 31st, 2021.

Data Extraction and Synthesis: Data were extracted following the PRISMA methodology. We performed a two-stage meta-analysis that entailed an exploratory and a validation phase based on a "golden" or "gold" diagnostic standard, respectively. Pooled specificity, negative likelihood ratio, diagnostic odds ratio, and summary area under the ROC curve (sAUROC) were calculated.

Main Outcome and Measure: The accuracy of exclusion tests.

Findings: 31 datasets comprising a total of 4,242 patients fulfilling the predefined inclusion criteria were meta-analyzed. Pooled accuracy estimates (sAUROC) did not differ between the ARR (0.95, 95% CI: 0.92-0.98), the captopril challenge test (CCT) (0.92, 95% CI: 0.88-0.97), and the saline infusion test (SIT) (0.96, 95% CI: 0.94-0.99). Solid information could not be obtained for the fludrocortisone suppression test and the furosemide upright test, which were assessed in only one study each.

Conclusions and Relevance: The apparently high diagnostic accuracy of the CCT and

the SIT was due to the selection of patients with an elevated ARR and thus a high pre-test probability of unilateral PA; however, neither test furnished a diagnostic gain over the ARR. Therefore, the systematic use of these exclusion tests in clinical practice is not justified by available evidence.

INTRODUCTION

In about half of the hypertensive patients on medications, the control of high blood pressure (BP) values remains disappointing despite improvements in awareness and treatment in the last decades (1). A major reason for this poor high BP control is overlooking secondary hypertension (1,2), of which primary aldosteronism (PA) is the most common (3,4).

Currently, owing to poor clinical awareness, with ensuing "under suspicion", unduly complex diagnostic work-up (5,6), limited availability of invasive investigations for localizing unilateral PA (uPA), constrained surgical capacity, and uncertainties about clinical outcomes (5), PA remains markedly underdiagnosed and undertreated.

The strategy for case detection of PA relies on the aldosterone-to-renin ratio (ARR) with use of low cut-off values to maximize sensitivity (6,7), a strategy that generates many false positives. These false positives patients must be excluded from invasive subtyping procedures that are scarcely available and costly and, therefore, should be reserved for patients with a high prior probability of harboring surgically curable uPA.

For this selection, "confirmatory" tests, including the captopril challenge test (CCT), the fludrocortisone suppression test (FST), the saline infusion test (SIT), the oral sodium loading test (OLT), and the furosemide upright test (FUT), have been proposed (5,8). In reality, they serve as "exclusion" tests, because at the prevalence rate of PA encountered in ARR-positive patients, their negative predictive value exceeds their positive (confirmatory) predictive value (9).

The validation of these tests, which increase complexity and costs of the diagnostic work-up, and involve the risk of overlooking the angiotensin II-responsive PA patients (10–12), was affected in many studies by a vicious circle type of bias in that it relied on

using another arbitrarily chosen exclusion test, as reference (13). To date, only few studies followed the Standards for Reporting Diagnostic Accuracy (STARD) statement (14) and used an unambiguous diagnosis of uPA as reference; furthermore, many preselected patients based on a positive ARR result, which, by increasing the rate of PA, led to overestimating the accuracy, *i.e.* sensitivity and specificity, of these tests.

Given the lack of studies that conclusively proved the accuracy and diagnostic gain of exclusion tests over the ARR, we conceived this meta-analysis to determine if available studies justify the systematic use of these tests in the work-up of unilateral PA.

METHODS

This meta-analysis entailed two phases, based on a different level of diagnostic uncertainty: an exploratory phase that used a golden reference, and a validation phase based on a gold standard. The "gold" reference standard used in the validation phase of the exclusion test(s) included the "four corners" criteria (Table S1). If the latter was not available, in the exploratory phase, we used a "golden" standard (comprising documentation of excess aldosterone production, positive adrenal imaging and/or lateralization of aldosterone on AVS), as a suboptimal surrogate (Table S1) (15).

Data Sources

The meta-analysis was registered on PROSPERO (ID:343220) (15) and followed the preferred reporting items for systematic review and meta-analysis (PRISMA) statement (16). The details of our search of articles on exclusion tests of human PA published from January 1st, 1970 to December 31st, 2021 are reported in Supplemental Literature Search (15).

Study Selection

After removal of duplicates, articles eligibility was assessed by two younger (RZ, and TS), and three senior investigators (TMS, GR, and GPR). The study inclusion criteria were: (1) prospective or retrospective design; (2) uPA diagnosis established by the gold or a golden standard; (3) reported diagnostic accuracy of the tests; (4) sufficient data to construct a 2x2 diagnostic table.

Exclusion criteria comprised: (1) reviews, case reports, case-control studies; (2) duplicated data.

Data Extraction

Data extraction was performed using a predefined standardized form. The ARR cutoff values of the different studies are reported in Table 1, whenever available. When accuracy of ARR and/or exclusion test(s) was reported for different cut-off values, the value that provided the best combination of sensitivity and specificity was selected.

Quality Assessment

Given the limitations of the QUADAS-2 method, which only allows a categorical Y/N classification, to assess the quality of the studies, we developed a novel quantitative scoring method that comprised 9 items in 4 domains: 'study design and patient selection', 'index test', 'reference standard', and 'flow and timing' (Table S2) (15).

In this scoring method we used a 10-cm digital scale on which the 3 senior investigators put a mark according to his/her judgement of the article quality regarding each item. Divergences of scores exceeding 3 cm were resolved by consensus. Each item received a score from 0 to 10; for items that required a Y/N answer, 10 corresponded to YES and 0 to NO. The different scores were then summed up and generated the overall score, which ranged from a minimum of 0 to a maximum of 90. We determined beforehand

that for inclusion studies had to fulfill the following quality criteria: enrollment of consecutive newly-diagnosed hypertensive patients, pre-specified cut-off values for the index tests, the same gold or golden reference standard for uPA diagnosis in all patients, and appropriate follow-up (in those with the gold reference), and to reach an overall score of at least 25. The scores of the individual studies meta–analyzed are given in Table S3 and Figure S1 (15).

Data Synthesis and Analysis

The threshold effect was assessed by the Spearman correlation coefficient between Logit(sensitivity) and Logit (1-specificity) with p value > 0.05 indicating a non-threshold effect. I^2 was used to evaluate heterogeneity among studies. The random effects model was used when I^2 was > 30% (17).

Summary receiving operation characteristic (sROC) curves and area under the curve (sAUROC), pooled specificity, negative likelihood ratio (NLR), and diagnostic odds ratio (DOR) with 95% confidence intervals (CI) were computed (18).

Meta-regression was performed to identify covariates that affected heterogeneity.

A sensitivity analysis was performed to evaluate the quality and consistency of results by sequentially excluding each single study at a time (19) Potential publication bias was evaluated by the p value of Deeks' funnel plot (20).

All analyses were performed with Meta-Disc version 1.4 and STATA version 12.0 (Stata Corp, College Station, TX); statistical tests were two-sided, with a p value < 0.05 denoting statistical significance.

RESULTS

Study selection

We identified 3010 relevant articles through database search. After removal of duplicates (n=1387), and studies that, based on abstract (n=1,563) and on full-text reading (n=40), were not relevant a total of 1,623 remained, as detailed in Table S4. A total of 20 articles, entailing 31 separate datasets and 4,242 patients, were eligible for the meta-analysis: 11 were examined in the exploratory phase, and 20 in the validation phase (Figure S2) (15).

Study Characteristics and Quality Assessment

Notwithstanding our strict inclusion criteria, the overall scores of study quality showed a wide range (Table S3 and Figure S1) (15). All selected studies were conducted in tertiary referral centers located in 3 continents (Europe, Asia, and Oceania).

Their main characteristics are detailed in Table 1 and Table S5 (15): six studies assessed the diagnostic accuracy of ARR (21–26), 3 of CCT (27–29), 4 of SIT (30–33), and 7 of multiple tests using a head-to-head comparison (9,34–39): one study compared ARR and CCT(9); one ARR and SIT (34); one ARR, CCT, SIT, and FUT (36); one SIT and FST (39), and 3 CCT and SIT (35,37,38).

No studies assessed the diagnostic accuracy of OLT vs the ARR. Thus, on the whole 10 datasets on ARR (9,21–26,34,36), 9 on CCT (9,27–29,35–38), and 10 SIT (30–39) were analyzed.

Target population

The study population comprised a total of 4,242 patients, 16% of whom (n=677) had uPA (Table 1). Two studies comprised newly diagnosed hypertensives (9,21); 8 studies recruited patients with a high prior probability of PA (22–26,28,29,35); 7 studies

patients with a positive ARR (27,30–32,34,36); 2 studies included patients with positive ARR plus one-third/fourth of patients with negative ARR (37,38); one study recruited patients with a positive recumbent SIT (33).

Controls comprised mainly patients with primary (essential) hypertension (PH) except in 3 studies: one that examined patients with PH and bilateral adrenal hyperplasia (BAH) (36); one that used patients with PH plus UAH plus BAH (23), and one that used patients with non-PA (39). Interfering medications were withheld for at least two weeks, and switched to calcium channel antagonists and/or α-blockers in all but one study (Table S5 (15)) (22).

Reference index

The diagnosis of PA was established by a positive ARR result in 3 studies (21,26,35), by the ARR and/or exclusion test(s) in 4 studies (9,24,34,37), and by positive exclusion test(s) result in 13 studies (22,23,25,27-33,36,38,39).

