BMJ Open

Diagnostic Accuracy of Adrenal Imaging for the Subtype Diagnosis of Primary Aldosteronism: Systematic Review and Meta-Analysis

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-038489
Article Type:	Original research
Date Submitted by the Author:	12-Mar-2020
Complete List of Authors:	Zhou, Yaqiong; Chengdu Medical College, Cardiology Wang, Dan Jiang, Licheng Ran, Fei Chen, Sichao Zhou, Peng Wang, Peijian
Keywords:	Hypertension < CARDIOLOGY, Endocrine tumours < DIABETES & ENDOCRINOLOGY, Cardiology < INTERNAL MEDICINE

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Diagnostic Accuracy of Adrenal Imaging for the Subtype Diagnosis of Primary Aldosteronism: Systematic Review and Meta-Analysis

Yaqiong Zhou[12] , Dan Wang[1], Licheng Jiang[12], Fei Ran[1], Sichao Chen[1], Peng Zhou[12*], Peijian Wang [12*]

From the 1 Department of Cardiology, The First Affiliated Hospital of Chengdu Medical College, Sichuang Province, People's Republic of China.

From the 2 Key Laboratory of Aging and Vascular Homeostasis of Sichuan Higher Education Institutes, Chengdu, Sichuan, People's Republic of China.

*Corresponding author: Prof. Peijian Wang

Department of Cardiology, the First Affiliated Hospital of Chengdu Medical College, 299 baoguang avenue, Chengdu 610500, Sichuan Province, China. E-mail: wpjmed@aliyun.com

*Corresponding author: Prof. Peng Zhou

Department of Cardiology, the First Affiliated Hospital of Chengdu Medical College, 299 baoguang avenue, Chengdu 610500, Sichuan Province, China. E-mail: ap216@163.com.

Email: Yaqiong Zhou.1273868593@qq.com- Dan Wang. 2292029981@qq.com- Licheng Jiang. 65839350@qq.com- Fei Ran. 1126852537@qq.com- Sichao Chen. 1350821872@qq.com-Peng Zhou. ap216@163.com-Peijian Wang. wpjmed@aliyun.com

Funding: This systematic review and meta-analysis is funded by the National Natural Sciences Foundation of China (NSFC)（Grant NO. 81970262）.

Abstract

Objectives: This study aimed to evaluate the diagnostic accuracy of adrenal imaging for the subtype diagnosis of primary aldosteronism (PA).

Methods: Systematic searches of PubMed, EMBASE, and the Cochrane databases were performed for all studies that used computed tomography (CT) or magnetic resonance imaging (MRI) in determining unilateral PA and validated the results against invasive adrenal vein sampling (AVS). Summary diagnostic accuracies were assessed by using a bivariate random effects model and a generalized linear mixed model was performed for heterogeneity analysis.

Result: Twenty-five studies were identified, involving a total of 4669 subjects in the meta-analysis. The overall analysis revealed a pooled sensitivity of 68% (95% confidence interval [CI]: 61 to 74), specificity of 58% (95% CI: 50 to 65), a positive LR of 1.6 (95% CI: 1.4 to 1.9), negative likelihood ratio (LR) of 0.56 (95% CI: 0.47 to 0.67), and diagnostic odds ratio (DOR) of 3 (95% CI: 2 to 4) for adrenal imaging to identify unilateral PA. Sensitivity and DOR were higher in contrast-enhanced CT group versus traditional CT group (79% vs. 58% and 6 vs. 2, respectively). Diagnostic accuracy of PA patients with an age of 40 years or younger was reported in four studies, the overall sensitivity of imaging for correctly identifying unilateral PA was 71%, with a 79% specificity. Heterogeneity analysis revealed a significant impact of age on specificity and sample size on sensitivity of adrenal imaging.

Conclusion: CT/MRI using currently available technology is not a reliable alternative to invasive AVS without excellent sensitivity and specificity for correctly

identifying unilateral PA. Even in young patients (<40 years), almost one-third patients would have undergone unnecessary adrenalectomy based on imaging results alone. In those centers without AVS facilities, contrast-enhanced CT can be considered as an alternative method.

Key words: adrenal vein sampling, computed tomography, magnetic resonance imaging, primary aldosteronism, subtype

Strengths and limitations of this study

- This systematic review and meta-analysis conclude that CT/MRI has a poor sensitivity and specificity in detecting unilateral PA when used AVS as the reference standard.
- Even in young patients (<40 years), nearly one-third patients would have undergone unnecessary adrenalectomy based on imaging results alone.
- We recommend routinely referring all patients for AVS, regardless of age and imaging results if the centers have access to AVS.
- If centers without AVS facilities at hand, contrast-enhanced CT can be considered as an alternative method.

Introduction

Primary aldosteronism (PA) is one of the most common cause of endocrine hypertension with a prevalence of around 20% among patients with resistant hypertension, 10% in those with severe hypertension and 6% in those with uncomplicated hypertension[1]. Accumulating clinical and epidemiological evidence suggests that PA amplifies cardio and cerebrovascular complications beyond essential

hypertension, even after controlling for the elevated blood pressure[2,3]. Accordingly, an early diagnosis and specific treatment of affected patients are a key step for reversal of target-organ damage and prevention of cardio and cerebrovascular events.

Selection of the most appropriate therapeutic strategy for patients with PA requires the distinction between unilateral and bilateral form of PA. The former requires a unilateral adrenalectomy, mainly entail aldosterone producing adenoma (APA) and less commonly, unilateral adrenal hyperplasia, whereas the latter, also known as idiopathic hyperaldosteronism, is optimally treated with target medical therapy[3]. Regarding the differentiation of unilateral from bilateral subtype, all current clinical practice guidelines recommend adrenal vein sampling (AVS) as the standard procedure for subtype diagnosis[4,5]. However, several shortages of AVS have been reported, such as technical challenges, invasive nature, poorly standardized procedure, high cost and lack of availability. Thus, it is urgent to explore alternative diagnostic methods without sacrificing accuracy.

Adrenal imaging with computed tomography (CT) or magnetic resonance imaging (MRI) is recommended as the first step for subtype classification owing to its ease of performance and relative accessibility[6]. By now, numerous studies have evaluated the diagnostic performance of CT/MRI in subtype diagnosis of PA, but the results have been inconsistent. Moreover, all of these studies were limited by small sample sizes in single center, which limited the credibility of results. In this context, systematic review and meta-analysis have the benefit to increase the sample size generating more precise results, which has been widely applied in clinical studies[7,8].

In 2009, one systematic review reported that CT/MR-based diagnoses were discordant with AVS results in 37.8% of PA patients[9]. However, the conclusions may not be reliable because of the potential for bias and concerns regarding the comparability of the included studies. Moreover, several additional studies were reported after this systematic review. We thus performed a comprehensive meta-analysis of all available studies to evaluate the diagnostic value of adrenal imaging (CT/MRI) for subtype classification of PA.

Method

Search Strategy

The study followed the guidelines specified in the preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P)[10]. We searched the PUBMED, EMBASE, and Cochrane Library databases from inception to February 2020 using the following terms in combination, both as MeSH or Emtree terms and text words: "primary aldosteronism", "adrenal vein sampling", "hyperaldosteronism", "computed tomography", "magnetic resonance imaging", "diagnostic", " subtype". To reflect modern practice, we decided to limit publication date of studies after 2000. We searched articles published in English language and the references of relevant studies were also searched. All studies were carefully examined to exclude overlapping or potential duplicates data. Trials in abstract form without a published manuscript were excluded.

Eligibility Criteria

We included a study if: 1) it used CT or MRI as a diagnostic test for PA

subtyping; 2) it used AVS with or without adrenocorticotropic hormone (ACTH) stimulation as the standard of reference and the cutoff value for the lateralization index (LI) should be 2.0 or greater as the criterion for unilateral disease; and 3) absolute numbers of true positive, true negative, false positive, and false negative results were provided or could be derived. Identified studies had to be independent. In case of multiple reports on the same population or subpopulation, the most recent or comprehensive information was used.

Data Abstraction and Quality Assessment

Data extraction of the eligible studies was performed by 2 independent investigators (Z.Y.Q and W.P.J) using a standardized data extraction form. The form included the following characteristics of each trial: first author's name and year of publication; study population characteristics, including sample size (absolute numbers of true-positive, true-negative, false-positive, and false-negative), geographical location, mean age, sex and AVS method (with or without ACTH stimulation); diagnostic text characteristics, including imaging methodology, contrast administered or not. Differences between reviewers were resolved by discussion and consensus whenever necessary.

The methodological quality of identified studies was assessed by 2 independent reviewers (Z.Y.Q and W.P.J) using the modified Quality Assessment of Diagnostic Accuracy Studies –2 criteria (QUADAS-2)[11], and discrepancies were resolved by discussion and consensus.

Statistical Analysis

Measures of diagnostic accuracy are reported as point estimates with 95% confidence intervals (CIs). Sensitivity, specificity, positive likelihood ratio (+LR) and negative likelihood ratio (-LR) were calculated based on the true positive, true negative, false positive, and false negative rates for each trial. There is consensus that a +LR >10 and a -LR <0.1 provide reliable evidence of satisfactory diagnostic performance [12]. The ratio of +LR to -LR was combined in a single global accuracy measure, the diagnostic odds ratio (DOR). Summary sensitivity, specificity, +LR, -LR and DOR were assessed by using a bivariate random effects model. A hierarchical summary receiver operating characteristic curves (ROC) was performed[13], presenting the point estimates for each trial and the pooled characteristics, including the 95% prediction region and the 95% confidence region.

Sources of statistical heterogeneity were explored by using the bivariate generalized linear mixed model[14], which was assessed using the I^2 statistic, applying the following interpretation for I^2: <50%=low heterogeneity; 50% to75%=moderate heterogeneity; >75%=high heterogeneity. We assessed the following covariates, which were selected a priori: sample size (divided by 100 patients), average age of patients (divided by a median age of 53years), AVS method (with or without ACTH stimulation), year of publication (divided by 2010) and imaging methodology (contrast administered or not). The potential publication bias was examined by using the Deeks test [15]. Cohen κ test was assessed for the inter-rater reliability between 2 observers for quality assessment. Statistical analyses were performed using Stata version13.0 (StataCorp LP, TX, USA) and Review Manager version 5.3. Statistical

tests were 2 sided and used a significance level of P<0.05.

Result

Study Selection

After removal of the 548 duplicates, the systematic review retrieved 1022 references that were screened according to title or abstract for possible inclusion (Figure 1). Sixty studies were identified as potentially eligibility and full text was retrieved for detailed evaluation. Thirty-five studies were excluded for the following reasons: data to compute diagnostic accuracies were not provided or could not be derived (18 papers), reporting on the same population (4 papers), no systematic AVS was performed (6 papers) and no values for true positive and false negative observations were reported (7 papers). Finally, twenty-five articles were deemed eligible and analyzed in our meta-analysis[16-40]. (Figure 1).

Study Characteristics

Overall, a total of 4669 patients (mean age 51 years; 54% male) from twenty-five articles were included. The sizes of the identified studies ranged from 35 to 1591, with the largest study recruiting over 1000 participants[20]. Five studies including 724 participants underwent cross-sectional imaging either CT or MRI, and the remaining 20 studies including 3945 patients only used CT scan (8 studies administered contrast material). Fifteen studies used reference standard AVS with ACTH stimulation, 7 studies without ACTH stimulation and the remaining 3 studies provide the above two ways of performing AVS. Further details of the eligible and analyzed studies are shown in Table 1.

Table1: Baseline characteristics of included studies.

ACTH: adrenocorticotropic hormone; CT: computed tomography; MRI: magnetic resonance imaging; NA: not available.

First Author	Publication Year	Location	Male (%)	Age (yrs)	Sample	Imaging methodology	Contrast used	AVS method
Sicheng[16]	2019	China	61(50)	48.5	122	CT	yes	without ACTH
Campbell[17]	2019	USA	45(61)	55.6	74	CT or MRI	no	with ACTH
Daisuke[19]	2019	Japan	NA	56	317	CT	no	with ACTH
Davis[18]	2019	Canada	201(59)	52.1	342	CT or MRI	yes	with ACTH
Umakoshi[20]	2018	Japan	762(47.9)	53	1591	CT	yes	with ACTH
Nanba[21]	2017	USA	87(59)	54	147	CT	no	with or without ACTH
Kamemura[22]	2017	Japan	177(45)	54	393	CT	no	with ACTH
Limin[24]	2016	China	NA	46	394	CT	no	without ACTH
Pedersen[23]	2016	Denmark	24(54)	51	45	CT or MRI	no	without ACTH
Kocjan[25]	2016	Slovenia	46(69)	56	67	CT	no	with ACTH
Asmar[26]	2015	USA	148(63)	55	235	CT or MRI	no	with ACTH
Riester[28]	2014	Germany	NA	35	28	CT or MRI	no	without ACTH
Sze[27]	2014	UK	42(56)	50.5	75	CT	yes	with ACTH
Kűpers[32]	2012	France	53(61)	46	87	CT	no	with ACTH
Lau[31]	2012	UK	24(64)	51.8	39	CT	no	with ACTH
Burton[29]	2012	UK	NA	50.9	40	CT	Yes	without ACTH
Salem[30]	2012	UK	16(44)	44.7	38	CT	no	without ACTH
Eun[40]	2012	Korea	45(52)	50.7	86	CT	yes	with ACTH
Gabrielle[33]	2011	France	NA	52	58	CT	yes	without ACTH
Mathur[34]	2010	USA	63(55)	50.6	114	CT	no	with or without ACTH
Mulatero[35]	2008	Italy	NA	52.4	70	CT	yes	with or without ACTH
Minami[36]	2008	Japan	12(34)	54	35	CT	no	with ACTH
Nwariaku[37]	2006	USA	27(67)	51	40	CT	yes	with ACTH
William	2004	USA	163(84)	53	194	CT	no	with ACTH
Magill[39]	2001	USA	27(71)	51	38	CT	no	with ACTH

Quality assessment

Overall, the identified studies showed excellent quality in terms of applicability and risk of bias (Figure 2), with an inter-observer agreement of κ=0.93. Risk of bias was high in one study regarding blinding the results of the reference standard from the investigators of the index test[24]. Risk of bias regarding flow and timing was unclear in eight studies [20-23,25,31,35,37] because the time interval between the index test and the

reference standard was unclear and was high in one study [38]because not all patients receive the same reference standard.

Overall analysis

Using the bivariate model, statistical heterogeneity was found for sensitivity (I^2 =86.9%; *P*= 0.001), specificity (I^2 =86.4%; *P*= 0.001), positive LR (I^2 = 75.7%; *P* = 0.001), negative LR (I^2 =78.9%; *P* = 0.001), and diagnostic odds ratio (I^2 = 100.0%; *P* =0.001), indicating high between-study heterogeneity for all pooled measures.

In the overall analysis, the pooled sensitivity, specificity, positive LR, negative LR, and diagnostic odds ratio for adrenal imaging were 68% (95% confidence interval [CI]: 61 to 74), 58% (95% CI: 50 to 65), 1.6 (95% CI: 1.4 to 1.9) , 0.56 (95% CI: 0.47 to 0.67) , and 3 (95% CI: 2 to 4) , respectively. (Figure 3, 4).

Subgroup analyses

Subgroup analysis, stratified by the imaging methodology, found more favorable specificity (61%) and negative LR (0.54) for CT scan compared with the specificity (45%) and negative LR (0.68) for CT/MRI. Notably, subgroup analysis showed an increase in sensitivity (79%) and diagnostic odds ratio (6) when contrast material is administered during CT scan.

There were 4 studies reported diagnostic accuracy of PA patients with an age of 40 years or younger. Using the bivariate model, the pooled sensitivity, specificity, positive LR, negative LR, and diagnostic odds ratio were 71% (95% CI:54 to 84), 79% (95% CI: 37 to 96) (Figure S1), 3.4 (95% CI: 0.2 to 14.7), 0.37 (95% CI: 0.2 to 0.68), and 9 (95% CI: 1 to 64), respectively. Summary estimates for pooled measures

of diagnostic accuracy are shown in Table 2.

Table 2: Pooled summary results by subgroups.

CI: confidence interval; DOR: diagnostic odds ratio; LR: likelihood ratio; other abbreviations as in Table 1.

	No.of Studies	Sensitivity (95% CI)	Specificity (95% CI)	Positive LR (95% CI)	Negative LR (95% CI)	DOR (95% CI)
Total	25	68(61-74)	58(50-65)	1.6(1.4-1.9)	0.56(0.47-0.67)	3(2-4)
Stratified by age						
< 40years	**4**	**71(54-84)**	**79(39-96)**	**3.4(0.2-14.7)**	**0.37(0.2-0.68)**	**9(1-64)**
Stratified by imaging methodology						
CT	20	67(59-75)	61(54-67)	1.7(1.4-2.0)	0.54(0.43-0.67)	3(2-5)
contrast CT	**8**	**79(69-86)**	**60(47-73)**	**2.0(1.5-2.7)**	**0.35(0.25-0.49)**	**6(3-9)**
nocontrast CT	12	58(50-66)	62(54-70)	1.5(1.3-1.9)	0.68(0.56-0.81)	2(2-3)
CT /MRI	5	69(62-76)	45(27-64)	1.3(1.0-1.7)	0.68(0.50-0.91)	2(1-3)
Stratified by publication date						
≤2010	6	60(47-71)	61(53-68)	1.5(1.1-2.1)	0.66(0.46-0.96)	2(1-5)
>2010	19	70(62-76)	56(47-65)	1.6(1.3-1.9)	0.54(0.44-0.66)	3(2-4)

Heterogeneity analyses

Adding pre-premised covariates to the bivariate generalized linear mixed model showed a significant interaction between the smaller sample size and higher sensitivity (I^2=77%; *P*=0.01) of CT/MRI for the detection of unilateral forms of PA. Age was the only covariate with a negative effect on specificity (I^2=67%; *P*=0.05) (Figure S2).

Publication bias

Using the Deeks test, there was no evidence of publication bias (*P*=0.60 for the all patient evaluation).

Discussion

The accurate differentiation of unilateral from bilateral PA is critical for optimal clinical management. Although AVS is the "gold standard" test for subtype

diagnosis[6], numerous studies have investigated the underlying diagnostic value of CT/MRI for subtype diagnosis due to several insurmountable shortages of AVS. The present meta-analysis, involving 4669 individuals from 25 studies demonstrated that CT/MRI has a poor sensitivity (68 %) and specificity (58 %) in the subtype classification when used AVS as the reference standard and the cutoff value for LI equal to or greater than 2.0 as the criterion for unilateral PA.

In the subtype diagnosis of PA, AVS was initially used in the 1960s. Subsequently, CT was adopted as the primary method for distinguishing unilateral from bilateral form of PA. Owing to less invasive nature, less cost and widely availability, many physicians prefer to perform CT/MRI as the first and sometimes the only investigation of subtype diagnosis of PA. But its sensitivity and specificity vary widely. The reported sensitivities ranged from 29%[3] to 94%[40] and the reported specificities ranged from 18%[17] to 87%[32]. Although the sensitivities were reported to exceed 80% in 5 studies[17,25,33,35,40] , relatively poor specificities were reported with only one study showing a specificity of 72%[19]. Similarly, 3 studies reported the sensitivities to be over 80%[28,29,32], but the specificities were reported to be lower than 76%[29]. The present meta-analysis showed that the pooled sensitivity was 68% and specificity was 58%, which means that decision of treatment based on the presence of unilateral disease on CT/MRI alone could result in inappropriate unilateral adrenalectomy in 42% patients. Whereas based on CT/MRI alone would miss the possibility of a potentially curative procedure by surgery in 32% patients. However, failure of early diagnosis and specific treatment of PA place these patients at higher

risk of irreversible renal and cardiovascular damage. Our results suggest that CT/MRI do not have satisfactory diagnostic performance in classifying subtypes of PA.

Contrast materials can improve the visibility of internal body structures imaged by CT and MRI scans. They make certain structures or tissues appear in a different way and thus help in distinguishing a contrast-enhanced area of the body from the surrounding tissue. In the present meta-analysis, eight studies including 530 participants underwent contrast-enhanced CT[16,27,29,31,33,35,37,40]. Notably, subgroup analysis showed a favorable increase in sensitivity (79%) in contrast-enhanced CT group, which indicates that contrast-enhanced examination, may be more reliable if centers without AVS facilities at hand currently have no other way than to rely on CT scan.

Given the nonfunctioning adrenocortical adenomas ("incidentaloma") are relatively uncommon in young people(<40 years), the 2009 guidelines for managing PA contended that younger patients with unequivocal biochemical diagnosis of PA and a clear-cut unilateral cortical adenoma on adrenal CT scan proceed directly to surgery and AVS procedure may be skipped[41]. Among studies included in the present meta-analysis, there were four studies reported diagnostic accuracy of CT/MRI in identifying unilaterally PA in patients<40 years old. By combining these four studies, our results demonstrated that although the sensitivity (71%) and specificity (79%) were improved in patients aged <40 years, but the diagnostic performance was still unsatisfactory since 29% patients would have undergone unnecessary adrenalectomy based on imaging results alone. In 2016, the updated clinical practice guidelines were

published and suggested the age cutoff for sparing AVS was 35 years[6]. Regarding patients aged <35 years, several retrospective studies have evaluated the diagnostic value of CT. The reported rate of concordance between CT and AVS ranges from 59% to 90% [17,20,24]. Based on these data, it seems that CT still cannot replace AVS in patients aged <35 years. However, due to the lack of numbers of false positive and true negative, we did not perform pooled analyses. Further studies are needed to clarify the diagnostic value of CT in patients aged <35 years.

As mentioned above, although adrenal imaging is not a reliable method to differentiate subtype of PA, this does not mean that CT/MRI must be wrong and should not be used as a basis for clinical management, including adrenalectomy. There is no doubt that the improvement of outcomes for the patient is much more relevant as an end point to assess the clinical value of a diagnostic test. However, in the SPARTACUS trial, a prospective, randomized, controlled study, treatment effect of primary aldosteronism was compared based on CT or AVS showed no significant difference in outcomes such as blood pressure, antihypertensive medication and quality of life for patients after 1 year of follow-up[42]. The results demonstrated that CT-based decision-making was a valid strategy and not inferior to the decision of AVS-based.

The present meta-analysis has several limitations. First, the minority of study participants underwent cross-sectional imaging either CT or MRI, but absolute numbers were not provided or could not be derived based on the imaging methodology used, which limited our ability to identify which imaging methodology

(CT or MRI) can provide more accurate diagnostic performance. Second, the AVS method of the majority of study used (sequential sampling during ACTH infusion) may limit the generalizability of our findings to other ways of performing AVS. Third, the possibility of selection bias that is present in all meta-analysis cannot be overlooked.

Conclusion

Based upon these analyses, we conclude that CT/MRI has a poor sensitivity (68 %) and specificity (58 %) in the detection of unilateral PA when used AVS as the reference standard. Even in young patients (<40 years), 29% would have undergone unnecessary adrenalectomy based on imaging results alone. We therefore recommend routinely referring all patients for AVS, regardless of age and imaging results if the centers have access to AVS. In those centers without AVS facilities at hand , Contrast-enhanced CT can be considered as an alternative method.

References:

1. S M, J B, D T,et.al. Prevalence and Clinical Manifestations of Primary Aldosteronism Encountered in Primary Care Practice. *Journal of the American College of Cardiology* 2017;69:1811-1820.

2. C C, G C, E N,et.al. Cardiovascular outcomes in patients with primary aldosteronism after treatment. *Archives of internal medicine* 2008;168:80-85.

3. PR P, M G, MD S,et.al. Primary Aldosteronism: Cardiovascular Risk, Diagnosis, and Management. *Cardiology in review* 2020;28:84-91.

4. TA W, M R. MANAGEMENT OF ENDOCRINE DISEASE: Diagnosis and management of primary aldosteronism: the Endocrine Society guideline 2016 revisited. *European journal of endocrinology* 2018;179:R19-R29.

5. JB B, AF T, RJ A. Primary Aldosteronism. 2018:823-835.

6. JW F, RM C, F M,et.al. The Management of Primary Aldosteronism: Case Detection, Diagnosis, and Treatment: An Endocrine Society Clinical Practice Guideline. 2016:1889-1916.

7. C G, PS T, J V,et.al. Impact of point-of-care panel tests in ambulatory care: a systematic review and meta-analysis

Social media and mobile health technology for cancer screening: a systematic review and meta-analysis protocol. *BMJ open* 2020;10:e032132.

8. Ruco A, Dossa F, Tinmouth J,et.al. Social media and mobile health technology for cancer screening: a systematic review and meta-analysis protocol. *BMJ open* 2020:e035411.

9. Kempers MJE, Lenders JWM, van Outheusden L,et.al. Systematic review: diagnostic procedures to differentiate unilateral from bilateral adrenal abnormality in primary aldosteronism. *Annals of internal medicine* 2009;151:329-337.

10. A L, DG A, J T,et.al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Annals of internal medicine* 2009;151:W65-94.

11. PF W, AW R, ME W,et.al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of internal medicine* 2011;155:529-536.

12. WL D, F B, LM B,et.al. Conducting systematic reviews of diagnostic studies: didactic guidelines. *BMC medical research methodology* 2002;2:9.

13. H C, H G, Y Z. Bivariate random effects meta-analysis of diagnostic studies using generalized linear mixed models. *Medical decision making : an international journal of the Society for Medical Decision Making* 2010;30:499-508.

14. TH H, LR A, HC vH,et.al. Multivariate random effects meta-analysis of diagnostic tests with multiple thresholds. *BMC medical research methodology* 2009;9:73.

15. JJ D, P M, L I. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *Journal of clinical epidemiology* 2005;58:882-893.

16. Li S, Ren Y, Zhu Y,et.al. THE POTENTIAL CLINICAL APPLICATION OF A LOWER BILATERAL ADRENAL LIMB WIDTH RATIO (L/RW) IN PATIENTS WITH BILATERAL PRIMARY HYPERALDOSTERONISM. *Endocrine practice : official journal of the American College of Endocrinology and the American Association of Clinical Endocrinologists* 2019;25:830-835.

17. Campbell RA, Young DS, Shaver CN,et.al. Influence of Adrenal Venous Sampling on Management in Patients with Primary Aldosteronism Independent of Lateralization on Cross-Sectional Imaging. *Journal of the American College of Surgeons* 2019;229:116-124.

18. Sam D, Kline GA, So B,et.al. Discordance Between Imaging and Adrenal Vein Sampling in Primary Aldosteronism Irrespective of Interpretation Criteria. *The Journal of clinical endocrinology and metabolism* 2019;104:1900-1906.

19. Aono D, Kometani M, Karashima S,et.al. Primary aldosteronism subtype discordance between computed tomography and adrenal venous sampling. *Hypertension research : official journal of the Japanese Society of Hypertension* 2019;42:1942-1950.

20. Umakoshi H, Tsuiki M, Takeda Y,et.al. Significance of Computed Tomography and Serum Potassium in Predicting Subtype Diagnosis of Primary Aldosteronism. *The Journal of clinical endocrinology and metabolism* 2018;103:900-908.

21. Nanba AT, Nanba K, Byrd JB,et.al. Discordance between imaging and immunohistochemistry in unilateral primary aldosteronism. *Clinical endocrinology* 2017;87:665-672.

22. Kamemura K, Wada N, Ichijo T,et.al. Significance of adrenal computed tomography in predicting laterality and indicating adrenal vein sampling in primary aldosteronism. *Journal of human hypertension* 2017;31:195-199.

23. Pedersen M, Karlsen MA, Ankjærgaard KL,et.al. Primary hyperaldosteronism diagnosed with adrenal vein sampling. Characteristics and follow-up after adrenalectomy in a Danish study. *Scandinavian journal of clinical and laboratory investigation* 2016;76:45-50.

24. L Z, Y Z, H Z,et.al. Comparison between adrenal venous sampling and computed tomography in the diagnosis of primary aldosteronism and in the guidance of adrenalectomy. *Medicine* 2016;95:e4986.

25. Kocjan T, Janez A, Stankovic M,et.al. A NEW CLINICAL PREDICTION CRITERION ACCURATELY DETERMINES A SUBSET OF PATIENTS WITH BILATERAL PRIMARY ALDOSTERONISM BEFORE ADRENAL VENOUS SAMPLING. *Endocrine practice : official journal of the American College of Endocrinology and the American Association of Clinical Endocrinologists* 2016;22:587-594.

26. Asmar M, Wachtel H, Yan Y,et.al. Reversing the established order: Should adrenal venous sampling precede cross-sectional imaging in the evaluation of primary aldosteronism? *Journal of surgical oncology* 2015;112:144-148.

27. WC S, LM S, JH L,et.al. Diagnosing unilateral primary aldosteronism - comparison of a clinical prediction score, computed tomography and adrenal venous sampling. *Clinical endocrinology* 2014;81:25-30.

28. A R, E F, C D,et.al. Age below 40 or a recently proposed clinical prediction score cannot bypass adrenal venous sampling in primary aldosteronism. *The Journal of clinical endocrinology and metabolism* 2014;99:E1035-1039.

29. TJ B, IS M, K B,et.al. Evaluation of the sensitivity and specificity of (11)C-metomidate positron emission tomography (PET)-CT for lateralizing aldosterone secretion by Conn's adenomas. *The Journal of clinical endocrinology and metabolism* 2012;97:100-109.

30. Salem V, Hopkins TG, El-Gayar H,et.al. Adrenal venous sampling as a diagnostic procedure for primary hyperaldosteronism: experience from a tertiary referral centre. *Hormones (Athens, Greece)* 2012;11:151-159.

31. JH L, WC S, RH R,et.al. A prospective evaluation of postural stimulation testing, computed tomography and adrenal vein sampling in the differential diagnosis of primary aldosteronism. *Clinical endocrinology* 2012;76:182-188.

32. EM K, L A, A R,et.al. A clinical prediction score to diagnose unilateral primary aldosteronism. *The Journal of clinical endocrinology and metabolism* 2012;97:3530-3537.

33. Sarlon-Bartoli G, Michel N, Taieb D,et.al. Adrenal venous sampling is crucial before an adrenalectomy whatever the adrenal-nodule size on computed tomography. *Journal of hypertension* 2011;29:1196-1202.

34. Mathur A, Kemp CD, Dutta U,et.al. Consequences of adrenal venous sampling in primary hyperaldosteronism and predictors of unilateral adrenal disease. *Journal of the American College of Surgeons* 2010;211:384-390.

35. Mulatero P, Bertello C, Rossato D,et.al. Roles of clinical criteria, computed tomography scan, and adrenal vein sampling in differential diagnosis of primary aldosteronism subtypes
Role for adrenal venous sampling in primary aldosteronism. *The Journal of clinical endocrinology and metabolism* 2008;93:1366-1371.

36. Minami I, Yoshimoto T, Hirono Y,et.al. Diagnostic accuracy of adrenal venous sampling in comparison with other parameters in primary aldosteronism. *Endocrine journal* 2008;55:839-846.

37. FE N, BS M, R A,et.al. Primary hyperaldosteronism: effect of adrenal vein sampling on surgical outcome. *Archives of surgery (Chicago, Ill : 1960)* 2006;141:497-502; discussion 502-493.

38. Young WF, Stanson AW, Thompson GB,et.al. Role for adrenal venous sampling in primary aldosteronism. 2004:1227-1235.

39. Magill SB, Raff H, Shaker JL,et.al. Comparison of adrenal vein sampling and computed tomography

in the differentiation of primary aldosteronism. *The Journal of clinical endocrinology and metabolism* 2001;86:1066-1071.

40. EM O, KE L, K Y,et.al. Value of adrenal venous sampling for lesion localization in primary aldosteronism. *World journal of surgery* 2012;36:2522-2527.

41. Zeiger MA, Thompson GB, Duh QY,et.al. The American Association of Clinical Endocrinologists and American Association of Endocrine Surgeons medical guidelines for the management of adrenal incidentalomas. 2009:1-20.

42. T D, A P, LJS K,et.al. Adrenal vein sampling versus CT scan to determine treatment in primary aldosteronism: an outcome-based randomised diagnostic trial. *The lancet Diabetes & endocrinology* 2016;4:739-746.

Figure legend

Figure 1: Flow diagram of the review process.

Thirty-five studies were excluded from the final analysis due to incomplete or unsystematic data.

Figure 2: Assessment of methodological quality of included studies using the QUADAS-2 Criteria.

Stacked bars represent the proportion of studies with a high (red), or unclear (yellow) or low (green) risk of bias and applicability concerns.

QUADAS-2: Quality Assessment of Diagnostic Accuracy Studies –2 criteria

Figure 3: Forest plots of sensitivity and specificity of adrenal imaging compared with invasive adrenal vein sampling.

Horizontal lines are the 95% confidence intervals (CIs)

Figure 4: Hierarchical SROC plot showing average sensitivity and specificity estimate of the study results with 95% confidence region.

The 95% prediction region represents the confidence region for a forecast of the true sensitivity (SENS) and specificity (SPEC) in a future study. AUC: area under the curve; SROC: summary receiver-operating characteristic.

