MANAGEMENT OF ENDOCRINE DISEASE

Imaging for the diagnosis of malignancy in incidentally discovered adrenal masses: a systematic review and meta-analysis

Jacqueline Dinnes[1,*], Irina Bancos[2,3,*], Lavinia Ferrante di Ruffano[1], Vasileios Chortis[2], Clare Davenport[1], Susan Bayliss[1], Anju Sahdev[4], Peter Guest[5], Martin Fassnacht[6,7], Jonathan J Deeks[1,†] and Wiebke Arlt[2,8,†]

[1]Institute of Applied Health Research, [2]Institute of Metabolism and Systems Research, University of Birmingham, Birmingham, UK, [3]Division of Endocrinology, Metabolism, Nutrition and Diabetes, Mayo Clinic, Rochester, Minnesota, USA, [4]Department of Imaging, St Bartholomew's Hospital, Barts Health, London, UK, [5]Department of Radiology, Queen Elizabeth Hospital, University Hospital Birmingham NHS Foundation Trust, Birmingham, UK, [6]Department of Internal Medicine I, Division of Endocrinology and Diabetes, University Hospital Würzburg, University of Würzburg, Würzburg, Germany, [7]Comprehensive Cancer Center Mainfranken, University of Würzburg, Würzburg, Germany, and [8]Centre for Endocrinology, Diabetes and Metabolism, Birmingham Health Partners, Birmingham, UK,

*(J Dinnes and I Bancos contributed equally to the research and are joint first authors)

†(J J Deeks and W Arlt are joint last authors)

Correspondence should be addressed to W Arlt

Email

w.arlt@bham.ac.uk

Abstract

Objective: Adrenal masses are incidentally discovered in 5% of CT scans. In 2013/2014, 81 million CT examinations were undertaken in the USA and 5 million in the UK. However, uncertainty remains around the optimal imaging approach for diagnosing malignancy. We aimed to review the evidence on the accuracy of imaging tests for differentiating malignant from benign adrenal masses.

Design: A systematic review and meta-analysis was conducted.

Methods: We searched MEDLINE, EMBASE, Cochrane CENTRAL Register of Controlled Trials, Science Citation Index, Conference Proceedings Citation Index, and ZETOC (January 1990 to August 2015). We included studies evaluating the accuracy of CT, MRI, or ^{18}F-fluoro-deoxyglucose (FDG)-PET compared with an adequate histological or imaging-based follow-up reference standard.

Results: We identified 37 studies suitable for inclusion, after screening 5469 references and 525 full-text articles. Studies evaluated the accuracy of CT ($n = 16$), MRI ($n = 15$), and FDG-PET ($n = 9$) and were generally small and at high or unclear risk of bias. Only 19 studies were eligible for meta-analysis. Limited data suggest that CT density >10 HU has high sensitivity for detection of adrenal malignancy in participants with no prior indication for adrenal imaging, that is, masses with ≤10 HU are unlikely to be malignant. All other estimates of test performance are based on too small numbers.

Conclusions: Despite their widespread use in routine assessment, there is insufficient evidence for the diagnostic value of individual imaging tests in distinguishing benign from malignant adrenal masses. Future research is urgently needed and should include prospective test validation studies for imaging and novel diagnostic approaches alongside detailed health economics analysis.

Introduction

An incidentally discovered adrenal mass is a frequent occurrence, serendipitously discovered in around 5% of cross-sectional abdominal imaging carried out for purposes other than a suspected adrenal problem (1, 2, 3). Due to the increasingly widespread use of cross-sectional imaging, adrenal incidentalomas represent a significant challenge to health care budgets. The rates of computed tomography (CT) scans carried out in the USA soared from 3 million per annum in 1980 to 81.2 million in 2014 (4). Concurrently, in the UK, 5 million CT scans were undertaken in 2012/2013, increasing from 1 million in 1996/1997 (www.england.nhs.uk/statistics/statistical-work-areas/diagnostics-waiting-times-and-activity/imaging-and-radiodiagnostics-annual-data/). The use of repeated and multiple modality imaging in adrenal incidentalomas represents a major challenge to health care budgets and a burden to patients affected. Therefore, evidence-based guidance on the use of imaging in adrenal incidentalomas is urgently needed.

Prevalence of adrenal incidentalomas increases with age (3% at 40 years, 10% at 70 years) (5), and is very low in children (<0.5%) (6). A key consideration for the diagnostic workup of adrenal incidentalomas is whether the adrenal mass is hormone-producing, requiring exclusion of pheochromocytoma, Cushing syndrome, and, in hypertensive patients, primary aldosteronism. Second, and usually perceived as most important by the affected patient, the possibility of malignancy has to be considered.

In patients with a history of extra-adrenal malignancy, the detection of a new adrenal mass raises suspicion of metastasis, but also requires careful exclusion of other causes. In cancer patients, the likelihood of an adrenal nodule being malignant is approximately 20%; eventually, only 70% of adrenal lesions surgically removed on the basis of imaging results are confirmed as metastasis by histology (7, 8, 9).

While the detection of adrenal metastasis is a rarity in adrenal incidentaloma patients who do not have a history of extra-adrenal malignancy, the discovery of an adrenocortical carcinoma (ACC) is not uncommon. Larger clinical and surgical adrenal incidentaloma series report an ACC prevalence of 1.4–12% (2, 10, 11, 12), with variability mostly driven by referral bias. Radiological studies describe lower rates of malignant and functionally active adrenal tumors, but usually lack uniform endocrine evaluation and an optimal reference standard such that malignant lesions could be missed (3).