The exploratory phase comprised 6 studies in 133 patients with uPA (23,24,26,29,32,36). The validation phase entailed in 14 studies in 544 patients with uPA (9,21,22,25,27,28,30,31,33–35,37–39).

The studies on the ARR were 10, equally split into the exploratory (23,24,26,34,36) and the validation (9,21,22,25) phase (Table 2). The mean ARR partition value was 48.1 ng/mIU (39.4 ng/dL/ng/mL/h) with a range from 19.1 to 96.4 ng/mIU (corresponding to 15.7 to 79.0 ng/dL/ng/mL/h) (9,21–26,34,36). ROC curve-based Youden index (YI) analysis for the diagnosis of uPA was used to determine the optimal cut-off value only in 5 studies (9,21,23,36).

The studies on the CCT were 9 (9,27–29,35–38): two in the exploratory phase (29,36), and 7 studies in the validation phase (9,27,28,35,37,38) (Table 2). All used a captopril

dose of 50 mg except one (35) that used 25 mg; patients were kept in a sitting position for 1-2 hours after drug challenge. The test readout differed: 5 studies used post-captopril ARR cutoff values (ranging from 15.6 to 81.6 ng/mIU [12.8 to 66.9 ng/dL/ng/mL/h]) (9,28,29,36); 4 used the post-captopril PAC values (with cutoffs ranging between 13.0 and 19.0 ng/d) (27,35,37,38). Seven studies selected the cutoffs by ROC curve and YI analysis using the diagnosis of uPA as category (9,28,29,36–38).

The studies on the SIT were 10 (30–39); all were performed in the supine position except 2 (33,39) that used a seated position. The post-saline PAC cutoff value was chosen according to ROC curve and YI analysis for uPA diagnosis only in 5 studies; it ranged from 3.0 to 15.2 ng/dL, with a mean value of 8.1 ng/dL (30,36–39). Three studies belonged to the exploratory phase (32,34,36) and 7 to the validation phase (30,31,33,35,37–39) (Table 2).

We could find only one eligible study for the FST (39) and the FUT (36). For FST, uPA was diagnosed by the gold standard, and the cut-off value of post-fludrocortisone PAC was 162 pmol/L (5.8 ng/dL); for the FUT, uPA was diagnosed by the golden standard, and the cut-off value of post-furosemide PRA was 0.55 ng/mL/h (Table 2).

Meta-Analysis

I^2 values of pooled specificity, NLR, DOR for ARR, CCT, and SIT were all > 30%, denoting a heterogeneous non-threshold effect for all tests. No evidence for a diagnostic threshold was detected. Hence, the heterogeneity analysis for ARR, CCT and SIT was performed by non-threshold effects using the random effects model.

Given the scope of the exclusion tests, the results are herein reported as specificity, NLR, and sAUROC, as prevalence-dependent indexes, and as DOR, as a prevalence-independent overall measure of diagnostic accuracy.

For ARR, the pooled specificity, NLR, and DOR were 0.90 (95% CI: 0.85-0.94), 0.12 (95% CI: 0.06-0.24), and 67.61 (95% CI: 27.05-168.99), respectively (Figure 1). The sAUROC was 0.95 (95% CI: 0.92-0.98) (Figure 2).

For CCT, the pooled specificity, NLR, and DOR were 0.83 (95% CI: 0.79-0.87), 0.19 (95% CI: 0.12-0.31), and 36.89 (95% CI: 16.16-84.17), respectively (Figure 3). The sAUROC was 0.92 (95% CI: 0.88-0.97) (Figure 2).

For SIT, the pooled specificity, NLR, and DOR were 0.87 (95% CI: 0.81-0.892), 0.11 (95% CI: 0.07-0.17), and 99.3 (95% CI: 40.9-241.79), respectively (Figure 4). The sAUROC was 0.96 (95% CI: 0.94-0.99) (Figure 2).

Overall, these results would indicate a similarly high accuracy (sAUROC) for ARR, CCT, and SIT, without significant differences among tests ($p = 0.328$ for ARR vs CCT; $p = 0.566$ for ARR vs SIT; $p = 0.125$ for CCT vs SIT). Due to the limited available studies no conclusion was feasible for FST and FUT.

Meta-Regression

A multivariable meta-regression analysis, including continent, populations (patients with suspected PA or with positive ARR), cut-off value (chosen by ROC curve or not), and reference standard for uPA diagnosis (gold or golden), was performed. The post-CCT readout variable (PAC or ARR) and posture (seated or supine) were also considered when exploring the heterogeneity of CCT and SIT results, respectively. For ARR, the selection of the populations (patients with suspected PA or with positive ARR) partly explained the heterogeneity ($p= 0.024$) (Table S6) (15). For CCT, the readout variable had an impact on heterogeneity ($p= 0.026$) (Table S7) (15). For SIT, none of these factors accounted for the heterogeneity (Table S8) (15).

Sensitivity analysis and publication bias

A sensitivity analysis performed by omitting each single study, showed no significant difference in the pooled results for ARR, CCT, and SIT (Figure S3-5) (15).

Neither Deek's funnel plot nor Deek's test showed evidence of publication bias (p= 0.73 for ARR, p= 0.81 for CCT, p= 0.76 for SIT) (Figure S6) (15).

DISCUSSION

We meta-analyzed the diagnostic accuracy of the most popular tests for the screening and exclusion of PA following the STARD recommendation (14), by a novel quantitative approach to select eligible studies. This method was developed, because we found that by the classical QUADAS-2 scale, the categorical assessment as Y/N was inadequate to depict the real quality of the studies. We determined beforehand to examine only studies that met predefined quality criteria in that they used a solid reference index, and were not tautologically biased by attempting to validate an exclusion test employing another non infallible test. Moreover, considering that the unilateral surgically curable forms of PA comprise the only PA subtype that can be unambiguously diagnosed, we decided to meta-analyze studies that used, as reference, the "gold" diagnosis of uPA confirmed by biochemical cure after surgery. However, in an exploratory survey, we also used a less certain (golden) diagnosis of uPA, entailing results of AVS, and imaging. This methodological approach by no means implies that it is not important to diagnose also bilateral PA, as these patients can benefit by target treatment with mineralocorticoid receptor antagonists and/or the upcoming aldosterone synthase inhibitors (40).

Based on these criteria involving a different level of diagnostic certainty, our meta-analysis involved two-phases: an exploratory and a validation phase that comprised

studies that used a golden and a gold reference, respectively.

The first most important findings were that: i) the ARR, when carefully performed in a standardized way in referral centers (6), provided a high accuracy for identification of uPA; ii) neither CCT nor SIT furnished an additional diagnostic gain (Figure 2).

It is worth mentioning that in 2003 Tanabe et al reported that plasma aldosterone concentration and plasma renin activity oscillates when measured repeatedly in patients with PA. They concluded that '*the renin/aldosterone profile in PA is not always abnormal due in part to conditions for blood sampling*'. This important study showed that most PAC values were higher than 15 ng/dl and exhibited a magnitude-related variability so that the higher values remained pathological even after ascillation, and that the ARR oscillation was large part due to the PRA values variability that showed no magnitude-related (41). However, in a sub-study of the PAPY Study (4), Seccia et al found that, notwithstanding some oscillation of PAC and PRA, the ARR was quite reproducible within-patient when performed under carefully standardized conditions (7), thus indicating its diagnostic usefulness for detecting PA.

The usefulness of the ARR and the lack of diagnostic gain of exclusion tests over it is in keeping with previous investigations: the largest study that prospectively examined with a standardized protocol over two thousand newly-diagnosed referred hypertensive patients, 4% of whom received a diagnosis of uPA by the gold criterion, reported identical results (9). Likewise, a prospective Japanese study of 102 patients with an elevated ARR (>20 ng/dL/ng/mL/h), where the accuracy of the ARR for discriminating uPA from PH and IHA patients was compared to CCT, SIT and FUT (36), came to the same conclusions. Different conclusions were reached in another study, which suggested post-SIT PAC as the test with the highest accuracy, i.e. a sensitivity of 97% and a specificity of 92% in 104 consecutive patients with suspected PA (34). However,

315 the study design entailed only a *golden* standard; moreover, the PAC cut-off value,
316 measured by liquid chromatography tandem mass spectrometry (LC-MS/MS), was 83
317 pmol/L (3.0 ng/dL), which corresponds to the lowest limit of detection for reliable PAC
318 measurement of their assay. In the PAPY study that used Youden index analysis to
319 determine post-SIT cut-off value with radioimmunoassay (RIA), the optimal cutoff was
320 more than two- fold higher (6.75 ng/dL) (37), closer to that (5 ng/dL) found by others
321 (22). While LC-MS/MS might give lower PAC values than the immunoassays owing
322 to the lack of antibody cross-reactivity with other steroids (42), 3 ng/dl value is a far
323 too low PAC value for an exclusion test. In fact, studies that used uPA as the gold
324 reference found that the diagnostic performance of such low cutoff was hampered by
325 the huge overlap of post-SIT PAC values between uPA patients and PH, and/or bilateral
326 PA patients (37,43). This means that a PAC partition value of 3.ng/dL would likely
327 identify all uPA patients, thus providing a very high sensitivity, but would also generate
328 many false positive ARR results, which certainly is an annoying outcome for an
329 exclusion test. However, according to a US multicenter study of a sizable cohort of
330 patients with untreated normotension and different stages of hypertension the
331 sensitivity of the baseline ARR for detecting PA would be poor (44), thus implying the
332 need of using low partition values. Undoubtedly suppression of PAC, and thus of the
333 ARR, by means of the exclusion tests would further worsen the current under detection
334 of curable PA.