Figure S1: Forest plots of sensitivity and specificity of adrenal imaging compared with invasive adrenal vein sampling in young patients <40 years.

Abbreviations as in Figure 3.

Figure S2: Graphical presentation of the generalized linear mixed model exploring the impact of selected variables on sensitivity and specificity of adrenal imaging.

Contributors

WPJ is the guarantor. ZYQ and WD drafted the manuscript. All authors contributed to the development of the selection criteria, the risk of bias assessment strategy and data extraction criteria. CSC and JLC developed the search strategy. WPJ provided statistical expertise. ZP provided expertise on primary aldosteronism. All authors read, provided feedback and approved the final manuscript.

Funding: This systematic review and meta-analysis is funded by the National Natural Sciences Foundation of China (NSFC)（Grant NO. 81970262）.

Competing interests: None declared

Patient and public involvement statement: Not required

Ethics approval: Not required.

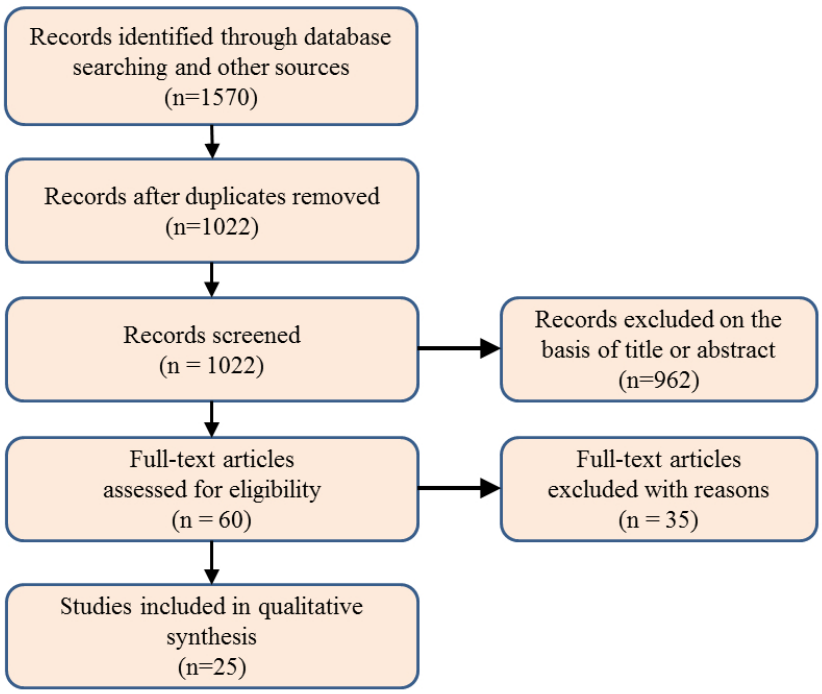


Figure 1: Flow diagram of the review process. / Thirty-five studies were excluded from the final analysis due to incomplete or unsystematic data.

254x190mm (96 x 96 DPI)

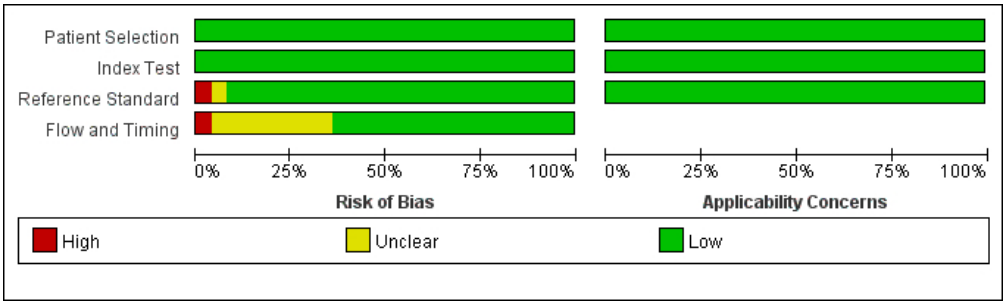


Figure 2: Assessment of methodological quality of included studies using the QUADAS-2 Criteria. / Stacked bars represent the proportion of studies with a high (red), or unclear (yellow) or low (green) risk of bias and applicability concerns.
QUADAS-2: Quality Assessment of Diagnostic Accuracy Studies –2 criteria

239x71mm (72 x 72 DPI)

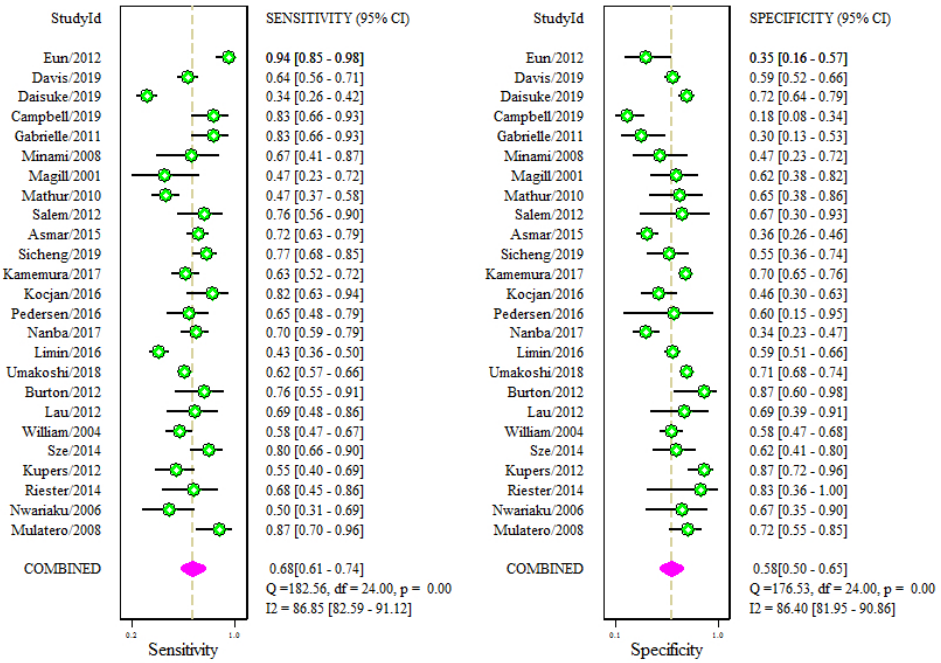


Figure 3: Forest plots of sensitivity and specificity of adrenal imaging compared with invasive adrenal vein sampling. / Horizontal lines are the 95% confidence intervals (CIs).

301x219mm (72 x 72 DPI)

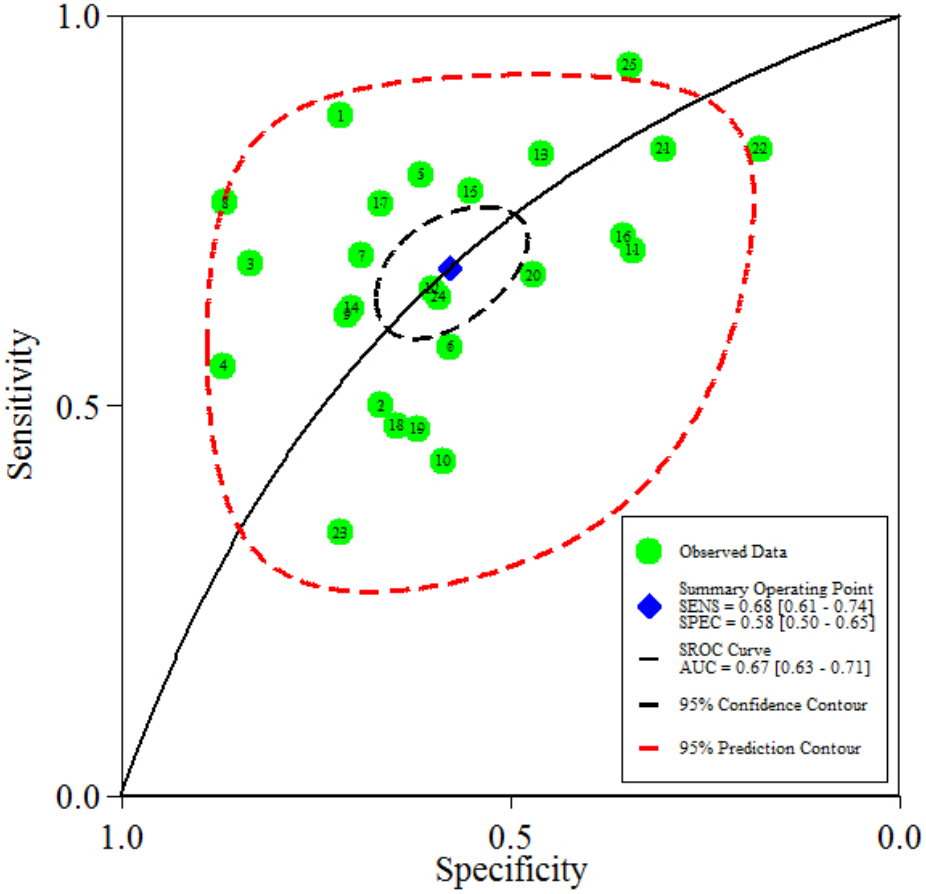
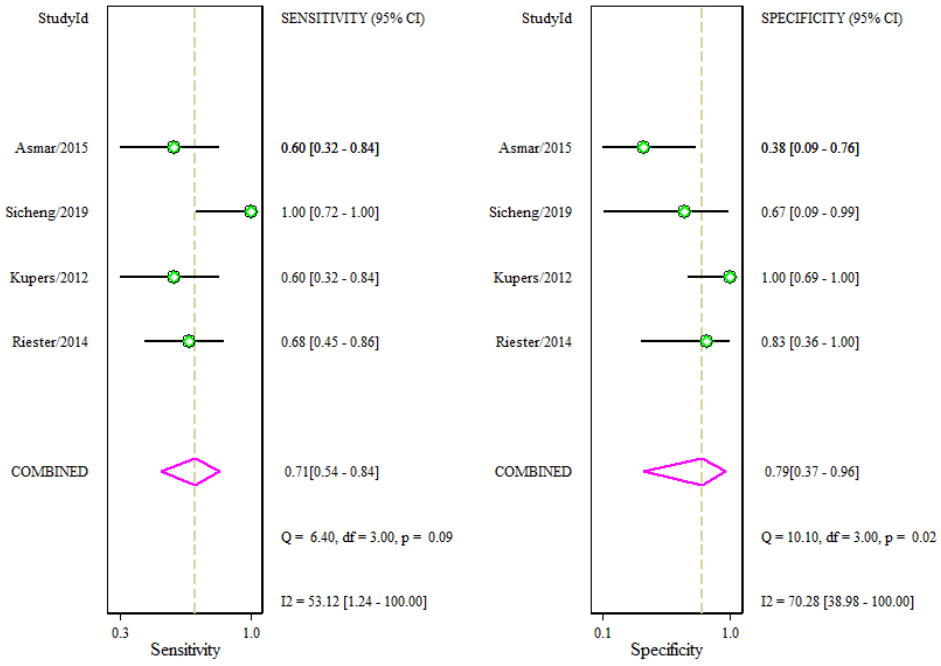
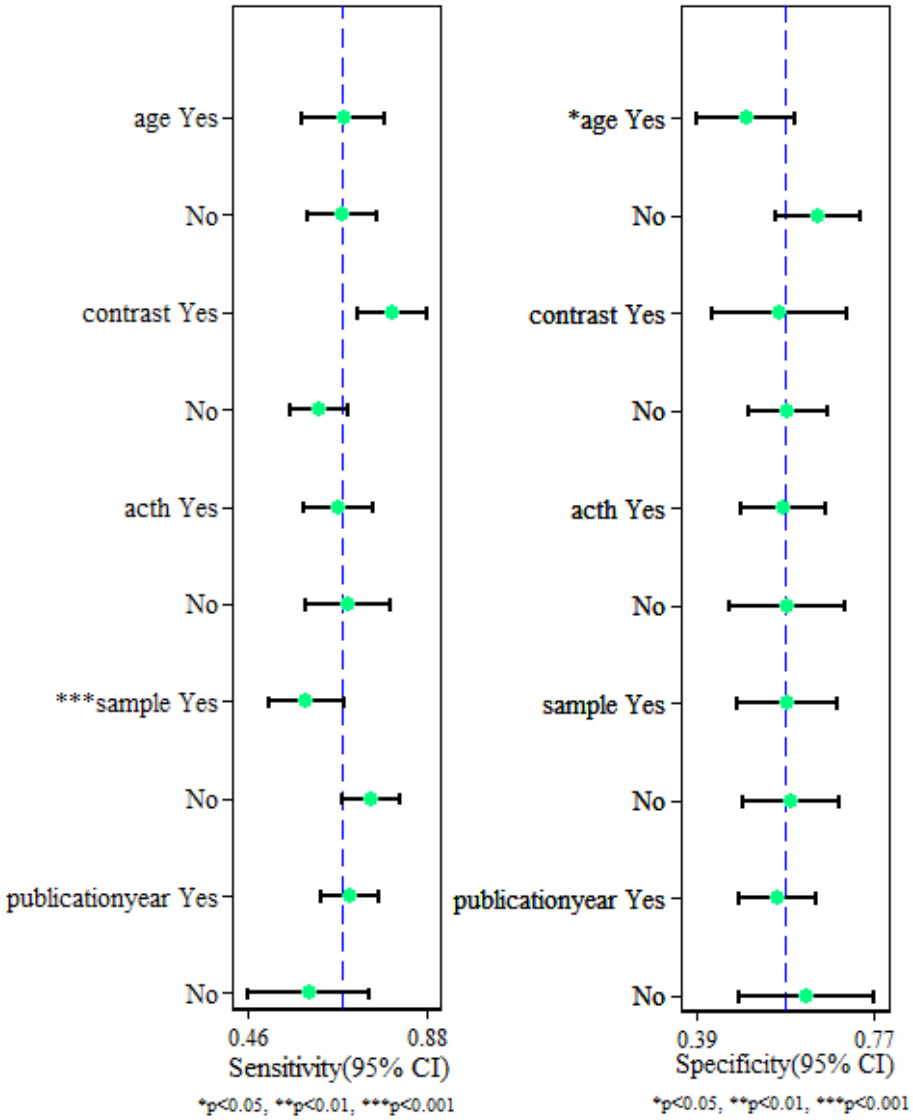


Figure 4: Hierarchical SROC plot showing average sensitivity and specificity estimate of the study results with 95% confidence region. /The 95% prediction region represents the confidence region for a forecast of the true sensitivity (SENS) and specificity (SPEC) in a future study. AUC: area under the curve; SROC: summary receiver-operating characteristic.

219x219mm (72 x 72 DPI)



301x219mm (72 x 72 DPI)



172x219mm (72 x 72 DPI)

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2-3
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3-4
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	4-5
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	none
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	5
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	5
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	6
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	6
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	6
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	5
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	7
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	7

PRISMA 2009 Checklist

Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	7

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	7
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	7
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	7-8
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	8-9
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	10
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	9
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	9-10
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	10
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	10-11
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	11-13
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	13-14
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	14
FUNDING			

PRISMA 2009 Checklist

Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	18

BMJ Open

Diagnostic Accuracy of Adrenal Imaging for the Subtype Diagnosis of Primary Aldosteronism: Systematic Review and Meta-Analysis

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-038489.R1
Article Type:	Original research
Date Submitted by the Author:	25-Aug-2020
Complete List of Authors:	Zhou, Yaqiong; The First Affiliated Hospital of Chengdu Medical College, Cardiology; Key Laboratory of Aging and Vascular Homeostasis of Sichuan Higher Education Institutes Wang, Dan; The First Affiliated Hospital of Chengdu Medical College, Cardiology; Key Laboratory of Aging and Vascular Homeostasis of Sichuan Higher Education Institutes Jiang, Licheng; The First Affiliated Hospital of Chengdu Medical College, Cardiology; Key Laboratory of Aging and Vascular Homeostasis of Sichuan Higher Education Institutes Ullah, Hameed; the First Affiliated Hospital of Xi'an Jiaotong University, Cardiology Ran, Fei; The First Affiliated Hospital of Chengdu Medical College, Cardiology; Key Laboratory of Aging and Vascular Homeostasis of Sichuan Higher Education Institutes Chen, Sichao ; The First Affiliated Hospital of Chengdu Medical College, Cardiology Zhou, Peng; The First Affiliated Hospital of Chengdu Medical College, Cardiology; Key Laboratory of Aging and Vascular Homeostasis of Sichuan Higher Education Institutes Wang, Peijian; The First Affiliated Hospital of Chengdu Medical College, Cardiology; Key Laboratory of Aging and Vascular Homeostasis of Sichuan Higher Education Institutes
Primary Subject Heading:	Diabetes and endocrinology
Secondary Subject Heading:	Cardiovascular medicine, Diabetes and endocrinology
Keywords:	Hypertension < CARDIOLOGY, Endocrine tumours < DIABETES & ENDOCRINOLOGY, Cardiology < INTERNAL MEDICINE

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Diagnostic Accuracy of Adrenal Imaging for the Subtype Diagnosis of Primary Aldosteronism: Systematic Review and Meta-Analysis

Yaqiong Zhou[12] , Dan Wang[1], Licheng Jiang[12], Hameed Ullah[3], Fei Ran[1], Sichao Chen[1], Peng Zhou[12*], Peijian Wang [12*]

*Corresponding author: Prof. Peijian Wang

Department of Cardiology, the First Affiliated Hospital of Chengdu Medical College, 299 baoguang avenues, Chengdu 610500, Sichuan Province, China. E-mail: wpjmed@aliyun.com

*Corresponding author: Prof. Peng Zhou

Department of Cardiology, the First Affiliated Hospital of Chengdu Medical College, 299 baoguang avenues, Chengdu 610500, Sichuan Province, China. E-mail: ap216@163.com.

Email: Yaqiong Zhou.1273868593@qq.com- Dan Wang.2292029981@qq.com- Licheng Jiang. 65839350@qq.com- Hameed Ullah. 807287144@qq.com- Fei Ran 1126852537@qq.com- Sichao Chen. 1350821872@qq.com-Peng Zhou. ap216@163.com-Peijian Wang. wpjmed@aliyun.com

Funding: This systematic review and meta-analysis are funded by the National Natural Sciences Foundation of China (NSFC)（Grant NO. 81970262）.

Abstract

Objectives: The accurate subtype classification of primary aldosteronism (PA) is critical in assessing optimal treatment options. This study aimed to evaluate the diagnostic accuracy of adrenal imaging for unilateral PA classification.

Methods: Systematic searches of PubMed, EMBASE, and the Cochrane databases were performed from 1st January 2000 to 1st February 2020 for all studies that used computed tomography (CT) or magnetic resonance imaging (MRI) in determining unilateral PA and validated the results against invasive adrenal vein sampling (AVS). Summary diagnostic accuracies were assessed by using a bivariate random effects model. A generalized linear mixed model was performed for heterogeneity analysis.

Result: 25 studies were identified, involving a total of 4669 subjects in the meta-analysis. The overall analysis revealed a pooled sensitivity of 68% (95% confidence interval [CI]: 61 to 74), specificity of 57% (95% CI: 50 to 65), a positive likelihood ratio (LR) of 1.6 (95% CI: 1.4 to 1.9), negative LR of 0.56 (95% CI: 0.47 to 0.68), and diagnostic odds ratio (DOR) of 3 (95% CI: 2 to 4) for CT/MRI to identify unilateral PA. Sensitivity and DOR were higher in contrast-enhanced CT group versus traditional CT group [77% (95% CI: 66 to 85 vs. 58% (95% CI: 50 to 66) and 5 (95% CI: 3 to 7) vs. 2 (95% CI: 1 to 3), respectively]. Diagnostic accuracy of PA patients aged ⩽40 years was reported in 4 studies, the overall sensitivity of imaging for correctly identifying unilateral PA was 71%, with a 79% specificity. Heterogeneity analysis revealed a significant impact of sample size on sensitivity of adrenal imaging.

Conclusion: CT/MRI is not a reliable alternative to invasive AVS without excellent sensitivity and specificity for correctly identifying unilateral PA. Even in young patients (⩽40 years), almost one-fourth patients would have undergone unnecessary adrenalectomy based on imaging results alone.

Keywords: adrenal vein sampling, computed tomography, magnetic resonance imaging, primary aldosteronism, subtype

Strengths and limitations of this study

- This study is the first meta-analysis to synthesize the evidence regarding the diagnostic value of adrenal imaging for PA classification and demonstrated that CT/MRI is not a reliable alternative to invasive AVS even in young patients (⩽40 years).
- The main methodological limitations of this systematic review and meta-analysis are the exclusion of unpublished high-quality trials and foreign-language publications.
- Another potential limitation is that we might encounter different AVS methods (with or without adrenocorticotropic hormone stimulation) and the criteria of lateralization, which might also affect the results of diagnostic accuracy.

Introduction

Primary aldosteronism (PA) is one of the most common causes of endocrine hypertension, with a prevalence of around 20% among patients with resistant hypertension, 10% in those with severe hypertension, and 6% in those with uncomplicated hypertension[1]. Accumulating clinical and epidemiological evidence suggest that PA amplifies cardiovascular and cerebrovascular complications beyond essential hypertension prior to treatment, even after controlling the elevated blood pressure[2,3]. However, patients with the unilateral resected PA and post-adrenalectomy biochemically cured patients had slightly better risk profiles than matched essential hypertensive patients. Patients with bilateral PA whose plasma renin activity is not suppressed after 6 months of mineralocorticoid receptor antagonists (MRA) therapy have the same risk profiles as essential hypertensive patients; those whose renin activity remains suppressed have 4-fold higher risk profiles than controls and titration of MRA therapy to raise renin might reduce this excess risk[4,5]. Accordingly, early diagnosis and specific treatment of affected patients are a key step for reversal of target-organ damage and prevention of cardiovascular and cerebrovascular events.

The selection of the most appropriate therapeutic strategy for patients with PA requires the distinction between the unilateral and bilateral forms of PA. The former requires a unilateral adrenalectomy, mainly entail aldosterone-producing adenoma

(APA) and less commonly, unilateral adrenal hyperplasia, in contrast the latter, also known as idiopathic hyperaldosteronism, is optimally treated with target medical therapy[3]. Regarding the differentiation of unilateral from bilateral subtype, all current clinical practice guidelines recommend adrenal vein sampling (AVS) as the standard procedures for subtype diagnosis[6,7]. However, several shortages of AVS have been reported, such as technical challenges, invasive nature, poorly standardized procedures, high cost, and lack of availability. Thus, it is urgent to explore alternative diagnostic methods without sacrificing accuracy.

Adrenal imaging with computed tomography (CT) or magnetic resonance imaging (MRI) is recommended as the first step for subtype classification owing to its ease of performance and relative accessibility[8]. By now, numerous studies have evaluated the diagnostic performance of CT/MRI in subtype diagnosis of PA, but the results have been inconsistent. Moreover, all of these studies were limited by small sample sizes in a single-center, which limited the credibility of results. In this context, systematic review and meta-analysis have the benefit of increasing the sample size generating more precise results, which has been widely applied in clinical studies[9]. In 2009, one systematic review reported that CT/MR-based diagnoses were discordant with AVS results in 37.8% of PA patients[10]. However, the conclusions may not be reliable because of the potential for bias and concerns regarding the comparability of the included studies. Moreover, several additional studies were reported after this systematic review. We thus performed a comprehensive meta-analysis of all available studies to evaluate the diagnostic value of adrenal imaging (CT/MRI) for subtype classification of PA.

Method

Search Strategy

The study followed the guidelines specified in the preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) [11]. We searched the PUBMED, EMBASE, and Cochrane Library databases from 1st January 2000 to 1st February 2020, using the following terms in combination, both as MeSH or Emtree terms and text words: "primary aldosteronism" "adrenal vein sampling" and

"hyperaldosteronism". Electronic search strategy for PubMed was shown in supplementary Table S1. To reflect modern practice, we decided to limit publication date of studies after 1st January 2000. We searched articles published in English language and, the references of relevant studies were also searched. All studies were carefully examined to exclude overlapping or potential duplicates data. Trials in abstract form without a published manuscript also excluded.

Eligibility Criteria

We included a study if: 1) it used CT or MRI as a diagnostic test for PA subtyping; 2) it used AVS as the standard of reference. Successful AVS should be determined by calculating the selectivity index (SI), defined as the adrenal/peripheral vein cortisol ratio and unilateral PA should be determined by calculating the lateralization index (LI), defined as the aldosterone/cortisol ratio between the dominant and the nondominant adrenal gland; and 3) absolute numbers of true-positive, true-negative, false-positive, and false-negative results were provided or could be derived. Identified studies had to be independent. In the case of multiple reports on the same population or subpopulation, the most recent or comprehensive information's used.

Data Abstraction and Quality Assessment

Data extraction of the eligible studies was performed by 2 independent investigators (Z.Y.Q and W.P.J) using a standardized data extraction form. The form included the following characteristics of each trial: first author's name and year of publication; study population characteristics, including sample size, geographical location, mean age and sex; AVS characteristics, including SI, LI, and with/without adrenocorticotropic hormone (ACTH) stimulation; diagnostic test characteristics, including imaging methodology and contrast administered or not; Differences between reviewers were resolved by discussion and consensus whenever necessary.

The methodological quality of identified studies was assessed by 2 independent reviewers (Z.Y.Q and W.P.J) using the modified Quality Assessment of Diagnostic Accuracy Studies –2 (QUADAS-2) criteria[12] and discrepancies were resolved by discussion and consensus.

Statistical Analysis

Measures of diagnostic accuracy are reported as point estimates with 95% confidence intervals (CIs). Sensitivity, specificity, positive likelihood ratio (+LR) and negative likelihood ratio (-LR) were modeled based on the true-positive, true-negative, false-positive, and false-negative rates for each trial. There is consensus that a +LR >10 and a -LR <0.1 provide reliable evidence of satisfactory diagnostic performance[13]. The ratio of +LR to -LR was combined in a single global accuracy measure, the diagnostic odds ratio (DOR). Summary sensitivity, specificity, +LR, -LR and DOR been assessed by using a bivariate random-effects model. The approach assumes bivariate normal distributions for the logit transformations of sensitivity and specificity from individual studies. The parameters of the bivariate models are estimated in a single model to incorporate the possible correlation between sensitivity and specificity. These bivariate models can be analyzed using linear mixed model techniques that are now widely available in statistical packages, such as *STATA gllamm*[14,15]. A hierarchical summary receiver operating characteristic curves (ROC) was performed[16], presenting the point estimates for each trial and the pooled characteristics, including the 95% prediction region and the 95% confidence region.

Sources of statistical heterogeneity explored by using the bivariate generalized linear mixed model[17,18], which assessed by using the I^2 statistic, applying the following interpretation for I^2: <50%=low heterogeneity; 50% to75%=moderate heterogeneity; >75%=high heterogeneity.

Several studies demonstrated that MRI has poorer resolution and slower acquisition, with the risk of respiratory artifacts and is inferior to adrenal CT in PA subtype evaluation[19-22]. Contrast materials can improve the visibility of adrenal structures imaged by CT and MRI scans and might have a positive effect on diagnosis accuracy[23]. Thus, imaging methods and contrast materials were thought as confounders for subgroup analyses. Moreover, a large sample number may represent experienced interventional radiologists and the credibility of the included studies. Thus, the sample size thought of as another confounder for subgroup analyses. Given that different cutoff for LI criteria and AVS procedure (with or without ACTH

stimulation) might also affect the results of diagnosis accuracy[22]; we also performed subgroup analyses stratified by these parameters. Thus, subgroup analyses were performed by the following factors: imaging methodology (CT or CT/MRI), contrast used or not, AVS procedure (with or without ACTH stimulation), the cutoff value for LI (2 or 4) and sample size (divided by 100 subjects).

The potential publication bias was examined by using the Deeks test[24]. Cohen κ test was assessed for the inter-rater reliability between 2 observers for quality assessment. If had no matched, a third reviewer been involved for disagreements and, final decisions determined by consensus. Statistical analyses performed using Stata version13.0 (StataCorp LP, TX, USA) and Review Manager version 5.3.

Result

Study Selection

After removal of the 548 duplicates, the systematic review retrieved 1022 references that were screened according to title or abstract for possible inclusion. Among them 962 studies excluded for the following reasons: 489 studies were not relevant; 280 studies were reviews or practice guidelines; 92 studies did not include humans; 101studies were case/letter report. After screening, 60 studies were identified as potentially eligibility, and full text was retrieved for detailed evaluation. 35 studies were excluded for the following reasons: data to compute diagnostic accuracies were not provided or could not be derived (18 papers), reporting on the same population (4 papers), no comparison of CT/MRI and AVS results in individual patients (6 papers), and no values for true-positive and false-negative observations (7 papers). Finally, 25 articles were deemed eligible and analyzed in our meta-analysis [19-22,25-45](Figure 1).

Study Characteristics

Overall, a total of 4669 patients (mean age 51 years; 54% male) from 25 articles were included. The sizes of the identified studies ranged from 35 to 1591, with the largest study recruiting over 1000 participants[27]. 5 studies including 724 participants underwent cross-sectional imaging either CT or MRI, and the remaining 20 studies including 3945 patients only used CT scan (8 studies administered contrast material). 17 studies performed AVS with ACTH stimulation, 7 studies without ACTH

stimulation and the remaining 1 study provided the above two methods. The 2016 Endocrine Society Guideline recommends more strict criteria LI (2.0 or greater under

unstimulated conditions and/or 4 for ACTH stimulation) and SI (2.0 or greater under unstimulated conditions and/or 3 for ACTH stimulation) [8]. In the meta-analysis, 1 included study used less-permissive criteria for LI[26]and 6 included studies for SI[22, 26, 37-40]. The threshold of SI was not accessible in 4 studies[34,35,20,41]and LI in 2 studies[20,41]. Further details of the eligible and analyzed studies are shown in Table 1.

Table 1: Study Characteristics

Study,Year	Location	Male (%)	Age (years)	Sample	Imaging methodology	Contrast used	AVS characteristics	TP	FP	FN	TN
Li et al. 2019	China	61(50)	48.5	122	CT	yes	without ACTH SI≥2 LI≥2	72	13	21	16
			<35					11	1	0	2
Campbell et al.2019	USA	45(61)	55.6	74	CT or MRI	no	with ACTH, SI≥5, LI≥4	29	32	6	7
Aono et al. 2019	Japan	NA	56	317	CT	no	with ACTH, SI>2, LI≥2	52	46	102	117
			<35					3	2	0	0
Sam et al.2019	Canada	201(59)	52.1	342	CT or MRI	yes	with ACTH,SI>2, LI≥4	101	75	57	109
Umakoshi et al.2018	Japan	762(47.9)	53	1591	CT	yes	with ACTH,SI>5,LI>4	297	322	185	787
			<35					27	3	0	0
Nanba et al.2017	USA	87(59)	54	147	CT	no	with ACTH, SI≥5, LI≥4	58	42	25	22
Kamemura et al.2017	Japan	177(45)	54	393	CT	no	with ACTH,SI>5,LI>4	60	88	36	209
Zhu et al. 2016	China	NA	46	394	CT	no	without ACTH, SI≥3, LI≥2	87	79	116	112
			<40					30	35	1	0
Pedersen et al.2016	Denmark	24(54)	51	45	CT or MRI	no	without ACTH. SI /LI NA	26	2	14	3
Kocjan et al. 2016	Slovenia	46(69)	56	67	CT	no	with ACTH,SI>5,LI>4	23	21	5	18
Asmar et al. 2015	USA	148(63)	55	235	CT or MRI	no	with ACTH, SI≥5, LI≥4	96	65	38	36
			<40					9	5	6	3
Riester et al. 2014	Germany	NA	35	28	CT or MRI	no	without ACTH, SI≥2, LI≥4	15	1	7	5
Candy Sze et al. 2014	UK	42(56)	50.5	75	CT	yes	with ACTH, SI≥5, LI≥4	39	10	10	16
Küpers et al.2012	France	53(61)	46	87	CT	no	with ACTH, SI≥2, LI≥4	26	5	23	33
			<40					9	0	6	10
Lau et al. 2012	UK	24(64)	51.8	39	CT	no	with ACTH, SI≥5, LI≥4	18	4	8	9
Burton et al. 2012	UK	NA	50.9	40	CT	Yes	without ACTH,LI≥4	19	2	6	13
Salem et al.2012	UK	16(44)	44.7	38	CT	no	without ACTH, LI≥4	22	3	7	6
Oh et al. 2012	Korea	45(52)	50.7	86	CT	yes	with ACTH, SI≥3, LI>4	59	15	4	8
Sarlon-Bartoli et al.2011	France	NA	52	58	CT	yes	without ACTH, SI≥1, LI>2	29	16	6	7
Mathur et al. 2010	USA	63(55)	50.6	114	CT	no	with ACTH, SI>2, LI≥4	46	6	51	11
Mulatero et al. 2008	Italy	NA	52.4	70	CT	yes	65 with ACTH; 5 without ACTH ,SI>2, LI≥4	27	11	4	28
Minami et al.2008	Japan	12(34)	54	35	CT	no	with ACTH SI /LI NA	12	9	6	8
Nwariaku et al. 2006	USA	27(67)	51	40	CT	yes	with ACTH, SI≥3, LI>4	14	4	14	8
Young et al. 2004	USA	163(84)	53	194	CT	no	with ACTH, SI>5, LI>4	57	48	42	47
Magill et al.2001	USA	27(71)	51	38	CT	no	with ACTH, SI≥3, LI≥4	8	8	9	13

adrenocorticotropic hormone; CT: computed tomography; MRI: magnetic resonance imaging; NA: not available; SI: selectivity index; LI: lateralization index; TP: true-positive; TN: true-negative; FP: false-positive; FN: false-negative.