An adrenal incidentaloma is most frequently noted on CT or MRI scans carried out for other purposes. Both imaging modalities can assess the lipid content in the adrenal mass, which serves as the basis for differentiating between a benign (high lipid content) and a potentially malignant (low lipid content) adrenal mass. However, at least a third of benign adrenal adenomas have been shown to be lipid-poor (13, 14). This lack of specificity causes many patients to undergo multiple scans and imaging modalities, often followed by surgery, with histology ultimately revealing a benign mass that would not have required surgery in 30–55% of patients (2, 15).

In addition to the general radiological criteria of size of the mass and its appearance (heterogeneity, borders, invasion) (13, 16), multiple imaging parameters are employed for the differential diagnosis of adrenal incidentaloma. These include unenhanced CT with assessment of tumor density, contrast-enhanced timed washout CT studies, MRI chemical shift analysis, and, more recently, ^{18}F-fluoro-deoxyglucose (FDG)-PET (FDG-PET) in combination with CT (PET-CT).

However, despite their widespread use in the workup of adrenal incidentalomas, the optimal choice, sequence and performance of imaging tests to distinguish benign from malignant adrenal masses is unclear (17), and clinical practice remains more expert-based than evidence-based. Individually, published reports are often unconvincing due to small sample sizes, heterogeneity of included populations and different imaging techniques or cut-offs as well as poor reference standards. Due to this, many patients with adrenal tumors undergo multiple scans, annual follow-up imaging and even unnecessary surgery (2), with previous guidelines and reviews requesting annual follow-up imaging for up to 2 years in most adrenal incidentaloma patients not undergoing surgery (16, 17, 18).

We have carried out a systematic review and meta-analysis of the diagnostic performance of imaging tests in incidentally discovered adrenal masses, with the aim of facilitating evidence-based recommendations on the effective use of imaging in adrenal incidentalomas. With advances in the evaluation of diagnostic test accuracy increasing the awareness of potential sources of bias (19, 20, 21, 22), as well as summarizing study findings, we provide insights into the validity and applicability of the available evidence base and identify current limitations.

Methods

This review follows methods as set out in the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy (23) and reporting standards set in the

Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) statement (24). This paper reports on the accuracy of CT, MRI, and FDG-PET or PET-CT at commonly used thresholds for the diagnosis of malignant adrenal masses in individuals with incidentally identified lesions, including those identified in individuals with known malignancy.

Data sources and searches

MEDLINE (Ovid), MEDLINE In Process (Ovid), EMBASE (Ovid), Cochrane CENTRAL Register of Controlled Trials and Cochrane Database of Systematic Reviews, Science Citation Index, and Conference Proceedings Citation Index (Web of Science) and ZETOC (British Library) databases were searched by an Information Specialist (SB) for titles published between 1990 and 13 August 2015. Studies published before 1990 were not considered to be representative of current imaging technologies. The full search strategy as designed for MEDLINE is available in Supplementary Table 1, see section on supplementary data given at the end of this article. The reference lists of included studies and relevant systematic reviews were reviewed for additional eligible studies.

Study selection

We considered all studies of CT, MRI, or FDG-PET in adult participants with incidentally identified adrenal masses for inclusion. These included both patients in whom imaging for any indication other than an adrenal mass led to the detection of an adrenal mass (true adrenal incidentalomas) and patients with an adrenal mass detected by imaging carried out for staging or follow-up of extra-adrenal malignancy. Studies that did not report the original indication for imaging are reported, but were not included in the meta-analyses.

The target condition of interest was the detection of adrenocortical carcinoma (ACC) or adrenal metastases from an extra-adrenal primary malignancy. We included all studies with reference standards where i) at least 50% of participants with ACC or a malignant adrenal mass had a histologically proven reference standard diagnosis (obtained either through adrenalectomy or adrenal biopsy) and ii) at least 50% of those with a benign adrenal mass had their final diagnosis reached by either histology or imaging-based follow-up of any duration.

In collaboration with clinical and radiological experts from the European Society of Endocrinology (ESE) and European Network for the Study of Adrenal Tumors (ENSAT) Clinical Practice Guideline Committee for the management of adrenal incidentalomas, we selected five commonly used diagnostic imaging thresholds for inclusion: (i) non-contrast CT: tumor density measured in Hounsfield units (HU) >10; (ii) contrast-enhanced CT washout studies: absolute percentage washout (APW) and/or relative percentage washout (RPW) at any washout percentage or delay time on enhanced CT; (iii) MRI chemical shift analysis: loss of signal intensity between in and out of phase images (including both qualitative and quantitative estimates of signal loss); and, for FDG-PET or PET-CT, (iv) the maximum standardized uptake value (SUV_{max}); and (v) the ratio of SUV_{max} in the adrenal gland compared with the liver (adrenal liver ratio (ALR)).

We excluded studies where more than half of participants presented with endocrine symptoms, or were otherwise suspected of hormone excess, and those concerned with the diagnosis of adrenomedullary tumors; pheochromocytomas can usually be detected by measuring plasma or urinary metanephrines and their imaging characteristics overlap with those observed in adrenocortical malignancy and adrenal metastases. Therefore, studies with more than 30% pheochromocytomas in the disease-positive group were excluded, unless data could be disaggregated to allow their exclusion from the analysis. We also excluded studies in pediatric populations, sample size <10, data collection before 1990, and with insufficient data presented to allow the construction of a 2×2 diagnostic contingency table. Non-English language studies and studies only reported in conference abstracts were excluded.