335 It is worth underlining that the studies on exclusion tests showed a prominent
336 heterogeneity, whose source could not be entirely revealed by a meta-regression, owing
337 to several reasons. One, likely the most important, regards use of PA or uPA diagnosis,
338 as reference. Another one was, as alluded above, the biochemical methods for
339 measuring aldosterone: currently, there are three methods for measuring aldosterone:

LC-MS/MS, and the immunoassays (RIA, and chemiluminescence (CLIA)). Although expected to provide identical results, in reality, they exhibit significant inter-assay variabilities (45). A source of heterogeneity also comprises the patient's preparation: a two-week withdrawal of interfering drugs is too short, particularly for β-blockers and RAS inhibitors that, in our experience, affect renin levels for more than 4 weeks (6).

A further source of variation relates to the calculation of the ARR, that can be performed with renin measured as PRA (by RIA) or as direct renin concentration by CLIA or by LC-MS/MS. These assays provide the results in different units of measure and, although values can be easily converted by the available ARR-APP (46), this was not systematically exploited in published studies: in fact, we had to exclude one study from this meta-analysis because of obviously wrong renin data.

The thresholds also differed among studies: 45% of the investigators used arbitrarily chosen cutoff values and only 55% of them determined their cutoffs by a rigorous ROC curve and Youden index analysis. Further heterogeneity originated from the choice of the readout of this test. For example, for the CCT the readout was either PAC or the ARR, or both.

A preselection bias was also evident, at least in some studies: the ARR was not done in consecutive hypertensive patients but mostly in patients with suspected PA with few exceptions (4,9,22,28,29,35). Not unexpectedly, when performed in patients selected for a higher prior probability of PA, mostly in patients with positive ARR, the exclusion tests provided high sensitivity and specificity. Accordingly, the uPA/controls ratio differed by 26.5-fold (from 2.8% to 73.5%, 32.9 ± 24.2%), and by 5.9-fold (from 11.8% to 69.7%, 37.6 ± 20.2%) for SIT.

Finally, factors as race, serum K^+ levels, salt intake at testing, might have also

contributed to heterogeneity, although they did not emerge at meta-regression.

Another important methodological flaw needs to be mentioned: 48% of the studies performed an entire work-up only in the cases with a positive result of the ARR and/or the exclusion test, and not in those with negative results. This verification bias, by leaving the diagnosis uncertain in the negative cases, might have contributed to overestimating the test performance.

Of note, the FST, which has been proposed as the most reliable exclusion test for PA (38), is supported by only a single-center study (39); the same applies to the FUT. Thus, these tests did not lend themselves to a meta-analysis.

Notwithstanding the ineludible limitations intrinsic to the study heterogeneity discussed above, this meta-analysis has major strengths that comprise a painstaking selection of the eligible studies based on a novel quantitative analysis of their quality, the evaluation of the performance of each test using uPA as reference, and the fact that data were examined according to the level of diagnostic certainty.

In summary, the present investigation reveals that studies of exclusion tests for PA are markedly heterogenous. Even when restricted to the studies that met the tightest quality criteria, our meta-analysis showed no evidence to support the systematic use of exclusion tests in clinical practice. Importantly, albeit seemingly highly sensitive and specific, the exclusion tests did not provide any diagnostic gain over a well performed ARR (Figure 2). As these tests contribute to the under-detection of PA, are time-consuming, increase the costs and complexity of the diagnostic work-up of PA, and are not free of risks, because of the need to keep patients on the “switch” antihypertensive treatment, their usefulness should be proven in a large outcome-based prospective study comparing head-to-head strategies “with and without exclusion tests” before their

388 systematic use can be recommended. Of note, while this manuscript was under
389 evaluation, another meta-analysis, albeit carried out with a different methodology and
390 a less stringent selection of the studies, reached similar conclusions (47).

391 Finally, exclusion tests are based on the premise of excess production of aldosterone
392 autonomous from angiotensin II, whilst human aldosterone-producing adenoma (APA)
393 were consistently found to express the angiotensin type I receptor, which mediates the
394 secretagogue action of on aldosterone. Moreover, angiotensin II-induced aldosterone
395 secretion from APA strips and cells ex vivo has been demonstrated (48). Thus, relying
396 on exclusion tests may preclude the chance of long-term surgical cure to patients with
397 angiotensin II-responsive uPA.

CONTRIBUTIONS

GPR conceived the study. RZ, and TS did the literature search and analyzed the data. RZ, TS, TMS, and GR contributed to study protocol and key data interpretation. RZ, and DG performed the analysis. RZ, TS, TMS, BC and GPR wrote the manuscript. RZ, TS, TMS, GR, and GPR critically revised the manuscript.

DECLARATION OF INTEREST

We declare no competing interests.

DATA AVAILABILITY

Some or all data generated or analyzed during this study are included in this published article or in the data repositories listed in References.

REFERENCE

1. **Olsen MH, Angell SY, Asma S, Boutouyrie P, Burger D, Chirinos JA, Damasceno A, Delles C, Gimenez-Roqueplo AP, Hering D, López-Jaramillo P, Martinez F, Perkovic V, Rietzschel ER, Schillaci G, Schutte AE, Scuteri A, Sharman JE, Wachtell K, Wang JG.** A call to action and a lifecourse strategy to address the global burden of raised blood pressure on current and future generations: the Lancet Commission on hypertension. *Lancet* 2016;388(10060):2665–2712.

2. **Azizi M, Sapoval M, Gosse P, Monge M, Bobrie G, Delsart P, Midulla M, Mounier-Véhier C, Courand PY, Lantelme P, Denolle T, Dourmap-Collas C, Trillaud H, Pereira H, Plouin PF, Chatellier G.** Optimum and stepped care standardised antihypertensive treatment with or without renal denervation for resistant hypertension (DENERHTN): A multicentre, open-label, randomised controlled trial. *Lancet* 2015;385:1957–1965.

3. **Xu Z, Yang J, Hu J, Song Y, He W, Luo T, Cheng Q, Ma L, Luo R, Fuller P, Cai J, Li Q, Yang S, Group and for the CPAS (CONPASS), Group.** Primary Aldosteronism in Patients in China With Recently Detected Hypertension. *JACC* 2020;75(16):DOI: 10.1016/j.jacc.2020.02.052.

4. **Rossi GP, Bernini G, Caliumi C, Desideri G, Fabris B, Ferri C, Ganzaroli C, Giacchetti G, Letizia C, Maccario M, Mallamaci F, Mannelli M, Mattarello M-JJ, Moretti A, Palumbo G, Parenti G, Porteri E, Semplicini A, Rizzoni D, Rossi E, Boscaro M, Pessina AC, Mantero F, PAPY Study**

Investigators. A Prospective Study of the Prevalence of Primary Aldosteronism in 1,125 Hypertensive Patients. *J. Am. Coll. Cardiol.* 2006;48(11):2293–2300.

5. **Funder JW, Carey RM, Mantero F, Murad MH, Reincke M, Shibata H, Stowasser M, Young WF.** The management of primary aldosteronism: Case detection, diagnosis, and treatment: An endocrine society clinical practice guideline. *J. Clin. Endocrinol. Metab.* 2016;101(5):1889–1916.

6. **Rossi GP, Bisogni V, Bacca AV, Belfiore A, Cesari M, Concistrè A, Del Pinto R, Fabris B, Fallo F, Fava C, Ferri C, Giacchetti G, Grassi G, Letizia C, Maccario M, Mallamaci F, Maiolino G, Manfellotto D, Minuz P, Monticone S, Morganti A, Muiesan ML, Mulatero P, Negro A, Parati G, Pengo MF, Petramala L, Pizzolo F, Rizzoni D, Rossitto G, Veglio F, Seccia TM.** The 2020 Italian Society of Arterial Hypertension (SIIA) practical guidelines for the management of primary aldosteronism. *Int. J. Cardiol. Hypertens.* 2020;5:e100029.

7. **Rossi GP, Seccia TM, Palumbo G, Belfiore A, Bernini G, Caridi G, Desideri G, Fabris B, Ferri C, Giacchetti G, Letizia C, MacCario M, Mallamaci F, Mannelli M, Patalano A, Rizzoni D, Rossi E, Pessina AC, Mantero F.** Within-patient reproducibility of the aldosterone:renin ratio in primary aldosteronism. *Hypertension* 2010;55(1):83–89.

8. **Nishikawa T, Omura M, Satoh F, Shibata H, Takahashi K, Tamura N, Tanabe A.** Guidelines for the diagnosis and treatment of primary aldosteronism -The Japan Endocrine Society 2009-. *Endocr. J.* 2011;58(9):711–721.