Quality assessment

Overall, the identified studies showed excellent quality in terms of applicability and risk of bias. Out of 175 QUADAS-2 items (25 articles×7 items), the 2 reviewers agreed on 172 (98%) with an inter-rater agreement of κ=0.9. Figure 2 summarizes the

QUADAS-2 assessment, and supplementary Table S2 displays each of the 25 individual QUADAS-2 evaluations. Risk of bias from reference standard was high in 2 studies[30,34] and it was unclear in 3 studies[20,35,41] because it was not clear whether the reference standard was interpreted blind to the adrenal imaging results or whether the cutoff values of SI and LI for correctly classifying the target condition. The risk of bias regarding flow and timing was unclear in 7 studies[27-29,31,36,40,42] because the time interval between the index test and the reference standard was unclear and it was high in 2 studies[20,43]because not all patients receive the same reference standard.

Overall analysis

Using the bivariate model, statistical heterogeneity was found for sensitivity (I^2 =86.9%; *P*= 0.001), specificity (I^2 =86.9%; *P*= 0.001), positive LR (I^2 = 76.3%; *P* = 0.001), negative LR (I^2 =79.2%; *P* = 0.001), and DOR (I^2 = 100.0%; *P* =0.001), indicating high between-study heterogeneity for all pooled measures, which might compromise the credibility of the study.

In the overall analysis, the pooled sensitivity, specificity, positive LR, negative LR, and DOR for adrenal imaging were 68% (95% confidence interval [CI]: 61 to 74), 57% (95% CI: 50 to 65), 1.6 (95% CI: 1.4 to 1.9), 0.56 (95% CI: 0.47 to 0.68) , and 3 (95% CI: 2 to 4) , respectively (Figure 3, 4).

Subgroup analyses

Subgroup analysis, stratified by the imaging methodology, found more favorable specificity (61%, 95%CI: 54-67) and negative LR (0.54, 95% CI: 0.43-0.67) for CT scan compared with the specificity (45%; 95% CI: 27-64) and negative LR (0.68, 95% CI: 0.50-0.91) for CT/MRI. Notably, subgroup analysis showed an increase in sensitivity and DOR when contrast material has administered during CT scan versus traditional CT group [77% (95% CI: 66 to 85 vs. 58% (95% CI: 50 to 66) and 5 (95% CI: 3 to 7) vs. 2 (95% CI: 1 to 3), respectively].

Subgroup analysis based on AVS procedure (with or without ACTH stimulation) revealed a slight decrease in sensitivity [66% (95% CI: 57 to 73 vs. 70% (95% CI: 58 to 79)] when ACTH has administered during AVS procedure. In a further stratified analysis by the LI demonstrated that specificity and specificity were higher when LI

was ⩾4 versus LI was ⩾2 [69% (95% CI: 62 to 75 vs. 61% (95% CI: 37 to 80) and 59% (95% CI: 50 to 68) vs. 54% (95% CI: 46 to 75), respectively].

There were 4 studies reported diagnostic accuracy of PA patients with an age of 40 years or younger. Using the bivariate model, the pooled sensitivity, specificity, positive LR, negative LR, and diagnostic odds ratio were 71% (95% CI: 54 to 84), 79% (95% CI: 37 to 96) (Figure S1), 3.4 (95% CI: 0.2 to 14.7), 0.37 (95% CI: 0.2 to 0.68), and 9 (95% CI: 1 to 64), respectively. Summary estimates for pooled measures of diagnostic accuracy have shown in Table 2.

Table 2: Pooled summary results by subgroups.

Subgroups	No.of Studies	Sensitivity (95% CI) with I^2	Specificity (95% CI) with I^2	Positive LR (95% CI) with I^2	Negative LR 95% CI) with I^2	DOR (95% CI) with I^2
Total	25	68(61-74) 86.9%	57(50-65) 86.9%	1.6(1.4-1.9) 76.3%	0.56(0.47-0.68) 79.2%	3(2-4) 100%
Age						
⩽40years	**4**	**71(54-84) 53.1%**	**79(39-96) 70.1%**	**3.4(0.2-14.7) 50.5%**	**0.37(0.2-0.68) 41.5%**	**9(1-64) 99.1%**
>40years	21	68(60-75) 87.4%	57(49-64) 87.4%	1.6(1.3-1.8) 76.5%	0.57(0.46-0.70) 79.4%	3(2-4) 100%
Cutoff values of LI						
LI⩾2	4	61(37-80) 95.5%	54(38-68) 88.3%	1.3(1.1-1.6) 76.1%	0.73(0.5-1.09) 91.3%	2(1-3) 100%
LI⩾4	18	69(62-75) 76.4%	59(50-68) 89.0%	1.7(1.4-2.1) 77.4%	0.52(0.43-0.64) 58.5%	3(2-5) 100%
AVS procedure						
With ACTH	17	66(57-73) 86.5%	56(46-65) 90.5%	1.5(1.3-1.7) 74.7%	0.62(0.52-0.73) 71.1%	2(2-3) 100%
Without ACTH	7	70(58-79) 90.3%	60(45-74) 79.2%	1.8(1.2-2.6) 81.2%	0.5(0.33-0.75) 91.1%	4(2-7) 100%
Imaging methodology						
CT	20	67(59-74) 88.6%	60(53-67) 82.5%	1.7(1.4-2.0) 73.7%	0.55(0.44-0.68) 83.8%	3(2-4) 100%
contrast CT	**8**	**77(66-85) 86.4%**	**60(49-69) 82.2%**	**1.9(1.6-2.3) 76.6%**	**0.38(0.28-0.53) 78.1%**	**5(3-7) 100%**
nocontrast CT	12	58(49-66) 83.8%	60(51-68) 81.8%	1.5(1.2-1.8) 55.5%	0.7(0.57-0.85) 72.6%	2(1-3) 99.9%
CT /MRI	5	69(62-76) 30%	45(27-64) 87.9%	1.3(1.0-1.7) 38.7%	0.68(0.50-0.91) 12%	2(1-3) 81.5%
Sample size						
⩾100	10	59(51-67) 90.6%	58(49-66) 92.1%	1.4(1.2-1.7) 82.9%	0.7(0.59-0.84) 84.8%	2(1-3) 100%
<100	15	74(67-81) 71.3%	60(47-71) 78.1%	1.9(1.4-2.4) 52.4%	0.43(0.33-0.54) 37.5%	4(3-7) 97.3%

LR: likelihood ratio; DOR: diagnostic odds ratio; other abbreviations as in Table 1.

Meta-regression analysis

Adding pre-premised covariates to the bivariate generalized linear mixed model showed a significant interaction between the smaller sample size and higher sensitivity (I^2=77%; *P*=0.01) of CT/MRI for the detection of unilateral forms of PA (Figure S2).

Publication bias

Neither Deeks' Funnel Plot nor Deeks test (t=0.46, *P*=0.65) showed evidence of publication bias (Figure S3).

Discussion

Main findings

The accurate differentiation of unilateral from bilateral PA is critical for optimal clinical management. Although AVS is the "gold standard" test for subtype diagnosis[8], numerous studies have investigated the underlying diagnostic value of CT/MRI for subtype diagnosis due to several insurmountable shortages of AVS. The present meta-analysis, involving 4669 individuals from 25 studies, demonstrated that CT/MRI has a poor sensitivity (68 %) and specificity (57 %) in the subtype classification when used AVS as the reference standard.

In the subtype diagnosis of PA, AVS was initially used in the 1960s. Subsequently, CT adopted as the primary method for distinguishing unilateral from the bilateral form of PA. Owing to less invasive nature, less cost and, widely availability, many physicians prefer to perform CT/MRI as the first and sometimes the only investigation of subtype diagnosis of PA. But its sensitivity and specificity vary widely. The reported sensitivities ranged from 29%[3] to 94%[45] and the reported specificities ranged from 18%[21] to 87%[37]. Although the sensitivities reported to exceed 80% in 5 studies[21, 31, 38, 40, 45], relatively poor specificities reported with only 1 study showing a specificity of 72%[26]. Similarly, 3 studies reported the sensitivities to be over 80%[33, 34, 37], but the specificities reported to be lower than 76%[34]. The present meta-analysis showed that the pooled sensitivity was 68% and specificity was 57%, which means that decision of treatment based on the presence of unilateral disease on CT/MRI alone could result in inappropriate unilateral adrenalectomy in 43% patients. Whereas, based on CT/MRI alone would miss the possibility of a potentially curative procedure by surgery in 32% of patients. However, failure of early diagnosis and specific treatment of PA place these patients at higher risk of irreversible renal and cardiovascular damage. Our results suggest that CT/MRI does not have satisfactory diagnostic performance in classifying subtypes of PA.

Contrast materials can improve the visibility of internal body structures imaged

by CT and MRI scans. They make certain structures appear in a differently way and thus help in distinguishing a contrast-enhanced area of the body from the surrounding tissue[46]. In the present meta-analysis, 8 studies including 530 participants underwent contrast-enhanced CT[25, 32, 34, 36, 38, 40, 42, 45].Notably, subgroup analysis showed a favorable increase in sensitivity (77%) in contrast-enhanced CT group. However, the diagnostic performance was still unsatisfactory since 23% of patients would miss the possibility of a potentially curative procedure by adrenalectomy.

Given the nonfunctioning adrenocortical adenomas ("incidentaloma") are relatively uncommon in young people(≤40 years), the 2009 guidelines for managing PA contended that younger patients with an unequivocal biochemical diagnosis of PA and a clear-cut unilateral cortical adenoma on adrenal CT scan proceed directly to surgery and AVS procedure may be skipped[47]. Among studies included in the present meta-analysis, there were 4 studies reported diagnostic accuracy of CT/MRI in identifying unilaterally PA in patients≤40 years. By combining these 4 studies, our results demonstrated that although the sensitivity (71%) and specificity (79%) were improved, while the diagnostic performance was still unsatisfactory since 21% of patients would have undergone unnecessary adrenalectomy based on imaging results alone. In 2016, the updated clinical practice guidelines were published and suggested the age cutoff for sparing AVS was 35 years[8]. Regarding patients aged ≤35 years, several retrospective studies have evaluated the diagnostic value of CT. The reported rate of concordance between CT and AVS ranges from 59% to 90% [21, 27, 30]. Based on these data, it seems that CT still cannot replace AVS in patients aged ≤35 years. However, due to the lack of numbers of false-positive and true-negative, we did not perform pooled analysis. Further studies are needed to clarify the diagnostic value of CT in patients aged ≤35 years.

Stimulation with ACTH during AVS procedure was introduced in 1979 and remains popular at many centers. Today, AVS procedure, with or without ACTH stimulation, is still a controversial debate. Moreover, different cutoff values of LI are used in different centers during the different AVS procedure[48]. Subgroup analyses stratified by these factors were performed. Analysis by AVS procedure (with or

without ACTH stimulation) revealed that there was no significant difference between the two groups in diagnostic accuracies of CT/MRI to identify unilateral PA. Analysis by the LI demonstrated that discordance between CT/MRI and AVS was common irrespective of LI threshold and stricter thresholds for determining lateralization on AVS would result in lower rates of discordance between adrenal imaging and AVS.

As mentioned above, although adrenal imaging is not a reliable method to differentiate subtype of PA, this does not mean that CT/MRI must be wrong and should not be used as a basis for clinical management. If centers without AVS facilities currently, what should a physician do? The past few years have witnessed a rapidly growing interest in testing the utility of hybrid steroids, such as 18-oxocortisol/18-hydroxycortisol, for PA subtype and the results demonstrated that levels of 18-oxocortisol/18-hydroxycortisol plus an adenoma on CT/MRI might be of more assistance in those centers without AVS facilities especially in Japan and China, given their very high percentage of KCNJ5 mutations[49-51]. What will hopefully very substantially reduce or replace lateralization by AVS, perhaps the possibility of multi-steroid fingerprints in peripheral blood samples that distinguish unilateral from bilateral PA with a high degree of accuracy.

Limitations

The present meta-analysis has several limitations. First, the minority of studies participants underwent cross-sectional imaging either CT or MRI, but absolute numbers were not provided or could not be derived based on the specific imaging methodology used, which limited our ability to identify which imaging methodology (CT or MRI) can provide more accurate diagnostic performance. Second, the possibility of selection bias that is present in all meta-analysis cannot be overlooked. Also in addition, there was great heterogeneity of included studies, which might compromise the credibility.

Conclusion

Based on these analyses, we conclude that CT/MRI has poor sensitivity (68 %) and specificity (57 %) in the detection of unilateral PA when used AVS as the

reference standard. Even in young patients (⩽40 years), 21% would have undergone unnecessary adrenalectomy based on imaging results alone. Therefore, on behalf of these we recommend routinely referring all patients for AVS, regardless of age and imaging results, if the centers have access to AVS.

References:

1. Monticone S, Burrello J, Tizzani D*, et al.* Prevalence and Clinical Manifestations of Primary Aldosteronism Encountered in Primary Care Practice. *J Am Coll Cardiol* 2017;69:1811-20.

2. Catena C, Colussi G, Nadalini E, *et al.* Cardiovascular outcomes in patients with primary aldosteronism after treatment. *Arch Intern Med* 2008;168: 80-5.

3. Pillai PR, Griffith M, Schwarcz MD*, et al.* Primary Aldosteronism: Cardiovascular Risk, Diagnosis, and Management. *Cardiol Rev* 2020;28:84-91.

4. Hundemer GL, Curhan GC, Yozamp N*, et al.* Cardiometabolic outcomes and mortality in medically treated primary aldosteronism: a retrospective cohort study. *Lancet Diabetes Endocrinol* 2018;6:51-59.

5. Funder JW. Primary Aldosteronism: Where Are We Now? Where to from Here? *Horm Metab Res* 2020;52:459-466.

6. Williams TA, Reincke M. MANAGEMENT OF ENDOCRINE DISEASE: Diagnosis and management of primary aldosteronism: the Endocrine Society guideline 2016 revisited. *Eur J Endocrinol* 2018;179:R19-R29.

7. Byrd JB, Turcu AF, Auchus RJ. Primary Aldosteronism: Practical Approach to Diagnosis and Management. *Circulation* 2018;138:823-835.

8. Funder JW, Carey RM, Mantero F*, et al.* The Management of Primary Aldosteronism: Case Detection, Diagnosis, and Treatment: An Endocrine Society Clinical Practice Guideline; *J Clin Endocrinol Metab* 2016;101:1889-916.

9. Ruco A, Dossa F, Tinmouth J*, et al.* Social media and mobile health technology for cancer screening: a systematic review and meta-analysis protocol. *BMJ Open* 2020;10:e035411.

10. Kempers JE, Lenders JW , Outheusden L*, et al.* Systematic review: diagnostic procedures to differentiate unilateral from bilateral adrenal abnormality in primary aldosteronism. *Ann Intern Med* 2009; 151: 329-37.

11. Liberati A, Altman DG, Tetzlaff J*, et al.* The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ann Intern Med* 2009; 151: W65-94.

12. Whiting PF, Rutjes A, Westwood M*, et al.* QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; 155: 529-36.

13. Devillé WL, Buntinx F, Bouter LM*, et al.* Conducting systematic reviews of diagnostic studies: didactic guidelines. *BMC Med Res Methodol* 2002; 2: 9.

14. Reitsma JB, Glas AS, Rutjes AW*, et al.* Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;58:982-90.

15. Chu H, Guo H, Zhou Y. Bivariate random effects meta-analysis of diagnostic studies using generalized linear mixed models. *Med Decis Making* 2010;30:499-508.

16. Zwinderman AH, Bossuyt PM. We should not pool diagnostic likelihood ratios in systematic

reviews. *Stat Med*. 2008;27:687–97.

17. Hamza TH, Arends LR Houwelingen H, *et al.* Multivariate random effects meta-analysis of diagnostic tests with multiple thresholds. *BMC Med Res Methodol* 2009;9:73.

18. Lijmer JG, Bossuyt PM, Heisterkamp S H. Exploring sources of heterogeneity in systematic reviews of diagnostic tests. *Stat Med* 2002;21:1525-37.

19. Asmar M, Wachtel H, Yan Y, *et al.* Reversing the established order: Should adrenal venous sampling precede cross-sectional imaging in the evaluation of primary aldosteronism? *J Surg Oncol* 2015; 112: 144-48.

20. Pedersen M, Karlsen MA, Ankjærgaard K L, *et al.* Primary hyperaldosteronism diagnosed with adrenal vein sampling. Characteristics and follow-up after adrenalectomy in a Danish study. *Scand J Clin Lab Invest* 2016; 76: 45-50.

21. Campbell RA, Young DS, Shaver CN, *et al.* Influence of Adrenal Venous Sampling on Management in Patients with Primary Aldosteronism Independent of Lateralization on Cross-Sectional Imaging. *J Am Coll Surg* 2019; 229: 116-24.

22. Sam D, Kline GA, So B, *et al.* Discordance Between Imaging and Adrenal Vein Sampling in Primary Aldosteronism Irrespective of Interpretation Criteria. *J Clin Endocrinol Metab* 2019; 104: 1900-06.

23. Nandra G, Duxbury O, Patel P, *et al.* Technical and Interpretive Pitfalls in Adrenal Imaging; 2020.

24. 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;58:882-93.

25. Li S, Ren Y, Zhu Y, *et al.* THE POTENTIAL CLINICAL APPLICATION OF A LOWER BILATERAL ADRENAL LIMB WIDTH RATIO (L/RW) IN PATIENTS WITH BILATERAL PRIMARY HYPERALDOSTERONISM. *Endocr Pract* 2019; 25: 830-35.

26. Aono D, Kometani M, Karashima S, *et al.* Primary aldosteronism subtype discordance between computed tomography and adrenal venous sampling. *Hypertens Res* 2019; 42: 1942-50.

27. Umakoshi H, Tsuiki M, Takeda Y, *et al.* Significance of Computed Tomography and Serum Potassium in Predicting Subtype Diagnosis of Primary Aldosteronism. *J Clin Endocrinol Metab* 2018; 103: 900-08.

28. Nanba AT, Nanba K, Byrd JB, *et al.* Discordance between imaging and immunohistochemistry in unilateral primary aldosteronism. *Clin Endocrinol* 2017; 87: 665-72.

29. Kamemura K, Wada N, Ichijo T, *et al.* Significance of adrenal computed tomography in predicting laterality and indicating adrenal vein sampling in primary aldosteronism. *J Hum Hypertens* 2017; 31: 195-99.

30. Zhu LM, Zhang Y, Zhang H, *et al.* Comparison between adrenal venous sampling and computed tomography in the diagnosis of primary aldosteronism and in the guidance of adrenalectomy. *Medicine* 2016; 95: e4986.

31. Kocjan T, Janez A, Stankovic M, *et al.* A NEW CLINICAL PREDICTION CRITERION ACCURATELY DETERMINES A SUBSET OF PATIENTS WITH BILATERAL PRIMARY ALDOSTERONISM BEFORE ADRENAL VENOUS SAMPLING. *Endocr Pract* 2016; 22: 587-94.

32. Sze WC, Soh LM, Lau JH, *et al.* Diagnosing unilateral primary aldosteronism - comparison of a clinical prediction score, computed tomography and adrenal venous sampling. *Clin Endocrinol* 2014; 81: 25-30.

33. Riester A, Fischer E, Degenhart C,et.al. Age below 40 or a recently proposed clinical prediction score cannot bypass adrenal venous sampling in primary aldosteronism. *J Clin Endocrinol Metab*

2014;99:E1035-1039.

34. Burton TJ, Mackenzie IS, Balan K, *et al.* Evaluation of the sensitivity and specificity of (11)C-metomidate positron emission tomography (PET)-CT for lateralizing aldosterone secretion by Conn's adenomas. *J Clin Endocrinol Metab* 2012; 97: 100-9.

35. Salem V, Hopkins T G, El-Gayar H, *et al.* Adrenal venous sampling as a diagnostic procedure for primary hyperaldosteronism: experience from a tertiary referral centre. *Hormones* 2012; 11: 151-59.

36. Lau JH, Sze WC, Reznek RH, *et al.* A prospective evaluation of postural stimulation testing, computed tomography and adrenal vein sampling in the differential diagnosis of primary aldosteronism. *Clin Endocrinol* 2012; 76: 182-8.

37. Küpers E M, Amar L, Raynaud A, *et al.* A clinical prediction score to diagnose unilateral primary aldosteronism. *J Clin Endocrinol Metab* 2012; 97: 3530-7.

38. Sarlon-Bartoli G, Michel N, Taieb D, *et al.* Adrenal venous sampling is crucial before an adrenalectomy whatever the adrenal-nodule size on computed tomography. *J Hypertens* 2011; 29: 1196-202.

39. Mathur A, Kemp CD, Dutta U, *et al.* Consequences of adrenal venous sampling in primary hyperaldosteronism and predictors of unilateral adrenal disease. *J Am Coll Surg* 2010; 211: 384-90.

40. Mulatero P, Bertello C, Rossato D, *et al.* Roles of clinical criteria, computed tomography scan, and adrenal vein sampling in differential diagnosis of primary aldosteronism subtypes. *J Clin Endocrinol Metab* 2008; 93: 1366-71.

41. Minami I, Yoshimoto T, Hirono Y, *et al.* Diagnostic accuracy of adrenal venous sampling in comparison with other parameters in primary aldosteronism. *Endocr J* 2008; 55: 839-46.

42. Nwariaku F E, Miller B S, Auchus R, *et al.* Primary hyperaldosteronism: effect of adrenal vein sampling on surgical outcome. *Arch Surg* 2006; 141: 497-502.

43. Young WF, Stanson AW, Thompson G B, *et al.* Role for adrenal venous sampling in primary aldosteronism. *Surgery* 2004;136:1227-35.

44. Magill SB, Raff H, Shaker JL, *et al.* Comparison of adrenal vein sampling and computed tomography in the differentiation of primary aldosteronism. *J Clin Endocrinol Metab* 2001; 86: 1066-71.

45. Oh EM, Lee KE, Yoon K, *et al.* Value of adrenal venous sampling for lesion localization in primary aldosteronism. *World J Surg* 2012; 36: 2522-7.

46. Mendiratta-Lala M, Avram A, Turcu AF, *et al.* Adrenal Imaging. *Endocrinol Metab Clin North Am* 2017;46:741-759.

47. Zeiger M A, Thompson G B, Duh Q Y, *et al.* The American Association of Clinical Endocrinologists and American Association of Endocrine Surgeons medical guidelines for the management of adrenal incidentalomas. *Endocr Pract* 2009;15 Suppl 1:1-20.

48. Laurent I, Astère M, Zheng F, *et al.* Adrenal venous sampling with or without adrenocorticotropic hormone stimulation: A meta-analysis. *J Clin Endocrinol Metab* 2019;104:1060–1068

49. Taguchi R, Yamada M, Nakajima Y, *et al.* Expression and mutations of KCNJ5 mRNA in Japanese patients with aldosterone-producing adenomas. *J Clin Endocrinol Metab* 2012;97:1311-9.

50. Wang B, Li X, Zhang X, *et al.* Prevalence and characterization of somatic mutations in Chinese aldosterone-producing adenoma patients. *Medicine* 2015;94:e708.

51. Satoh F, Morimoto R, Ono Y, *et al.* Measurement of peripheral plasma 18-oxocortisol can discriminate unilateral adenoma from bilateral diseases in patients with primary aldosteronism; *Hypertension* 2015;65:1096-102.

Figure legend

Figure 1: Flow diagram of the review process.

Figure 2: Assessment of methodological quality of included studies using the QUADAS-2 Criteria.

Stacked bars represent the proportion of studies with a high (red), or unclear (yellow) or low (green) risk of bias and applicability concerns.

QUADAS-2: Quality Assessment of Diagnostic Accuracy Studies –2 criteria

Figure 3: Forest plots of sensitivity and specificity of adrenal imaging compared with AVS.

Horizontal lines are the 95% confidence intervals (CIs)

Figure 4: Hierarchical SROC plot showing average sensitivity and specificity estimate of the study results with 95% confidence region.

The 95% prediction region represents the confidence region for a forecast of the true sensitivity (SENS) and specificity (SPEC) in a future study.

AUC: area under the curve; SROC: summary receiver-operating characteristic.

Figure S1: Forest plots of sensitivity and specificity of adrenal imaging compared with AVS in young patients (<40 years).

Figure S2: Graphical presentation of the generalized linear mixed model exploring the impact of selected variables on sensitivity and specificity of adrenal imaging.

Figure S3: Deeks' funnel plot for checking the publication bias. DOR: diagnostic odds ratio.

Author affiliations

[1]Department of Cardiology, The First Affiliated Hospital of Chengdu Medical College, Sichuan Province, People's Republic of China.

[2]Key Laboratory of Aging and Vascular Homeostasis of Sichuan Higher Education Institutes, Chengdu, Sichuan, People's Republic of China.

[3]Department of Cardiology, the First Affiliated Hospital of Xi'an Jiaotong University, Shaanxi Province, People's Republic of China.

Contributors

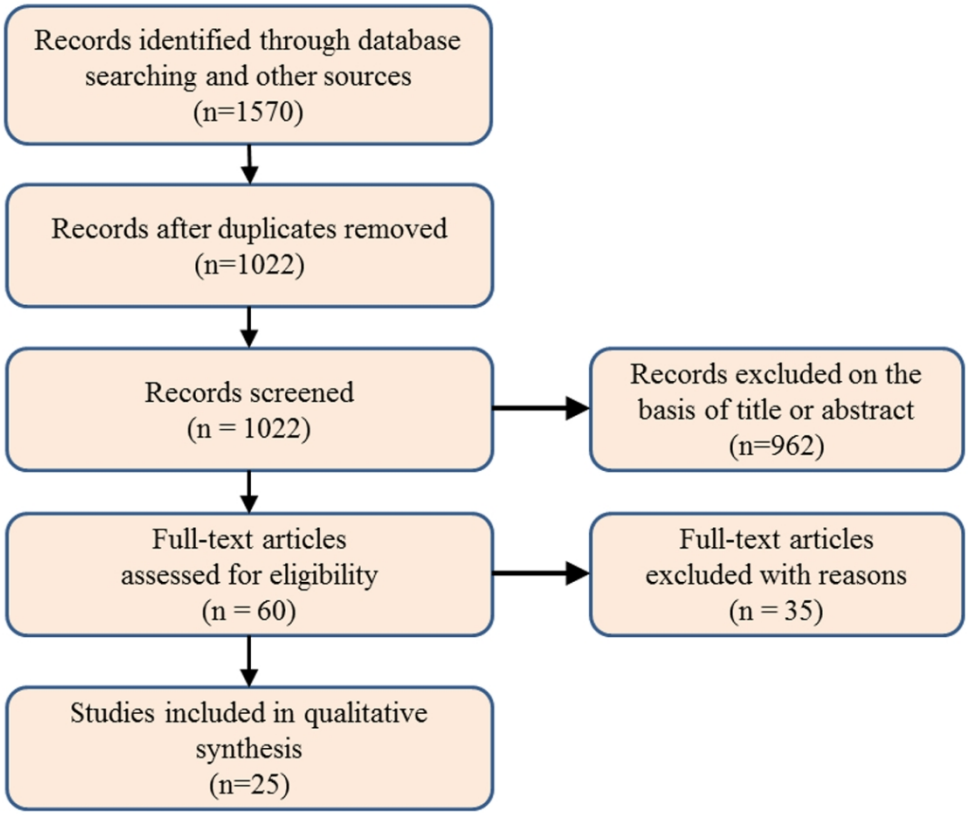
WPJ is the guarantor. All authors provided substantial contribution to conception and design of the project; drafted and revised the manuscript. ZYQ led the literature search, and completed the study selection, data extraction, and critical appraisal with WPJ. JLC and HU accept responsibility for the integrity of the data analyses. RF led the drafting of all sections of the article in consultation with all of the coauthors. WD, CSC and ZP provided substantial contributions to the background, critical appraisal of prior studies and interpretation of meta-analysis findings. ZP provided substantial contribution to the methods section

Funding: This systematic review and meta-analysis are funded by the National Natural Sciences Foundation of China (NSFC)（Grant NO. 81970262）.

Competing Interests: None declared

Patient and Public Involvement: No patient involvement.

Ethics Approval: Not required.

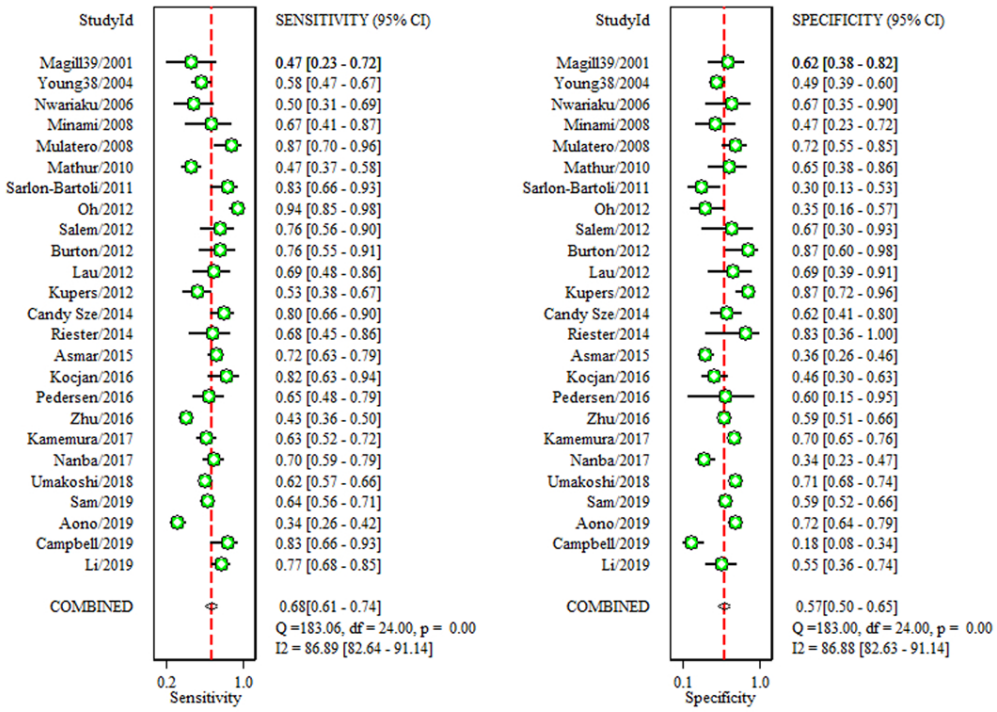


Flow diagram of the review process.

