Diagnostic accuracy of adrenal imaging for subtype diagnosis in primary aldosteronism: systematic review and meta-analysis

Yaqiong Zhou, Dan Wang, Licheng Jiang, Fei Ran, Sichao Chen, Peng Zhou, Peijian Wang

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 1 January 2000 to 1 February 2020, for all studies that used CT or 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% 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 four 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.

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. 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. 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 fourfold higher-risk profiles than controls, and titration of mineralocorticoid receptor antagonist therapy to raise renin might reduce this excess risk. 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 and, less commonly, unilateral adrenal hyperplasia. In contrast, the latter, also known as idiopathic hyperaldosteronism, is optimally treated with target medical therapy.\(^8\) 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 standardised procedures, high cost and lack of availability. Thus, it is urgent to explore alternative diagnostic methods without sacrificing accuracy.

Adrenal imaging with CT or 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 Reviews and Meta-Analyses Protocols.\(^11\) We searched the PubMed, EMBASE and Cochrane Library databases from 1 January 2000 to 1 February 2020, using the following terms in combination, as both Medical Subject Headings (MeSH) or Emtree terms and text words: “primary aldosteronism”, “adrenal vein sampling” and “hyperaldosteronism”. The electronic search strategy for PubMed is shown in online supplemental table S1. To reflect modern practice, we decided to limit the publication date to after 1 January 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 lateralisation 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 two independent investigators (YZ and PW) using a standardised 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 two independent reviewers (YZ and PW) 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 one domain, then it was judged as a low-quality study. If a study was judged to be 'unclear' in one 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% 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 OR (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.
Sources of statistical heterogeneity were explored by subgroup analyses, sensitivity analysis and meta-regression analysis, which involved the I² statistic; the following interpretation was applied for I²: <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. Contrast materials can improve the visibility of adrenal structures imaged by CT and MRI scans and might have a positive effect on diagnosis accuracy. 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. The Cohen κ test was employed to assess the inter-rater reliability between two 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 V.13.0 (StataCorp) and Review Manager V.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 (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. 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 (eight studies administered contrast material). Seventeen studies performed AVS with ACTH stimulation, seven studies performed it without ACTH stimulation, and the remaining one 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 four 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 five 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). In the meta-analysis, one included study used less permissive criteria for the LI and six included studies used less permissive criteria for the

Figure 1 Flow diagram of the review process. AVS, adrenal vein sampling.
The threshold of the SI was not accessible in four studies, and it was not accessible for the LI in two studies. Further details about the eligible and analysed studies are shown in Table 1 and online supplemental table S2.

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 two reviewers, with an inter-rater agreement of κ=0.9. Figure 2 summarises the QUADAS-2 assessment, and online supplemental table S3 displays each of the 25 individual QUADAS-2 evaluations.

The risk of bias from the reference standard was high in three studies, and it was unclear in five studies 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 six studies because the time interval between the index test and the reference standard was unclear; it was high in one study because not all patients had the same reference standard (Figure 2). Finally, 13 studies were considered to be high-quality studies, 7 were considered to be low-quality studies and 5 were unclear-quality studies.

Overall analysis

Using the bivariate model, statistical heterogeneity was found for sensitivity (I²=86.9%; p=0.001), specificity (I²=86.9%; p=0.00), the positive LR (I²=76.3%; p=0.00), the negative LR (I²=79.2%; p=0.00) and DOR (I²=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 and 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 with the traditional CT group (77% vs 58%). There was low heterogeneity detected on sensitivity in the CT/MRI group (I²=30%). However, the heterogeneity was high in all other groups, regardless of sensitivity or specificity (I² >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 vs LI was ≥2. However, a large degree of heterogeneity was observed in all groups (I² 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 (I²=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 (I² 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 (I² 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 four 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%) (online supplemental figure S1). Summary estimates for pooled measures of diagnostic accuracy are shown in Table 2.

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 (online supplemental figure S2).