9. **Maiolino G, Rossitto G, Bisogni V, Cesari M, Seccia TM, Plebani M, Rossi GP, Semplicini A, Ganzaroli C, Pessina AC, Mantero F, Armanini D, Opocher G, Mattarello MY, Giacchetti G, Ronconi V, Boscaro M, Rossi E, Bernini G, Moretti A, Ferri C, Desideri G, Andronico G, Rizzoni D, Porteri E, Palumbo G, Letizia C, Caliumi C, Fabris B, Mannelli M, Parenti G, Maccario M, Ghigo E, Mallamaci F, Zoccali C, Belfiore A, PAPY Study Investigators the PS, Semplicini A, Ganzaroli C, Pessina AC, Mantero F, Armanini D, Opocher G, Mattarello MY, Giacchetti G, Ronconi V, Boscaro M, Rossi E, Bernini G, Moretti A, Ferri C, Desideri G, Andronico G, Rizzoni D, Porteri E, Palumbo G, Letizia C, Caliumi C, Fabris B, Mannelli M, Parenti G, Maccario M, Ghigo E, Mallamaci F, Zoccali C, Belfiore A, Ros1. Maiolino, G. et al. Quantitative value of aldosterone-renin ratio for detection of aldosterone-producing adenoma: The Aldosterone-Renin Ratio for Primary Aldosteronism (AQUARR) study. Journal of the American Heart Association 6, (2017).sitto G, Bisogni V, Cesari M, Seccia TM, Plebani M, Rossi GP, Semplicini A, Ganzaroli C, Pessina AC, Mantero F, Armanini D, Opocher G, Mattarello MY, Giacchetti G, Ronconi V, Boscaro M, Rossi E, Bernini G, Moretti A, Ferri C, Desideri G, Andronico G, Rizzoni D, Porteri E, Palumbo G, Letizia C, Caliumi C, Fabris B, Mannelli M, Parenti G, Maccario M, Ghigo E, Mallamaci F, Zoccali C, Belfiore A.** Quantitative value of aldosterone-renin ratio for detection of aldosterone-producing adenoma: The Aldosterone-Renin Ratio for Primary Aldosteronism (AQUARR) study. *J.*

Am. Heart Assoc. 2017;6(5):e005574.

10. **Phillips JL, Walther MM, Pezzullo JC, Rayford W, Choyke PL, Berman AA, Linehan WM, Doppman JL, Jr JRGJ.** Predictive value of preoperative tests in discriminating bilateral adrenal hyperplasia from an aldosterone-producing adrenal adenoma. *J. Clin. Endocrinol. Metab.* 2000;85(12):4526–4533.

11. **Irony I, Kater CE, Biglieri EG, Shackleton CH.** Correctable subsets of primary aldosteronism. Primary adrenal hyperplasia and renin responsive adenoma. *Am. J. Hypertens.* 1990;3:576–582.

12. **Gordon RD, Gomez-Sanchez CE, Hamlet SM, Tunny TJ, Klemm SA.** Angiotensin-responsive aldosterone-producing adenoma masquerades as idiopathic hyperaldosteronism (IHA: adrenal hyperplasia) or low-renin essential hypertension. *J. Hypertens Suppl.* 1987;5(5):S103–S106.

13. **Rossi GP, Seccia TM, Pessina AC.** Adrenal gland: A diagnostic algorithm - The holy grail of primary aldosteronism. *Nat. Rev. Endocrinol.* 2011;7(12):697–699.

14. **Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, Lijmer JG, Moher D, Rennie D, De Vet HCW, Kressel HY, Rifai N, Golub RM, Altman DG, Hooft L, Korevaar DA, Cohen JF.** STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. *Clin. Chem.* 2015;61(12):1446–1452.

15. **Zhu R, Shagjaa T, Rossitto G, Caroccia B, Seccia TM, Gregori D, Rossi GP.**

Exclusion tests in Unilateral Primary Aldosteronism (ExcluPA) Study. doi:10.25430/researchdata.cab.unipd.it.00000666.

16. **Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, John PA.** The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *Br. Med. J.* 2009;339:b2700.

17. **Higgins JPT, Thompson SG, Deeks JJ, Altman DG.** Measuring inconsistency in meta-analyses. *Br. Med. J.* 2003;327(7414):557–560.

18. **Devillé WL, Buntinx F, Bouter LM, Montori VM, De Vet HCW, Van Der Windt DAWM, Bezemer PD.** Conducting systematic reviews of diagnostic studies: Didactic guidelines. *BMC Med. Res. Methodol.* 2002;2:1–13.

19. **Rutter CM, Gatsonis CA.** A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat. Med.* 2001;20(19):2865–2884.

20. **Deeks JJ, Macaskill P, Irwig L.** The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J. Clin. Epidemiol.* 2005;58(9):882–893.

21. **Bernini G, Moretti A, Orlandini C, Berti P, Miccoli P, Bardini M, Taurino C, Bernini M, Salvetti A.** Plasma and urine aldosterone to plasma renin activity ratio in the diagnosis of primary aldosteronism. *J. Hypertens.* 2008;26(5):981–988.

22. **Burrello J, Monticone S, Buffolo F, Lucchiari M, Tetti M, Rabbia F, Mengozzi G, Williams TA, Veglio F, Mulatero P.** Diagnostic accuracy of

aldosterone and renin measurement by chemiluminescent immunoassay and radioimmunoassay in pr imar y aldosteronism. *J. Hypertens.* 2016;34(5):920–927.

23. **Ducher M, Mounier-Véhier C, Baguet JP, Tartière JM, Sosner P, Régnier-Le Coz S, Perez L, Fourcade J, Jabourek O, Lejeune S, Stolz A, Fauvel JP.** Aldosterone-to-renin ratio for diagnosing aldosterone-producing adenoma: A multicentre study. *Arch. Cardiovasc. Dis.* 2012;105(12):623–630.

24. **Giacchetti G, Ronconi V, Lucarelli G, Boscaro M, Mantero F.** Analysis of screening and confirmatory tests in the diagnosis of primary aldosteronism: Need for a standardized protocol. *J. Hypertens.* 2006;24(4):737–745.

25. **Vorselaars WMCM, Valk GD, Vriens MR, Westerink J, Spiering W.** Case detection in primary aldosteronism: High-diagnostic value of the aldosterone-to-renin ratio when performed under standardized conditions. *J. Hypertens.* 2018;36(7):1585–1591.

26. **Weickert MO, Schöfl-Siegert B, Arafat AM, Pfeiffer AFH, Möhlig M, Schöfl C.** A reverse postural test as a screening tool for aldosteroneproducing adenoma: A pilot study. *Endocrine* 2009;36(1):75–82.

27. **Kim JH, Park KS, Hong AR, Shin CS, Kim SY, Kim SW.** Diagnostic role of captopril challenge test in Korean subjects with high aldosterone-to-renin ratios. *Endocrinol. Metab.* 2016;31(2):277–283.

28. **Wu VC, Chang HW, Liu KL, Lin YH, Chueh SC, Lin WC, Ho YL, Huang JW, Chiang CK, Yang SY, Chen YM, Wang SM, Huang KH, Hsieh B Sen,**

Wu KD. Primary aldosteronism: Diagnostic accuracy of the losartan and captopril tests. *Am. J. Hypertens.* 2009;22(8):821–827.

29. **Wu VC, Kuo CC, Chang HW, Tsai CT, Lin CY, Lin LY, Lin YH, Wang SM, Huang KH, Fang CC, Ho YL, Liu KL, Chang CC, Chueh SC, Lin SL, Yen RF, Wu KD.** Diagnosis of primary aldosteronism: Comparison of post-captopril active renin concentration and plasma renin activity. *Clin. Chim. Acta* 2010;411(9–10):657–663.

30. **Fuss CT, Brohm K, Kurlbaum M, Hannemann A, Kendl S, Fassnacht M, Deutschbein T, Hahner S, Kroiss M.** Confirmatory testing of primary aldosteronism with saline infusion test and LC-MS/MS. *Eur. J. Endocrinol.* 2021;184(1):167–178.

31. **Mulatero P, Milan A, Fallo F, Regolisti G, Pizzolo F, Fardella C, Mosso L, Marafetti L, Veglio F, Maccario M.** Comparison of confirmatory tests for the diagnosis of primary aldosteronism. *J. Clin. Endocrinol. Metab.* 2006;91(7):2618–2623.

32. **Vivien M, Deberles E, Morello R, Haddouche A, Guenet D, Reznik Y.** Evaluation of Biochemical Conditions Allowing Bypass of Confirmatory Testing in the Workup of Primary Aldosteronism: A Retrospective Study in a French Hypertensive Population. *Horm. Metab. Res.* 2019;51(3):172–177.

33. **Zhang D, Chen T, Tian H, Li Y, Mo D, Zhang T, Wang W, Zhang G, Liu Y, Tang L, Zhu Y, Yang L, Ren Y.** Exploration Of The Seated Saline Suppression Test For The Diagnosis Of Primary Aldosteronism In The Chinese Population.

Endocr. Pract. 2020;26(8):891–899.

34. **Fries CM, Bae YJ, Rayes N, Sandner B, Isermann B, Stumvoll M, Fagotto V, Reincke M, Bidlingmaier M, Mandy V, Kratzsch J, Fenske WK.** Prospective evaluation of aldosterone LC-MS/ MS-specific cutoffs for the saline infusion test. *Eur. J. Endocrinol.* 2020;183(2):191–201.

35. **Meng X, Li Y, Wang X, Li J, Liu Y, Yu Y.** Evaluation of the saline infusion test and the captopril challenge test in Chinese patients with primary aldosteronism. *J. Clin. Endocrinol. Metab.* 2018;103(3):853–860.

36. **Okamoto R, Taniguchi M, Onishi Y, Kumagai N, Uraki J, Fujimoto N, Fujii E, Yano Y, Ogura T, Ito M.** Predictors of confirmatory test results for the diagnosis of primary hyperaldosteronism in hypertensive patients with an aldosterone-to-renin ratio greater than 20. The SHRIMP study. *Hypertens. Res.* 2019;42(1):40–51.