101x90mm (300 x 300 DPI)

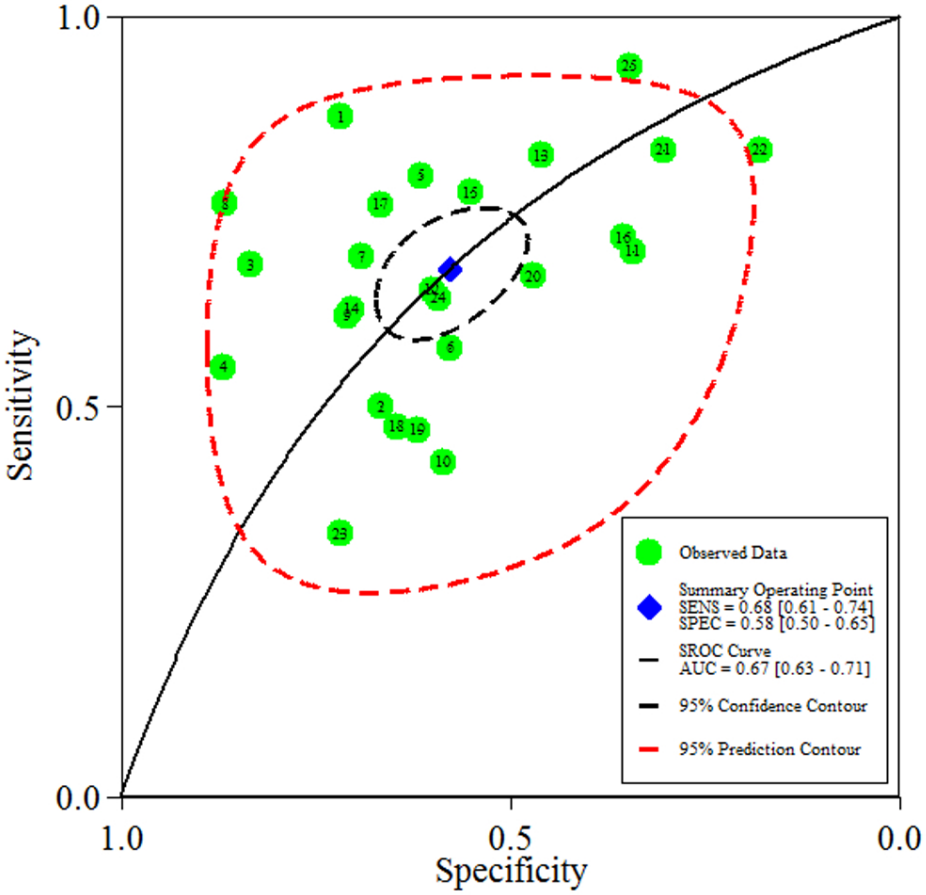
Assessment of methodological quality of included studies using the QUADAS-2 Criteria./Stacked bars represent the proportion of studies with a high (red), or unclear (yellow) or low (green) risk of bias and applicability concerns. QUADAS-2: Quality Assessment of Diagnostic Accuracy Studies –2 criteria

90x51mm (300 x 300 DPI)



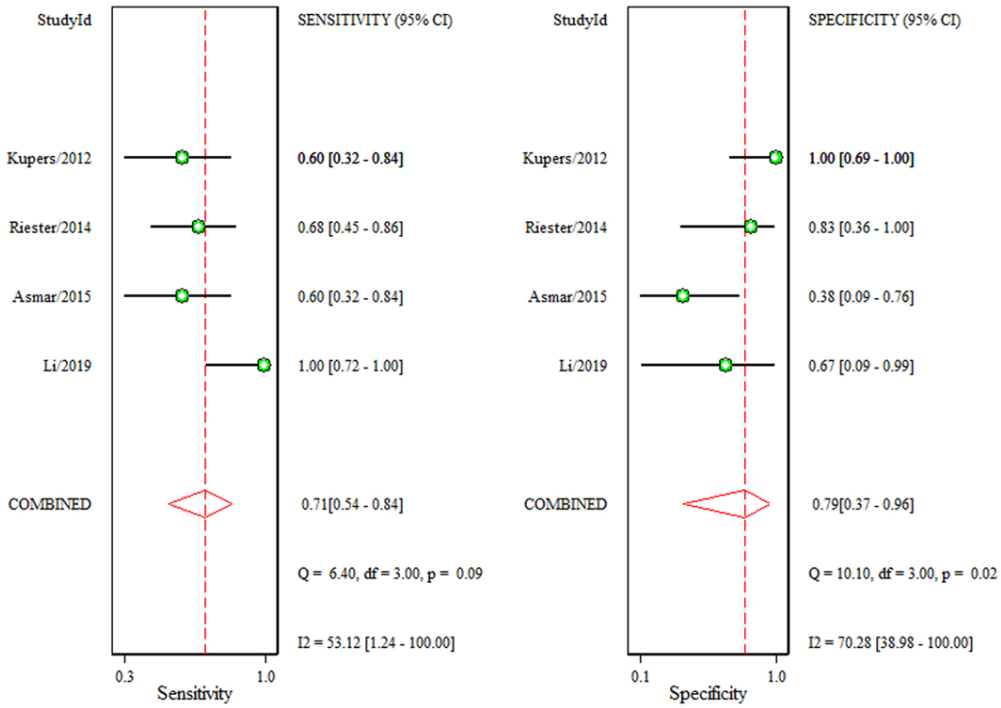
Forest plots of sensitivity and specificity of adrenal imaging compared with AVS. /Horizontal lines are the 95% confidence intervals (CIs)

90x69mm (300 x 300 DPI)

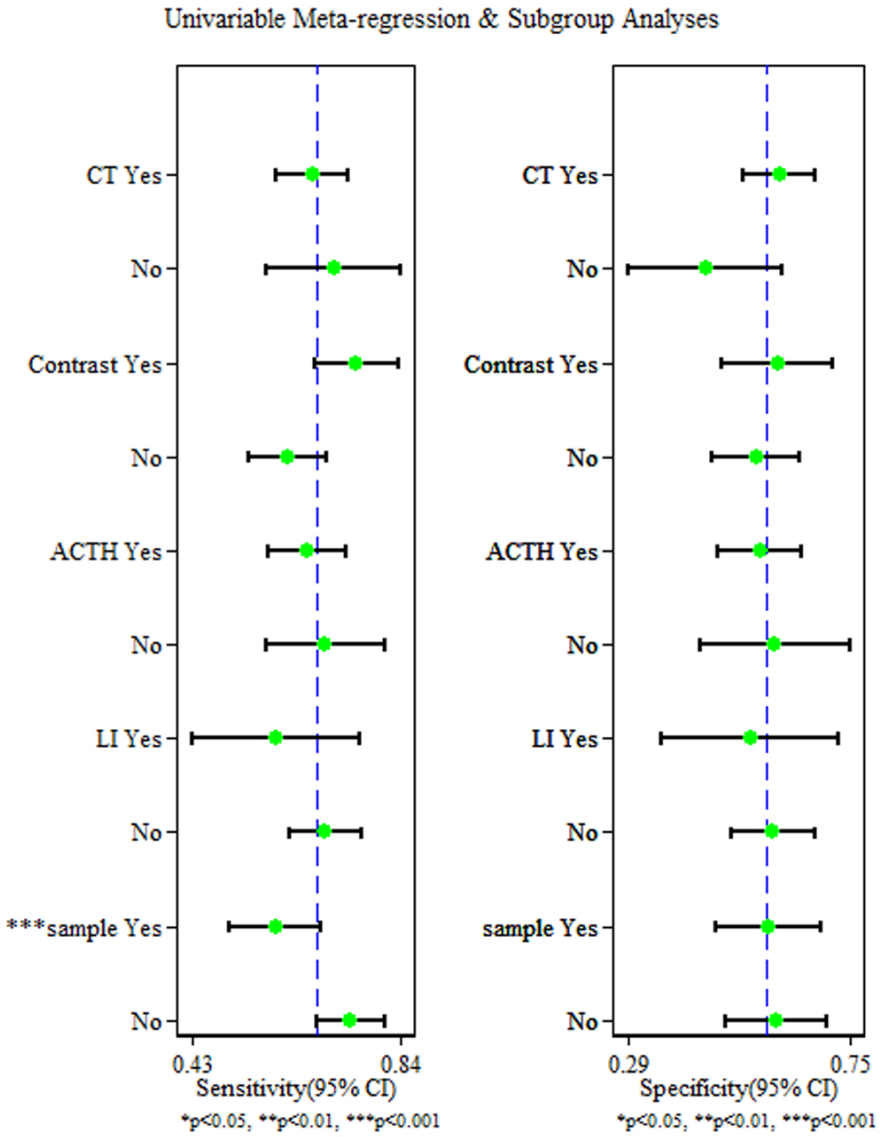


Hierarchical SROC plot showing average sensitivity and specificity estimate of the study results with 95% confidence region. /The 95% prediction region represents the confidence region for a forecast of the true sensitivity (SENS) and specificity (SPEC) in a future study. AUC: area under the curve; SROC: summary receiver-operating characteristic.

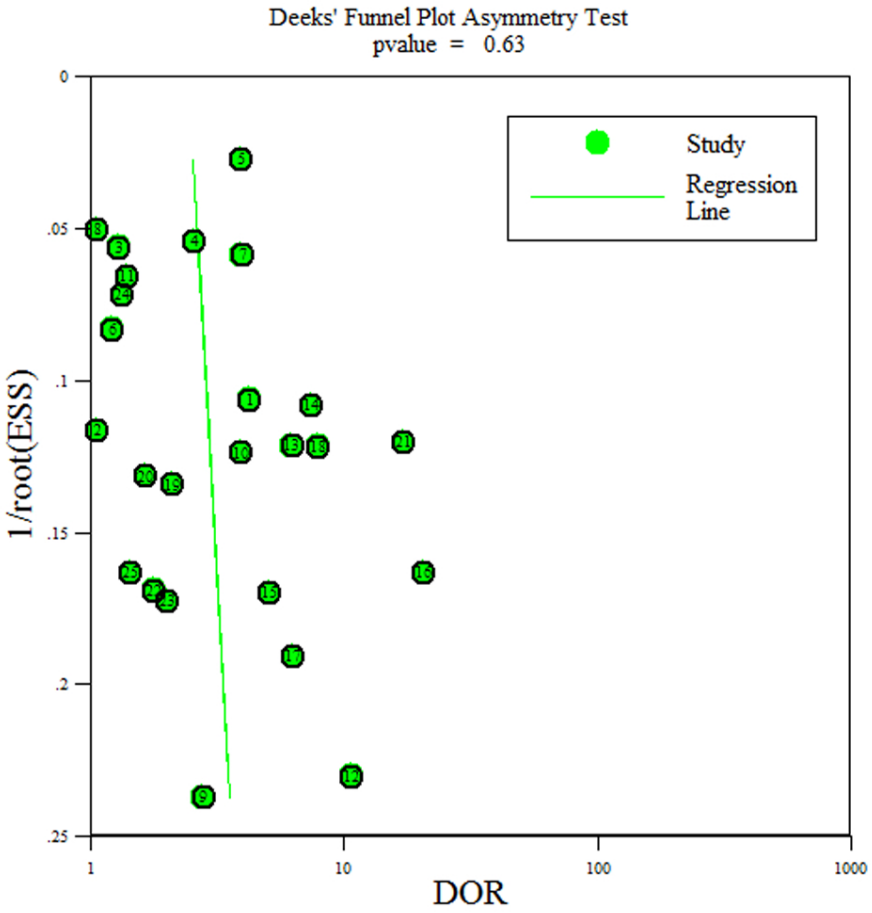
90x90mm (300 x 300 DPI)



90x69mm (300 x 300 DPI)



90x114mm (300 x 300 DPI)



90x90mm (300 x 300 DPI)

Search strategies for PubMed

#1. ((aldosteronism) OR (primary hyperaldosteronism) OR (primary aldosteronism) OR (conn syndrome) OR (conn) OR (aldosterone-producing adenoma) OR (APA) OR (idiopathic hyperaldosteronism) OR (IHA) OR (primary adrenal hyperplasia) OR (PAH) OR (bilateral adrenal hyperplasia) OR (BAH))

#2. ((adrenal venous sampling) OR (AVS) OR (adrenal vein sampling) OR (adrenal vein) OR (venous sampling) OR (vein sampling) OR (adrenal venous))

#3. ("2000/01/01"[Date - Entry] : "2020/01/01"[Date - Entry])

#4. #1 AND #2 AND #3

Table S1: The quality assessment for each study -QUADAS-2 results

Study	Risk of bias				Applicability Concerns		
	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Li et al	☺	☺	☺	☺	☺	☺	☺
Campbell et al	☺	☺	☺	☺	☺	☺	☺
Aono et al	☺	☺	☺	☺	☺	☺	☺
Sam et al et al	☺	☺	☺	☺	☺	☺	☺
Umakoshi et al	☺	☺	☺	?	☺	☺	☺
Nanba et al	☺	☺	☺	?	☺	☺	☺
Kamemura et al	☺	☺	☺	?	☺	☺	☺
Zhu et al	☺	☺	☹	☺	☺	☺	☺
Pedersen et al	☺	☺	?	☹	☺	☺	?
Kocjan et al	☺	☺	☺	?	☺	☺	☺
Asmar et al	☺	☺	☺	☺	☺	☺	☺
Riester et al	☺	☺	☺	☺	☺	☺	☺
Candy Sze et al	☺	☺	☺	☺	☺	☺	☺
Kűpers et al	☺	☺	☺	☺	☺	☺	☺
Lau et al	☺	☺	☺	?	☺	☺	☺
Burton et al	☺	☺	☹	☺	☺	☺	☹
Salem et al	☺	☺	?	☺	☺	☺	☹
Oh et al	☺	☺	☺	☺	☺	☺	☺
Sarlon-Bartoli et al	☺	☺	☺	☺	☺	☺	☺
Mathur et al	☺	☺	☺	☺	☺	☺	☺
Mulatero et al	☺	☺	☺	?	☺	☺	☺
Minami et al	☺	☺	?	☺	☺	☺	☺
Nwariaku et al	☺	☺	☺	?	☺	☺	☺
Young et al	☺	☺	☺	☹	☺	☺	☺
Magill et al	☺	☺	☺	☺	☺	☺	☺

☺= low risk, ? = unclear risk, ☹=high risk

□ Table S3: The quality assessment for each study by 2 reviewers.

Study	Risk of bias				Applicability Concerns		
	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Li et al	☺/☺	☺/☺	☺/☺	☺/☺	☺/☺	☺/☺	☺/☺
Campbell et al	☺/☺	☺/☺	☺/☺	☺/☺	☺/☺	☺/☺	☺/☺
Aono et al	☺/☺	☺/☺	☺/☺	☺/☺	☺/☺	☺/☺	☺/☺
Sam et al	☺/☺	☺/☺	☺/☺	☺/☺	☺/☺	☺/☺	☺/☺
Umakoshi et al	☺/☺	☺/☺	☺/☺	?/?	☺/☺	☺/☺	☺/☺
Nanba et al	☺/☺	☺/☺	☺/☺	?/?	☺/☺	☺/☺	☺/☺
Kamemura et al	☺/☺	☺/☺	☺/☺	?/?	☺/☺	☺/☺	☺/☺
Zhu et al	☺/☺	☺/☺	☹/☹	☺/☺	☺/☺	☺/☺	☺/☺
Pedersen et al	☺/☺	☺/☺	?/?	?/☹	☺/☺	☺/☺	?/?
Kocjan et al	☺/☺	☺/☺	☺/☺	?/?	☺/☺	☺/☺	☺/☺
Asmar et al	☺/☺	☺/☺	☺/☺	☺/☺	☺/☺	☺/☺	☺/☺
Riester et al	☺/☺	☺/☺	☺/☺	☺/☺	☺/☺	☺/☺	☺/☺
Candy Sze et al	☺/☺	☺/☺	☺/☺	☺/☺	☺/☺	☺/☺	☺/☺
Kűpers et al	☺/☺	☺/☺	☺/☺	☺/☺	☺/☺	☺/☺	☺/☺
Lau et al	☺/☺	☺/☺	☺/☺	?/?	☺/☺	☺/☺	☺/☺
Burton et al	☺/☺	☺/☺	☹/?	☺/☺	☺/☺	☺/☺	☹/☹
Salem et al	☺/☺	☺/☺	?/☹	☺/☺	☺/☺	☺/☺	?/☹
Oh et al	☺/☺	☺/☺	☺/☺	☺/☺	☺/☺	☺/☺	☺/☺
Sarlon-Bartoli et al	☺/☺	☺/☺	☺/☺	☺/☺	☺/☺	☺/☺	☺/☺
Mathur et al	☺/☺	☺/☺	☺/☺	☺/☺	☺/☺	☺/☺	☺/☺
Mulatero et al	☺/☺	☺/☺	☺/☺	?/?	☺/☺	☺/☺	☺/☺
Minami et al	☺/☺	☺/☺	?/?	☺/☺	☺/☺	☺/☺	☺/☺
Nwariaku et al	☺/☺	☺/☺	☺/☺	?/?	☺/☺	☺/☺	☺/☺
Young et al	☺/☺	☺/☺	☺/☺	☹/☹	☺/☺	☺/☺	☺/☺
Magill et al	☺/☺	☺/☺	☺/☺	☺/☺	☺/☺	☺/☺	☺/☺

☺= low risk, ? = unclear risk, ☹=high risk.

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2-3
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3-4
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	4-5
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	none
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	5
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	5
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	6
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	6
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	6
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	5
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	7
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	7

PRISMA 2009 Checklist

Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	7

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	7
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	7
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	7-8
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	8-9
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	10
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	9
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	9-10
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	10
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	10-11
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	11-13
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	13-14
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	14
FUNDING			

PRISMA 2009 Checklist

Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	18

BMJ Open

Diagnostic Accuracy of Adrenal Imaging for Subtype Diagnosis in Primary Aldosteronism: Systematic Review and Meta-Analysis

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-038489.R2
Article Type:	Original research
Date Submitted by the Author:	19-Oct-2020
Complete List of Authors:	Zhou, Yaqiong; The First Affiliated Hospital of Chengdu Medical College, Cardiology; Key Laboratory of Aging and Vascular Homeostasis of Sichuan Higher Education Institutes Wang, Dan; The First Affiliated Hospital of Chengdu Medical College, Cardiology; Key Laboratory of Aging and Vascular Homeostasis of Sichuan Higher Education Institutes Jiang, Licheng; The First Affiliated Hospital of Chengdu Medical College, Cardiology; Key Laboratory of Aging and Vascular Homeostasis of Sichuan Higher Education Institutes Ran, Fei; The First Affiliated Hospital of Chengdu Medical College, Cardiology; Key Laboratory of Aging and Vascular Homeostasis of Sichuan Higher Education Institutes Chen, Sichao ; The First Affiliated Hospital of Chengdu Medical College, Cardiology Zhou, Peng; The First Affiliated Hospital of Chengdu Medical College, Cardiology; Key Laboratory of Aging and Vascular Homeostasis of Sichuan Higher Education Institutes Wang, Peijian; The First Affiliated Hospital of Chengdu Medical College, Cardiology; Key Laboratory of Aging and Vascular Homeostasis of Sichuan Higher Education Institutes
Primary Subject Heading:	Diabetes and endocrinology
Secondary Subject Heading:	Cardiovascular medicine, Diabetes and endocrinology
Keywords:	Hypertension < CARDIOLOGY, Endocrine tumours < DIABETES & ENDOCRINOLOGY, Cardiology < INTERNAL MEDICINE

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Diagnostic Accuracy of Adrenal Imaging for Subtype Diagnosis in Primary Aldosteronism: Systematic Review and Meta-Analysis

Yaqiong Zhou[12] , Dan Wang[1], Licheng Jiang[12], Fei Ran[1], Sichao Chen[1], Peng Zhou[12*], Peijian Wang [12*]

From the 1 Department of Cardiology, The First Affiliated Hospital of Chengdu Medical College, Sichuan Province, People's Republic of China.

From the 2 Key Laboratory of Aging and Vascular Homeostasis of Sichuan Higher Education Institutes, Chengdu, Sichuan, People's Republic of China.

*Corresponding author: Prof. Peijian Wang

Department of Cardiology, the First Affiliated Hospital of Chengdu Medical College, 299 baoguang avenues, Chengdu 610500, Sichuan Province, China. E-mail: wpjmed@aliyun.com

*Corresponding author: Prof. Peng Zhou

Department of Cardiology, the First Affiliated Hospital of Chengdu Medical College, 299 baoguang avenues, Chengdu 610500, Sichuan Province, China. E-mail: ap216@163.com.

Email: Yaqiong Zhou.1273868593@qq.com- Dan Wang.2292029981@qq.com- Licheng Jiang. 65839350@qq.com- Hameed Ullah. 807287144@qq.com- Fei Ran 1126852537@qq.com- Sichao Chen. 1350821872@qq.com-Peng Zhou. ap216@163.com-Peijian Wang. wpjmed@aliyun.com

Funding: This systematic review and meta-analysis are funded by the National Natural Sciences Foundation of China (NSFC)（Grant NO. 81970262）.

Abstract

Objectives: Accurate subtype classification in primary aldosteronism (PA) is critical in assessing the optimal treatment options. This study aimed to evaluate the diagnostic accuracy of adrenal imaging for unilateral PA classification.

Methods: Systematic searches of PubMed, EMBASE, and the Cochrane databases were performed from January 1, 2000, to February 1, 2020, for all studies that used computed tomography (CT) or magnetic resonance imaging (MRI) in determining unilateral PA and validated the results against invasive adrenal vein sampling (AVS). Summary diagnostic accuracies were assessed using a bivariate random-effects model. Subgroup analyses, meta-regression and sensitivity analysis were performed to explore the possible sources of heterogeneity.

Result: A total of 25 studies, involving a total of 4669 subjects, were identified. The overall analysis revealed a pooled sensitivity of 68% (95% confidence interval [CI]: 61 to 74) and specificity of 57% (95% CI: 50 to 65) for CT/MRI in identifying unilateral PA. Sensitivity was higher in the contrast-enhanced (CT) group versus the traditional CT group [77% (95% CI: 66 to 85) vs. 58% (95% CI: 50 to 66)]. Subgroup analysis stratified by screening test for PA showed that the sensitivity of the aldosterone-to-renin ratio (ARR) group was higher than that of the non-ARR group [78% (95% CI: 69 to 84) vs. 66% (95% CI: 58 to 72)]. The diagnostic accuracy of PA patients aged ≤40 years was reported in 4 studies, and the overall sensitivity was 71%, with 79% specificity. Meta-regression revealed a significant impact of sample size on sensitivity and of age and study quality on specificity.

Conclusion: CT/MRI is not a reliable alternative to invasive AVS without excellent sensitivity or specificity for correctly identifying unilateral PA. Even in young patients (≤40 years), 21% of patients would have undergone unnecessary adrenalectomy based on imaging results alone.

Keywords: adrenal vein sampling, computed tomography, magnetic resonance imaging, primary aldosteronism, subtype

Strengths and limitations of this study

- This study is the first meta-analysis to synthesize the evidence regarding the

diagnostic value of adrenal imaging for PA classification and demonstrated that CT/MRI is not a reliable alternative to invasive AVS even in young patients (≤40 years).

- The main methodological limitations of this systematic review and meta-analysis are the exclusion of unpublished high-quality trials and foreign-language publications.
- Another potential limitation is that we encountered different AVS methods and large variation in the lateralization criteria, which might also have affected the results for diagnostic accuracy.

Introduction

Primary aldosteronism (PA) is one of the most common causes of endocrine hypertension, with a prevalence of approximately 20% in patients with resistant hypertension, 10% in those with severe hypertension, and 6% in those with uncomplicated hypertension[1]. Accumulating clinical and epidemiological evidence suggests that PA amplifies cardiovascular and cerebrovascular complications beyond essential hypertension prior to treatment, even after controlling elevated blood pressure[2, 3]. However, patients with unilateral resected PA have slightly better risk profiles than matched essential hypertensive patients. Patients with bilateral PA whose plasma renin activity is not suppressed have the same risk profiles as essential hypertensive patients; those whose renin activity remains suppressed have 4-fold higher risk profiles than controls, and titration of mineralocorticoid receptor antagonist (MRA) therapy to raise renin might reduce this excess risk[4, 5]. Accordingly, early diagnosis and specific treatment of affected patients are key steps for the reversal of target-organ damage and prevention of cardiovascular and cerebrovascular events.

Selection of the most appropriate therapeutic strategy for patients with PA requires a distinction between unilateral and bilateral forms of PA. The former requires a unilateral adrenalectomy, mainly entailing aldosterone-producing adenoma (APA) and, less commonly, unilateral adrenal hyperplasia. In contrast, the latter, also known as idiopathic hyperaldosteronism, is optimally treated with target medical

therapy[3]. Regarding the differentiation of the unilateral and bilateral subtypes, all current clinical practice guidelines recommend adrenal vein sampling (AVS) as the standard procedure for subtype diagnosis [6, 7]. However, several shortcomings of AVS have been reported, such as its technical challenges, invasive nature, poorly standardized procedures, high cost, and lack of availability. Thus, it is urgent to explore alternative diagnostic methods without sacrificing accuracy.

Adrenal imaging with computed tomography (CT) or magnetic resonance imaging (MRI) is recommended as the first step for subtype classification given the ease of performance and relative accessibility[8]. By now, numerous studies have evaluated the diagnostic performance of CT/MRI in subtype diagnosis of PA, but the results have been inconsistent. Moreover, all these studies were limited by small sample sizes in a single centre, which limited the credibility of the results. In this context, systematic reviews and meta-analyses have the benefit of increasing the sample size, generating more precise results, which have been widely applied in clinical studies[9]. In 2009, one systematic review reported that CT/MR-based diagnoses were discordant with AVS results in 37.8% of PA patients[10]. However, the conclusions may not be reliable because of the potential for bias and concerns regarding the comparability of the included studies. Moreover, several additional studies were reported after this systematic review. We thus performed a comprehensive meta-analysis of all the available studies to evaluate the diagnostic value of adrenal imaging (CT/MRI) for subtype classification of PA.

Method

Search Strategy

The study followed the guidelines specified in the Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P)[11]. We searched the PUBMED, EMBASE, and Cochrane Library databases from January 1, 2000, to February 1, 2020, using the following terms in combination, as both MeSH or Emtree terms and text words: "primary aldosteronism", "adrenal vein sampling" and "hyperaldosteronism". The electronic search strategy for PubMed is shown in supplementary Table S1. To reflect modern practice, we decided to limit the

publication date to after January 1, 2000. We searched articles published in English, and the references of relevant studies were also searched. All studies were carefully examined to exclude overlapping or potential duplicate data.

Eligibility Criteria

We included a study if: 1) it used CT or MRI as a diagnostic test for PA subtyping; 2) it used AVS as the standard of reference. Successful AVS should be determined by calculating the selectivity index (SI), defined as the adrenal/peripheral vein cortisol ratio. Unilateral PA should be determined by calculating the lateralization index (LI), defined as the aldosterone/cortisol ratio between the dominant and the non-dominant adrenal gland; and 3) absolute numbers of true-positive, true-negative, false-positive, and false-negative results were provided or could be derived. Identified studies had to be independent. In the case of multiple reports on the same population or subpopulation, the most recent or comprehensive information was used.

Data Extraction and Quality Assessment

Data extraction from the eligible studies was performed by 2 independent investigators (Z.Y.Q and W.P.J) using a standardized data extraction form. The form included the following characteristics of each trial: first author's name and year of publication; study population characteristics, including sample size, geographical location, mean age and sex; diagnostic criteria characteristics, including screening test and confirmatory test for PA; AVS characteristics, including with/without adrenocorticotropic hormone (ACTH) stimulation, SI and LI; diagnostic test characteristics, including imaging methodology and whether contrast was administered. Differences between reviewers were resolved by discussion and consensus when necessary.

The methodological quality of the identified studies was assessed by 2 independent reviewers (Z.Y.Q and W.P.J) using the modified Quality Assessment of Diagnostic Accuracy Studies –2 (QUADAS-2) criteria. If a study was judged as “low” on all domains relating to bias or applicability, then it was judged to be a high-quality study. If a study was judged to be "high" and/or "unclear" in more than 1

domain, then it was judged as a low-quality study. If a study was judged to be "unclear" in 1 domain, it was considered an unclear-quality study[12]. Discrepancies were resolved by discussion and consensus.

Statistical Analysis

Measures of diagnostic accuracy are reported as point estimates with 95% confidence intervals (CIs). Sensitivity, specificity, the positive likelihood ratio (+LR) and the negative likelihood ratio (-LR) were modelled based on the true-positive, true-negative, false-positive, and false-negative rates for each trial[13]. The ratio of +LR to -LR was combined in a single global accuracy measure, the diagnostic odds ratio (DOR). Summary sensitivity, specificity, +LRs, -LRs and DORs were assessed using a bivariate random-effects model. The approach assumes bivariate normal distributions for the logit transformations of sensitivity and specificity from the individual studies. These bivariate models can be analysed using linear mixed model techniques that are now widely available in statistical packages, such as STATA gllamm[14, 15]. A hierarchical summary receiver operating characteristic (ROC) curve analysis was performed, yielding point estimates for each trial and pooled characteristics, including the 95% prediction region and the 95% confidence region.

Sources of statistical heterogeneity were explored by subgroup analyses, sensitivity analysis and meta-regression analysis[16], which involved the I^2 statistic; the following interpretation was applied for I^2: <50%=low heterogeneity, 50% to 75%=moderate heterogeneity, and >75%=high heterogeneity.

Several studies demonstrated that MRI has poorer resolution and slower acquisition than CT, with a risk of respiratory artefacts and that MRI is inferior to adrenal CT in PA subtype evaluation[17-20]. Contrast materials can improve the visibility of adrenal structures imaged by CT and MRI scans and might have a positive effect on diagnosis accuracy[21]. Thus, imaging methods and contrast materials were thought to be confounders for subgroup analyses. Moreover, a large sample size may represent experienced interventional radiologists and support the credibility of the included studies. Thus, a small sample size was thought of as another confounder for subgroup analyses. The different diagnostic criteria for PA, the AVS procedure

(with or without ACTH stimulation), different cut-offs for the LI criteria, and methodological quality might also affect the results for diagnosis accuracy[20], Therefore, we also performed subgroup analyses stratified by these parameters. Thus, subgroup analyses were performed by the following factors: imaging methodology (CT or CT/MRI), contrast use, AVS procedure (with or without ACTH stimulation), cut-off value for the LI (2 or 4), diagnostic criteria for PA, sample size (divided by 100 subjects), and methodological quality (high-quality, low-quality and unclear-quality).

Potential publication bias was examined using the Deeks test[22]. The Cohen κ test was employed to assess the inter-rater reliability between 2 observers for quality assessment. If there was not agreement, a third reviewer was involved to resolve disagreements, and final decisions were determined by consensus. Statistical analyses were performed using Stata version 13.0 (StataCorp LP, TX, USA) and Review Manager version 5.3.

Results

Study Selection

After removal of 548 duplicates, the systematic review generated 1022 references that were screened according to titles and abstracts for possible inclusion. Among them, 962 studies were excluded for the following reasons: 489 studies were not relevant; 280 studies were reviews or practice guidelines; 92 studies did not include humans; and 101 studies were case reports/letters. After screening, 60 studies were identified as being potentially eligible, and their full texts were retrieved for detailed evaluation. A total of 35 studies were excluded for the following reasons: data to compute diagnostic accuracy were not provided or could not be derived (25 papers), reporting on the same population (4 papers) and no comparison of CT/MRI and AVS results in individual patients (6 papers). Finally, 25 articles were deemed eligible and analysed in our meta-analysis[17-20, 23-43] (Figure 1).

Study Characteristics

Overall, a total of 4669 patients (mean age of 51 years; 54% male) from 25 articles were included. The sample sizes of the identified studies ranged from 35 to

1591, with the largest study recruiting over 1000 participants[41]. Five studies including 724 participants underwent cross-sectional imaging either CT or MRI, and the remaining 20 studies including 3945 patients only used CT scans (8 studies administered contrast material). Seventeen studies performed AVS with ACTH stimulation, 7 studies performed it without ACTH stimulation, and the remaining 1 study provided the above two methods. The aldosterone-to-renin ratio (ARR) was used as a screening tool for PA in 21 of the included articles, and an ARR >20 was commonly used as the threshold for a positive PA screening. The remaining 4 studies did not use the ARR as a screening test for PA. In 12 articles, a salt-loading test was performed to confirm the diagnosis of PA. Eight studies used additional options, including the fludrocortisone suppression test, captopril challenge test, upright-furosemide loading test and postural stimulation test, as a confirmatory test for PA. The diagnosis of PA was not confirmed in the remaining 5 studies by one of the confirmatory tests. The 2016 Endocrine Society Guideline recommends more strict criteria for the LI (2.0 or greater under unstimulated conditions and/or 4 for ACTH stimulation) and SI (2.0 or greater under unstimulated conditions and/or 3 for ACTH stimulation)[8]. In the meta-analysis, 1 included study used less permissive criteria for the LI [42], and 6 included studies used less permissive criteria for the SI[20, 27-30, 42]. The threshold of the SI was not accessible in 4 studies[18, 23, 33, 34], and it was not accessible for the LI in 2 studies [1, 26]. Further details about the eligible and analysed studies are shown in Table 1and Table S2.