Title and abstract screening and full-text inclusion was carried out independently by two reviewers (I B, J Di,). Any disagreements were resolved through discussion or referral to a third reviewer (C D, V C, L F R).

Data extraction and quality assessment

Data extraction was carried out independently by at least two authors (I B, J Di, L F R, V C, C D) using a standardized and piloted data extraction form. Details of the study design, participants, lesion characteristics, index test(s) or test combinations and index test positivity thresholds, reference standards, and 2×2 diagnostic contingency table data were extracted. Any malignant masses detected in addition to ACC or adrenal metastases (malignant pheochromocytomas, other malignant medullary tumors or other malignancy) were considered disease positive, as

their clinical management is sufficiently similar. If study data could not be fully disaggregated, the malignant group could include up to 10% benign masses and up to 10% of the benign group could include medullary tumors (pheochromocytomas, neuroblastoma, ganglioneuroma, or schwannoma). Discrepancies in data extraction were resolved by consensus or by a third reviewer.

We considered the risk of bias and concerns about the applicability of findings related to the patients, tests, reference standard, and execution of each study, using the QUADAS-2 checklist (19), tailored to the review topic. Three authors (I B and V C plus J Di or L F R) independently rated each study with disagreement resolved by consensus.

Patient selection was regarded at risk of bias if consecutive or random selection was not used, patients were selected according to presence of adrenalectomy data, or patients were inappropriately excluded based on previous lesion assessments. Test and reference standard implementation were considered at risk of bias when each was undertaken with knowledge of the other, or when test thresholds were not prespecified, and when final diagnoses of malignancy were not all based on histology or tumor sampling was inadequate, or benignity was assumed without histology or <12 months imaging follow-up. Non-blinded interpretation of other imaging tests added to bias in test interpretation. Risk of bias in the execution of the study was considered when reference standards were not undertaken in all patients, when participants were excluded from analyses, when the reference standards used in malignant or benign cases varied, or when there was no follow-up of suspected benign cases within 6 months.

Concerns about applicability were noted for participants when <90% were recruited with incidentally discovered adrenal tumors or having known or prior malignancy; for tests, when inadequate detail of the test measure was given to allow replication or standard thresholds were not used; and when the reference standard did not allow full disaggregation of the tumor types into malignant and benign.

Data synthesis and analysis

Data synthesis focused on estimating the accuracy of each test for diagnosis of malignancy for separate clinical pathways for (i) adrenal incidentaloma, that is, investigation of an adrenal tumor detected by imaging carried out for an indication other than suspected adrenal disease and for (ii) history of extra-adrenal malignancy, that is, imaging evaluation or staging in patients with known or prior non-adrenal malignancy. It was considered possible that the accuracy of each test may differ between these clinical pathways. Each study was characterized according to whether the majority (>50%) or nearly all (>90%) individuals were recruited in each pathway, and separate analyses were undertaken for each group. Studies that did not meet these criteria or where the reasons for imaging could not be ascertained were excluded from the analysis. For analysis of MRI chemical shift we restricted inclusion to studies using 1.5 Tesla machines, which were the majority.

Estimates of sensitivity and specificity and 95% CIs for the detection of malignancy were calculated using the binomial exact method when there was only one study, or when there were no false negatives or false positives. Otherwise, the bivariate hierarchical model was used to obtain meta-analytical estimates of average sensitivity and specificity (25). Where possible, the model included terms for random effects for sensitivity and specificity and their correlation, but was simplified when inadequate numbers of studies were available (26).

Results

Characteristics of included studies

A total of 5496 unique references were identified and screened for inclusion. Of these, 525 full-text papers were reviewed and 37 studies were included (Fig. 1A) (7, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62). Studies were primarily excluded due to lack of test accuracy data (167 studies), inadequate reference standards (93 studies), and ineligible populations (86 studies). A further 17 studies did not present their data in accordance with our review question and it was not possible to disaggregate their results to allow their inclusion (i.e. >30% pheochromoytomas in the malignant group (*n*=8), >10% medullary tumors (*n*=6) or any malignant mass (*n*=2) in the benign group, and >10% benign masses in the malignant group (*n*=1)) and 11 studies were excluded as they did not use any of our preselected diagnostic thresholds (Supplementary Table 2).