37. **Rossi GP, Belfiore A, Bernini G, Desideri G, Fabris B, Ferri C, Giacchetti G, Letizia C, Maccario M, Mallamaci F, Mannelli M, Palumbo G, Rizzoni D, Rossi E, Agabiti-Rosei E, Pessina AC, Mantero F.** Comparison of the captopril and the saline infusion test for excluding aldosterone-producing adenoma. *Hypertension* 2007;50(2):424–431.

38. **Song Y, Yang S, He W, Hu J, Cheng Q, Wang Y, Luo T, Ma L, Zhen Q, Zhang S, Mei M, Wang Z, Qing H, Bruemmer D, Peng B, Li Q.** Confirmatory Tests for the Diagnosis of Primary Aldosteronism:A Prospective Diagnostic Accuracy Study. *Hypertension* 2018;71(1):118–124.

39. **Stowasser M, Ahmed AH, Cowley D, Wolley M, Guo Z, McWhinney BC, Ungerer JP, Gordon RD.** Comparison of seated with recumbent saline suppression testing for the diagnosis of primary aldosteronism. *J. Clin. Endocrinol. Metab.* 2018;103(11):4113–4124.

40. **Lenzini L, Zanotti G, Bonchio M, Rossi GP.** Aldosterone Synthase Inhibitors for Cardiovascular Diseases: A Comprehensive Review of Preclinical, Clinical and In Silico Data. *Pharmacol. Res.* 2020;163(August 2020):105332.

41. **Tanabe A, Naruse M, Takagi S, Tsuchiya K, Imaki T, Takano K.** Variability in the renin/aldosterone profile under random and standardized sampling conditions in primary aldosteronism. *J. Clin. Endocrinol. Metab.* 2003;88(6):2489–2494.

42. **Eisenhofer G, Durán C, Cannistraci CV, Peitzsch M, Williams TA, Riester A, Burrello J, Buffolo F, Prejbisz A, Beuschlein F, Januszewicz A, Mulatero P, Lenders JWM, Reincke M.** Use of Steroid Profiling Combined With Machine Learning for Identification and Subtype Classification in Primary Aldosteronism. *JAMA Netw. open* 2020;3(9):e2016209.

43. **Rossi GP, Belfiore A, Bernini G, Desideri G, Fabris B, Ferri C, Giacchetti G, Letizia C, Maccario M, Mallamaci F, Mannelli M, Montemurro D, Palumbo G, Rizzoni D, Rossi E, Semplicini A, Agabiti-Rosei E, Pessina AC, Mantero F.** Prospective evaluation of the saline infusion test for excluding primary aldosteronism due to aldosterone-producing adenoma. *J. Hypertens.* 2007;25:1433–1442.

44. **Brown JM, Siddiqui M, Calhoun DA, Carey RM, Hopkins PN, Williams GH, Vaidya A.** The unrecognized prevalence of primary aldosteronism: A cross-sectional study. *Ann. Intern. Med.* 2020;173(1):10–20.

45. **Brown JM, Auchus RJ, Honzel B, Luther JM, Yozamp N, Vaidya A.** Recalibrating Interpretations of Aldosterone Assays Across the Physiologic Range: Immunoassay and Liquid Chromatography-Tandem Mass Spectrometry Measurements Under Multiple Controlled Conditions. *J. Endocr. Soc.* 2022;6(6):bvac049.

46. **Rossi GP, Bisogni V.** A useful tool to improve the case detection rate of primary aldosteronism: The aldosterone -renin ratio (ARR)-App. *J. Hypertens.* 2016;34(5):1019–1021.

47. **Leung AA, Symonds CJ, Hundemer GL, Ronksley PE, Lorenzetti DL, Pasieka JL, Harvey A, Kline GA.** Performance of Confirmatory Tests for Diagnosing Primary Aldosteronism: A Systematic Review and Meta-Analysis. *Hypertension* 2022;79(8):1835–1844.

48. **Caroccia B, Vanderriele PE, Seccia TM, Piazza M, Lenzini L, Prisco S, Torresan F, Domenig O, Iacobone M, Poglitsch M, Rossi GP.** Aldosterone and cortisol synthesis regulation by angiotensin-(1-7) and angiotensin-converting enzyme 2 in the human adrenal cortex. *J. Hypertens.* 2021;39(8):1577–1585.

49. **Williams TA, Lenders JWM, Mulatero P, Burrello J, Rottenkolber M, Adolf C, Satoh F, Amar L, Quinkler M, Deinum J, Beuschlein F, Kitamoto**

KK, Pham U, Morimoto R, Umakoshi H, Prejbisz A, Kocjan T, Naruse M, Stowasser M, Nishikawa T, Young WF, Gomez-Sanchez CE, Funder JW, Reincke M, Williams TA, Auchus RJ, Bartsch DK, Baudrand R, Björklund P, Brown MJ, Carey RM, Catena C, Connell JM, Dekkers T, Fahey TJ, Fallo F, Fardella CE, Giacchetti G, Giraudo G, Hellman P, Januszewicz A, Kitamoto KK, Kline GA, Mantero F, Miller BS, Plouin PF, Prejbisz A, Rump CL, Sechi LA, Veglio F, Widimský J, Willenberg HS. 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(9). doi:10.1016/S2213-8587(17)30135-3.

LEGENDS

Figure 1. Forest Plots of Specificity, NLR, and DOR for the Aldosterone-to-renin Ratio in the Exploratory and the Validation Phase.

Figure 2. Summary Area Under the Operating Characteristics Curve (sAUROC) for the Aldosterone-Renin Ratio (ARR, panel A), the Captopril Challenge Test (CCT, panel B) and the Saline Infusion Test (SIT, panel C). Please note the similar sAUROC for the 3 tests, indicating the lack of diagnostic gain with application of the CCT and SIT over the ARR

Figure 3. Forest Plots of Specificity, NLR, and DOR for the Captopril Challenge Test in the Exploratory and the Validation Phase.

Figure 4. Forest Plots of Specificity, NLR, and DOR for the Saline Infusion Test in the Exploratory and the Validation phase.

652 **Table 1.** Main characteristics of the studies that were meta-analyzed.

Author, year	Country	Population	Dates	Index test(s)	Assay	uPA	Controls	PA diagnosis	uPA diagnosis
Bernini 2008 (21)	Italy	New diagnosed PT	1998-2003	ARR	PAC by RIA (DiaSorin); PRA by RIA (DiaSorin)	30	100	Baseline PAC > 35 ng/dL and PRA < 0.5 ng/mL/h	Gold reference (Biochemical cure after surgery)
Burrello 2016 (22)	Germany	Suspected PA	2014	ARR	PAC by RIA amd CLIA (DiaSorin); PRA by RIA (DiaSorin); DRC by CLIA (DiaSorin)	5	75	Pos. SIT [PAC > 5 ng/dL (> 38.7 pmol/L)], or Pos. CCT [ARR > 30 ng/dL/ng/mL/h (832.2pmol/L/ng/mL/h) and ADRR > 3.7 ng/dL/mU/L (102.6 pmol/L/mU/L)]	Gold reference (Biochemical cure after surgery)
Ducher 2012 (23)	France	Suspected PA	2006-2007	ARR	N.A.	12	167	An outcome committee	Golden reference (Pathology after surgery)
Fries 2020 (34)	Germany	Pos. ARR (cutoff N.A.)	2016-2019	ARR, SIT	PAC by CL-MS/MS (Chromsystems); DRC by CLIA (DiaSorin)	9	67	Pos. ARR [PAC > 550 pmol/L (20 ng/dL), s-k^+ ↓, and PRA↓) or Pos. SIT [PAC > 140 pmol/L (5 ng/dL)]	Golden reference (AVS, LI ≥ 4)
Fuss 2021 (30)	Germany	Pos. ARR (cutoff > 20 ng/dL/ng/mL/h)	2009-2018	SIT	PAC by RIA (Siemens) or CLIA (IDS-iSYS) or LC-MS/MS (SCIEX); DRC by RIA (Cisbio) or CLIA (IDS-iSYS)	56	84	Pos. SIT	Gold reference (Biochemical cure after surgery)
Giacchetti 2006 (24)	Italy	Suspected PA	1996-2000	ARR	PAC by RIA (Biodata); UA by RIA (DiaSorin); PRA by RIA (Radim)	26	96	At least two of the following: (a) PAC↑, UA↑; (b) upright PRA↓ (≤ 1.0 ng/mL/h); (c) Pos. SIT (PAC ≥ 10 ng/dL);	Golden reference (Pathology after surgery)