Study,Year	Location	Male (%)	Age (years)	Sample	Imaging methodology	Contrast used	AVS characteristics	screening test for PA (ng/ml)	confirmatory test for PA
Li et al. 2019	China	61(50)	48.5	122	CT	yes	without ACTH SI≥2 LI≥2	ARR	Captopril test or salt-loading test
Campbell et al.2019	USA	45(61)	55.6	74	CT or MRI	no	with ACTH, SI≥5, LI≥4	not ARR	no
Aono et al. 2019	Japan	NA	56	317	CT	no	with ACTH, SI>2, LI≥2	ARR>20	Captopril test, furosemide upright test or salt-loading test
Sam et al.2019	Canada	201(59)	52.1	342	CT or MRI	yes	with ACTH,SI>2, LI≥4	ARR	no
Umakoshi et al.2018	Japan	762(47.9)	53	1591	CT	yes	with ACTH,SI>5,LI>4	ARR	Captopril test, furosemide upright test or salt-loading test
Nanba et al.2017	USA	87(59)	54	147	CT	no	with ACTH, SI≥5, LI≥4	ARR	Salt-loading test
Kamemura et al.2017	Japan	177(45)	54	393	CT	no	with ACTH,SI>5,LI>4	ARR	Captopril test, furosemide upright test or Salt-loading test
Zhu et al. 2016	China	NA	46	394	CT	no	without ACTH, SI≥3, LI≥2	ARR	Fludrocortisone test or salt-loading test
Pedersen et al.2016	Denmark	24(54)	51	45	CT or MRI	no	without ACTH. SI /LI NA	ARR	Fludrocortisone test or postural test
Kocjan et al. 2016	Slovenia	46(69)	56	67	CT	no	with ACTH,SI>5,LI>4	ARR	Salt-loading test
Asmar et al. 2015	USA	148(63)	55	235	CT or MRI	no	with ACTH, SI≥5, LI≥4	ARR	no
Riester et al. 2014	Germany	NA	35	28	CT or MRI	no	without ACTH, SI≥2, LI≥4	ARR	Salt-loading test
Candy Sze et al. 2014	UK	42(56)	50.5	75	CT	yes	with ACTH, SI≥5, LI≥4	not ARR	salt-loading test
Küpers et al.2012	France	53(61)	46	87	CT	no	with ACTH, SI≥2, LI≥4	ARR	Salt-loading test
Lau et al. 2012	UK	24(64)	51.8	39	CT	no	with ACTH, SI≥5, LI≥4	not ARR	salt-loading test
Burton et al. 2012	UK	NA	50.9	40	CT	Yes	without ACTH,LI≥4	not ARR	no
Salem et al.2012	UK	16(44)	44.7	38	CT	no	without ACTH, LI≥4	ARR	Salt-loading test
Oh et al. 2012	Korea	45(52)	50.7	86	CT	yes	with ACTH, SI≥3, LI>4	ARR	Salt-loading test
Sarlon-Bartoli et al.2011	France	NA	52	58	CT	yes	without ACTH, SI≥1, LI>2	ARR	no
Mathur et al. 2010	USA	63(55)	50.6	114	CT	no	with ACTH, SI>2, LI≥4	ARR	Captopril test, posture test or salt-loading test
Mulatero et al. 2008	Italy	NA	52.4	70	CT	yes	65 with ACTH; 5 without ACTH ,SI>2, LI≥4	ARR	Salt-loading test
Minami et al.2008	Japan	12(34)	54	35	CT	no	with ACTH SI /LI NA	ARR	Salt-loading test
Nwariaku et al. 2006	USA	27(67)	51	40	CT	yes	with ACTH, SI≥3, LI>4	ARR	Captopril test or posture test
Young et al. 2004	USA	163(84)	53	194	CT	no	with ACTH, SI>5, LI>4	ARR	Salt-loading test
Magill et al.2001	USA	27(71)	51	38	CT	no	with ACTH, SI≥3, LI≥4	ARR	Salt-loading test

Table 1: Study Characteristics

ACTH: adrenocorticotropic hormone; CT: computed tomography; MRI: magnetic resonance imaging; NA: not available; SI: selectivity index; LI: lateralization index; ARR: aldosterone-to-renin ratio

Quality assessment

Overall, the identified studies were of excellent quality in terms of applicability and risk of bias. Out of 175 QUADAS-2 items (25 articles×7 items), the 172 (98%)

were agreed on by the 2 reviewers, with an inter-rater agreement of κ=0.9. Figure 2 summarizes the QUADAS-2 assessment, and supplementary Table S3 displays each of the 25 individual QUADAS-2 evaluations.

The risk of bias from the reference standard was high in 3 studies[28, 38, 42] , and it was unclear in 5 studies[18, 20, 25, 26, 30] because it was not clear whether the reference standard was interpreted without knowledge of the adrenal imaging results or whether the cut-off values of the SI and LI correctly classified the target condition. The risk of bias regarding flow and timing was unclear in 6 studies[20, 24, 25, 30, 39, 41] because the time interval between the index test and the reference standard was unclear; it was high in 1 study [23] because not all patients had the same reference standard (Figure 2). Finally, 13 studies were considered to be high-quality studies[17, 19, 27, 29, 31-38, 43], 7 were considered low-quality studies[20, 23, 25, 28, 30, 38, 42] and 5 were unclear-quality studies [18, 24, 26, 39, 41].

Overall analysis

Using the bivariate model, statistical heterogeneity was found for sensitivity (I^2 =86.9%; *P*= 0.001), specificity (I^2 =86.9%; *P*=0.00), the positive LR (I^2 = 76.3%; *P*= 0.00), the negative LR (I^2=79.2%; *P*=0.00) and DOR (I^2 =100.0%; *P*=0.00), indicating high between-study heterogeneity for all pooled measures, which might compromise the credibility of the study.

In the overall analysis, the pooled sensitivity, specificity, positive LR, negative LR and DOR for adrenal imaging were 68% (95% CI: 61 to 74), 57% (95% CI: 50 to 65), 1.6 (95% CI: 1.4 to 1.9), 0.56 (95% CI: 0.47 to 0.68), and 3 (95% CI: 2 to 4), respectively (Figures 3,4).

Subgroup analyses

Subgroup analysis, stratified by the imaging methodology, found more favourable specificity (60%) for CT than CT/MRI (45%). Notably, subgroup analysis showed an increase in sensitivity when contrast material was administered during the CT scan, compared to the traditional CT group (77% vs. 58%). There was low heterogeneity detected on sensitivity in the CT/MRI group (I^2=30%). However, the heterogeneity was high in all other groups, regardless of sensitivity or specificity (I^2 >

75%).

Subgroup analysis based on AVS procedure (with or without ACTH stimulation) revealed a slight decrease in sensitivity when ACTH was administered during the AVS procedure (66% vs. 70%). Sensitivity and specificity were higher when LI was ≥4 versus LI was ≥2. However, a large degree of heterogeneity was observed in all groups (I^2 all > 75%).

Subgroup analysis stratified by screening test showed that the sensitivity of the ARR group was higher than that of the non-ARR group (78% vs. 66%). The heterogeneity was high (I^2 =87.7%) in the ARR subgroup, whereas it disappeared (0%) in the non-ARR subgroup. Regarding specificity, the heterogeneity was high for both groups (86.2% vs. 89.1%). Subgroup analysis stratified by the confirmatory test for PA demonstrated an increase in sensitivity (71% vs. 57%) and a slight decrease in specificity (60% vs. 66%) for the salt-loading test group compared with additional options group, with significant heterogeneity observed in all the above groups (I^2 all > 50%).

Subgroup analysis based on methodological quality (high-quality, low-quality and unclear-quality) revealed that there was low heterogeneity for sensitivity in all the above groups (I^2 all< 50%). The diagnostic pooled sensitivity for the high-quality group was the highest, followed by the unclear-quality group and the low-quality group (78% vs. 62% vs. 48%). The unclear-quality group had the highest specificity, followed by the high-quality group and the low-quality group (69% vs. 62% vs. 51%). Regarding specificity, heterogeneity was significantly decreased but still high in all the groups.

There were 4 studies that reported diagnostic accuracy in PA patients with an age of 40 years or younger. Using the bivariate model, the pooled sensitivity and specificity were 71% (95% CI: 54 to 84) and 79% (95% CI: 37 to 96), respectively, with moderate heterogeneity (53.1% vs. 70.1%) (Figure S1). Summary estimates for pooled measures of diagnostic accuracy are shown in Table 2.

Subgroups	No.of Studies	Sensitivity (95% CI)	Sensitivity with I^2	Specificity (95% CI)	Specificity with I^2
Total	25	68(61-74)	86.9%	57(50-65)	86.9%
Age					
≤40years	**4**	**71(54-84)**	**53.1%**	**79(39-96)**	**70.1%**
>40years	21	68(60-75)	87.4%	57(49-64)	87.4%
AVS procedure					
With ACTH	17	66(57-73)	86.5%	56(46-65)	90.5%
Without ACTH	7	70(58-79)	90.3%	60(45-74)	79.2%
Cutoff values of LI					
LI≥2	4	61(37-80)	95.5%	54(38-68)	88.3%
LI≥4	18	69(62-75)	76.4%	59(50-68)	89.0%
Screening test for PA					
ARR	21	66(58-72)	87.7%	58(50-65)	86.2%
Not ARR	**4**	**78(69-84)**	**0%**	**59(29-84)**	**89.1%**
Confirmatory test for PA					
salt-loading test	12	71(62-80)	78.6%	60(49-70)	74.9%
additional options	8	57(46-67)	90.6%	66(60-72)	66%
no	5	72(64-79)	55.8%	42(24-63)	90.3%
Imaging methodology					
CT	20	67(59-74)	88.6%	60(53-67)	82.5%
contrast CT	8	77(66-85)	86.4%	60(49-69)	82.2%
nocontrast CT	12	58(49-66)	83.8%	60(51-68)	81.8%
CT /MRI	5	69(62-76)	30%	45(27-64)	87.9%
Quality of studies					
high-quality studies	13	78(73-83)	48.6%	51(39-63)	78.6%
unclear-quality studies	6	62(58-65)	0%	62(54-70)	85.1%
low-quality studies	6	44(38-50)	46.4%	69(60-78)	69.2%
Sample size					
≥100	10	59(51-67)	90.6%	58(49-66)	92.1%
<100	15	74(67-81)	71.3%	60(47-71)	78.1%
Outlier excluded	20	65(60-70)	77.1%	59(52-66)	85.2%

Table 2: Pooled summary results by subgroups

AVS: adrenal vein sampling; ACTH: adrenocorticotropic hormone; CT: computed tomography; MRI: magnetic resonance imaging; PA: primary aldosteronism ; LI: lateralization index; ARR: aldosterone-to-renin ratio.

Meta-regression analysis

Results of meta-regression analysis showed that the sample size was the only covariate with a negative effect on sensitivity. Additionally, there was a significant

interaction between lower age, as well as high methodological quality, and higher specificity of CT/MRI for the detection of unilateral forms of PA (Figure S2).

Sensitivity analysis

Goodness-of-fit and bivariate normality analyses (Figures S3a, S3b) showed that the bivariate model was moderately robust. Influence analysis and outlier detection identified 4 outliers (Figures S3c, S3d). After we excluded these outliers, the overall results did not change significantly, which suggested that the results of this study were statistically reliable (Table 2).

Publication bias

Neither Deeks' Funnel Plot nor Deeks test (t=0.46, *P*=0.65) showed evidence of publication bias (Figure S4).

Discussion

Main findings

The accurate differentiation of unilateral and bilateral PA is critical for optimal clinical management. Although AVS is the "gold standard" test for subtype diagnosis [8], numerous studies have investigated the underlying diagnostic value of CT/MRI for subtype diagnosis due to several insurmountable shortcomings of AVS. The present meta-analysis, involving 4669 individuals from 25 studies, demonstrated that CT/MRI has poor sensitivity (68%) and specificity (57%) in subtype classification when AVS was used as the reference standard.

In the subtype diagnosis of PA, AVS was initially used in the 1960s. Subsequently, CT was adopted as the primary method for distinguishing the unilateral and bilateral forms of PA. Owing to its less invasive nature, lower cost and wide availability, many physicians prefer to perform CT/MRI as the first, and sometimes only, investigation of PA subtype. However, its sensitivity and specificity vary widely. The reported sensitivities ranged from 29%[3] to 94%[31], and the reported specificities ranged from 18%[19] to 87%[30]. Although the sensitivities reportedly exceeded 80% in 5 studies[19, 27, 29, 31, 37], relatively poor specificities were reported, with only 1 study showing a specificity of 72%[42]. Similarly, 3 studies reported the sensitivities to be over 80% [30, 34, 35], but the specificities were reported to be lower

than 76%[34]. The present meta-analysis showed that the pooled sensitivity was 68% and the specificity was 57%, which means that treatment decisions based on the presence of unilateral disease on CT/MRI alone could result in inappropriate unilateral adrenalectomy in 43% of patients. Basing this decision on CT/MRI alone would miss the possibility of a potentially curative procedure by surgery in 32% of patients. However, failure to make an early diagnosis and provide specific treatment for PA places these patients at higher risk of irreversible renal and cardiovascular damage. Our results suggest that CT/MRI does not have satisfactory diagnostic performance in classifying the subtypes of PA.

Stimulation with ACTH during the AVS procedure was introduced in 1979 and remains popular at many centres. Today, the AVS procedure, with or without ACTH stimulation, is still controversial[44]. The present meta-analysis revealed that there was no significant difference between the two AVS procedures (with or without ACTH stimulation) in terms of the diagnostic accuracy of CT/MRI to identify unilateral PA. In theory, the application of more stringent lateralization criteria (a condition that is more likely to capture true cases of unilateral PA) would result in increased sensitivity and decreased specificity of CT/MRI to identify unilateral PA. However, our analysis demonstrated that stricter thresholds for determining lateralization on AVS would result in higher sensitivity and specificity, which is not completely consistent with the theoretical situation.

Although the overall analysis suggested that CT/MRI does not have satisfactory diagnostic performance in classifying the subtypes of PA, the results should be interpreted with caution because of significant heterogeneity due to several underlying confounders. First, the screening test and confirmatory test for PA may influence the results. In our meta-analysis, some patients did not undergo a screening test and confirmatory test for PA, which is the diagnostic reference standard test, and some of them might not have PA. Generally, inadvertently including patients without PA should increase the specificity of CT/MRI in identifying unilateral PA, as these subjects would not show lateralization on AVS or a unilateral aldosteronoma on CT/MRI. However, although the difference in the screening test was responsible for

the heterogeneity of sensitivity and specificity to some extent, according to our analysis, there is no evidence to indicate that the confirmatory test influences the specificity of CT/MRI.

Second, meta-regression analysis showed that the heterogeneity of specificity may partly be due to age. Given that nonfunctioning adrenocortical adenomas (“incidentaloma”) are relatively uncommon in young people (≤40 years), the 2009 guidelines for managing PA contended that younger patients with an unequivocal biochemical diagnosis of PA and a clear-cut unilateral adenoma on adrenal CT scan should proceed directly to surgery, whereas the AVS procedure may be skipped[45]. Among studies included in the present meta-analysis, 4 reported the diagnostic accuracy of CT/MRI in identifying unilateral PA in patients ≤40 years. By combining these 4 studies, our results demonstrated that although the sensitivity (71%) and specificity (79%) were improved, the diagnostic performance was still unsatisfactory because 21% of patients would have undergone unnecessary adrenalectomy based on imaging results alone. In 2016, the updated clinical practice guidelines were published and suggested that the age cut-off for sparing AVS be 35 years[8]. Regarding patients aged ≤35 years, several retrospective studies have evaluated the diagnostic value of CT. The reported rate of concordance between CT and AVS ranges from 59% to 90% [19, 38, 41]. Based on these data, it still seems that CT cannot replace AVS in patients aged ≤35 years. However, due to the lack of numbers of false-positives and true-negatives, we did not perform a pooled analysis. Further studies are needed to clarify the diagnostic value of CT in patients aged ≤35 years.

As mentioned above, although adrenal imaging is not a reliable method to differentiate subtypes of PA, it does not mean that CT/MRI must be wrong and should not be used as a basis for clinical management. In centres without AVS facilities currently, what should a physician do? In the past few years, there has been rapidly growing interest in testing the utility of hybrid steroids, such as 18-oxocortisol/18-hydroxycortisol, for PA subtypes, and the results demonstrated that levels of 18-oxocortisol/18-hydroxycortisol plus an adenoma on CT/MRI might be of more assistance in those centres without AVS facilities, especially in Japan and

China, given their very high percentage of KCNJ5 mutations[46-48]. It is hoped that perhaps the possibility of multi-steroid fingerprints in peripheral blood samples that distinguish unilateral from bilateral PA with a high degree of accuracy can substantially reduce or replace the use of lateralization by AVS.

Limitations

The present meta-analysis has several limitations. First, there was great heterogeneity among the included studies, which might have compromised the credibility. The results of the subgroup analyses and meta-regression suggested that the screening test for PA, age, study quality, sample size, and other unknown factors may also contribute to the aforementioned heterogeneity. However, the results from the subgroup analyses and sensitivity analysis all confirmed the robustness of our meta-analysis's results. Second, a minority of study participants underwent cross-sectional imaging with either CT or MRI, but absolute numbers were not provided or could not be derived based on the specific imaging methodology used, which limited our ability to identify which imaging methodology can provide more accurate diagnostic performance. In addition, the possibility of selection bias that is present in all meta-analysis cannot be overlooked.

Conclusion

Based on these analyses, we conclude that CT/MRI has poor sensitivity (68%) and specificity (57%) in the detection of unilateral PA when AVS is used as the reference standard. Even in young patients (≤40 years), 21% would have undergone unnecessary adrenalectomy based on imaging results alone. Given these findings, we recommend routinely referring all patients for AVS, regardless of age and imaging results, if the centre has access to AVS. However, due to significant heterogeneity, our study should be interpreted with caution, and further high-quality studies with larger sample sizes are needed.

Author affiliations

[1]Department of Cardiology, The First Affiliated Hospital of Chengdu Medical College, Sichuan Province, People's Republic of China.

[2]Key Laboratory of Aging and Vascular Homeostasis of Sichuan Higher Education Institutes, Chengdu, Sichuan, People's Republic of China.

Contributors

WPJ is the guarantor. All authors provided substantial contribution to conception and design of the project; drafted and revised the manuscript. ZYQ led the literature search, and completed the study selection, data extraction, and critical appraisal with WPJ. JLC accepts responsibility for the integrity of the data analyses. RF led the drafting of all sections of the article in consultation with all of the coauthors. WD, CSC and ZP provided substantial contributions to the background, critical appraisal of prior studies and interpretation of meta-analysis findings. ZP provided substantial contribution to the methods section.

Competing Interests: None declared

Funding: This systematic review and meta-analysis are funded by the National Natural Sciences Foundation of China (NSFC)（Grant NO. 81970262）.

Data availability statement: All data relevant to the study are included in the article or uploaded as supplementary information.

Patient and Public Involvement: No patient involvement.

Ethics Approval: Not required.

References:

1. Monticone S, Burrello J, Tizzani D*, et al.* Prevalence and Clinical Manifestations of Primary Aldosteronism Encountered in Primary Care Practice. *J Am Coll Cardiol* 2017;69:1811-20.
2. Catena C, Colussi G, Nadalini E, *et al.* Cardiovascular outcomes in patients with primary aldosteronism after treatment. *Arch Intern Med* 2008;168: 80-5.
3. Pillai PR, Griffith M, Schwarcz MD*, et al.* Primary Aldosteronism: Cardiovascular Risk, Diagnosis, and Management. *Cardiol Rev* 2020;28:84-91.
4. Hundemer GL, Curhan GC, Yozamp N*, et al.* Cardiometabolic outcomes and mortality in medically treated primary aldosteronism: a retrospective cohort study. *Lancet Diabetes Endocrinol* 2018;6:51-59.
5. Funder JW. Primary Aldosteronism: Where Are We Now? Where to from Here? *Horm Metab Res* 2020;52:459-466.
6. Williams TA, Reincke M. MANAGEMENT OF ENDOCRINE DISEASE: Diagnosis and management of primary aldosteronism: the Endocrine Society guideline 2016 revisited. *Eur J Endocrinol* 2018;179:R19-R29.

7. Byrd JB, Turcu AF, Auchus RJ. Primary Aldosteronism: Practical Approach to Diagnosis and Management. *Circulation* 2018;138:823-835.

8. Funder JW, Carey RM, Mantero F*, et al.* The Management of Primary Aldosteronism: Case Detection, Diagnosis, and Treatment: An Endocrine Society Clinical Practice Guideline; *J Clin Endocrinol Metab* 2016;101:1889-916.

9. Ruco A, Dossa F, Tinmouth J*, et al.* Social media and mobile health technology for cancer screening: a systematic review and meta-analysis protocol. *BMJ Open* 2020;10:e035411.

10. Kempers JE, Lenders JW , Outheusden L*, et al.* Systematic review: diagnostic procedures to differentiate unilateral from bilateral adrenal abnormality in primary aldosteronism. *Ann Intern Med* 2009; 151: 329-37.

11. Liberati A, Altman DG, Tetzlaff J*, et al.* The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ann Intern Med* 2009; 151: W65-94.

12. Whiting PF, Rutjes A, Westwood M*, et al.* QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; 155: 529-36.

13. Devillé WL, Buntinx F, Bouter LM*, et al.* Conducting systematic reviews of diagnostic studies: didactic guidelines. *BMC Med Res Methodol* 2002; 2: 9.

14. Reitsma JB, Glas AS, Rutjes AW*, et al.* Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;58:982-90.

15. Chu H, Guo H, Zhou Y. Bivariate random effects meta-analysis of diagnostic studies using generalized linear mixed models. *Med Decis Making* 2010;30:499-508.

16. Lijmer JG, Bossuyt PM, Heisterkamp S H. Exploring sources of heterogeneity in systematic reviews of diagnostic tests. *Stat Med* 2002;21:1525-37.

17. Asmar M, Wachtel H, Yan Y*, et al.* Reversing the established order: Should adrenal venous sampling precede cross-sectional imaging in the evaluation of primary aldosteronism? *J Surg Oncol* 2015; 112: 144-48.

18. Pedersen M, Karlsen MA, Ankjærgaard K L*, et al.* Primary hyperaldosteronism diagnosed with adrenal vein sampling. Characteristics and follow-up after adrenalectomy in a Danish study. *Scand J Clin Lab Invest* 2016; 76: 45-50.

19. Campbell RA, Young DS, Shaver CN*, et al.* Influence of Adrenal Venous Sampling on Management in Patients with Primary Aldosteronism Independent of Lateralization on Cross-Sectional Imaging. *J Am Coll Surg* 2019; 229: 116-24.

20. Sam D, Kline GA, So B*, et al.* Discordance Between Imaging and Adrenal Vein Sampling in Primary Aldosteronism Irrespective of Interpretation Criteria. *J Clin Endocrinol Metab* 2019; 104: 1900-06.

21. Nandra G, Duxbury O, Patel P*, et al.* Technical and Interpretive Pitfalls in Adrenal Imaging. *Radiographics* 2020;40:1041-1060.

22. 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;58:882-93.

23. Magill SB, Raff H, Shaker JL*, et al.* Comparison of adrenal vein sampling and computed tomography in the differentiation of primary aldosteronism. *J Clin Endocrinol Metab* 2001; 86: 1066-71.

24. Young WF, Stanson AW, Thompson G B*, et al.* Role for adrenal venous sampling in primary aldosteronism. *Surgery* 2004;136:1227-35.

25. Nwariaku F E, Miller B S, Auchus R*, et al.* Primary hyperaldosteronism: effect of adrenal vein sampling on surgical outcome. *Arch Surg* 2006; 141: 497-502.

26. Minami I, Yoshimoto T, Hirono Y*, et al.* Diagnostic accuracy of adrenal venous sampling in comparison with other parameters in primary aldosteronism. *Endocr J* 2008; 55: 839-46.

27. Mulatero P, Bertello C, Rossato D*, et al.* Roles of clinical criteria, computed tomography scan, and adrenal vein sampling in differential diagnosis of primary aldosteronism subtypes. *J Clin Endocrinol Metab* 2008; 93: 1366-71.

28. Mathur A, Kemp CD, Dutta U*, et al.* Consequences of adrenal venous sampling in primary hyperaldosteronism and predictors of unilateral adrenal disease. *J Am Coll Surg* 2010; 211: 384-90.

29. Sarlon-Bartoli G, Michel N, Taieb D*, et al.* Adrenal venous sampling is crucial before an adrenalectomy whatever the adrenal-nodule size on computed tomography. *J Hypertens* 2011; 29: 1196-202.

30. Küpers E M, Amar L, Raynaud A*, et al.* A clinical prediction score to diagnose unilateral primary aldosteronism. *J Clin Endocrinol Metab* 2012; 97: 3530-7.

31. Oh EM, Lee KE, Yoon K*, et al.* Value of adrenal venous sampling for lesion localization in primary aldosteronism. *World J Surg* 2012; 36: 2522-7.

32. Lau JH, Sze WC, Reznek RH*, et al.* A prospective evaluation of postural stimulation testing, computed tomography and adrenal vein sampling in the differential diagnosis of primary aldosteronism. *Clin Endocrinol* 2012; 76: 182-8.

33. Salem V, Hopkins T G, El-Gayar H*, et al.* Adrenal venous sampling as a diagnostic procedure for primary hyperaldosteronism: experience from a tertiary referral centre. *Hormones* 2012; 11: 151-59.

34. Burton TJ, Mackenzie IS, Balan K*, et al.* Evaluation of the sensitivity and specificity of (11)C-metomidate positron emission tomography (PET)-CT for lateralizing aldosterone secretion by Conn's adenomas. *J Clin Endocrinol Metab* 2012; 97: 100-9.

35. Riester A, Fischer E, Degenhart C,et.al. Age below 40 or a recently proposed clinical prediction score cannot bypass adrenal venous sampling in primary aldosteronism. *J Clin Endocrinol Metab* 2014;99:E1035-1039.

36. Sze WC, Soh LM, Lau JH*, et al.* Diagnosing unilateral primary aldosteronism - comparison of a clinical prediction score, computed tomography and adrenal venous sampling. *Clin Endocrinol* 2014; 81: 25-30.

37. Kocjan T, Janez A, Stankovic M*, et al.* A NEW CLINICAL PREDICTION CRITERION ACCURATELY DETERMINES A SUBSET OF PATIENTS WITH BILATERAL PRIMARY ALDOSTERONISM BEFORE ADRENAL VENOUS SAMPLING. *Endocr Pract* 2016; 22: 587-94.

38. Zhu LM, Zhang Y, Zhang H*, et al.* Comparison between adrenal venous sampling and computed tomography in the diagnosis of primary aldosteronism and in the guidance of adrenalectomy. *Medicine* 2016; 95: e4986.

39. Kamemura K, Wada N, Ichijo T*, et al.* Significance of adrenal computed tomography in predicting laterality and indicating adrenal vein sampling in primary aldosteronism. *J Hum Hypertens* 2017; 31: 195-99.

40. Nanba AT, Nanba K, Byrd JB*, et al.* Discordance between imaging and immunohistochemistry in unilateral primary aldosteronism. *Clin Endocrinol* 2017; 87: 665-72.

41. Umakoshi H, Tsuiki M, Takeda Y*, et al.* Significance of Computed Tomography and Serum Potassium in Predicting Subtype Diagnosis of Primary Aldosteronism. *J Clin Endocrinol Metab* 2018; 103: 900-08.

42. Aono D, Kometani M, Karashima S*, et al.* Primary aldosteronism subtype discordance between computed tomography and adrenal venous sampling. *Hypertens Res* 2019; 42: 1942-50.

43. Li S, Ren Y, Zhu Y*, et al.* THE POTENTIAL CLINICAL APPLICATION OF A LOWER BILATERAL ADRENAL LIMB WIDTH RATIO (L/RW) IN PATIENTS WITH BILATERAL PRIMARY HYPERALDOSTERONISM. *Endocr Pract* 2019; 25: 830-35.

44. Laurent I, Astère M, Zheng F*, et al.* Adrenal venous sampling with or without adrenocorticotropic hormone stimulation: A meta-analysis. *J Clin Endocrinol Metab* 2019;104:1060–1068

45. Zeiger M A, Thompson G B, Duh Q Y*, et al.* The American Association of Clinical Endocrinologists and American Association of Endocrine Surgeons medical guidelines for the management of adrenal incidentalomas. *Endocr Pract* 2009;15 Suppl 1:1-20.

46. Taguchi R, Yamada M, Nakajima Y*, et al.* Expression and mutations of KCNJ5 mRNA in Japanese patients with aldosterone-producing adenomas. *J Clin Endocrinol Metab* 2012;97:1311-9.

47. Wang B, Li X, Zhang X*, et al.* Prevalence and characterization of somatic mutations in Chinese aldosterone-producing adenoma patients. *Medicine* 2015;94:e708.

48. Satoh F, Morimoto R, Ono Y*, et al.* Measurement of peripheral plasma 18-oxocortisol can discriminate unilateral adenoma from bilateral diseases in patients with primary aldosteronism; *Hypertension* 2015;65:1096-102.

Figure legend

Figure 1: Flow diagram of the review process.

Figure 2: Assessment of methodological quality of included studies using the QUADAS-2 Criteria.

Stacked bars represent the proportion of studies with a high (red), or unclear (yellow) or low (green) risk of bias and applicability concerns.

QUADAS-2: Quality Assessment of Diagnostic Accuracy Studies –2 criteria

Figure 3: Forest plots of sensitivity and specificity of adrenal imaging compared with AVS.

Horizontal lines are the 95% confidence intervals (CIs)

Figure 4: Hierarchical SROC plot showing average sensitivity and specificity estimate of the study results with 95% confidence region.

The 95% prediction region represents the confidence region for a forecast of the true sensitivity (SENS) and specificity (SPEC) in a future study.

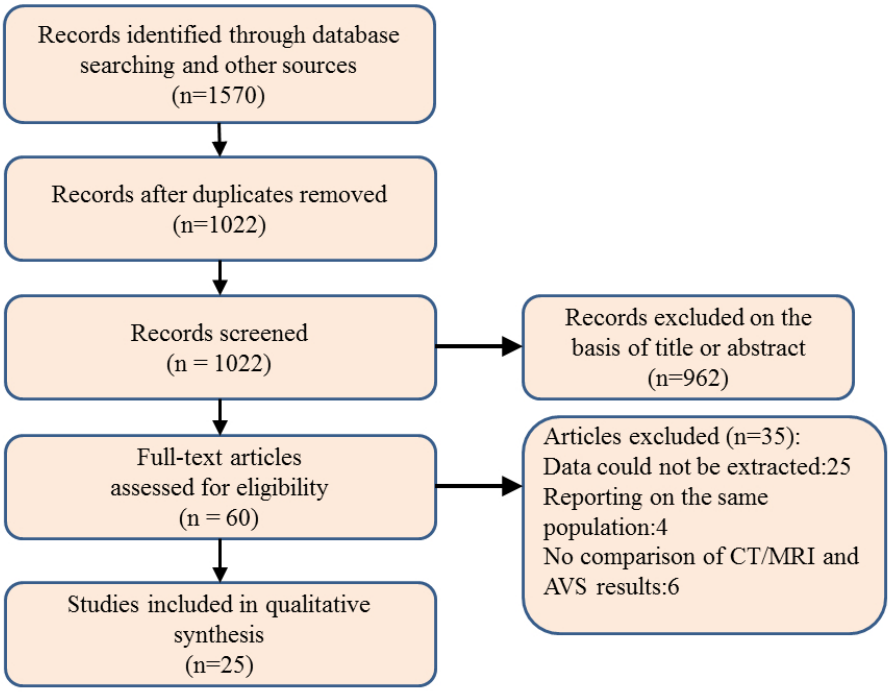
AUC: area under the curve; SROC: summary receiver-operating characteristic.

Figure S1: Forest plots of sensitivity and specificity of adrenal imaging compared with AVS in young patients (≤40 years).

Figure S2: Graphical presentation of the generalized linear mixed model exploring the impact of selected variables on sensitivity and specificity of adrenal imaging.

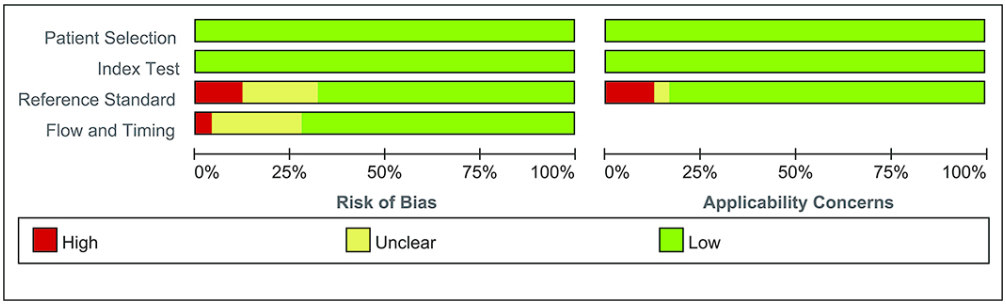
Figure S3 Graphs for sensitivity analyses: a goodness of fit, b bivariate normality, c influence analysis, and d outlier detection.

Figure S4: Deeks' funnel plot for checking the publication bias. DOR: diagnostic odds ratio.



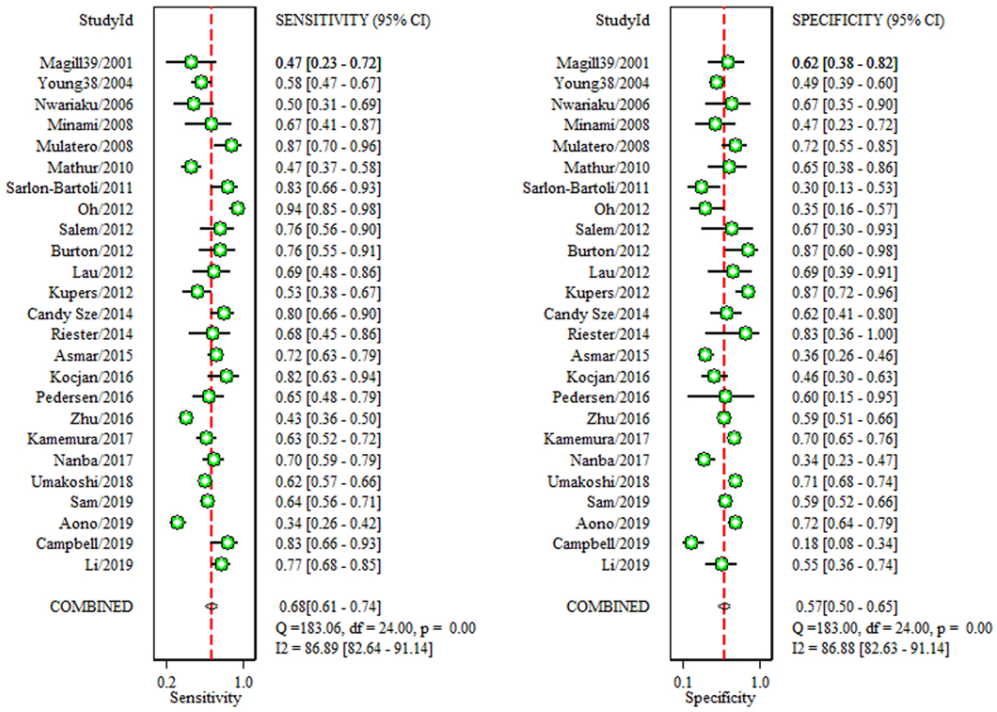
Flow diagram of the review process.