Sensitivity analysis

Goodness-of-fit and bivariate normality analyses (online supplemental figure S3A,B) showed that the bivariate model was moderately robust. Influence analysis and outlier detection identified four outliers (online supplemental figure S3C,D). 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 (online supplemental figure S4).
| 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 | |
| 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 | |
| Umakoshi et al 2019 | 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 | |
| Riester et al 2014 | Germany | NA | 35 | 28 | CT or MRI | no | without ACTH, SI ≥2, LI ≥4 | ARR | Salt-loading test |
| Sze et al 2014 | UK | 42 (56) | 50.5 | 75 | CT | yes | with ACTH, SI ≥5, LI ≥4 | not ARR | salt-loading test |
| Kupers 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 |

Continued...
| Study, year          | Location | Male (%) | Age (years) | Sample | Imaging methodology | Contrast used | AVS characteristics | Screening test for PA (ng/ml) | Confirmatory test for PA |
|----------------------|----------|----------|-------------|--------|---------------------|---------------|---------------------|------------------------------|--------------------------|
| 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; five 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; ARR, aldosterone-to-renin ratio; AVS, adrenal vein sampling; LI, laterisation index; NA, not available; PA, primary aldosteronism; SI, selectivity index.
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, 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% to 94%, and the reported specificities ranged from 18% to 87%. Although the sensitivities reportedly exceeded 80% in five studies, relatively poor specificities were reported, with only one study showing a specificity of 72%. Similarly, three studies reported the sensitivities to be over 80%, but the specificities were reported to be lower than 76%. 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. 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 lateralisation 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 lateralisation 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 lateralisation 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 non-functioning 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

---

**Table 2** Pooled summary results by subgroups

| Subgroups          | No of studies | Sensitivity (95% CI) | Sensitivity with I², % | Specificity (95% CI) | Specificity with I², % |
|--------------------|---------------|----------------------|------------------------|----------------------|------------------------|
| **Total**          | 25            | 68 (61 to 74)        | 86.9%                  | 57 (50 to 65)        | 86.9                   |
| **Age**            |               |                      |                        |                      |                        |
| ≤40 years          | 4             | 71 (54 to 84)        | 53.1%                  | 79 (39 to 96)        | 70.1                   |
| >40 years          | 21            | 68 (60 to 75)        | 87.4%                  | 57 (49 to 64)        | 87.4                   |
| **AVS procedure**  |               |                      |                        |                      |                        |
| With ACTH          | 17            | 66 (57 to 73)        | 86.5%                  | 56 (46 to 65)        | 90.5                   |
| Without ACTH       | 7             | 70 (58 to 79)        | 90.3%                  | 60 (45 to 74)        | 79.2                   |
| **Cutoff values of LI** |         |                      |                        |                      |                        |
| LI≥2               | 4             | 61 (37 to 80)        | 95.5%                  | 54 (38 to 68)        | 88.3                   |
| LI≥4               | 18            | 69 (62 to 75)        | 76.4%                  | 59 (50 to 68)        | 89.0                   |
| **Screening test for PA** |       |                      |                        |                      |                        |
| ARR                | 21            | 66 (58 to 72)        | 87.7%                  | 58 (50 to 65)        | 86.2                   |
| Not ARR            | 4             | **78 (69 to 84)**    | 0%                     | 59 (29 to 84)        | **89.1**               |
| **Confirmatory test for PA** |     |                      |                        |                      |                        |
| Salt-loading test  | 12            | 71 (62 to 80)        | 78.6%                  | 60 (49 to 70)        | 74.9                   |
| Additional options | 8             | 57 (46 to 67)        | 90.6%                  | 66 (60 to 72)        | 66                     |
| No                 | 5             | 72 (64 to 79)        | 55.8%                  | 42 (24 to 63)        | 90.3                   |
| **Imaging methodology** |         |                      |                        |                      |                        |
| CT                 | 20            | 67 (59 to 74)        | 88.6%                  | 60 (53 to 67)        | 82.5                   |
| Contrast CT        | 8             | 77 (66 to 85)        | 86.4%                  | 60 (49 to 69)        | 82.2                   |
| Nocontrast CT      | 12            | 58 (49 to 66)        | 83.8%                  | 60 (51 to 68)        | 81.8                   |
| CT/MRI             | 5             | 69 (62 to 76)        | 30%                    | 45 (27 to 64)        | 87.9                   |
| **Quality of studies** |       |                      |                        |                      |                        |
| High-quality studies | 13        | 78 (73 to 83)        | 48.6%                  | 51 (39 to 63)        | 78.6                   |
| Unclear-quality studies | 6         | 62 (58 to 65)        | 0%                     | 62 (54 to 70)        | 85.1                   |
| Low-quality studies | 6            | 44 (38 to 50)        | 46.4%                  | 69 (60 to 78)        | 69.2                   |
| **Sample size**    |               |                      |                        |                      |                        |
| ≥100               | 10            | 59 (51 to 67)        | 90.6%                  | 58 (49 to 66)        | 92.1                   |
| <100               | 15            | 74 (67 to 81)        | 71.3%                  | 60 (47 to 71)        | 78.1                   |
| Outlier excluded   | 20            | 65 (60 to 70)        | 77.1%                  | 59 (52 to 66)        | 85.2                   |