Summary study characteristics are presented in Table 1. CT was evaluated in 16 studies (non-contrast CT was evaluated in 13 studies, contrast-enhanced CT washout studies in 6 studies), MRI in 15 studies, and PET in 9 studies. Studies were generally small with a median
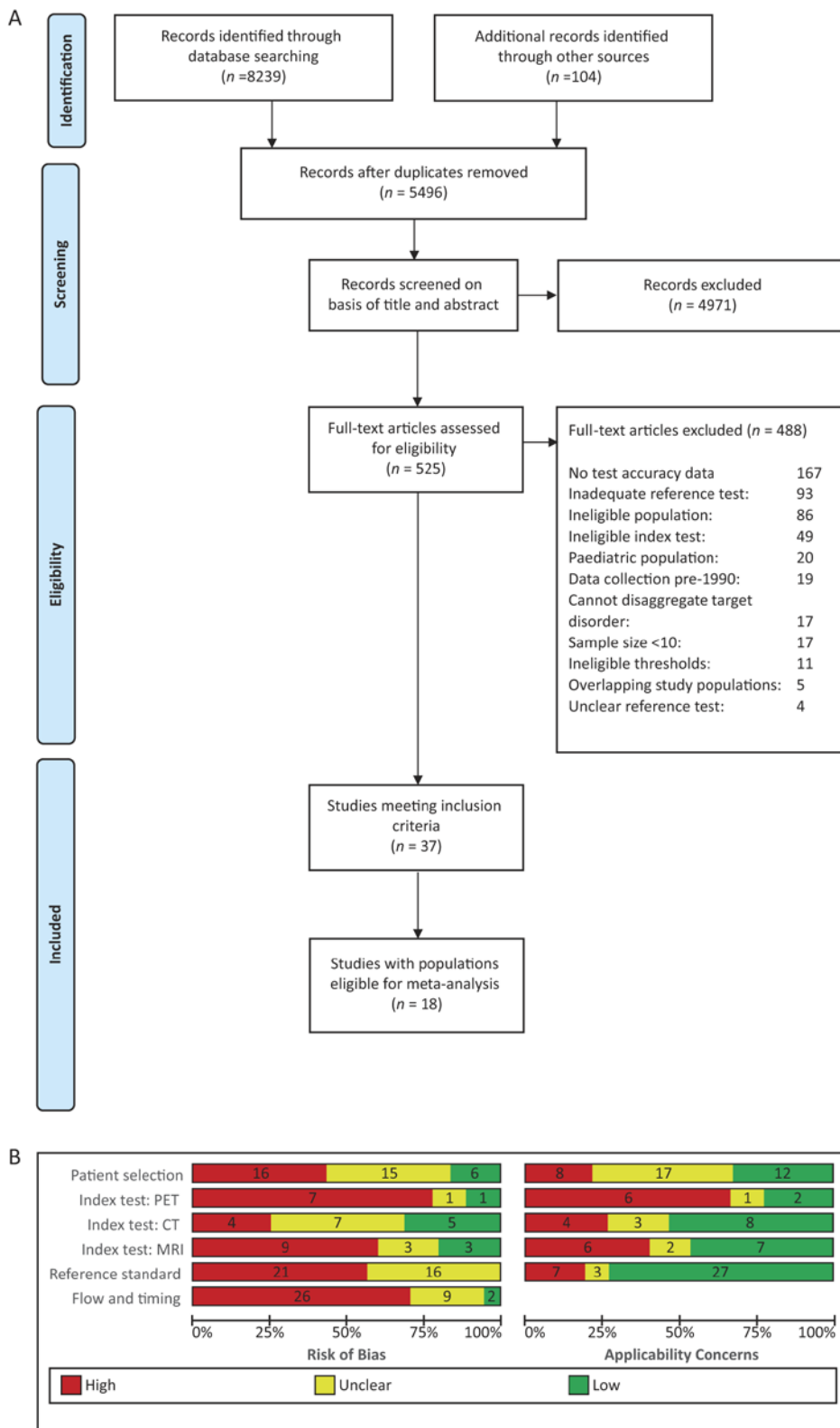
Figure 1

(A) PRISMA flow diagram (adapted from Moher 2009 (24)). (B) Summary risk of bias and concerns about applicability (based on adapted QUADAS-2 (19)).

Table 1 Summary of the characteristics of the 37 studies fulfilling the inclusion criteria.

Characteristic	Studies (n)	(%)
Study design		
Prospective case series	9	(24)
Retrospective case series	19	(51)
Diagnostic case–control (two-gate series)	4	(11)
Design unclear	5	(14)
Population characteristics		
Sample size (participants)	*50.4	†(12–181)
Sample size (lesions)	*52.3	†(14–146)
Prevalence of malignancy (%)	*38.1	†(13–74)
Mean age (years; 29 studies)	‡*55.8	†(44.1–66.7)
Female participants (%; 31 studies)	*49.4	†(6–87)
Mean tumor size (mm; 24 studies)	‡*41.9	†(22–68.1)
Mean % symptomatic participants (5 studies)	*36	†(26–47)
Confirmed hormone excess (%; 11 studies)	*36.3	†(0.02–88)
Index tests and thresholds		
CT	16	(43)
Non-enhanced tumor density	*13*	*(35)*
Contrast-enhanced washout studies	*6*	*(16)*
MRI	15	(41)
Chemical shift loss of signal intensity	*8*	*(22)*
Adrenal to liver ratio signal intensity	*8*	*(22)*
Adrenal to spleen ratio signal intensity	*5*	*(14)*
Adrenal to muscle ratio signal intensity	*2*	*(5)*
PET	4	(11)
SUV_{max}	*3*	*(8)*
SUV_{max} adrenal to liver ratio	*5*	*(14)*
PET-CT	5	(14)
SUV_{max}	*3*	*(8)*
ALR SUV_{max}	*4*	*(11)*
Population grouping for analysis		
Initial finding incidental in ≥90% included participants	3	(8)
Initial finding incidental in 50–90% included participants	4	(11)
Initial indication for imaging due to known cancer in ≥90% included participants	9	(24)
Initial indication for imaging due to known cancer in 50–90% included participants	2	(5)
Initial finding incidental in <50% OR <50% imaging indication known cancer	2	(5)
Population composition not reported	17	(46)
Reference standard		
Histology alone	16	(43)
Histology and imaging follow-up	15	(41)
Histology and imaging follow-up, plus other reference	5	(14)
Histology plus other reference	1	(3)

SUV_{max}, maximum standardized uptake value; ALR SUV_{max}, ratio of SUV_{max} in the adrenal gland compared with the liver.
*Mean; †Range; ‡Mean of reported means.

sample size of 45 (range 12–181) and less than a third were prospective in design (n=10, 27%). Mean prevalence of malignancy was 38% (range 13–74%). Most were conducted in Europe (n=15, 41%) and North America (n=12, 32%). Datasets for a single imaging test were included from most studies (n=34, 92%) compared with a reference standard of histology alone (n=14, 38%; excision or biopsy sample) or a mixed reference standard of histology or imaging follow-up (n=14, 38%). Reported follow-up periods ranged from 6 to 24 months. Of the papers

reporting participant recruitment dates (n=27, 76%), most were conducted between 2000–2005 (n=12, 32%) and 2005–2009 (n=9, 24%).