254x190mm (96 x 96 DPI)



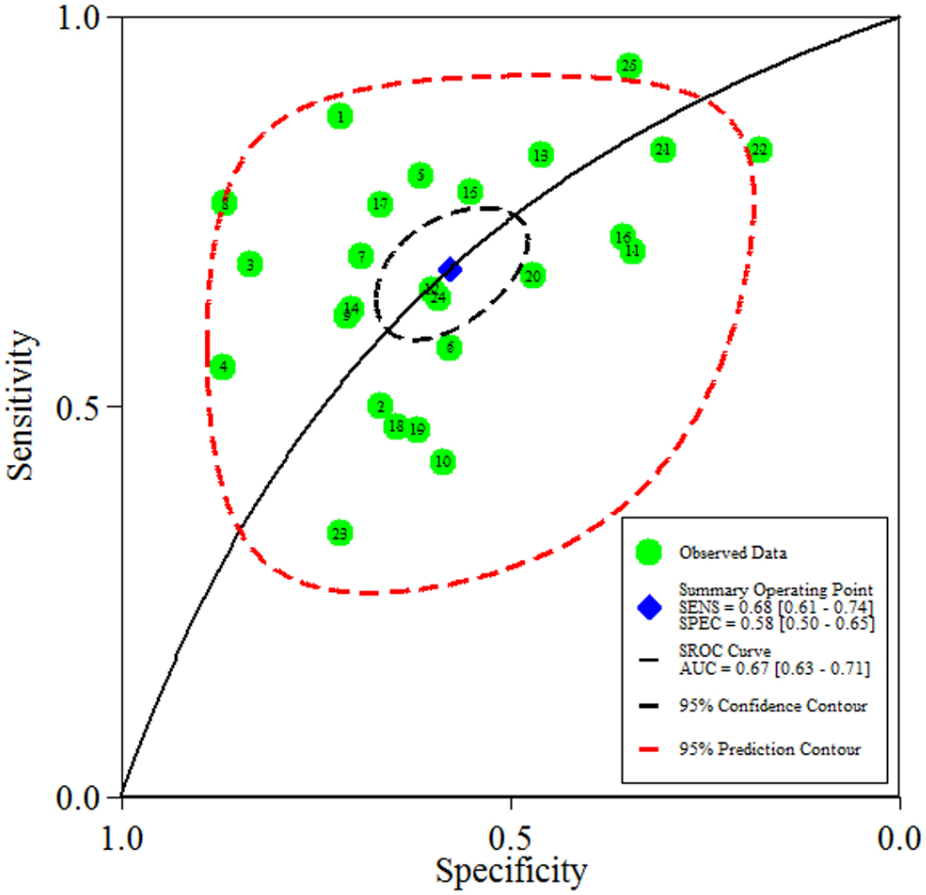
Assessment of methodological quality of included studies using the QUADAS-2 Criteria. /Stacked bars represent the proportion of studies with a high (red), or unclear (yellow) or low (green) risk of bias and applicability concerns.
QUADAS-2: Quality Assessment of Diagnostic Accuracy Studies –2 criteria

90x26mm (300 x 300 DPI)



Forest plots of sensitivity and specificity of adrenal imaging compared with AVS. /Horizontal lines are the 95% confidence intervals (CIs)

90x69mm (300 x 300 DPI)



Hierarchical SROC plot showing average sensitivity and specificity estimate of the study results with 95% confidence region. /The 95% prediction region represents the confidence region for a forecast of the true sensitivity (SENS) and specificity (SPEC) in a future study. AUC: area under the curve; SROC: summary receiver-operating characteristic.

90x90mm (300 x 300 DPI)

Search strategies for PubMed

#1. ((aldosteronism) OR (primary hyperaldosteronism) OR (primary aldosteronism) OR (conn syndrome) OR (conn) OR (aldosterone-producing adenoma) OR (APA) OR (idiopathic hyperaldosteronism) OR (IHA) OR (primary adrenal hyperplasia) OR (PAH) OR (bilateral adrenal hyperplasia) OR (BAH))

#2. ((adrenal venous sampling) OR (AVS) OR (adrenal vein sampling) OR (adrenal vein) OR (venous sampling) OR (vein sampling) OR (adrenal venous))

#3. ("2000/01/01"[Date - Entry] : "2020/01/01"[Date - Entry])

#4. #1 AND #2 AND #3

Table S2: original data

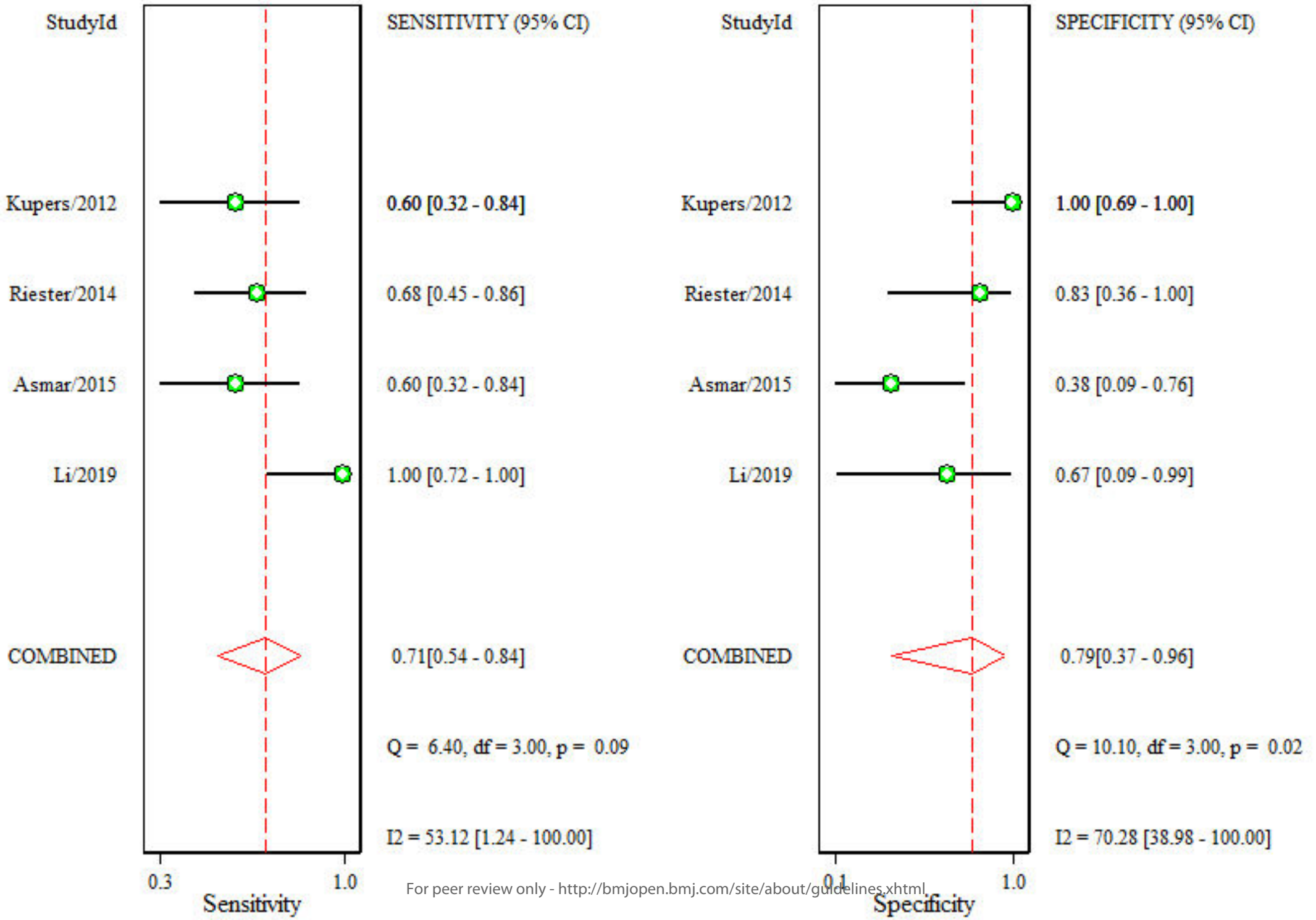
Study,Year	Age (years)	TP	FP	FN	TN
Li et al. 2019	48.5	72	13	21	16
	<35	11	1	0	2
Campbell et al.2019	55.6	29	32	6	7
Aono et al. 2019	56	52	46	102	117
	<35	3	2	0	0
Sam et al.2019	52.1	101	75	57	109
Umakoshi et al.2018	53	297	322	185	787
	<35	27	3	0	0
Nanba et al.2017	54	58	42	25	22
Kamemura et al.2017	54	60	88	36	209
Zhu et al. 2016	46	87	79	116	112
	<40	30	35	1	0
Pedersen et al.2016	51	26	2	14	3
Kocjan et al. 2016	56	23	21	5	18
Asmar et al. 2015	55	96	65	38	36
	<40	9	5	6	3
Riester et al. 2014	35	15	1	7	5
Candy Sze et al. 2014	50.5	39	10	10	16
Kűpers et al.2012	46	26	5	23	33
	<40	9	0	6	10
Lau et al. 2012	51.8	18	4	8	9
Burton et al. 2012	50.9	19	2	6	13
Salem et al.2012	44.7	22	3	7	6
Oh et al. 2012	50.7	59	15	4	8
Sarlon-Bartoli et al.2011	52	29	16	6	7
Mathur et al. 2010	50.6	46	6	51	11
Mulatero et al. 2008	52.4	27	11	4	28
Minami et al.2008	54	12	9	6	8
Nwariaku et al. 2006	51	14	4	14	8
Young et al. 2004	53	57	48	42	47
Magill et al.2001	51	8	8	9	13

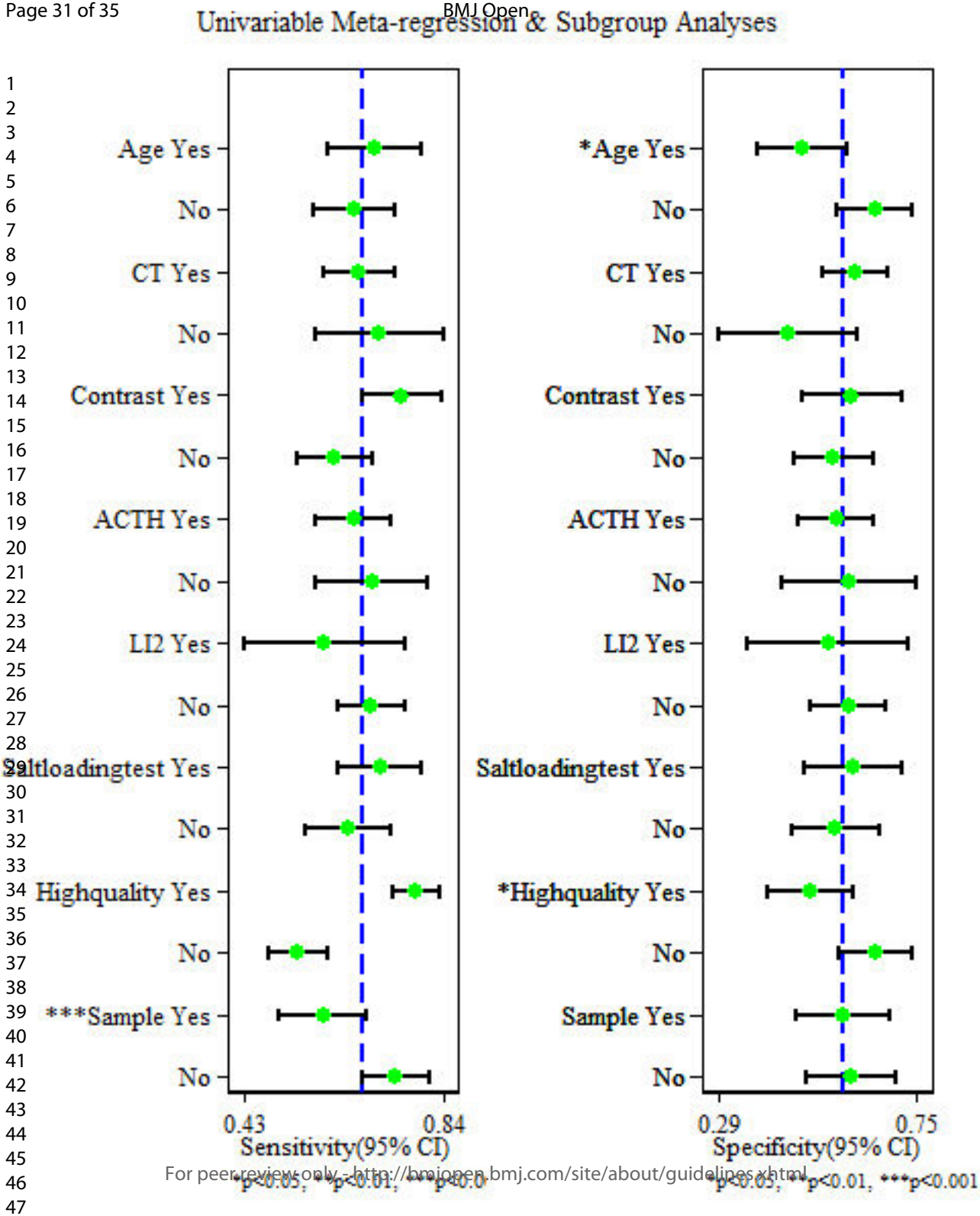
TP: true-positive; TN: true-negative; FP: false-positive; FN: false-negative.

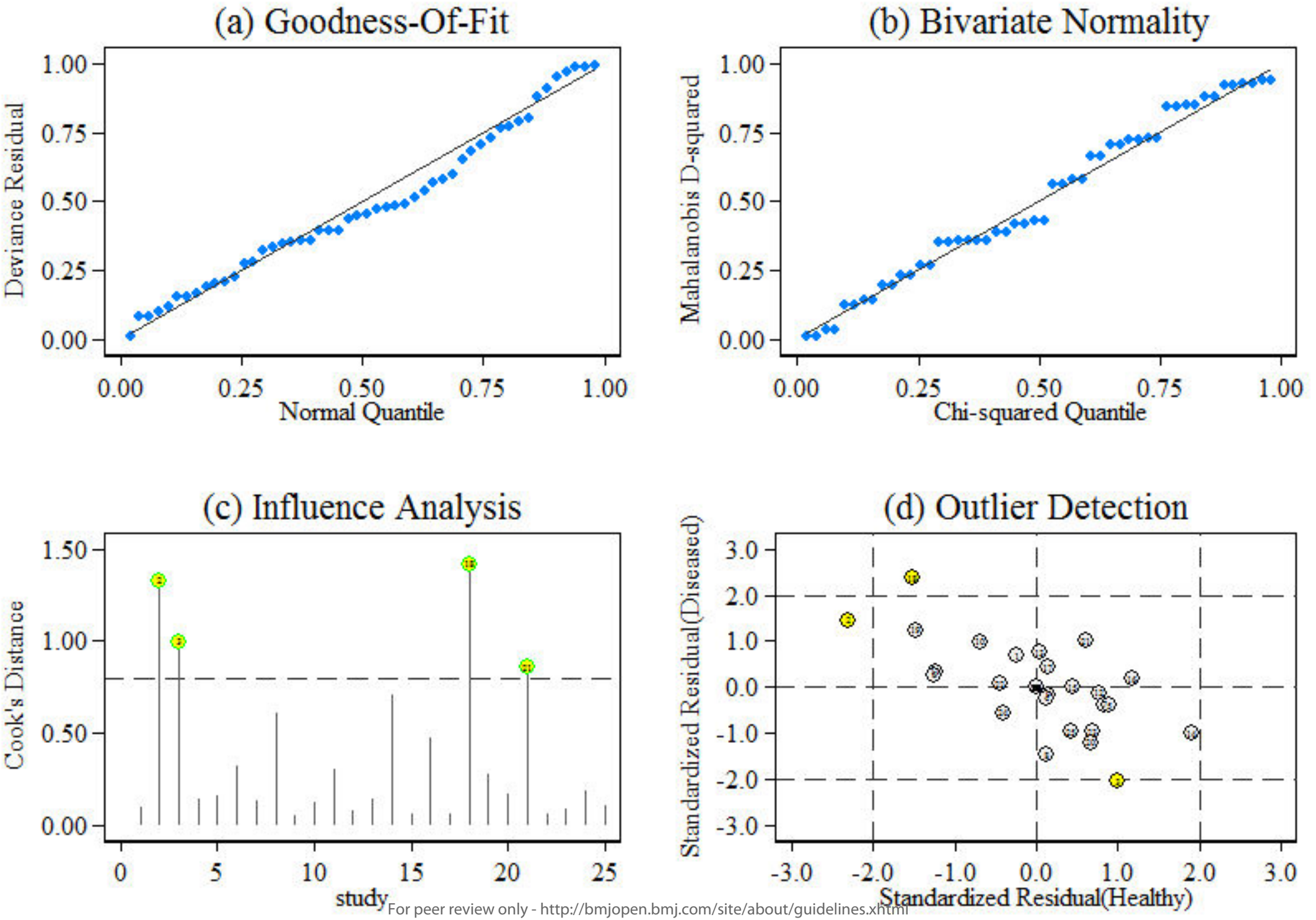
Table S3: The quality assessment for each study -QUADAS-2 results

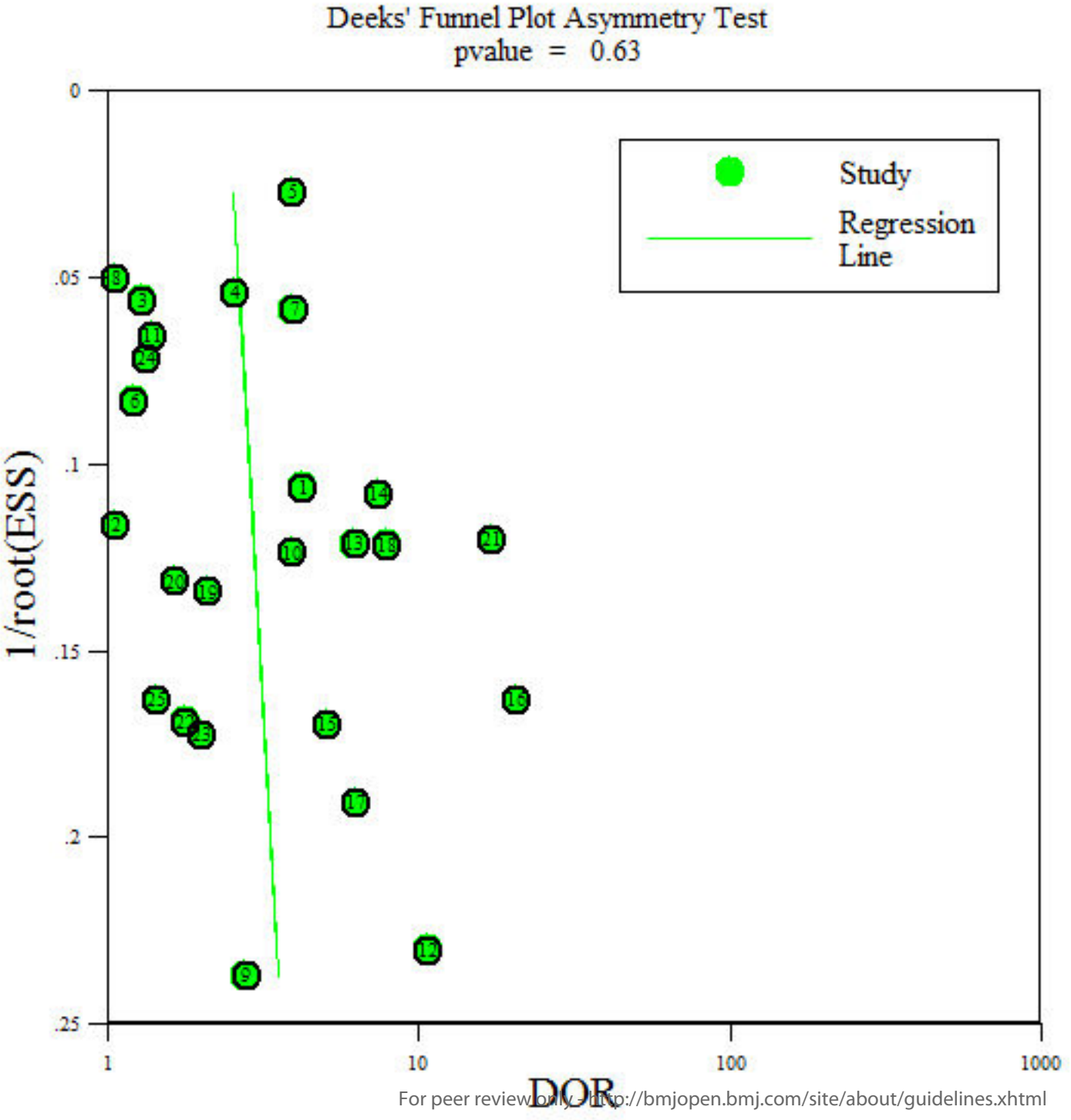
Study	Risk of bias				Applicability Concerns		
	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Li et al	☺	☺	☺	☺	☺	☺	☺
Campbell et al	☺	☺	☺	☺	☺	☺	☺
Aono et al	☺	☺	☹	☺	☺	☺	☹
Sam et al et al	☺	☺	?	?	☺	☺	☺
Umakoshi et al	☺	☺	☺	?	☺	☺	☺
Nanba et al	☺	☺	☺	☺	☺	☺	☺
Kamemura et al	☺	☺	☺	?	☺	☺	☺
Zhu et al	☺	☺	☹	☺	☺	☺	☹
Pedersen et al	☺	☺	?	☺	☺	☺	☺
Kocjan et al	☺	☺	☺	☺	☺	☺	☺
Asmar et al	☺	☺	☺	☺	☺	☺	☺
Riester et al	☺	☺	☺	☺	☺	☺	☺
Candy Sze et al	☺	☺	☺	☺	☺	☺	☺
Kűpers et al	☺	☺	?	?	☺	☺	☺
Lau et al	☺	☺	☺	☺	☺	☺	☺
Burton et al	☺	☺	☺	☺	☺	☺	☺
Salem et al	☺	☺	☺	☺	☺	☺	☺
Oh et al	☺	☺	☺	☺	☺	☺	☺
Sarlon-Bartoli et al	☺	☺	☺	☺	☺	☺	☺
Mathur et al	☺	☺	☹	☺	☺	☺	☺
Mulatero et al	☺	☺	☺	☺	☺	☺	☺
Minami et al	☺	☺	?	☺	☺	☺	☺
Nwariaku et al	☺	☺	?	?	☺	☺	☺
Young et al	☺	☺	☺	?	☺	☺	☺
Magill et al	☺	☺	☺	☹	☺	☺	☺

☺= low risk, ? = unclear risk, ☹=high risk






(a) Goodness-Of-Fit
Deviance Residual
Normal Quantile
0.00
0.25
0.50
0.75
1.00
(b) Bivariate Normality
Mahalanobis D-squared
Chi-squared Quantile
(c) Influence Analysis
Cook's Distance
study
0.00
0.50
1.00
1.50
0
5
10
15
20
25
(d) Outlier Detection
Standardized Residual(Diseased)
Standardized Residual(Healthy)
-3.0
-2.0
-1.0
0.0
1.0
2.0
3.0


Deeks' Funnel Plot Asymmetry Test
pvalue = 0.63
Study
Regression Line
1/root(ESS)
0
.05
.1
.15
.2
.25
1
10
100
1000
DOR

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2-3
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3-4
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	4-5
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	none
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	5
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	5
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	6
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	6
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	6
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	5
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	7
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	7

PRISMA 2009 Checklist

Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	7

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	7
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	7
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	7-8
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	8-9
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	10
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	9
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	9-10
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	10
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	10-11
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	11-13
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	13-14
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	14
FUNDING			

PRISMA 2009 Checklist

Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	18

BMJ Open

Diagnostic Accuracy of Adrenal Imaging for Subtype Diagnosis in Primary Aldosteronism: Systematic Review and Meta-Analysis

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-038489.R3
Article Type:	Original research
Date Submitted by the Author:	28-Jul-2020
Complete List of Authors:	Zhou, Yaqiong; Chengdu Medical College, Cardiology Wang, Dan; Chengdu Medical College, Cardiology Jiang, Licheng; Chengdu Medical College, Cardiology Ran, Fei; Chengdu Medical College, Cardiology Chen, Sichao ; Chengdu Medical College, Cardiology Zhou, Peng; Chengdu Medical College, Cardiology Wang, Peijian; Chengdu Medical College, Cardiology
Primary Subject Heading:	Diabetes and endocrinology
Secondary Subject Heading:	Cardiovascular medicine, Diabetes and endocrinology
Keywords:	Hypertension < CARDIOLOGY, Endocrine tumours < DIABETES & ENDOCRINOLOGY, Cardiology < INTERNAL MEDICINE

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Diagnostic Accuracy of Adrenal Imaging for Subtype Diagnosis in Primary Aldosteronism: Systematic Review and Meta-Analysis

Yaqiong Zhou[1] , Dan Wang[1], Licheng Jiang[1], Fei Ran[1], Sichao Chen[1], Peng Zhou[1*], Peijian Wang [1*]

From the 1 Department of Cardiology, Chengdu Medical College, Sichuan Province, People's Republic of China

*Corresponding author: Prof. Peijian Wang

Department of Cardiology, the First Affiliated Hospital of Chengdu Medical College, 299 baoguang avenues, Chengdu 610500, Sichuan Province, China. E-mail: wpjmed@aliyun.com

*Corresponding author: Prof. Peng Zhou

Department of Cardiology, the First Affiliated Hospital of Chengdu Medical College, 299 baoguang avenues, Chengdu 610500, Sichuan Province, China. E-mail: ap216@163.com.

Email: Yaqiong Zhou.1273868593@qq.com- Dan Wang.2292029981@qq.com- Licheng Jiang. 65839350@qq.com- Hameed Ullah. 807287144@qq.com- Fei Ran 1126852537@qq.com- Sichao Chen. 1350821872@qq.com-Peng Zhou. ap216@163.com-Peijian Wang. wpjmed@aliyun.com

Funding: This systematic review and meta-analysis are funded by the National Natural Sciences Foundation of China (NSFC)（Grant NO. 81970262）.

Abstract

Objectives: Accurate subtype classification in primary aldosteronism (PA) is critical in assessing the optimal treatment options. This study aimed to evaluate the diagnostic accuracy of adrenal imaging for unilateral PA classification.

Methods: Systematic searches of PubMed, EMBASE, and the Cochrane databases were performed from January 1, 2000, to February 1, 2020, for all studies that used computed tomography (CT) or magnetic resonance imaging (MRI) in determining unilateral PA and validated the results against invasive adrenal vein sampling (AVS). Summary diagnostic accuracies were assessed using a bivariate random-effects model. Subgroup analyses, meta-regression and sensitivity analysis were performed to explore the possible sources of heterogeneity.

Result: A total of 25 studies, involving a total of 4669 subjects, were identified. The overall analysis revealed a pooled sensitivity of 68% (95% confidence interval [CI]: 61 to 74) and specificity of 57% (95% CI: 50 to 65) for CT/MRI in identifying unilateral PA. Sensitivity was higher in the contrast-enhanced (CT) group versus the traditional CT group [77% (95% CI: 66 to 85) vs. 58% (95% CI: 50 to 66)]. Subgroup analysis stratified by screening test for PA showed that the sensitivity of the aldosterone-to-renin ratio (ARR) group was higher than that of the non-ARR group [78% (95% CI: 69 to 84) vs. 66% (95% CI: 58 to 72)]. The diagnostic accuracy of PA patients aged ≤40 years was reported in 4 studies, and the overall sensitivity was 71%, with 79% specificity. Meta-regression revealed a significant impact of sample size on sensitivity and of age and study quality on specificity.

Conclusion: CT/MRI is not a reliable alternative to invasive AVS without excellent sensitivity or specificity for correctly identifying unilateral PA. Even in young patients (≤40 years), 21% of patients would have undergone unnecessary adrenalectomy based on imaging results alone.

Keywords: adrenal vein sampling, computed tomography, magnetic resonance imaging, primary aldosteronism, subtype

Strengths and limitations of this study

- This study is the first meta-analysis to synthesize the evidence regarding the

diagnostic value of adrenal imaging for PA classification and demonstrated that CT/MRI is not a reliable alternative to invasive AVS even in young patients (≤40 years).

- The main methodological limitations of this systematic review and meta-analysis are the exclusion of unpublished high-quality trials and foreign-language publications.
- Another potential limitation is that we encountered different AVS methods and large variation in the lateralization criteria, which might also have affected the results for diagnostic accuracy.

Introduction

Primary aldosteronism (PA) is one of the most common causes of endocrine hypertension, with a prevalence of approximately 20% in patients with resistant hypertension, 10% in those with severe hypertension, and 6% in those with uncomplicated hypertension[1]. Accumulating clinical and epidemiological evidence suggests that PA amplifies cardiovascular and cerebrovascular complications beyond essential hypertension prior to treatment, even after controlling elevated blood pressure[2, 3]. However, patients with unilateral resected PA have slightly better risk profiles than matched essential hypertensive patients. Patients with bilateral PA whose plasma renin activity is not suppressed have the same risk profiles as essential hypertensive patients; those whose renin activity remains suppressed have 4-fold higher risk profiles than controls, and titration of mineralocorticoid receptor antagonist (MRA) therapy to raise renin might reduce this excess risk[4, 5]. Accordingly, early diagnosis and specific treatment of affected patients are key steps for the reversal of target-organ damage and prevention of cardiovascular and cerebrovascular events.

Selection of the most appropriate therapeutic strategy for patients with PA requires a distinction between unilateral and bilateral forms of PA. The former requires a unilateral adrenalectomy, mainly entailing aldosterone-producing adenoma (APA) and, less commonly, unilateral adrenal hyperplasia. In contrast, the latter, also known as idiopathic hyperaldosteronism, is optimally treated with target medical therapy[3]. Regarding the differentiation of the unilateral and bilateral subtypes, all current clinical

practice guidelines recommend adrenal vein sampling (AVS) as the standard procedure for subtype diagnosis [6, 7]. However, several shortcomings of AVS have been reported, such as its technical challenges, invasive nature, poorly standardized procedures, high cost, and lack of availability. Thus, it is urgent to explore alternative diagnostic methods without sacrificing accuracy.

Adrenal imaging with computed tomography (CT) or magnetic resonance imaging (MRI) is recommended as the first step for subtype classification given the ease of performance and relative accessibility[8]. By now, numerous studies have evaluated the diagnostic performance of CT/MRI in subtype diagnosis of PA, but the results have been inconsistent. Moreover, all these studies were limited by small sample sizes in a single centre, which limited the credibility of the results. In this context, systematic reviews and meta-analyses have the benefit of increasing the sample size, generating more precise results, which have been widely applied in clinical studies[9]. In 2009, one systematic review reported that CT/MR-based diagnoses were discordant with AVS results in 37.8% of PA patients[10]. However, the conclusions may not be reliable because of the potential for bias and concerns regarding the comparability of the included studies. Moreover, several additional studies were reported after this systematic review. We thus performed a comprehensive meta-analysis of all the available studies to evaluate the diagnostic value of adrenal imaging (CT/MRI) for subtype classification of PA.

Method

Patient and Public Involvement

Patient and public involvement

Search Strategy

The study followed the guidelines specified in the Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P)[11]. We searched the PUBMED, EMBASE, and Cochrane Library databases from January 1, 2000, to February 1, 2020, using the following terms in combination, as both MeSH or Emtree terms and text words: "primary aldosteronism", "adrenal vein sampling" and "hyperaldosteronism". The electronic search strategy for PubMed is shown in

supplementary Table S1. To reflect modern practice, we decided to limit the publication date to after January 1, 2000. We searched articles published in English, and the references of relevant studies were also searched. All studies were carefully examined to exclude overlapping or potential duplicate data.