ACTH, adrenocorticotropic hormone; ARR, aldosterone-to-renin ratio; AVS, adrenal vein sampling; LI, lateralisation index; PA, primary aldosteronism.
skipped. Among studies included in the present meta-analysis, four reported the diagnostic accuracy of CT/MRI in identifying unilateral PA in patients ≤40 years. By combining these four 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 adenaoma 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 lateralisation 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.

Contributors PW is the guarantor. All authors provided substantial contribution to conception and design of the project; drafted and revised the manuscript. YZ led the literature search, and completed the study selection, data extraction and critical appraisal with PW. LJ accepts responsibility for the integrity of the data analyses. FR led the drafting of all sections of the article in consultation with all of the coauthors. DW, SC and PZ provided substantial contributions to the background, critical appraisal of prior studies and interpretation of meta-analysis findings. PZ 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 statement Patients and the public were not involved in the design or conduct of the study.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. Extra data can be accessed via the Dryad data repository at http://datadryad.org/ with the doi: 10.5061/dryad.jm63xj8t.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iD Yaqiong Zhou http://orcid.org/0000-0002-9441-0291

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 G, 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, Schwarzw 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;6:51–9.
5 Funder JW. Primary aldosteronism: where are we now? Where to from here? Horm Metab Res 2020;52:459–66.
exploring sources of Chu H, Guo H, Zhou Y. Bivariate random effects meta-analysis of Liberati A, Altman DG, Tetzlaff J, Kempers MJE, Lenders JWM, van Outheusden L, Ruco A, Dossa F, Tinmouth J. Roles of clinical criteria, diagnostic test accuracy was assessed. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 2005;58:982–90. Ahrens W, Guo H, Zhou Y. Bivariate random effects meta-analysis of diagnostic studies using generalized linear mixed models. Med Decis Making 2010;30:499–508. Ljungcr JG, Bossuyt PMM, Heisterkamp SH. Exploring sources of heterogeneity in systematic reviews of diagnostic tests. Stat Med 2002;21:1525–37. 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;112:144–8. Pedersen M, Karlsen MA, Ankjærgaard KL, Asmar M, Wachtel H, Yan Y, et al. Adrenal venous sampling with or without adrenocorticotropic hormone stimulation: a meta-analysis. J Clin Endocrinol Metab 2019;104:1060–8. Laurent I, Astère M, Zheng F, et al. Adrenal venous sampling as a diagnostic procedure for primary hyperaldosteronism: experience from a tertiary referral centre. Hormones 2012;11:151–19. Burton TJ, Mackenzie IS, Balian K, et al. Evaluation of the sensitivity and specificity of [11C]metomidate positron emission tomography (PET)-CT for lateralizing aldosterone secretion by Conn's adenomas. J Clin Endocrinol Metab 2012;97:100–9. Kocjan T, Janez A, Stankovic M, et al. A new clinical prediction criterion accurately determines a subset of patients with bilateral primary aldosteronism before adrenal venous sampling. Endocr Pract 2016;22:587–94. Zhu L, 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. J Hum Hypertens 2016;31:195–9. Kamemura K, Wada N, Ichijo T, et al. Significance of adrenal computed tomography in primary aldosteronism and in the guidance of adrenalectomy. Medicine 2016;95:e4986. 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. Zhu L, 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. Kamemura K, Wada N, Ichijo T, et al. Significance of adrenal computed tomography in primary aldosteronism and in the guidance of adrenalectomy. Medicine 2016;95:e4986. Kamemura K, Wada N, Ichijo T, et al. Significance of adrenal computed tomography in primary aldosteronism and in the guidance of adrenalectomy. Medicine 2016;95:e4986. Kamemura K, Wada N, Ichijo T, et al. Significance of adrenal computed tomography in primary aldosteronism and in the guidance of adrenalectomy. Medicine 2016;95:e4986.