Where reported, study populations were highly varied, with only 7 studies (19%) including a majority of participants with purely incidental findings and 11 (29%) focusing primarily on participants with known extra-adrenal malignancy (>50% of population) (Table 1). Studies variously excluded masses with particular imaging characteristics including CT HU<10 (n=3), size <10 mm

(n=11), pheochromocytomas (n=15), or functioning masses (n=5). Patients with symptoms of hormone excess were explicitly included in five studies and confirmed hormone excess following imaging was identified in around a third of participants (mean 36%, range 0–88%; n=11 studies). Mean tumor size ranged from 19 mm (42) to 68 mm (55).

Study quality

The vast majority (84%) of studies were at high or unclear risk of bias across all quality domains assessed (Fig. 1A and Supplementary Figs 1, 2, 3). A third of studies (n=12) only included participants selected for adrenalectomy and therefore at higher risk of malignancy, and four adopted a case–control type approach with separate selection of those with confirmed malignancy and benign disease (33, 39, 49, 53). PET (Supplementary Fig. 3) and MRI evaluations (Supplementary Fig. 2) were at particular risk of bias due to retrospective selection of the diagnostic threshold (in 8/9 and 6/15 evaluations, respectively), potentially leading to inflated estimates of test accuracy. Test interpretation could have been influenced by the same observer interpreting more than one imaging test in the same study (affecting 14 of 40 evaluations). Test interpretation was blinded to the reference standard diagnosis in around half of all test evaluations (52%; 21/40) and differential verification was present in 62% (23/37). More than half of studies used an inadequate reference standard either due to the use of biopsy rather than full excision of malignant masses (n=15, 40%) or imaging follow-up of <12 months (n=6, 16%). Concerns around the applicability of study results were high (n=8) or unclear (n=17) due to varying or unclear indications for imaging in the included populations and due to the evaluation of a new threshold, not previously assessed in other studies (present or unclear in 21 of 40 test evaluations).

Results according to clinical pathway

Poor reporting of the clinical pathways leading to the conduct of the imaging tests resulted in exclusion of 19 of 37 eligible studies from analysis (described in Supplementary Table 3). Characteristics of the 18 studies eligible for analysis are provided according to clinical pathway in Table 2 and results of test performance are reported in Table 3, with raw data for all test evaluations provided in Supplementary Table 4.

Test performance in the investigation of incidentally detected tumors

Seven studies presented data on test performance (two for CT (27, 30), three for MRI (28, 31, 46), and two for PET-CT (29, 61)) in patient groups presenting with more than 50% (and two with >90%) incidentally detected tumors. Two studies evaluating tumor density >10 HU on non-contrast CT (27, 30), and one evaluating CT contrast-enhanced washout tests (27) showed high sensitivity and specificity. Only two (28, 31) of the three studies of MRI used 1.5 Tesla machines and reported slightly lower sensitivity and specificity than CT for measures of adrenal-liver and adrenal-spleen ratios and loss of signal intensity. The performance of PET for ALR and SUV_{max} measures was no better than CT.

The data suggest that CT density >10 HU has high sensitivity for the detection of malignancy, the 95% CI suggesting that this is above 90%. However, all other estimates of test performance are based on small numbers of studies with few patients, and 95% CIs are notably wide, indicating uncertainty in test performance for all other imaging markers. It is not possible to discern from the available data whether any test performs adequately or better than alternative tests.

Test performance in the investigation of tumors in participants with current or prior non-adrenal malignancy

Eleven studies presented data on test performance (five for CT (7, 33, 34, 35, 37), five for MRI (32, 34, 35, 36, 60), and three for PET-CT (8, 38, 62)) in patient groups presenting with more than 50% (and 9 with >90%) tumors detected in patients undergoing imaging following previous non-adrenal malignancy. The five studies evaluating CT density >10 HU on non-contrast CT (7, 33, 34, 35, 37) showed high sensitivity (93%) but variable specificity; CT contrast-enhanced washout tests were only reported in one study (33), which showed very low sensitivity (16%). Four (32, 34, 36, 60) of the five studies of MRI used 1.5 Tesla machines and reported high sensitivity (89–99%) for measures of adrenal-liver, adrenal-spleen, adrenal-muscle ratios and loss of signal intensity. Specificity varied (60–93%) but was high for most MRI measures. The performance of PET was similar to MRI for ALR and SUV_{max} measures.

Although more studies had evaluated CT, MRI, and PET in the pathway for follow-up of known malignancy than for incidentally discovered adrenal lesions, estimates of test performance are still based on too small numbers

Table 2 Characteristics of the 18 studies eligible for meta-analysis.