Eligibility Criteria

We included a study if: 1) it used CT or MRI as a diagnostic test for PA subtyping; 2) it used AVS as the standard of reference. Successful AVS should be determined by calculating the selectivity index (SI), defined as the adrenal/peripheral vein cortisol ratio. Unilateral PA should be determined by calculating the lateralization index (LI), defined as the aldosterone/cortisol ratio between the dominant and the non-dominant adrenal gland; and 3) absolute numbers of true-positive, true-negative, false-positive, and false-negative results were provided or could be derived. Identified studies had to be independent. In the case of multiple reports on the same population or subpopulation, the most recent or comprehensive information was used.

Data Extraction and Quality Assessment

Data extraction from the eligible studies was performed by 2 independent investigators (Z.Y.Q and W.P.J) using a standardized data extraction form. The form included the following characteristics of each trial: first author's name and year of publication; study population characteristics, including sample size, geographical location, mean age and sex; diagnostic criteria characteristics, including screening test and confirmatory test for PA; AVS characteristics, including with/without adrenocorticotropic hormone (ACTH) stimulation, SI and LI; diagnostic test characteristics, including imaging methodology and whether contrast was administered. Differences between reviewers were resolved by discussion and consensus when necessary.

The methodological quality of the identified studies was assessed by 2 independent reviewers (Z.Y.Q and W.P.J) using the modified Quality Assessment of Diagnostic Accuracy Studies –2 (QUADAS-2) criteria. If a study was judged as “low” on all domains relating to bias or applicability, then it was judged to be a high-quality study. If a study was judged to be "high" and/or "unclear" in more than 1 domain, then

it was judged as a low-quality study. If a study was judged to be "unclear" in 1 domain, it was considered an unclear-quality study[12]. Discrepancies were resolved by discussion and consensus.

Statistical Analysis

Measures of diagnostic accuracy are reported as point estimates with 95% confidence intervals (CIs). Sensitivity, specificity, the positive likelihood ratio (+LR) and the negative likelihood ratio (-LR) were modelled based on the true-positive, true-negative, false-positive, and false-negative rates for each trial[13]. The ratio of +LR to -LR was combined in a single global accuracy measure, the diagnostic odds ratio (DOR). Summary sensitivity, specificity, +LRs, -LRs and DORs were assessed using a bivariate random-effects model. The approach assumes bivariate normal distributions for the logit transformations of sensitivity and specificity from the individual studies. These bivariate models can be analysed using linear mixed model techniques that are now widely available in statistical packages, such as STATA gllamm[14, 15]. A hierarchical summary receiver operating characteristic (ROC) curve analysis was performed, yielding point estimates for each trial and pooled characteristics, including the 95% prediction region and the 95% confidence region.

Sources of statistical heterogeneity were explored by subgroup analyses, sensitivity analysis and meta-regression analysis[16], which involved the I^2 statistic; the following interpretation was applied for I^2: <50%=low heterogeneity, 50% to 75%=moderate heterogeneity, and >75%=high heterogeneity.

Several studies demonstrated that MRI has poorer resolution and slower acquisition than CT, with a risk of respiratory artefacts and that MRI is inferior to adrenal CT in PA subtype evaluation[17-20]. Contrast materials can improve the visibility of adrenal structures imaged by CT and MRI scans and might have a positive effect on diagnosis accuracy[21]. Thus, imaging methods and contrast materials were thought to be confounders for subgroup analyses. Moreover, a large sample size may represent experienced interventional radiologists and support the credibility of the included studies. Thus, a small sample size was thought of as another confounder for subgroup analyses. The different diagnostic criteria for PA, the AVS procedure (with or without

ACTH stimulation), different cut-offs for the LI criteria, and methodological quality might also affect the results for diagnosis accuracy[20], Therefore, we also performed subgroup analyses stratified by these parameters. Thus, subgroup analyses were performed by the following factors: imaging methodology (CT or CT/MRI), contrast use, AVS procedure (with or without ACTH stimulation), cut-off value for the LI (2 or 4), diagnostic criteria for PA, sample size (divided by 100 subjects), and methodological quality (high-quality, low-quality and unclear-quality).

Potential publication bias was examined using the Deeks test[22]. The Cohen κ test was employed to assess the inter-rater reliability between 2 observers for quality assessment. If there was not agreement, a third reviewer was involved to resolve disagreements, and final decisions were determined by consensus. Statistical analyses were performed using Stata version 13.0 (StataCorp LP, TX, USA) and Review Manager version 5.3.

Results

Study Selection

After removal of 548 duplicates, the systematic review generated 1022 references that were screened according to titles and abstracts for possible inclusion. Among them, 962 studies were excluded for the following reasons: 489 studies were not relevant; 280 studies were reviews or practice guidelines; 92 studies did not include humans; and 101 studies were case reports/letters. After screening, 60 studies were identified as being potentially eligible, and their full texts were retrieved for detailed evaluation. A total of 35 studies were excluded for the following reasons: data to compute diagnostic accuracy were not provided or could not be derived (25 papers), reporting on the same population (4 papers) and no comparison of CT/MRI and AVS results in individual patients (6 papers). Finally, 25 articles were deemed eligible and analysed in our meta-analysis[17-20, 23-43] (Figure 1).

Study Characteristics

Overall, a total of 4669 patients (mean age of 51 years; 54% male) from 25 articles were included. The sample sizes of the identified studies ranged from 35 to 1591, with the largest study recruiting over 1000 participants[41]. Five studies including 724

participants underwent cross-sectional imaging either CT or MRI, and the remaining 20 studies including 3945 patients only used CT scans (8 studies administered contrast material). Seventeen studies performed AVS with ACTH stimulation, 7 studies performed it without ACTH stimulation, and the remaining 1 study provided the above two methods. The aldosterone-to-renin ratio (ARR) was used as a screening tool for PA in 21 of the included articles, and an ARR >20 was commonly used as the threshold for a positive PA screening. The remaining 4 studies did not use the ARR as a screening test for PA. In 12 articles, a salt-loading test was performed to confirm the diagnosis of PA. Eight studies used additional options, including the fludrocortisone suppression test, captopril challenge test, upright-furosemide loading test and postural stimulation test, as a confirmatory test for PA. The diagnosis of PA was not confirmed in the remaining 5 studies by one of the confirmatory tests. The 2016 Endocrine Society Guideline recommends more strict criteria for the LI (2.0 or greater under unstimulated conditions and/or 4 for ACTH stimulation) and SI (2.0 or greater under unstimulated conditions and/or 3 for ACTH stimulation)[8]. In the meta-analysis, 1 included study used less permissive criteria for the LI [42], and 6 included studies used less permissive criteria for the SI[20, 27-30, 42]. The threshold of the SI was not accessible in 4 studies[18, 23, 33, 34], and it was not accessible for the LI in 2 studies [1, 26]. Further details about the eligible and analysed studies are shown in Table 1and Table S2.

Table 1: Study Characteristics

Study,Year	Location	Male (%)	Age (years)	Sample	Imaging methodology	Contrast used	AVS characteristics	screening test for PA (ng/ml)	confirmatory test for PA

Li et al. 2019	China	61(50)	48.5	122	CT	yes	without ACTH SI≥2 LI≥2	ARR	Captopril test or salt-loading test
Campbell et al.2019	USA	45(61)	55.6	74	CT or MRI	no	with ACTH, SI≥5, LI≥4	not ARR	no
Aono et al. 2019	Japan	NA	56	317	CT	no	with ACTH, SI>2, LI≥2	ARR>20	Captopril test, furosemide upright test or salt-loading test
Sam et al.2019	Canada	201(59)	52.1	342	CT or MRI	yes	with ACTH,SI>2, LI≥4	ARR	no
Umakoshi et al.2018	Japan	762(47.9)	53	1591	CT	yes	with ACTH,SI>5,LI>4	ARR	Captopril test, furosemide upright test or salt-loading test
Nanba et al.2017	USA	87(59)	54	147	CT	no	with ACTH, SI≥5, LI≥4	ARR	Salt-loading test
Kamemura et al.2017	Japan	177(45)	54	393	CT	no	with ACTH,SI>5,LI>4	ARR	Captopril test, furosemide upright test or Salt-loading test
Zhu et al. 2016	China	NA	46	394	CT	no	without ACTH, SI≥3, LI≥2	ARR	Fludrocortisone test or salt-loading test
Pedersen et al.2016	Denmark	24(54)	51	45	CT or MRI	no	without ACTH. SI /LI NA	ARR	Fludrocortisone test or postural test
Kocjan et al. 2016	Slovenia	46(69)	56	67	CT	no	with ACTH,SI>5,LI>4	ARR	Salt-loading test
Asmar et al. 2015	USA	148(63)	55	235	CT or MRI	no	with ACTH, SI≥5, LI≥4	ARR	no
Riester et al. 2014	Germany	NA	35	28	CT or MRI	no	without ACTH, SI≥2, LI≥4	ARR	Salt-loading test
Candy Sze et al. 2014	UK	42(56)	50.5	75	CT	yes	with ACTH, SI≥5, LI≥4	not ARR	salt-loading test
Küpers et al.2012	France	53(61)	46	87	CT	no	with ACTH, SI≥2, LI≥4	ARR	Salt-loading test
Lau et al. 2012	UK	24(64)	51.8	39	CT	no	with ACTH, SI≥5, LI≥4	not ARR	salt-loading test
Burton et al. 2012	UK	NA	50.9	40	CT	Yes	without ACTH,LI≥4	not ARR	no
Salem et al.2012	UK	16(44)	44.7	38	CT	no	without ACTH, LI≥4	ARR	Salt-loading test
Oh et al. 2012	Korea	45(52)	50.7	86	CT	yes	with ACTH, SI≥3, LI>4	ARR	Salt-loading test
Sarlon-Bartoli et al.2011	France	NA	52	58	CT	yes	without ACTH, SI≥1, LI>2	ARR	no
Mathur et al. 2010	USA	63(55)	50.6	114	CT	no	with ACTH, SI>2, LI≥4	ARR	Captopril test, posture test or salt-loading test
Mulatero et al. 2008	Italy	NA	52.4	70	CT	yes	65 with ACTH; 5 without ACTH ,SI>2, LI≥4	ARR	Salt-loading test
Minami et al.2008	Japan	12(34)	54	35	CT	no	with ACTH SI /LI NA	ARR	Salt-loading test
Nwariaku et al. 2006	USA	27(67)	51	40	CT	yes	with ACTH, SI≥3, LI>4	ARR	Captopril test or posture test
Young et al. 2004	USA	163(84)	53	194	CT	no	with ACTH, SI>5, LI>4	ARR	Salt-loading test
Magill et al.2001	USA	27(71)	51	38	CT	no	with ACTH, SI≥3, LI≥4	ARR	Salt-loading test

ACTH: adrenocorticotropic hormone; CT: computed tomography; MRI: magnetic resonance imaging; NA: not available; SI: selectivity index; LI: lateralization index; ARR: aldosterone-to-renin ratio

Quality assessment

Overall, the identified studies were of excellent quality in terms of applicability and risk of bias. Out of 175 QUADAS-2 items (25 articles×7 items), the 172 (98%) were agreed on by the 2 reviewers, with an inter-rater agreement of κ=0.9. Figure 2 summarizes the QUADAS-2 assessment, and supplementary Table S3 displays each of the 25 individual QUADAS-2 evaluations.

The risk of bias from the reference standard was high in 3 studies[28, 38, 42] , and it

was unclear in 5 studies[18, 20, 25, 26, 30] because it was not clear whether the reference standard was interpreted without knowledge of the adrenal imaging results or whether the cut-off values of the SI and LI correctly classified the target condition. The risk of bias regarding flow and timing was unclear in 6 studies[20, 24, 25, 30, 39, 41] because the time interval between the index test and the reference standard was unclear; it was high in 1 study [23] because not all patients had the same reference standard (Figure 2). Finally, 13 studies were considered to be high-quality studies[17, 19, 27, 29, 31-38, 43], 7 were considered low-quality studies[20, 23, 25, 28, 30, 38, 42] and 5 were unclear-quality studies [18, 24, 26, 39, 41].

Overall analysis

Using the bivariate model, statistical heterogeneity was found for sensitivity (I^2 =86.9%; *P*= 0.001), specificity (I^2 =86.9%; *P*=0.00), the positive LR (I^2 = 76.3%; *P*= 0.00), the negative LR (I^2=79.2%; *P*=0.00) and DOR (I^2 =100.0%; *P*=0.00), indicating high between-study heterogeneity for all pooled measures, which might compromise the credibility of the study.

In the overall analysis, the pooled sensitivity, specificity, positive LR, negative LR and DOR for adrenal imaging were 68% (95% CI: 61 to 74), 57% (95% CI: 50 to 65), 1.6 (95% CI: 1.4 to 1.9), 0.56 (95% CI: 0.47 to 0.68), and 3 (95% CI: 2 to 4), respectively (Figures 3,4).

Subgroup analyses

Subgroup analysis, stratified by the imaging methodology, found more favourable specificity (60%) for CT than CT/MRI (45%). Notably, subgroup analysis showed an increase in sensitivity when contrast material was administered during the CT scan, compared to the traditional CT group (77% vs. 58%). There was low heterogeneity detected on sensitivity in the CT/MRI group (I^2=30%). However, the heterogeneity was high in all other groups, regardless of sensitivity or specificity (I^2 > 75%).

Subgroup analysis based on AVS procedure (with or without ACTH stimulation) revealed a slight decrease in sensitivity when ACTH was administered during the AVS procedure (66% vs. 70%). Sensitivity and specificity were higher when LI was ≥4 versus LI was ≥2. However, a large degree of heterogeneity was observed in all groups (I^2 all > 75%).

Subgroup analysis stratified by screening test showed that the sensitivity of the

Subgroups	No.of Studies	Sensitivity (95% CI)	Sensitivity with I^2	Specificity (95% CI)	Specificity with I^2
Total	25	68(61-74)	86.9%	57(50-65)	86.9%
Age					
≤40years	**4**	**71(54-84)**	**53.1%**	**79(39-96)**	**70.1%**

ARR group was higher than that of the non-ARR group (78% vs. 66%). The heterogeneity was high (I^2 =87.7%) in the ARR subgroup, whereas it disappeared (0%) in the non-ARR subgroup. Regarding specificity, the heterogeneity was high for both groups (86.2% vs. 89.1%). Subgroup analysis stratified by the confirmatory test for PA demonstrated an increase in sensitivity (71% vs. 57%) and a slight decrease in specificity (60% vs. 66%) for the salt-loading test group compared with additional options group, with moderate to high heterogeneity observed in all the above groups (I^2 all > 50%).

Subgroup analysis based on methodological quality (high-quality, low-quality and unclear-quality) revealed that there was low heterogeneity for sensitivity in all the above groups (I^2 all< 50%). The diagnostic pooled sensitivity for the high-quality group was the highest, followed by the unclear-quality group and the low-quality group (78% vs. 62% vs. 48%). The unclear-quality group had the highest specificity, followed by the high-quality group and the low-quality group (69% vs. 62% vs. 51%). Regarding specificity, heterogeneity was decreased but still high in all the groups.

There were 4 studies that reported diagnostic accuracy in PA patients with an age of 40 years or younger. Using the bivariate model, the pooled sensitivity and specificity were 71% (95% CI: 54 to 84) and 79% (95% CI: 37 to 96), respectively, with moderate heterogeneity (53.1% vs. 70.1%) (Figure S1). Summary estimates for pooled measures of diagnostic accuracy are shown in Table 2.

Table 2: Pooled summary results by subgroups

>40years	21	68(60-75)	87.4%	57(49-64)	87.4%
AVS procedure					
With ACTH	17	66(57-73)	86.5%	56(46-65)	90.5%
Without ACTH	7	70(58-79)	90.3%	60(45-74)	79.2%
Cutoff values of LI					
LI≥2	4	61(37-80)	95.5%	54(38-68)	88.3%
LI≥4	18	69(62-75)	76.4%	59(50-68)	89.0%
Screening test for PA					
ARR	21	66(58-72)	87.7%	58(50-65)	86.2%
Not ARR	**4**	**78(69-84)**	**0%**	**59(29-84)**	**89.1%**
Confirmatory test for PA					
salt-loading test	12	71(62-80)	78.6%	60(49-70)	74.9%
additional options	8	57(46-67)	90.6%	66(60-72)	66%
no	5	72(64-79)	55.8%	42(24-63)	90.3%
Imaging methodology					
CT	20	67(59-74)	88.6%	60(53-67)	82.5%
contrast CT	8	77(66-85)	86.4%	60(49-69)	82.2%
nocontrast CT	12	58(49-66)	83.8%	60(51-68)	81.8%
CT /MRI	5	69(62-76)	30%	45(27-64)	87.9%
Quality of studies					
high-quality studies	13	78(73-83)	48.6%	51(39-63)	78.6%
unclear-quality studies	6	62(58-65)	0%	62(54-70)	85.1%
low-quality studies	6	44(38-50)	46.4%	69(60-78)	69.2%
Sample size					
≥100	10	59(51-67)	90.6%	58(49-66)	92.1%
<100	15	74(67-81)	71.3%	60(47-71)	78.1%
Outlier excluded	20	65(60-70)	77.1%	59(52-66)	85.2%

AVS: adrenal vein sampling; ACTH: adrenocorticotropic hormone; CT: computed tomography; MRI: magnetic resonance imaging; PA: primary aldosteronism ; LI: lateralization index; ARR: aldosterone-to-renin ratio.

Meta-regression analysis

Results of meta-regression analysis showed that the sample size was the only covariate with a negative effect on sensitivity. Additionally, there was a significant interaction between lower age, as well as high methodological quality, and higher specificity of CT/MRI for the detection of unilateral forms of PA (Figure S2).

Sensitivity analysis

Goodness-of-fit and bivariate normality analyses (Figures S3a, S3b) showed that the bivariate model was moderately robust. Influence analysis and outlier detection

identified 4 outliers (Figures S3c, S3d). After we excluded these outliers, the overall results did not change significantly, which suggested that the results of this study were statistically reliable (Table 2).

Publication bias

Neither Deeks' Funnel Plot nor Deeks test (t=0.46, *P*=0.65) showed evidence of publication bias (Figure S4).

Discussion

Main findings

The accurate differentiation of unilateral and bilateral PA is critical for optimal clinical management. Although AVS is the "gold standard" test for subtype diagnosis [8], numerous studies have investigated the underlying diagnostic value of CT/MRI for subtype diagnosis due to several insurmountable shortcomings of AVS. The present meta-analysis, involving 4669 individuals from 25 studies, demonstrated that CT/MRI has poor sensitivity (68%) and specificity (57%) in subtype classification when AVS was used as the reference standard.

In the subtype diagnosis of PA, AVS was initially used in the 1960s. Subsequently, CT was adopted as the primary method for distinguishing the unilateral and bilateral forms of PA. Owing to its less invasive nature, lower cost and wide availability, many physicians prefer to perform CT/MRI as the first, and sometimes only, investigation of PA subtype. However, its sensitivity and specificity vary widely. The reported sensitivities ranged from 29%[3] to 94%[31], and the reported specificities ranged from 18%[19] to 87%[30]. Although the sensitivities reportedly exceeded 80% in 5 studies[19, 27, 29, 31, 37], relatively poor specificities were reported, with only 1 study showing a specificity of 72%[42]. Similarly, 3 studies reported the sensitivities to be over 80% [30, 34, 35], but the specificities were reported to be lower than 76%[34]. The present meta-analysis showed that the pooled sensitivity was 68% and the specificity was 57%, which means that treatment decisions based on the presence of unilateral disease on CT/MRI alone could result in inappropriate unilateral adrenalectomy in 43% of patients. Basing this decision on CT/MRI alone would miss the possibility of a potentially curative procedure by surgery in 32% of patients. However, failure to make an early diagnosis

and provide specific treatment for PA places these patients at higher risk of irreversible renal and cardiovascular damage. Our results suggest that CT/MRI does not have satisfactory diagnostic performance in classifying the subtypes of PA.

Stimulation with ACTH during the AVS procedure was introduced in 1979 and remains popular at many centres. Today, the AVS procedure, with or without ACTH stimulation, is still controversial[44]. The present meta-analysis revealed that there was no significant difference between the two AVS procedures (with or without ACTH stimulation) in terms of the diagnostic accuracy of CT/MRI to identify unilateral PA. In theory, the application of more stringent lateralization criteria (a condition that is more likely to capture true cases of unilateral PA) would result in increased sensitivity and decreased specificity of CT/MRI to identify unilateral PA. However, our analysis demonstrated that stricter thresholds for determining lateralization on AVS would result in higher sensitivity and specificity, which is not completely consistent with the theoretical situation.

Although the overall analysis suggested that CT/MRI does not have satisfactory diagnostic performance in classifying the subtypes of PA, the results should be interpreted with caution because of moderate to high heterogeneity due to several underlying confounders. First, the screening test and confirmatory test for PA may influence the results. In our meta-analysis, some patients did not undergo a screening test and confirmatory test for PA, which is the diagnostic reference standard test, and some of them might not have PA. Generally, inadvertently including patients without PA should increase the specificity of CT/MRI in identifying unilateral PA, as these subjects would not show lateralization on AVS or a unilateral aldosteronoma on CT/MRI. However, although the difference in the screening test was responsible for the heterogeneity of sensitivity and specificity to some extent, according to our analysis, there is no evidence to indicate that the confirmatory test influences the specificity of CT/MRI.

Second, meta-regression analysis showed that the heterogeneity of specificity may partly be due to age. Given that nonfunctioning adrenocortical adenomas

(“incidentaloma”) are relatively uncommon in young people (≤40 years), the 2009 guidelines for managing PA contended that younger patients with an unequivocal biochemical diagnosis of PA and a clear-cut unilateral adenoma on adrenal CT scan should proceed directly to surgery, whereas the AVS procedure may be skipped[45]. Among studies included in the present meta-analysis, 4 reported the diagnostic accuracy of CT/MRI in identifying unilateral PA in patients ≤40 years. By combining these 4 studies, our results demonstrated that although the sensitivity (71%) and specificity (79%) were improved, the diagnostic performance was still unsatisfactory because 21% of patients would have undergone unnecessary adrenalectomy based on imaging results alone. In 2016, the updated clinical practice guidelines were published and suggested that the age cut-off for sparing AVS be 35 years[8]. Regarding patients aged ≤35 years, several retrospective studies have evaluated the diagnostic value of CT. The reported rate of concordance between CT and AVS ranges from 59% to 90% [19, 38, 41]. Based on these data, it still seems that CT cannot replace AVS in patients aged ≤35 years. However, due to the lack of numbers of false-positives and true-negatives, we did not perform a pooled analysis. Further studies are needed to clarify the diagnostic value of CT in patients aged ≤35 years.

As mentioned above, although adrenal imaging is not a reliable method to differentiate subtypes of PA, it does not mean that CT/MRI must be wrong and should not be used as a basis for clinical management. In centres without AVS facilities currently, what should a physician do? In the past few years, there has been rapidly growing interest in testing the utility of hybrid steroids, such as 18-oxocortisol/18-hydroxycortisol, for PA subtypes, and the results demonstrated that levels of 18-oxocortisol/18-hydroxycortisol plus an adenoma on CT/MRI might be of more assistance in those centres without AVS facilities, especially in Japan and China, given their very high percentage of KCNJ5 mutations[46-48]. It is hoped that perhaps the possibility of multi-steroid fingerprints in peripheral blood samples that distinguish unilateral from bilateral PA with a high degree of accuracy can substantially reduce or replace the use of lateralization by AVS.

Limitations

The present meta-analysis has several limitations. First, there was great heterogeneity among the included studies, which might have compromised the credibility. The results of the subgroup analyses and meta-regression suggested that the screening test for PA, age, study quality, sample size, and other unknown factors may also contribute to the aforementioned heterogeneity. However, the results from the subgroup analyses and sensitivity analysis all confirmed the robustness of our meta-analysis's results. Second, a minority of study participants underwent cross-sectional imaging with either CT or MRI, but absolute numbers were not provided or could not be derived based on the specific imaging methodology used, which limited our ability to identify which imaging methodology can provide more accurate diagnostic performance. In addition, the possibility of selection bias that is present in all meta-analysis cannot be overlooked.

Conclusion

Based on these analyses, we conclude that CT/MRI has poor sensitivity (68%) and specificity (57%) in the detection of unilateral PA when AVS is used as the reference standard. Even in young patients (≤40 years), 21% would have undergone unnecessary adrenalectomy based on imaging results alone. Given these findings, we recommend routinely referring all patients for AVS, regardless of age and imaging results, if the centre has access to AVS. However, due to moderate to high heterogeneity, our study should be interpreted with caution, and further high-quality studies with larger sample sizes are needed.

Author affiliations

[1]Department of Cardiology, Chengdu Medical College, Sichuan Province, People's Republic of China.

Contributors

WPJ is the guarantor. All authors provided substantial contribution to conception and design of the project; drafted and revised the manuscript. ZYQ led the literature search, and completed the study selection, data extraction, and critical appraisal with WPJ. JLC accepts responsibility for the integrity of the data analyses. RF led the

drafting of all sections of the article in consultation with all of the coauthors. WD, CSC and ZP provided substantial contributions to the background, critical appraisal of prior studies and interpretation of meta-analysis findings. ZP provided substantial contribution to the methods section.

Competing Interests: None declared

Funding: This systematic review and meta-analysis are funded by the National Natural Sciences Foundation of China (NSFC)（Grant NO. 81970262）.

Data availability statement: Extra data can be accessed via the Dryad data repository at http://datadryad.org/ with the doi: 10.5061/dryad.jm63xsj8t

Patient and Public Involvement: Patients and the public were not involved in the design or conduct of the study.

Ethics Approval: Not required.

References:

1. Monticone S, Burrello J, Tizzani D*, et al.* Prevalence and Clinical Manifestations of Primary Aldosteronism Encountered in Primary Care Practice. *J Am Coll Cardiol* 2017;69:1811-20.

2. Catena C, Colussi G, Nadalini E, *et al.* Cardiovascular outcomes in patients with primary aldosteronism after treatment. *Arch Intern Med* 2008;168: 80-5.

3. Pillai PR, Griffith M, Schwarcz MD*, et al.* Primary Aldosteronism: Cardiovascular Risk, Diagnosis, and Management. *Cardiol Rev* 2020;28:84-91.

4. Hundemer GL, Curhan GC, Yozamp N*, et al.* Cardiometabolic outcomes and mortality in medically treated primary aldosteronism: a retrospective cohort study. *Lancet Diabetes Endocrinol* 2018;6:51-59.

5. Funder JW. Primary Aldosteronism: Where Are We Now? Where to from Here? *Horm Metab Res* 2020;52:459-466.

6. Williams TA, Reincke M. MANAGEMENT OF ENDOCRINE DISEASE: Diagnosis and management of primary aldosteronism: the Endocrine Society guideline 2016 revisited. *Eur J Endocrinol* 2018;179:R19-R29.

7. Byrd JB, Turcu AF, Auchus RJ. Primary Aldosteronism: Practical Approach to Diagnosis and Management. *Circulation* 2018;138:823-835.

8. Funder JW, Carey RM, Mantero F*, et al.* The Management of Primary Aldosteronism: Case Detection, Diagnosis, and Treatment: An Endocrine Society Clinical Practice Guideline; *J Clin Endocrinol Metab* 2016;101:1889-916.

9. Ruco A, Dossa F, Tinmouth J*, et al.* Social media and mobile health technology for cancer screening: a systematic review and meta-analysis protocol. *BMJ Open* 2020;10:e035411.

10. Kempers JE, Lenders JW , Outheusden L*, et al.* Systematic review: diagnostic procedures to differentiate unilateral from bilateral adrenal abnormality in primary aldosteronism. *Ann Intern Med* 2009; 151: 329-37.

11. Liberati A, Altman DG, Tetzlaff J*, et al.* The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ann Intern Med* 2009; 151: W65-94.

12. Whiting PF, Rutjes A, Westwood M*, et al.* QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; 155: 529-36.

13. Devillé WL, Buntinx F, Bouter LM*, et al.* Conducting systematic reviews of diagnostic studies: didactic guidelines. *BMC Med Res Methodol* 2002; 2: 9.

14. Reitsma JB, Glas AS, Rutjes AW*, et al.* Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;58:982-90.

15. Chu H, Guo H, Zhou Y. Bivariate random effects meta-analysis of diagnostic studies using generalized linear mixed models. *Med Decis Making* 2010;30:499-508.

16. Lijmer JG, Bossuyt PM, Heisterkamp S H. Exploring sources of heterogeneity in systematic reviews of diagnostic tests. *Stat Med* 2002;21:1525-37.

17. Asmar M, Wachtel H, Yan Y*, et al.* Reversing the established order: Should adrenal venous sampling precede cross-sectional imaging in the evaluation of primary aldosteronism? *J Surg Oncol* 2015; 112: 144-48.

18. Pedersen M, Karlsen MA, Ankjærgaard K L*, et al.* Primary hyperaldosteronism diagnosed with adrenal vein sampling. Characteristics and follow-up after adrenalectomy in a Danish study. *Scand J Clin Lab Invest* 2016; 76: 45-50.

19. Campbell RA, Young DS, Shaver CN*, et al.* Influence of Adrenal Venous Sampling on Management in Patients with Primary Aldosteronism Independent of Lateralization on Cross-Sectional Imaging. *J Am Coll Surg* 2019; 229: 116-24.

20. Sam D, Kline GA, So B*, et al.* Discordance Between Imaging and Adrenal Vein Sampling in Primary Aldosteronism Irrespective of Interpretation Criteria. *J Clin Endocrinol Metab* 2019; 104: 1900-06.

21. Nandra G, Duxbury O, Patel P*, et al.* Technical and Interpretive Pitfalls in Adrenal Imaging. *Radiographics* 2020;40:1041-1060.

22. 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;58:882-93.

23. Magill SB, Raff H, Shaker JL*, et al.* Comparison of adrenal vein sampling and computed tomography in the differentiation of primary aldosteronism. *J Clin Endocrinol Metab* 2001; 86: 1066-71.

24. Young WF, Stanson AW, Thompson G B*, et al.* Role for adrenal venous sampling in primary aldosteronism. *Surgery* 2004;136:1227-35.

25. Nwariaku F E, Miller B S, Auchus R*, et al.* Primary hyperaldosteronism: effect of adrenal vein sampling on surgical outcome. *Arch Surg* 2006; 141: 497-502.

26. Minami I, Yoshimoto T, Hirono Y*, et al.* Diagnostic accuracy of adrenal venous sampling in comparison with other parameters in primary aldosteronism. *Endocr J* 2008; 55: 839-46.

27. Mulatero P, Bertello C, Rossato D*, et al.* Roles of clinical criteria, computed tomography scan, and adrenal vein sampling in differential diagnosis of primary aldosteronism subtypes. *J Clin Endocrinol Metab* 2008; 93: 1366-71.

28. Mathur A, Kemp CD, Dutta U*, et al.* Consequences of adrenal venous sampling in primary hyperaldosteronism and predictors of unilateral adrenal disease. *J Am Coll Surg* 2010; 211: 384-90.

29. Sarlon-Bartoli G, Michel N, Taieb D*, et al.* Adrenal venous sampling is crucial before an adrenalectomy whatever the adrenal-nodule size on computed tomography. *J Hypertens* 2011; 29: 1196-202.

30. Küpers E M, Amar L, Raynaud A*, et al.* A clinical prediction score to diagnose unilateral primary aldosteronism. *J Clin Endocrinol Metab* 2012; 97: 3530-7.

31. Oh EM, Lee KE, Yoon K*, et al.* Value of adrenal venous sampling for lesion localization in primary aldosteronism. *World J Surg* 2012; 36: 2522-7.

32. Lau JH, Sze WC, Reznek RH*, et al.* A prospective evaluation of postural stimulation testing, computed tomography and adrenal vein sampling in the differential diagnosis of primary aldosteronism. *Clin Endocrinol* 2012; 76: 182-8.

33. Salem V, Hopkins T G, El-Gayar H*, et al.* Adrenal venous sampling as a diagnostic procedure for primary hyperaldosteronism: experience from a tertiary referral centre. *Hormones* 2012; 11: 151-59.

34. Burton TJ, Mackenzie IS, Balan K*, et al.* Evaluation of the sensitivity and specificity of (11)C-metomidate positron emission tomography (PET)-CT for lateralizing aldosterone secretion by Conn's adenomas. *J Clin Endocrinol Metab* 2012; 97: 100-9.

35. Riester A, Fischer E, Degenhart C,et.al. Age below 40 or a recently proposed clinical prediction score cannot bypass adrenal venous sampling in primary aldosteronism. *J Clin Endocrinol Metab* 2014;99:E1035-1039.

36. Sze WC, Soh LM, Lau JH*, et al.* Diagnosing unilateral primary aldosteronism - comparison of a clinical prediction score, computed tomography and adrenal venous sampling. *Clin Endocrinol* 2014; 81: 25-30.

37. Kocjan T, Janez A, Stankovic M*, et al.* A NEW CLINICAL PREDICTION CRITERION ACCURATELY DETERMINES A SUBSET OF PATIENTS WITH BILATERAL PRIMARY ALDOSTERONISM BEFORE ADRENAL VENOUS SAMPLING. *Endocr Pract* 2016; 22: 587-94.