								(d) an adrenal mass by imaging	
Kim 2016 (27)	Korea	Pos. ARR (cutoff > 20 ng/dL/ng/mL/h)	2011-2014	CCT	PAC by RIA (TFB Inc.); PRA by RIA (TFB Inc.)	36	13	Pos. SIT (PAC ≥ 10 ng/dL)	Gold reference (Biochemical cure after surgery)
Maiolino 2017 (9)[a]	Italy	New diagnosed PT	2000-2005	ARR, CCT	PAC by RIA (Mirya); PRA by RIA (DiaSorin, or Radim)	51	991	Pos. ARR 40 ≥ ng/dL/ng/mL/h or Pos. CCT (ARR ≥ 30 ng/dL/ng/mL/h), or a logistic discriminant function ≥ 0.5	Gold reference (Biochemical cure after surgery)
Maiolino 2017 (9)[b]	Italy	New diagnosed PT	2012-2015	ARR, CCT	PAC by RIA (Mirya); PRA by RIA (DiaSorin, or Radim)	30	1028	Pos. ARR 40 ≥ ng/dL/ng/mL/h or Pos. CCT (ARR ≥ 30 ng/dL/ng/mL/h) or a logistic discriminant function ≥ 0.5	Gold reference (Biochemical cure after surgery)
Meng 2018 (35)	China	Suspected PA	2011-2016	CCT, SIT	PAC by RIA (Jiuding Bio); PRA by RIA (Northern Bio)	70	49	Pos. ARR > 30 ng/dL/ng/mL/h	Gold reference (Biochemical cure after surgery)
Mulatero 2006 (31)	Italy	Pos. bARR (cutoff chosen by each center)	2004	SIT	PAC by RIA (DiaSorin, or DCS California); PRA by RIA (DiaSorin)	18	31	Pos. FST (PAC > 5 ng/dL)	Gold reference (Biochemical cure after surgery)
Okamoto 2019 (36)	Japan	Pos. bARR (cutoff > 20 ng/dL/ng/mL/h)	2012-2018	ARR, CCT, SIT, FUT	N.A.	16	86	At least two of the following: Pos. CCT (ARR > 20 ng/dL/ng/mL/h), or Pos. SIT (PAC> 6.0 ng/dL), or Pos. FUT (PRA < 2.0 ng/mL/h)	Golden reference (AVS, LI ≥ 4)

Rossi 2007 (37)	Italy	Pos. bARR (cutoff ≥ 40 ng/dL/ng/mL/h) or pos. CCT (ARR ≥ 30 ng/dL/ng/mL/h) or a logistic discriminant function ≥ 0.50 + 1/4 not fulling the above criteria	2000-2005	CCT, SIT	PAC by RIA (Mirya); PRA by RIA (DiaSorin)	46	197	Pos. ARR 40 ≥ ng/dL/ng/mL/h or Pos. CCT (ARR ≥ 30 ng/dL/ng/mL/h) or a logistic discriminant function ≥ 0.5	Gold reference (Biochemical cure after surgery)
Song 2018 (38)	China	Pos. bARR (cutoff >37 ng/mIU) + 1/3 neg. bARR	2013-2016	CCT, SIT	PAC by CLIA (DiaSorin); DRC by CLIA (DiaSorin)	71	101	Pos. FST (PAC > 8 ng/dL)	Gold reference (Biochemical cure after surgery)
Stowasser 2018 (39)	Australia	Pos. bARR [cutoff >70 pmol/mIU (19ng/mIU)]by RIA or >55 pmol/mIU (15 ng/mIU)by HPLC-MS/MS)	2012-2017	SIT, FST	PAC by RIA kit or HPLC-MS/MS; DRC by CLIA (DiaSorin)	28	18	Pos. FST [PAC ⩾ 133 pmol/L (4.8 ng/dL)]	Gold reference (Biochemical cure after surgery)
Vivien 2019 (32)	France	Pos. bARR [> 64 pmol/mIU(18 ng/mIU)]	2010-2015	SIT	PAC by RIA (Immunotech); DRC by IRMA (Cisbio)	24	76	Pos. SIT (PAC > 5 ng/dL) or Pos. CCT (PAC suppressed < 30%)	Golden reference (AVS /CT/MRI)
Vorselaars 2018 (25)	Netherland	Suspected PA	2015-2017	ARR	PAC by RIA (Siemens); PRA by RIA (In-house)	10	217	Pos. SIT [PAC >280 pmol/L (10.1 ng/dL) and PRA > 100 fmol/L/s (0.3 ng/mL/h)]	Gold reference (Biochemical cure after surgery)
Weickert 2009 (26)	Germany	Suspected PA	2005-2006	ARR	PAC by RIA (Immunotech); PRA by RIA (DiaSorin)	7	22	Pos. ARR (PAC↑, UA↑, PRA↓， s-k^+↓, and PRA↓)	Golden reference

									(Pathology after surgery)
Wu 2009 (28)	Taipei	Suspected PA	2003-2006	CCT	PAC by RIA (Adaltis); PRA by RIA (Cisbio)	47	64	Pos. SIT (PAC > 10 ng/dL) or UA ≥ 12 μg/24h	Gold reference (Biochemical cure after surgery)
Wu 2010 (29)	Taipei	Suspected PA	2008	CCT	PAC by RIA (Adaltis); PRA by RIA (Stillwater)	39	63	Pos. SIT (PAC > 10 ng/dL)	Golden reference (AVS, LI ≥ 4 or scintigraphy)
Zhang 2020 (33)	China	Pos. SIT (PAC >11.2 ng/dL)	2018-2019	SIT	PAC by RIA (Jiuding Bio); PRA by RIA (Northern Bio)	46	20	Pos. SIT (PAC > 11.2 ng/dL)	Gold reference (Biochemical cure after surgery)

ARR, aldosterone-to-renin ratio; AVS, adrenal vein sampling; BP, blood pressure; bARR, baseline aldosterone-to-renin ratio; CCT, captopril challenge test; CLIA, chemiluminescence immunoassay; CT, computed tomography; FUT, furosemide upright test; LC-MS/MS , liquid chromatography tandem mass spectrometry; LI, lateralization index; MRI, magnetic resonance tomography; NA, not available; Neg., negative; P, prospective; PA, primary aldosteronism; PAC, plasma aldosterone concentration; PH, primary hypertension; Pos., positive; PRA, plasma renin activity; R, retrospective; RIA, radioimmunoassay; SIT, saline infusion test; UA, urinary aldosterone; uPA, unilateral primary aldosteronism. a and b represent the exploratory and the validation cohort in Maiolino's study, respectively. Biochemical cure was defined following the PASO study.(49) The ARR cutoff value was reported whenever available and expressed in ng/dl of PAC over ng/ml/h of PRA. To convert into ng of PAC over mIU of active renin (DRC) the ARR-App can be used.(46)

Table 2. 2x2 Table reporting true (TP) and false positive (FP) rate, and true (TN) and false negative (FN) rate, using the gold or golden diagnosis of uPA as reference.

Author, year	Procedure, dosage, time interval	Cut-off value (original/converted)	Reference (golden/gold)	Cut-off by ROC curve	TP	FP	FN	TN
ARR								
Bernini 2008 (21)	Morning, after upright for at least 2 h and seated for 5-15 min	ARR > 96.4 ng/mIU	Gold	Yes	29	13	1	87
Burrello 2016 (22)	Morning, after upright for at least 2 h and seated for at least 15 min	ARR ≥ 37.0 ng/mIU and PAC ≥ 10.0 ng/dL	Gold	No	5	1	0	74
Ducher 2012 (23)	Morning, after supine for 1 h	ARR ≥ 32.0 ng/ng / 20.2 ng/mIU	Golden	Yes	11	13	1	154
Fries 2020 (34)	Morning, after seated for 15 min	ARR ≥ 53.0 pmol/mIU / 19.1 ng/mIU	Golden	No	9	27	0	40
Giacchetti 2006 (24)	Morning, after upright for 2 h and seated for 5-15 min	ARR ≥ 40.0 ng/dL/ng/mL/h / 48.8 ng/mIU	Golden	No	26	15	0	81
Maiolino 2017 (9)[a]	Morning, after seated for 1 h	ARR ≥ 33.3 ng/dL/ng/mL/h / 40.6 ng/mIU	Gold	Yes	40	117	11	874
Maiolino 2017 (9)[b]	Morning, after supine for 1 h	ARR ≥ 30.9 ng/dL/ng/mL/h / 37.3 ng/mIU	Gold	Yes	29	64	1	964
Okamoto 2019 (36)	Morning, after seated for 15 min	ARR ≥ 52.8 ng/dL/ng/mL/h / 64.4 ng/mIU	Golden	Yes	12	22	4	64
Vorselaars 2018 (25)	Morning, after upright for at least 2 h and seated for 5-15 min	ARR > 7 pmol/fmol / 65.6 ng/mIU	Gold	No	10	23	0	194
Weickert 2009 (26)	Morning, after upright for 30 min	ARR ≥ 425 pg/ml/ ng/mL/h / 51.8 ng/mIU	Golden	No	7	4	0	18
CCT								
Kim 2016 (27)	50 mg 1.5 h seated	PAC ≥ 19.0 ng/dL	Gold	No	27	0	9	13
Maiolino 2017 (9)[a]	50 mg 2 h seated	ARR ≥ 13.9 ng/dL/ng/mL/h / 17.0 ng/mIU	Gold	Yes	40	120	51	871
Maiolino 2017 (9)[b]	50 mg 2 h seated	ARR ≥ 12.8 ng/dL/ng/mL/h / 15.6 ng/mIU	Gold	Yes	28	227	2	801
Meng 2018 (35)	25 mg 2 h seated	PAC ≥ 15.0 ng/dL	Gold	No	68	9	2	40