Reference	Index test(s)	Study design	Population	Exclusions HU (HU)	Exclusions Size (mm)	Exclusions Other	Pat./Les. (n)	I/KM/S (%)	HE (%[a])	Reference standard[b] (%)	Dis n; %	No. ACC	No. Mets	Threshold for malignancy
Studies investigating incidentally detected tumors (n=7)														
(27)	CT	BPC P	Imaging series, ≥50% incidental	No	<10, >120	Cysts, myelolipoma	50/50	74/0/26	NR	42/58/0	28; 56%	7	13	1. >10HU 2. APW <60% OR RPW <40% at 10' 3. APW <60% OR RPW <40% at 15'
(46)	MRI	NC R	Imaging series, ≥90% incidental	No	<10	No	59/66	100/0/0	NR	35/55/11	17; 26%	5	11	1. SII ≤23% (OP/IP dataset)[l]
(31)	MRI	NC P	Imaging series, ≥50% incidental	NR	NR	Functioning masses; Pheos (n=4) excluded by Bham team	30/30	66/33/0	0	63/37/0	8; 31%	4	3	1. ALR – qualitative* 2. SI – qualitative‡
(29)	PET	WPC R	Imaging series, ≥50% incidental	<10	NR	Pheos; prior cancer; ACC on CT; eventual washout of contrast >50% on CT	23/23	65/0/35	43	100/0/0	3; 13%	2	0	1. ALR maxSUV >1.6§ 2. maxSUV >3.4
(28)	MRI	NC R	Imaging series, ≥50% incidental	No	<10	Myelolipoma; cysts; artefacts on diffusion weighted imaging; lack of adequate reference	48/49	69/31/0	2	38/63/0	12; 24%	1	9	1. ASR ≥62 (ADC)[¶] 2. SII ≤23% (ADC)[l]
(62)	PET	WPC P	Imaging series, ≥90% incidental	<10	No	Functioning masses; washout on delayed enhanced CT, decrease of signal intensity on CS MRI	37/41	100/0/0	0	71/29/0	12; 29%	3	4	1. ALR maxSUV >1.8§ 2. maxSUV >3.28
(30)	CT	NC R	Imaging series, ≥90% incidental	NR	NR	None reported	52/52	100/0/0	25%	38/40/17	13; 25%	2	5	1. >10HU
Studies investigating tumors in participants with current or prior non-adrenal malignancy (n=11)														
(32)	MRI	NC P	Operable NSCLC, ≥90% known malignancy	NR	No	None reported	27/27	0/100/0	NR	100/4/0	4; 16%	0	5	1. ALR qualitative*
(33)	CT	WPC R	Imaging series, ≥90% known malignancy	No	No	All diagnoses other than adenoma and metastasis	36/40	0/100/0	NR	100/30/0	19; 48%	0	19	1. >10HU 2. APW at 15' <60% 3. RPW at 15' <40%
(38)	PET-CT	NC R	Imaging series, ≥50% known malignancy	NR	NR	Symptomatic tumors; Contraindications to PET;	15/15	0/53/47	NR	87/13/0	11; 73%	3	5	1. ALR maxSUV >1.8§ 2. maxSUV >6

Ref	Modality	Data	Population description			n/N	Dist.		Dist. (%)	%;		n	Thresholds
(7)	CT	WPC P	Adrenalectomy series, ≥90% known malignancy	No	≤60	Evidence of extra-adrenal tumor spread	42/44	0/100/0	0	100/0/0	31; 70%	0	31	>10HU
(62)	PET-CT	WPC R	Imaging series, ≥90% known malignancy	No	No	Functioning masses	85/104	0/100/0	NR	100/0/0	32; 31%	1	30	1. ALR maxSUV >1.53§ 2. maxSUV >5.2
(8)	PET-CT	NC R	Adrenalectomy series, ≥90% known malignancy	No	No	Functioning masses; no clinical suspicion of metastasis (based on CT findings)	39/39	0/100/0	0	100/0/0	29; 74%	0	28	1. ALR maxSUV >1.29§ 2. maxSUV >3.7
(34)	CT MRI	WPC P	Imaging series, ≥90% known malignancy	NR	<10	Pheos	33/37	0/100/0	NR	51/46/0	19; 51%	0	18	CT: >10HU MRI: ASR ≥75¶
(35)	CT MRI	WPC P	Operable NSCLC, ≥90% known malignancy	NR	<10	Pheos	32/32	0/100/0	NR	100/44/0	18; 56%	0	18	CT: >10HU MRI: ALR qualitative†
(36)	MRI	NC R	Imaging series, ≥50% known malignancy	No	<8	Myelolipoma; cysts; nonstandardised imaging protocol; Lack of adequate reference	36/37	NR/78/NR	NR	19/76/5	10; 28%	0	8	1. ALR >0.674†† 2. ASR >64.1‡‡ 3. AMR >70.7§§
(60)	MRI	NC P	Biopsy referrals, ≥90% known malignancy	NR	NR	None reported	68/68	0/100/0	NR	71/29/0	23; 34%	NR	NR	1. ALR ≥1.5** 2. ASR ≥55¶
(37)	CT	WPC R	Imaging series, ≥90% known malignancy	NR	NR	Grades 4 or 5 disease; bleeding tendency and coagulopathy	12/16	0/100/0	NR	93/0/7	6; 40%	0	6	1. >10HU

ACC, adrenocortical carcinoma; ADC, apparent diffusion coefficient; ALR, adrenal to liver ratio; AMR, adrenal to muscle ratio; APW, absolute percentage washout; ASR, adrenal to spleen ratio; BPC, between–person comparison (multiple index tests evaluated in partial study population); CS, chemical shift; CT, computed tomography; Dis, diseased; HE, hormone excess; HU, Hounsfield units; I, incidental; IP, in-phase; KM, known malignancy; les, lesions; mets, metastases; NC, non-comparative study; NR, not reported; OP, opposed phase; P, prospective data collection; Pat., patients; PET, positron emission tomography; R, retrospective data collection; RPW, relative percentage washout; S, symptomatic; SI, signal intensity; SII, signal intensity index; SUV_{max} maximum standardized uptake value; WPC, within–person comparison (multiple index tests evaluated in all study participants).