38. Zhu LM, Zhang Y, Zhang H*, et al.* Comparison between adrenal venous sampling and computed tomography in the diagnosis of primary aldosteronism and in the guidance of adrenalectomy. *Medicine* 2016; 95: e4986.

39. Kamemura K, Wada N, Ichijo T*, et al.* Significance of adrenal computed tomography in predicting laterality and indicating adrenal vein sampling in primary aldosteronism. *J Hum Hypertens* 2017; 31: 195-99.

40. Nanba AT, Nanba K, Byrd JB*, et al.* Discordance between imaging and immunohistochemistry in unilateral primary aldosteronism. *Clin Endocrinol* 2017; 87: 665-72.

41. Umakoshi H, Tsuiki M, Takeda Y*, et al.* Significance of Computed Tomography and Serum Potassium in Predicting Subtype Diagnosis of Primary Aldosteronism. *J Clin Endocrinol Metab* 2018; 103: 900-08.

42. Aono D, Kometani M, Karashima S*, et al.* Primary aldosteronism subtype discordance between computed tomography and adrenal venous sampling. *Hypertens Res* 2019; 42: 1942-50.

43. Li S, Ren Y, Zhu Y*, et al.* THE POTENTIAL CLINICAL APPLICATION OF A LOWER BILATERAL ADRENAL LIMB WIDTH RATIO (L/RW) IN PATIENTS WITH BILATERAL PRIMARY HYPERALDOSTERONISM. *Endocr Pract* 2019; 25: 830-35.

44. Laurent I, Astère M, Zheng F*, et al.* Adrenal venous sampling with or without adrenocorticotropic hormone stimulation: A meta-analysis. *J Clin Endocrinol Metab* 2019;104:1060–1068

45. Zeiger M A, Thompson G B, Duh Q Y*, et al.* The American Association of Clinical Endocrinologists and American Association of Endocrine Surgeons medical guidelines for the management of adrenal incidentalomas. *Endocr Pract* 2009;15 Suppl 1:1-20.

46. Taguchi R, Yamada M, Nakajima Y*, et al.* Expression and mutations of KCNJ5 mRNA in Japanese patients with aldosterone-producing adenomas. *J Clin Endocrinol Metab* 2012;97:1311-9.

47. Wang B, Li X, Zhang X*, et al.* Prevalence and characterization of somatic mutations in Chinese

aldosterone-producing adenoma patients. *Medicine* 2015;94:e708.

48. Satoh F, Morimoto R, Ono Y, *et al.* Measurement of peripheral plasma 18-oxocortisol can discriminate unilateral adenoma from bilateral diseases in patients with primary aldosteronism; *Hypertension* 2015;65:1096-102.

Figure legend

Figure 1: Flow diagram of the review process.

Figure 2: Assessment of methodological quality of included studies using the QUADAS-2 Criteria.

Stacked bars represent the proportion of studies with a high (red), or unclear (yellow) or low (green) risk of bias and applicability concerns.

QUADAS-2: Quality Assessment of Diagnostic Accuracy Studies –2 criteria

Figure 3: Forest plots of sensitivity and specificity of adrenal imaging compared with AVS.

Horizontal lines are the 95% confidence intervals (CIs)

Figure 4: Hierarchical SROC plot showing average sensitivity and specificity estimate of the study results with 95% confidence region.

The 95% prediction region represents the confidence region for a forecast of the true sensitivity (SENS) and specificity (SPEC) in a future study.

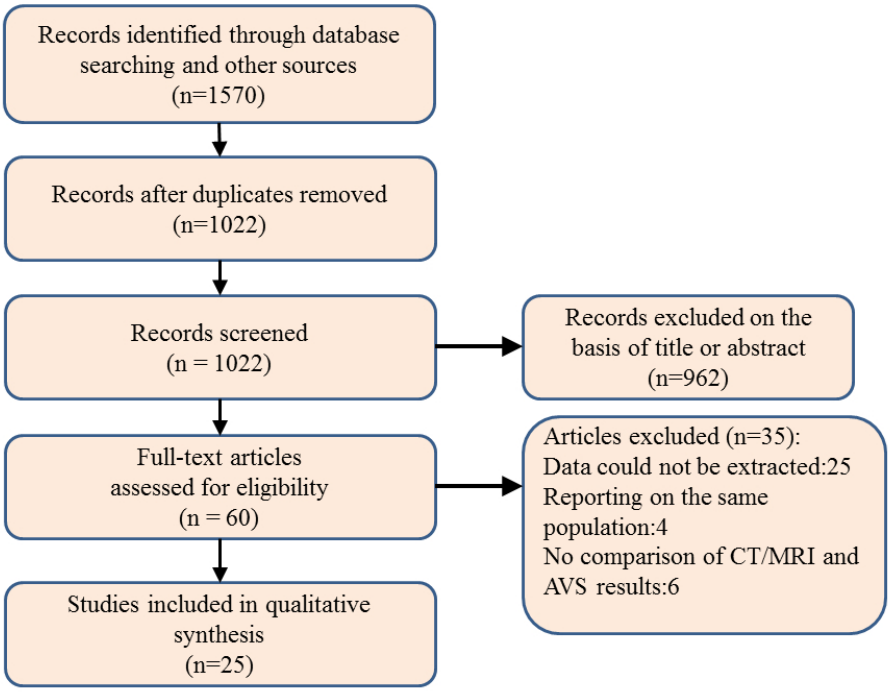
AUC: area under the curve; SROC: summary receiver-operating characteristic.

Figure S1: Forest plots of sensitivity and specificity of adrenal imaging compared with AVS in young patients (≤40 years).

Figure S2: Graphical presentation of the generalized linear mixed model exploring the impact of selected variables on sensitivity and specificity of adrenal imaging.

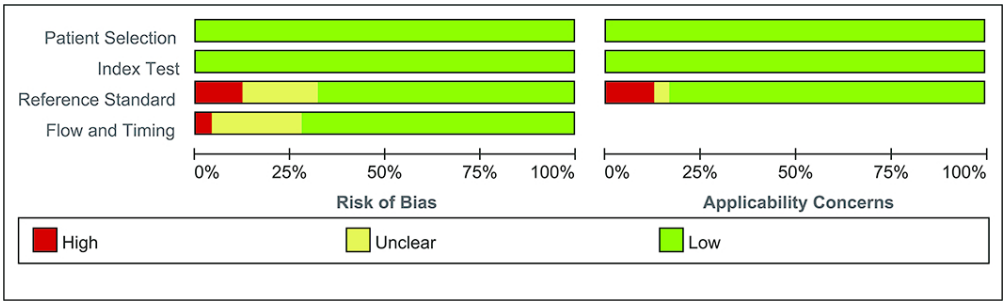
Figure S3 Graphs for sensitivity analyses: a goodness of fit, b bivariate normality, c influence analysis, and d outlier detection.

Figure S4: Deeks' funnel plot for checking the publication bias. DOR: diagnostic odds ratio.



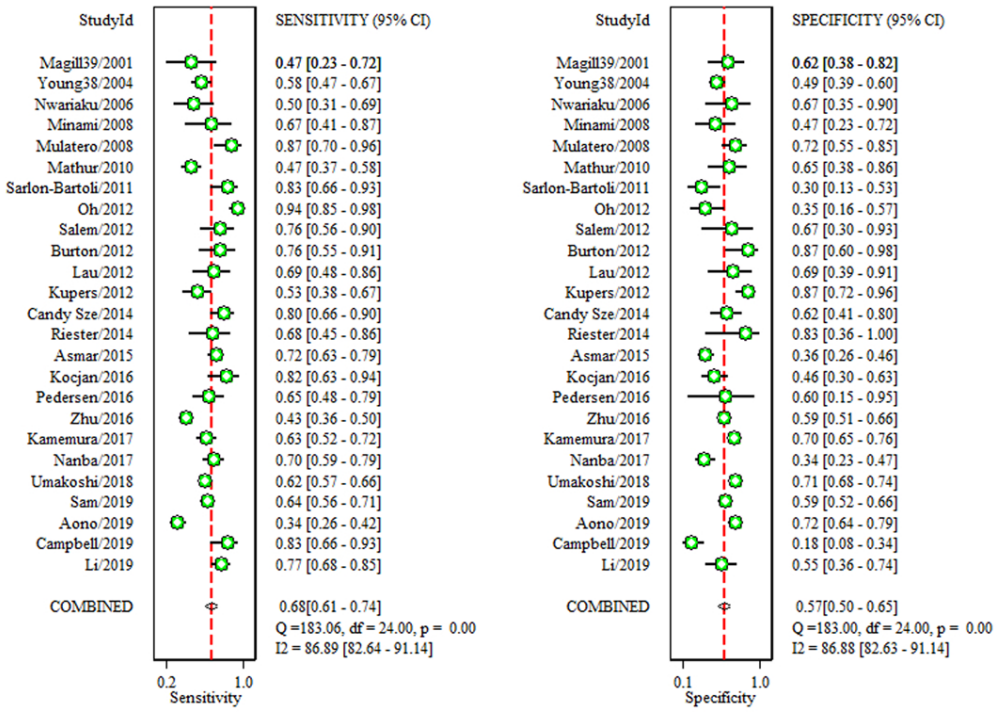
Flow diagram of the review process.

254x190mm (96 x 96 DPI)



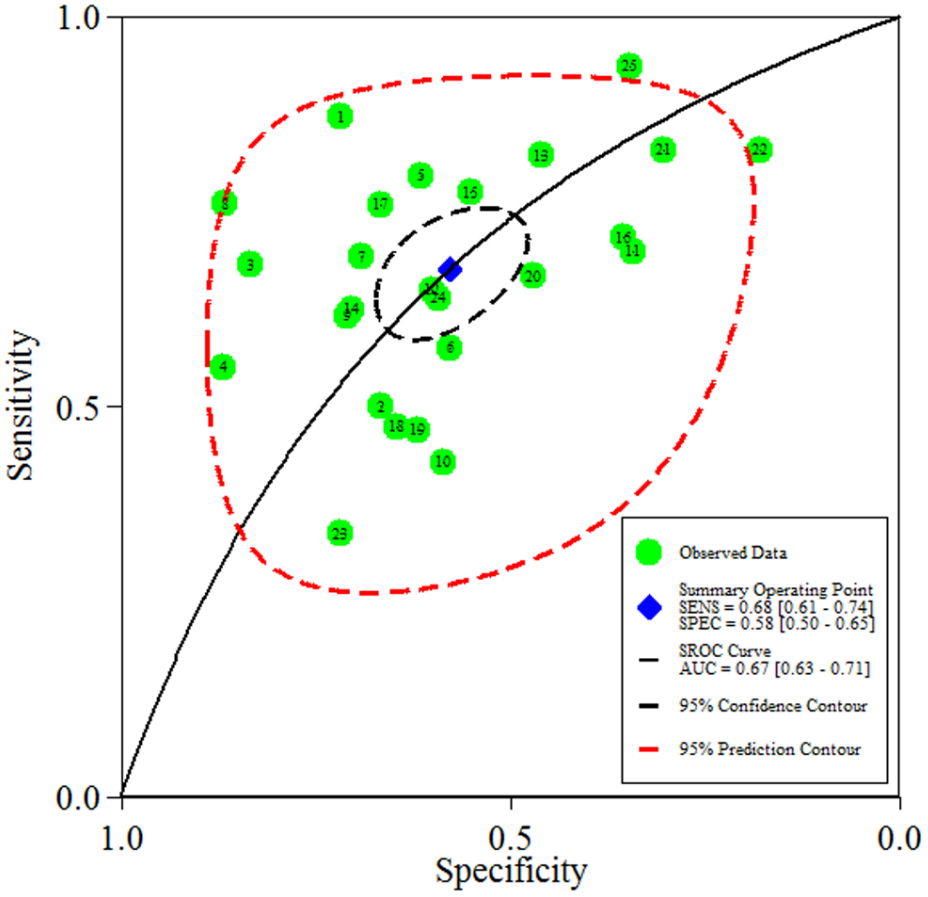
Assessment of methodological quality of included studies using the QUADAS-2 Criteria. /Stacked bars represent the proportion of studies with a high (red), or unclear (yellow) or low (green) risk of bias and applicability concerns.
QUADAS-2: Quality Assessment of Diagnostic Accuracy Studies –2 criteria

90x26mm (300 x 300 DPI)



Forest plots of sensitivity and specificity of adrenal imaging compared with AVS. /Horizontal lines are the 95% confidence intervals (CIs)

90x69mm (300 x 300 DPI)



Hierarchical SROC plot showing average sensitivity and specificity estimate of the study results with 95% confidence region. /The 95% prediction region represents the confidence region for a forecast of the true sensitivity (SENS) and specificity (SPEC) in a future study. AUC: area under the curve; SROC: summary receiver-operating characteristic.

90x90mm (300 x 300 DPI)

Search strategies for PubMed

#1. ((aldosteronism) OR (primary hyperaldosteronism) OR (primary aldosteronism) OR (conn syndrome) OR (conn) OR (aldosterone-producing adenoma) OR (APA) OR (idiopathic hyperaldosteronism) OR (IHA) OR (primary adrenal hyperplasia) OR (PAH) OR (bilateral adrenal hyperplasia) OR (BAH))

#2. ((adrenal venous sampling) OR (AVS) OR (adrenal vein sampling) OR (adrenal vein) OR (venous sampling) OR (vein sampling) OR (adrenal venous))

#3. ("2000/01/01"[Date - Entry] : "2020/01/01"[Date - Entry])

#4. #1 AND #2 AND #3

Table S2: original data

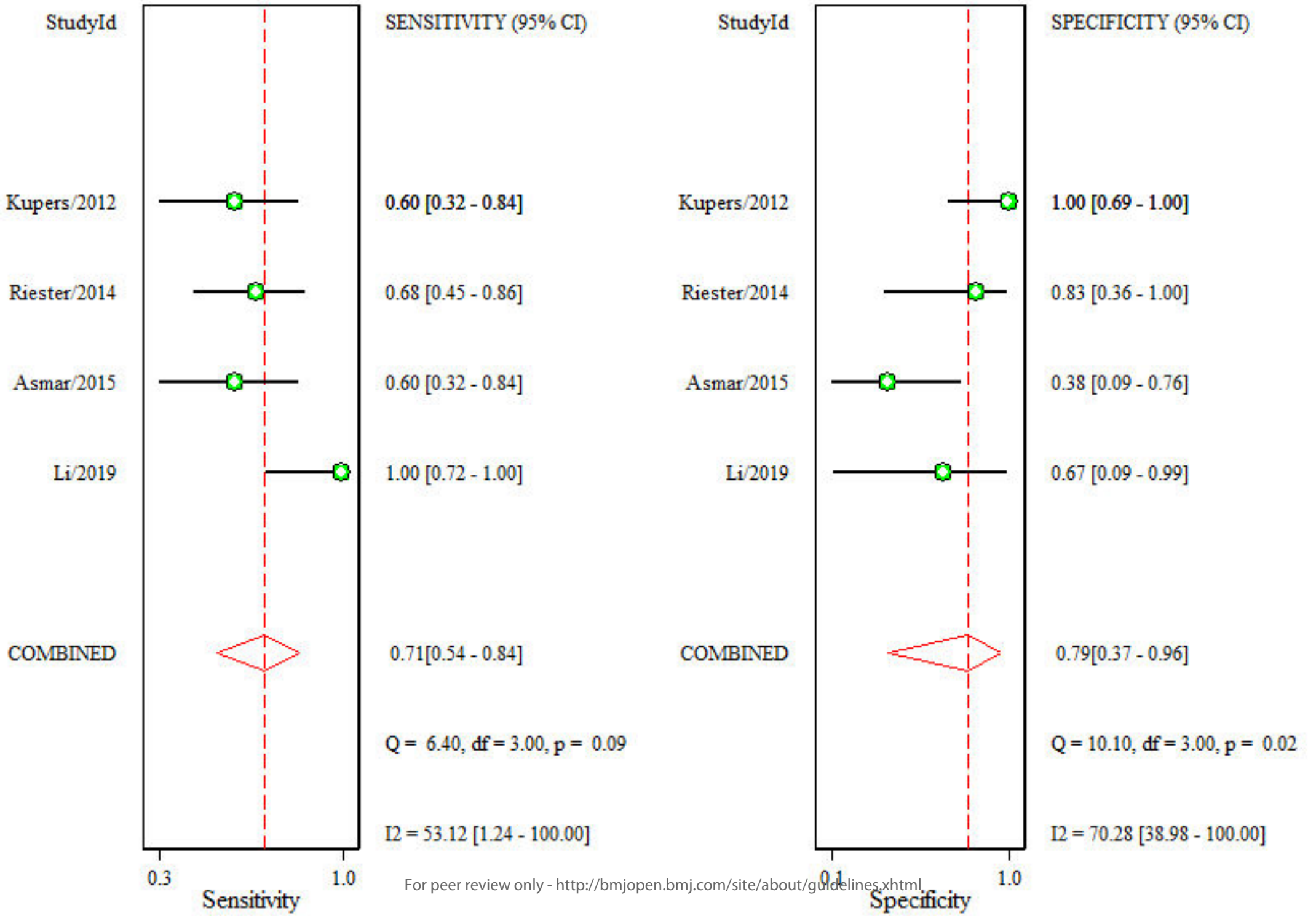
Study,Year	Age (years)	TP	FP	FN	TN
Li et al. 2019	48.5	72	13	21	16
	<35	11	1	0	2
Campbell et al.2019	55.6	29	32	6	7
Aono et al. 2019	56	52	46	102	117
	<35	3	2	0	0
Sam et al.2019	52.1	101	75	57	109
Umakoshi et al.2018	53	297	322	185	787
	<35	27	3	0	0
Nanba et al.2017	54	58	42	25	22
Kamemura et al.2017	54	60	88	36	209
Zhu et al. 2016	46	87	79	116	112
	<40	30	35	1	0
Pedersen et al.2016	51	26	2	14	3
Kocjan et al. 2016	56	23	21	5	18
Asmar et al. 2015	55	96	65	38	36
	<40	9	5	6	3
Riester et al. 2014	35	15	1	7	5
Candy Sze et al. 2014	50.5	39	10	10	16
Kűpers et al.2012	46	26	5	23	33
	<40	9	0	6	10
Lau et al. 2012	51.8	18	4	8	9
Burton et al. 2012	50.9	19	2	6	13
Salem et al.2012	44.7	22	3	7	6
Oh et al. 2012	50.7	59	15	4	8
Sarlon-Bartoli et al.2011	52	29	16	6	7
Mathur et al. 2010	50.6	46	6	51	11
Mulatero et al. 2008	52.4	27	11	4	28
Minami et al.2008	54	12	9	6	8
Nwariaku et al. 2006	51	14	4	14	8
Young et al. 2004	53	57	48	42	47
Magill et al.2001	51	8	8	9	13

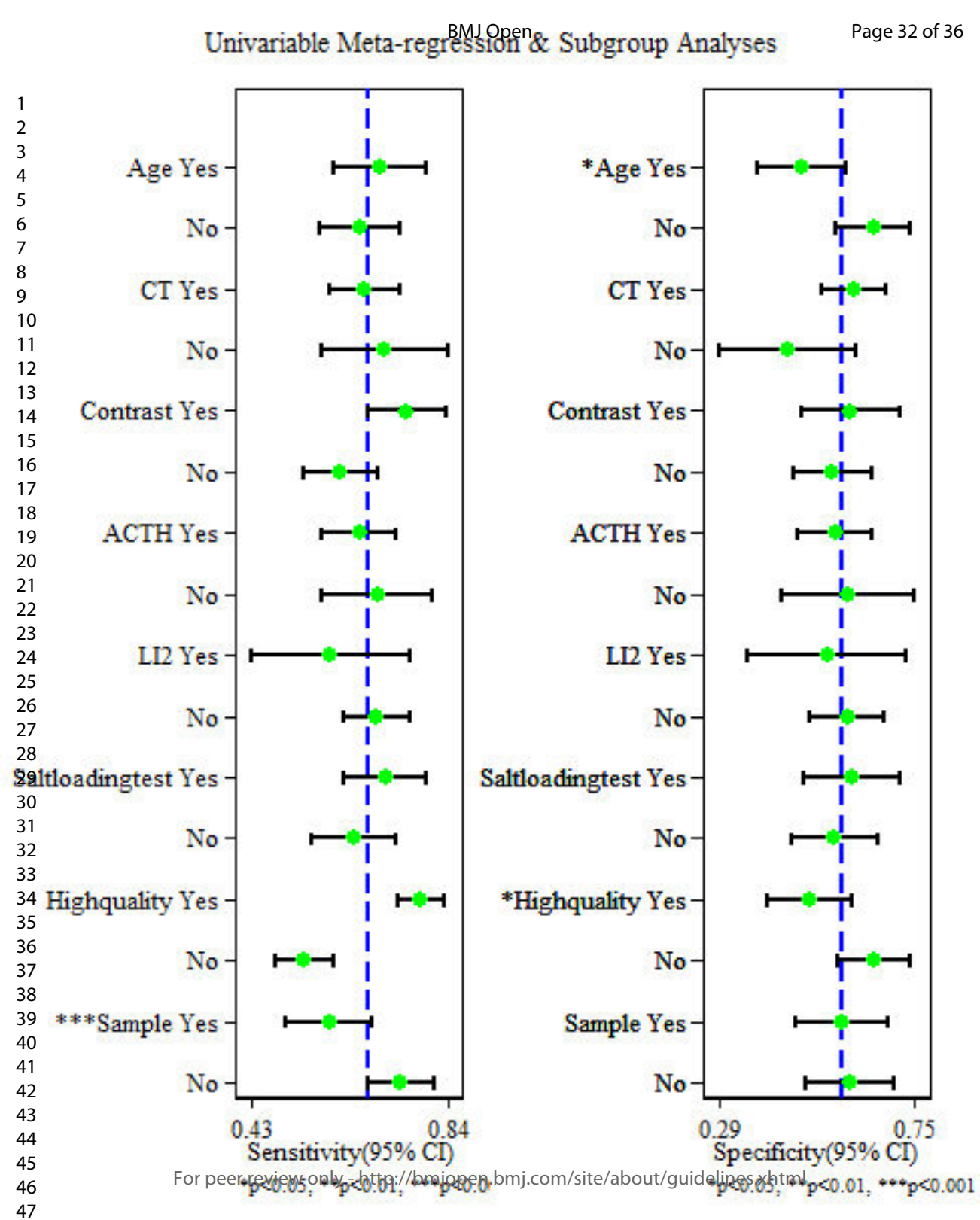
TP: true-positive; TN: true-negative; FP: false-positive; FN: false-negative.

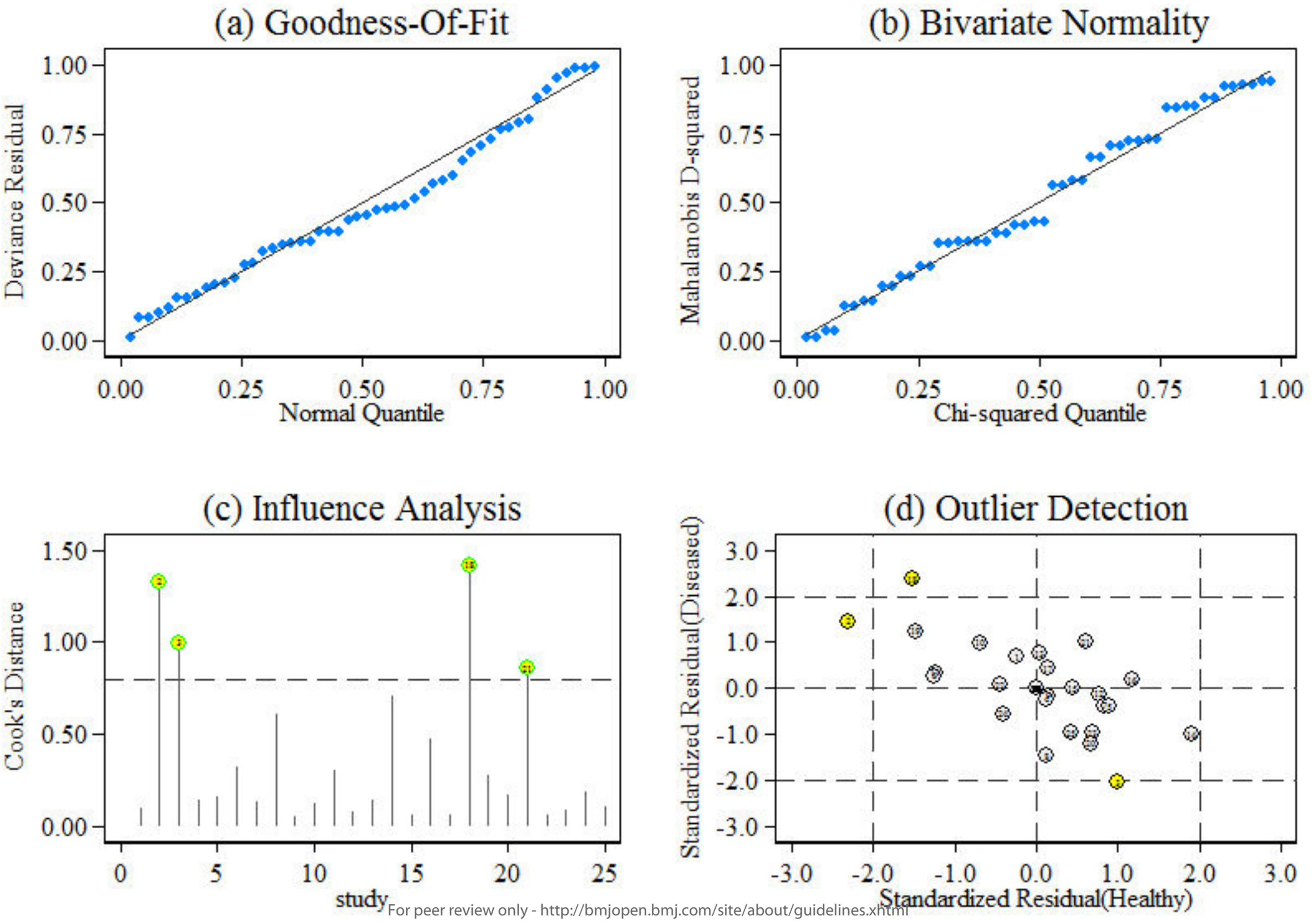
Table S3: The quality assessment for each study -QUADAS-2 results

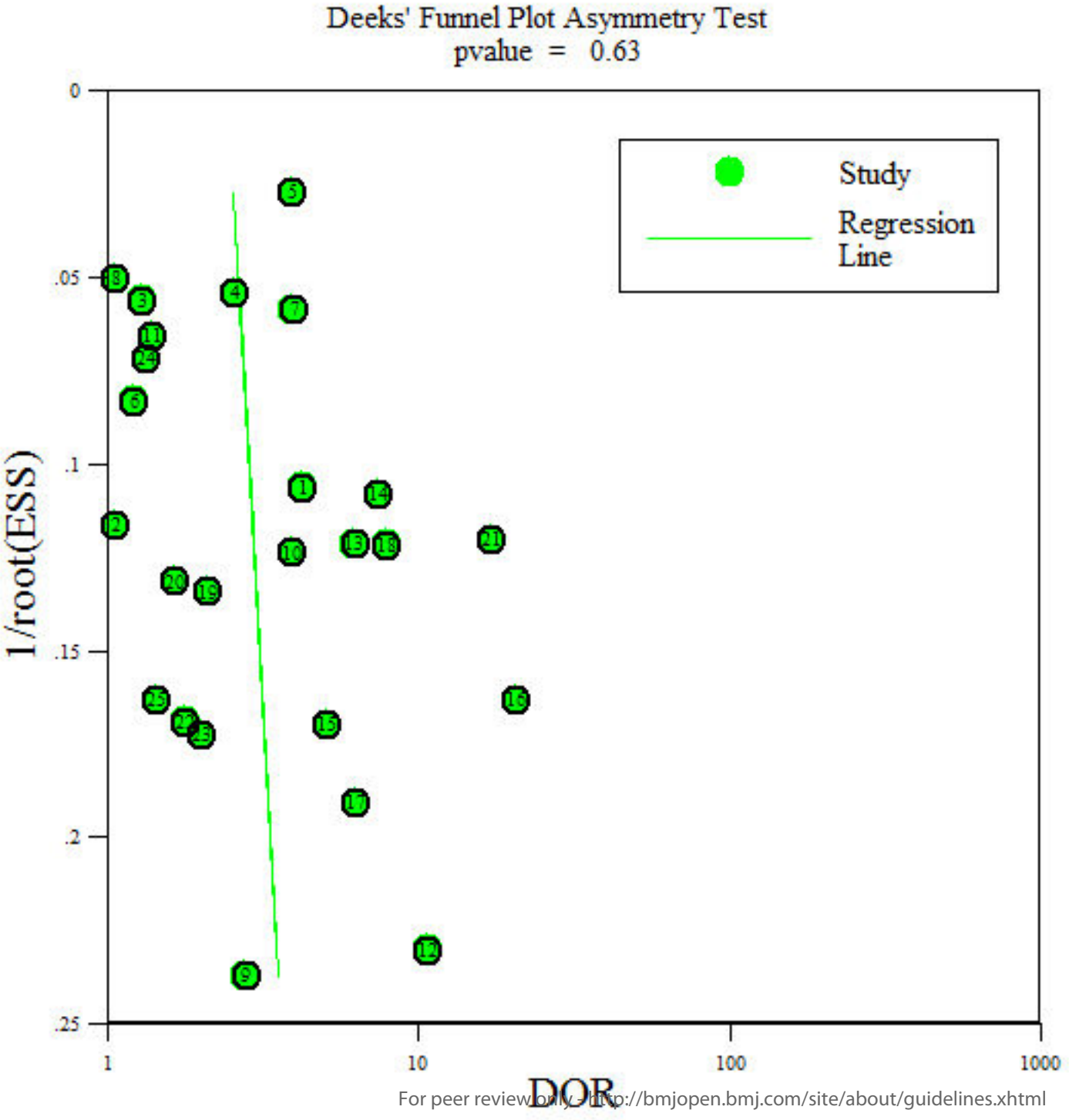
Study	Risk of bias				Applicability Concerns		
	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Li et al	☺	☺	☺	☺	☺	☺	☺
Campbell et al	☺	☺	☺	☺	☺	☺	☺
Aono et al	☺	☺	☹	☺	☺	☺	☹
Sam et al et al	☺	☺	?	?	☺	☺	☺
Umakoshi et al	☺	☺	☺	?	☺	☺	☺
Nanba et al	☺	☺	☺	☺	☺	☺	☺
Kamemura et al	☺	☺	☺	?	☺	☺	☺
Zhu et al	☺	☺	☹	☺	☺	☺	☹
Pedersen et al	☺	☺	?	☺	☺	☺	☺
Kocjan et al	☺	☺	☺	☺	☺	☺	☺
Asmar et al	☺	☺	☺	☺	☺	☺	☺
Riester et al	☺	☺	☺	☺	☺	☺	☺
Candy Sze et al	☺	☺	☺	☺	☺	☺	☺
Kűpers et al	☺	☺	?	?	☺	☺	☺
Lau et al	☺	☺	☺	☺	☺	☺	☺
Burton et al	☺	☺	☺	☺	☺	☺	☺
Salem et al	☺	☺	☺	☺	☺	☺	☺
Oh et al	☺	☺	☺	☺	☺	☺	☺
Sarlon-Bartoli et al	☺	☺	☺	☺	☺	☺	☺
Mathur et al	☺	☺	☹	☺	☺	☺	☺
Mulatero et al	☺	☺	☺	☺	☺	☺	☺
Minami et al	☺	☺	?	☺	☺	☺	☺
Nwariaku et al	☺	☺	?	?	☺	☺	☺
Young et al	☺	☺	☺	?	☺	☺	☺
Magill et al	☺	☺	☺	☹	☺	☺	☺

☺= low risk, ? = unclear risk, ☹=high risk






(a) Goodness-Of-Fit
Deviance Residual
Normal Quantile
0.00
0.25
0.50
0.75
1.00
(b) Bivariate Normality
Mahalanobis D-squared
Chi-squared Quantile
(c) Influence Analysis
Cook's Distance
study
0.00
0.50
1.00
1.50
0
5
10
15
20
25
(d) Outlier Detection
Standardized Residual(Diseased)
Standardized Residual(Healthy)
-3.0
-2.0
-1.0
0.0
1.0
2.0
3.0


Deeks' Funnel Plot Asymmetry Test
pvalue = 0.63
Study
Regression Line
1/root(ESS)
0
.05
.1
.15
.2
.25
1
10
100
1000
DOR

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2-3
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3-4
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	4-5
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	none
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	5
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	5
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	6
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	6
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	6
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	5
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	7
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	7

PRISMA 2009 Checklist

Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	7

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	7
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	7
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	7-8
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	8-9
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	10
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	9
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	9-10
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	10
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	10-11
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	11-13
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	13-14
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	14
FUNDING			

PRISMA 2009 Checklist

Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	18