Okamoto 2019 (36)	50 mg 1.5 h seated	ARR ≥ 42.2 ng/dL/ng/mL/h / ≥ 51.5 ng/mIU	Golden	Yes	12	16	4	70
Rossi 2007 (37)	50 mg 2 h seated	PAC ≥ 13.9 ng/dL	Gold	Yes	32	51	14	146
Song 2018 (38)	50 mg 2 h seated	PAC ≥ 13.0 ng/dL	Gold	Yes	68	5	3	96
Wu 2009 (28)	50 mg 1.5 h seated	ARR ≥ 23.9 ng/dL/ng/mL/h / 29.2 ng/mIU	Gold	Yes	39	8	8	56
Wu 2010 (29)	50 mg 1 h seated	ARR ≥ 66.9 ng/dL/ng/mL/h / 81.6 ng/mIU	Golden	Yes	28	6	11	57
SIT								
Fries 2020 (34)	2 L 4 h supine	PAC ≥ 83 pmol/L / 3.0 ng/dL	Golden	No	9	5	0	62
Fuss 2021 (30)	2 L 4 h supine	PAC ≥ 5.0 ng/dL	Gold	Yes	47	9	9	75
Meng 2017 (35)	2 L 4 h supine	PAC ≥ 10.0 ng/dL	Gold	No	69	10	1	39
Mulatero 2006 (31)	2 L 4 h supine	PAC ≥ 5.0 ng/dL	Gold	No	18	5	0	26
Okamoto 2019 (36)	2 L 4 h supine	PAC ≥ 15.2 ng/dL	Golden	Yes	14	9	2	77
Rossi 2007 (37)	2 L 4 h supine	PAC ≥ 6.8 ng/dL	Gold	Yes	38	49	8	148
Song 2018 (38)	2 L 4 h supine	PAC ≥ 10.0 ng/dL	Gold	Yes	68	4	3	97
Stowasser 2018 (39)	2 L 4 h seated	PAC ≥ 162 pmol/L / 5.8 ng/dL	Gold	Yes	26	1	2	17
Vivien 2019 (32)	2 L 4 h supine	PAC ≥ 5.7 ng/dL	Golden	No	22	4	2	72
Zhang 2020 (33)	2 L 4 h seated	PAC ≥ 12.9 ng/dL	Gold	No	41	2	5	18
FST								
Stowasser 2018 (39)	0.1 mg every 6 h	PAC > 162 pmol/L / 5.8 ng/dL	Gold	Yes	24	0	4	18
FUT								
Okamoto 2019 (36)	40 mg 2 h upright	PRA ≤ 0.55 ng/mL/h	Golden	Yes	13	23	3	63

ARR, aldosterone-to-renin ratio; CCT, captopril challenge test; FN, false negatives; FP, false positives; FST, fludrocortisone suppression test; FUT, furosemide upright test; PAC, plasma aldosterone concentration; PRA, plasma renin activity; ROC, receiver-operating characteristic curve; SIT, Saline infusion test; TN, true negatives; TP, true

665 positives. All units were converted to ng/mIU for ARR, ng/dL for PAC, ng/mL/h for PRA. To homogenize studies, all units of ARR, PAC, and PRA were converted with help
666 of ARR smartphone application (ARR-APP). a and b represent the exploratory and the validation cohort in Maiolino's study, respectively.

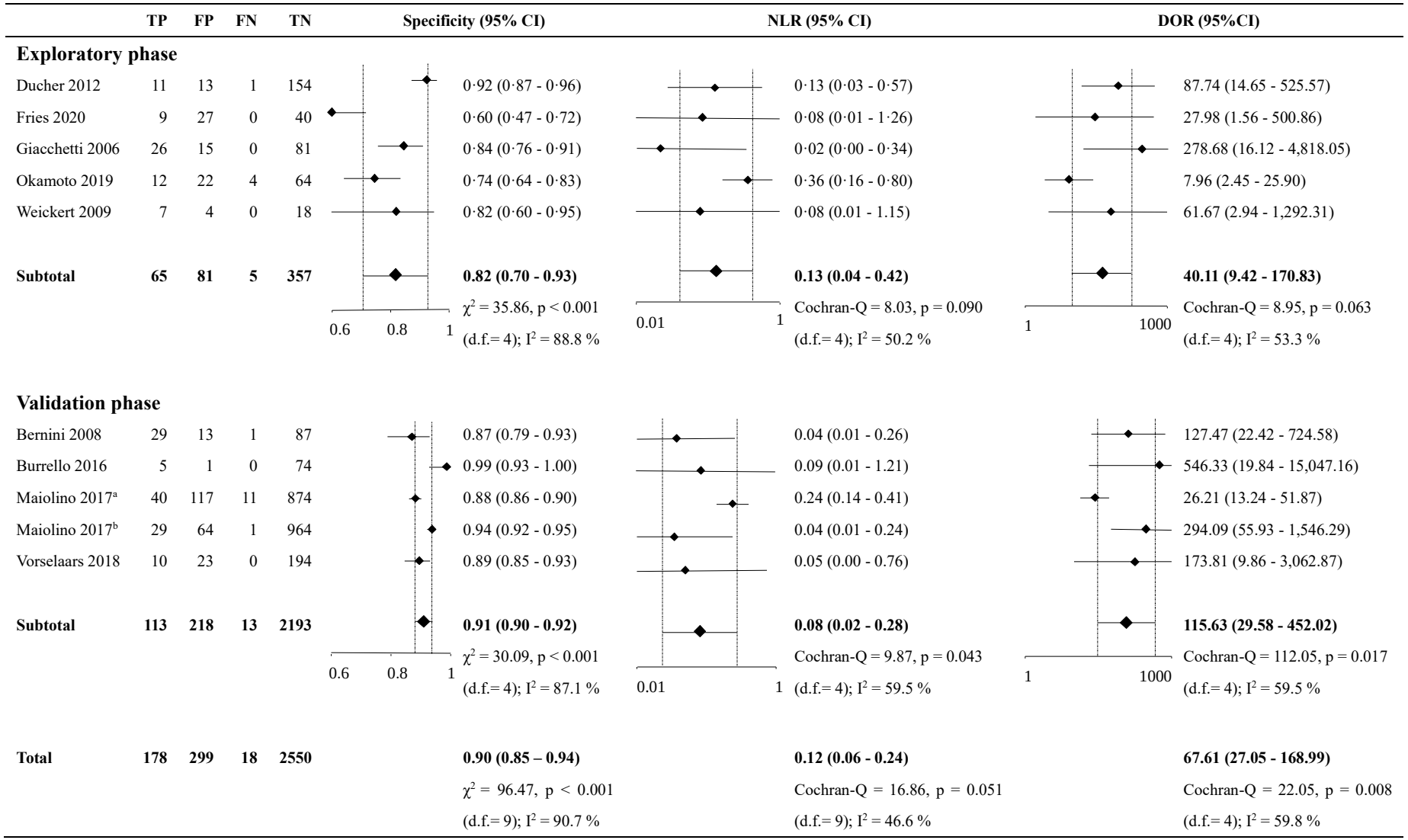
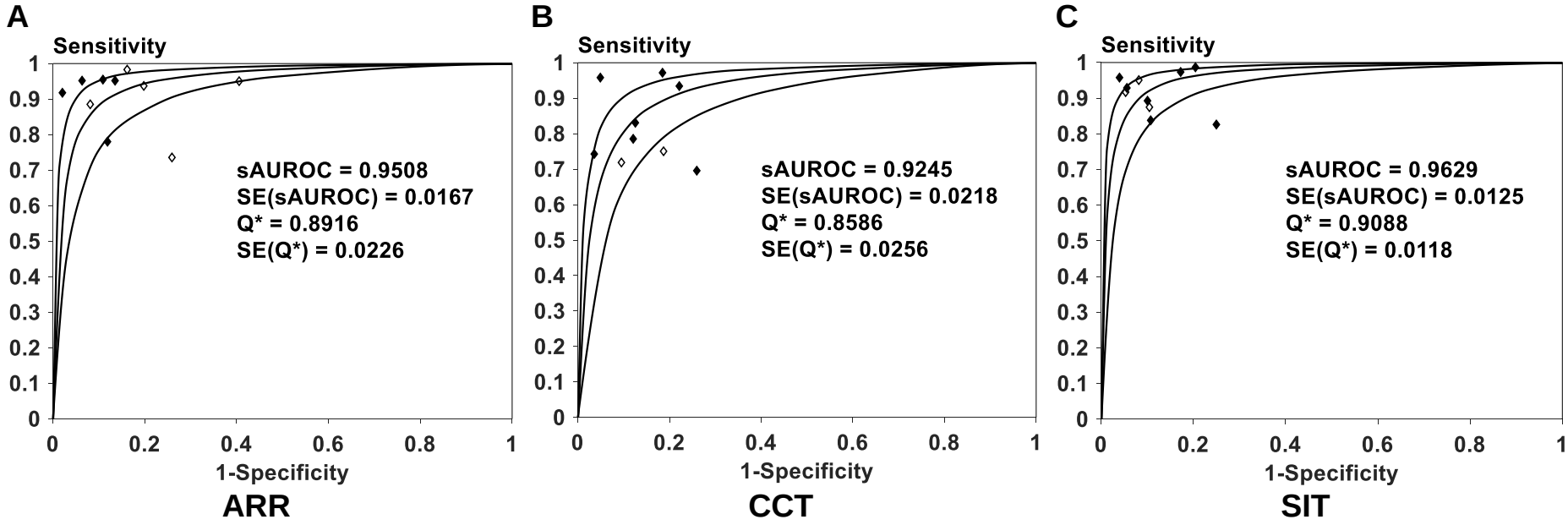


Figure 1. Forest Plots of Specificity, NLR, and DOR for the Aldosterone-to-renin Ratio in the Exploratory and the Validation Phase

DOR, diagnostic odds ratio; FN, false negatives; FP, false positives; NLR, negative likelihood ratio; TN, true negatives; TP, truth positives. Error bars on the plots represent the 95% confidence intervals. Square size is proportional to the weight of the study. *, $p < 0.05$; †, $p < 0.01$; ‡, < 0.001 the exploratory vs the validation phase. a and b represent the exploratory and the validation cohort in Maiolino's study, respectively.