*masses considered to be malignant if their signal was more intense than liver signal; †masses considered to be malignant if their signal was more intense than liver signal and inferior to kidney signal; ‡masses considered to be malignant if no loss of signal intensity observed on chemical shift; §ALR maxSUV – ratio of SUV_{max} in the adrenal gland compared with the liver; ª% confirmed; bHistol/FU/Other.

Formulae for calculating quantitative thresholds:
¶Signal intensity index=(SI adrenal IP) – (SI adrenal OP)]/(SI adrenal IP); ¶MRI adrenal to liver ratio=[(SI adrenal OP)/(SI adrenal IP/SI spleen IP)/(SI adrenal IP/SI spleen IP); **MRI adrenal to liver ratio=SI adrenal/SI liver; ††MRI adrenal to liver ratio=[(SI adrenal OP)/(SI adrenal IP/SI liver IP)] –1)×100%; ‡‡MRI adrenal to spleen ratio=[(SI adrenal OP)/(SI adrenal IP/SI Spleen IP) –1)×100%; §§MRI adrenal to muscle ratio=[(SI adrenal OP)/(SI adrenal IP/SI Muscle IP) –1)×100%].

Table 3 Test performance according to clinical pathway. Studies focusing on truly incidentally discovered adrenal masses (incidentaloma pathway) vs studies on adrenal masses discovered during follow-up monitoring for extra-adrenal malignancy (follow-up from previous malignancy pathway).

	≥ 50%*			≥90%**		
	Studies (n/N)	Sensitivity (95% CI)	Specificity (95% CI)	Studies (n/N)	Sensitivity (95% CI)	Specificity (95% CI)
Incidentaloma pathway						
CT non–contrast tumor density (>10 HU)	2 (41/102)	100% (91–100%)	72% (60–82%)	1 (13/52)	100% (75–100%)	72% (55–85%)
CT contrast enhanced washout (combination at 10 min)	1 (14/25)	93% (68–100%)	100% (69–100%)	0	–	–
CT contrast enhanced washout (combination at 15 min)	1 (13/25)	100% (75–100%)	92% (62–100%)	0	–	
MRI adrenal-liver ratio (1.5 Tesla only)	1 (8/26)	100% (63–100%)	44% (22–69%)	0	–	–
MRI adrenal-spleen ratio (1.5 Tesla only)	1 (12/49)	58% (28–85%)	86% (71–95%)	0	–	–
MRI loss of signal intensity (1.5 Tesla only)	2 (20/75)	86% (31–99%)	85% (73–93%)	0	–	–
PET ALR SUV$_{max}$	2 (15/64)	100% (78–100%)	96% (57–100%)	1 (12/41)	100% (74–100%)	100% (88–100%)
PET SUV$_{max}$	2 (15/64)	93% (65–99%)	73% (59–84%)	1 (12/41)	92% (62–100%)	72% (53–87%)
Follow-up from previous malignancy pathway						
CT non–contrast tumor density (>10 HU)	5 (93/168)	93% (79–98%)	71% (38–91%)	5 (93/168)	93% (79–98%)	71% (38–91%)
CT contrast enhanced washout (absolute at 15 min)	1 (19/40)	16% (3–40%)	86% (64–97%)	1 (19/40)	16% (3–40%)	86% (64–97%)
CT contrast enhanced washout (relative at 15 min)	1 (19/40)	16% (3–40%)	95% (76–100%)	1 (19/40)	16% (3–40%)	95% (76–100%)
MRI adrenal-liver ratio (1.5 Tesla only)	3 (37/129)	89% (74–96%)	60% (21–89%)	2 (27/93)	92% (55–99%)	39% (21–60%)
MRI adrenal-spleen ratio (1.5 Tesla only)	3 (52/142)	99% (69–100%)	84% (72–91%)	2 (42/105)	100% (92–100%)	79% (68–88%)
MRI adrenal-muscle ratio (1.5 Tesla only)	1 (10/37)	90% (55–100%)	93% (76–99%)	0	–	–
MRI loss of signal intensity (1.5 Tesla only)	1 (10/37)	90% (55–100%)	85% (66–96%)	0	–	–
PET ALR SUV$_{max}$	2 (45/117)	82% (41–97%)	96% (76–99%)	1 (34/102)	94% (80–99%)	94% (86–98%)
PET SUV$_{max}$	3 (72/156)	84% (62–94%)	90% (71–97%)	2 (61/141)	90% (80–96%)	87% (78–93%)

ALR SUV$_{max}$, ratio of SUV$_{max}$ in the adrenal gland compared with the liver; HU, Hounsfield units; n, number of cases; N, total population; PET, positron emission tomography; SUV$_{max}$, maximum standardized uptake value.
*refers to ≥50% with incidentaloma in studies in the incidentaloma pathway and ≥50% with current or prior non-adrenal malignancy in the follow-up from previous malignancy pathway; **refers to ≥90% with incidentaloma in studies in the incidentaloma pathway and ≥90% with current or prior non-adrenal malignancy in the follow-up from previous malignancy pathway.

of studies to be able to discern whether any test performs adequately or better than alternative tests from the available data.