Figure 2. Summary Area Under the Operating Characteristics Curve (sAUROC) for the Aldosterone-Renin Ratio (ARR, panel A), the Captopril Challenge Test (CCT, panel B) and the Saline Infusion Test (SIT, panel C).



ARR, aldosterone-to-renin ratios; CCT, captopril challenge test; SIT, saline infusion test. Solid diamond: study by the gold standard; empty diamond: study by the golden standard.

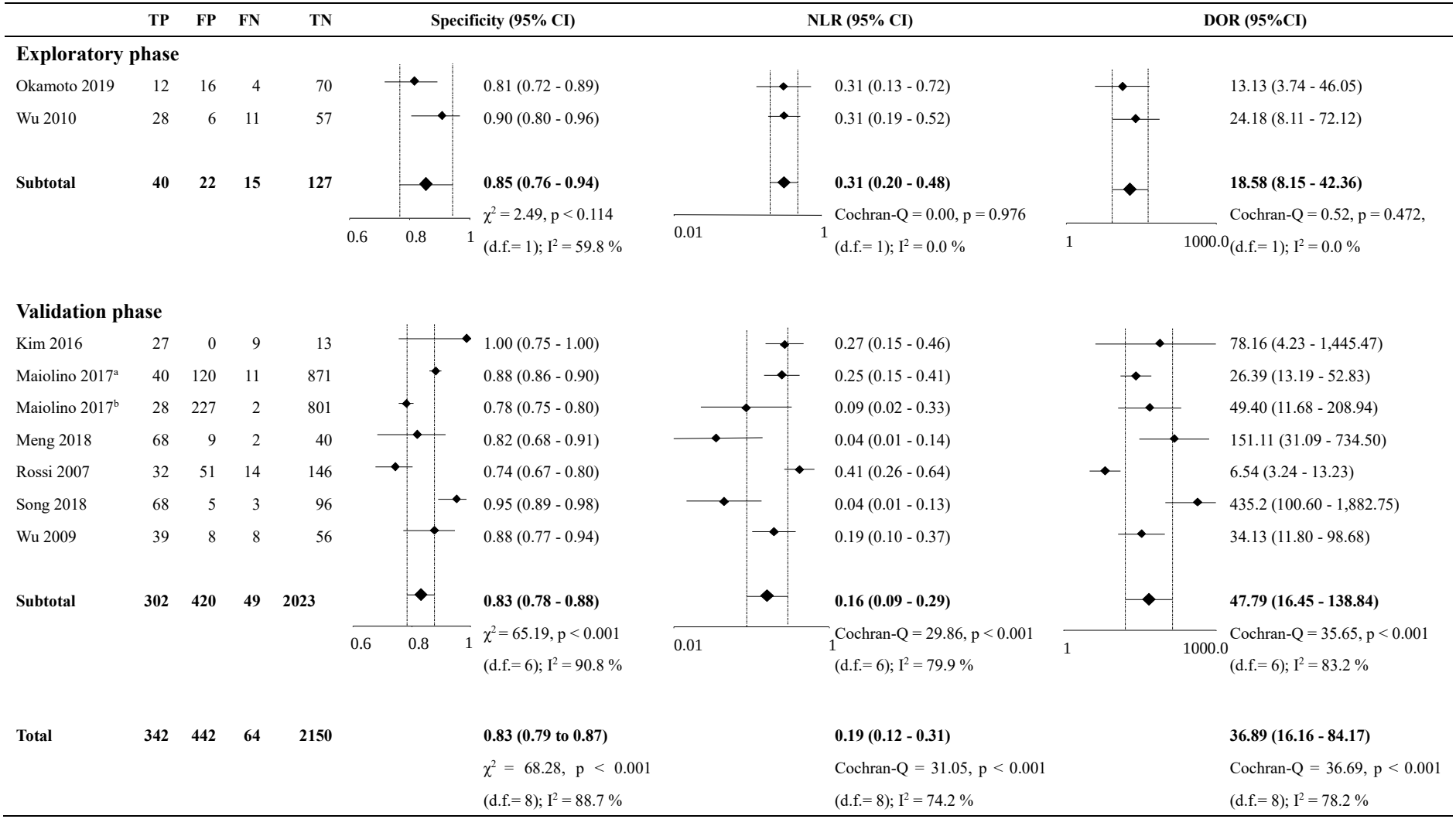


Figure 3. Forest Plots of Specificity, NLR, and DOR for the Captopril Challenge Test in the Exploratory and the Validation Phase

DOR, diagnostic odds ratio; FN, false negatives; FP, false positives; NLR, negative likelihood ratio; TN, true negatives; TP, truth positives. Error bars on the plots represent the 95% confidence intervals. Square size is proportional to the weight of the study. *, $p < 0.05$; †, $p < 0.01$; ‡, < 0.001 the exploratory vs the validation phase. a and b represent the exploratory and the validation cohort in Maiolino's study, respectively.

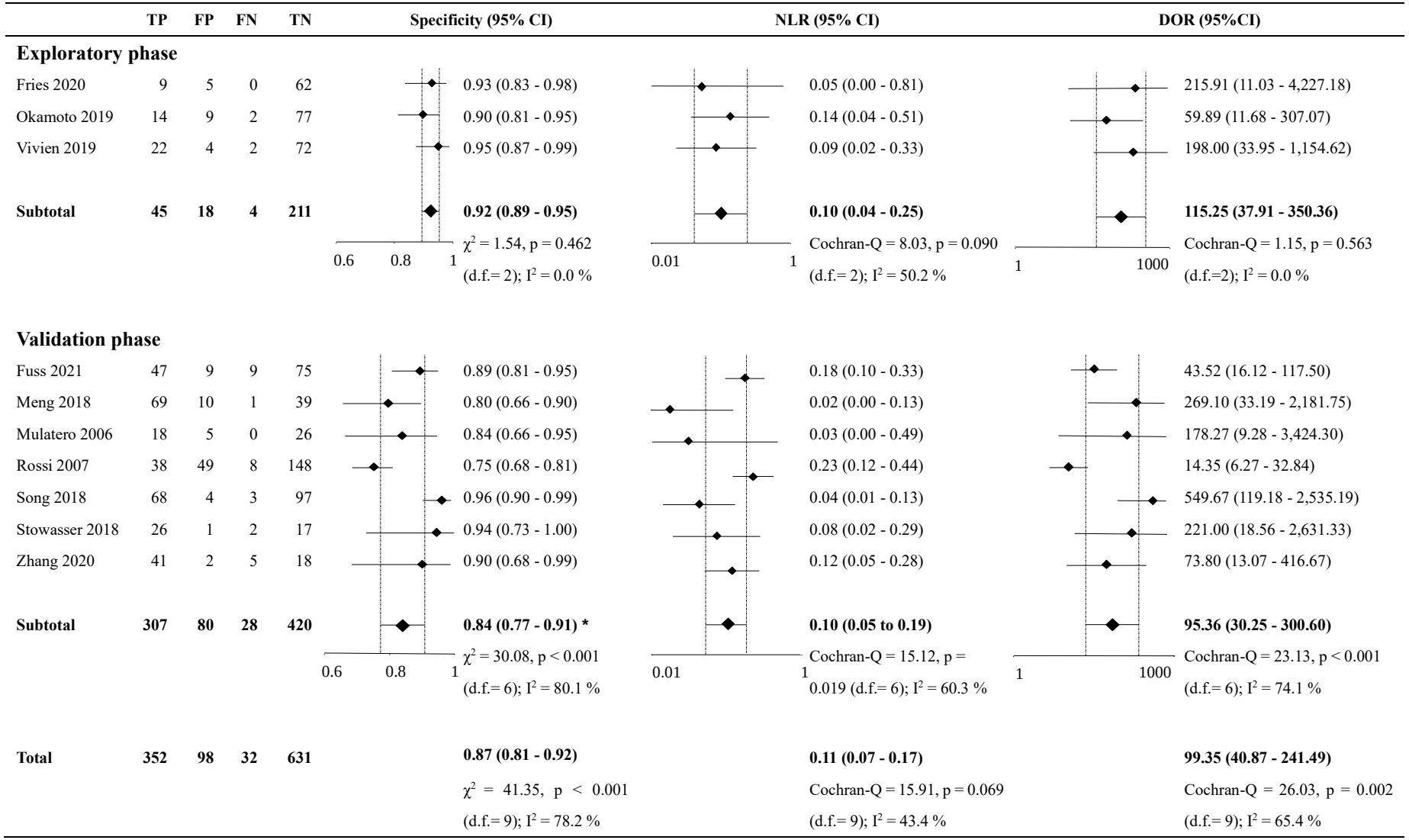


Figure 4. Forest Plots of Specificity, NLR, and DOR for the Saline Infusion Test in the Exploratory and the Validation Phase

DOR, diagnostic odds ratio; FN, false negatives; FP, false positives; NLR, negative likelihood ratio; TN, true negatives; TP, truth positives. Error bars on the plots represent the 95% confidence intervals. Square size is proportional to the weight of the study. *, $p < 0.05$; †, $p < 0.01$; ‡, < 0.001 the exploratory vs the validation phase.