Discussion

Our main finding cautiously suggests that in patients without known extra-adrenal malignancy, a non-contrast CT tumor density of 10 HU is a diagnostically relevant cut-off, albeit based only on data from two small studies. The sensitivity of >10 HU for detecting malignancy was high (100%; 95% CI: 91, 100%), however, the specificity was poor. Conversely, this means that an incidentally discovered adrenal mass with a non-contrast CT tumor density of ≤10 HU is unlikely to be malignant. Tumor density ≤10 HU was less conclusive for ruling out malignancy in patients with a history of extra-adrenal malignancy, however, with a pooled false-negative rate of 7%, although CIs were wide. With positive predictive values for detection of malignancy in the order of 70–80% in both populations, a considerable number of adrenal masses with tumor density >10 HU are likely to be benign.

These and all other pooled estimates have such wide CIs that no further conclusions can be drawn regarding the accuracy of imaging tests for the detection of malignancy in incidentally discovered adrenal masses.

Possible clinical explanations for this uncertainty include variability in the lipid content of adenomas, tissue heterogeneity, small size of metastatic lesions, or differences in selecting regions of interest for HU measurement. However, most of the uncertainty is due to small numbers of eligible studies and hence results from few patients available for analysis. Despite the availability of a significant number of studies addressing imaging characteristics in patients with an adrenal mass, more than 90% of full-text papers retrieved had to be excluded. Many had small sample sizes, mixed populations, inadequate reporting on imaging techniques and thresholds, as well as unacceptable reference standards for both malignant and benign masses. Even with our stringent eligibility criteria, included studies were characterized by heterogeneity in study populations, imaging tests and thresholds, and reference standards as well as poor methodological quality. Given differences in patient spectrum according to the indication for adrenal imaging and the potential impact on accuracy (63, 64, 65), our meta-analysis was further restricted to studies where a majority of participants had either incidentaloma or were undergoing imaging due to known malignancy, leading to the exclusion of another 50% of included studies. Heterogeneity in study conduct and poor methodological quality remained, further contributing to the lack of certainty in pooled estimates.

Our findings are disappointingly consistent with another systematic review of the literature on tests for adrenal incidentaloma published almost 15 years ago (66). Observed heterogeneity in tests and populations meant that no meta-analysis was undertaken and no clear conclusions could be drawn (66). Almost three-quarters (27/37) of the studies in our review were published in the interim period; however, methodological and reporting quality have not improved sufficiently to allow any new conclusions to be drawn. A more recent meta-analysis of FDG-PET (67) applied considerably less stringent inclusion criteria compared with our review, thereby including more studies (*n*=21); however, highly heterogeneous data limited the conclusions that could be drawn.

Our findings of poor quality and reporting of test accuracy studies are similar to findings from other fields (68, 69, 70). Introduction of the Standards for Reporting Diagnostic Accuracy (STARD) statement (71) has only led to small improvements in reporting (72) and our results indicate that greater awareness is required of methodological considerations in the design and delivery of multicenter studies in this field, as in many others, to improve reporting.

The strengths of this review include an in-depth comprehensive literature search, a focused review question, and stringent predefined reference standard. The limitations were derived from the heterogeneity and low quality of included studies. Unclear definitions of study populations, various and often data-driven thresholds, as well as different techniques for the same imaging tests, limit the interpretation and generalization of results. The weak conclusions derived from this systematic review and meta-analysis should be interpreted in relation to the low volume and poor quality of included studies (Fig. 1B).

Our results do not suggest that current imaging practice is inappropriate: small study numbers prevent us from providing substantive evidence to either support current practice or to prompt a need for a change in imaging practice. We suggest further studies are needed to answer the following key questions:

1. Do adrenal lesions with unenhanced CT tumor density ≤10 HU need additional imaging, in particular in patients with a history of extra-adrenal malignancy?
2. What is the best second-line imaging study that would accurately diagnose (or exclude) a malignant adrenal mass?
3. What additional factors influence decisions on imaging choice? (patient preference, radiation risks, costs)
4. How much tumor growth, and over what period of time, is indicative of a malignant adrenal mass?

In addition, future studies should include the systematic evaluation of alternative testing approaches and detailed analysis of health economics impact. All these questions can only be answered with larger multicenter studies, with prospective recruitment of consecutive series of participants in appropriately defined clinical pathways, and imaging test interpretation blinded to the reference standard diagnosis and to the result of any other imaging tests. Diagnostic thresholds for determining benignity or malignancy must be prespecified to avoid data-driven threshold selection and overestimation of test accuracy. The reliance on a histological reference standard leads to study populations with a high pretest probability of malignancy, however, imaging follow-up of those with indeterminate imaging characteristics needs to be long enough to ensure that malignant masses are not missed. Centralized radiological and pathology review would further help to strengthen the results. Future investigators must also meet the updated STARD recommendations (20) so that study conduct and quality can be judged appropriately.

In conclusion, current evidence on imaging tests and cut-offs in diagnosis of incidentally discovered adrenal mass is highly heterogeneous and disappointingly poor. Not surprisingly, many patients with adrenal incidentaloma undergo repeated multimodal imaging and even unnecessary adrenalectomy. With adrenal incidentalomas detected on 1 in 20 (1, 2, 3) of an ever-increasing number of cross-sectional abdominal imaging studies performed every year, the potential economic and health impact of unnecessary procedures and interventions could be significant. In this era of evidence-based medicine, and with advances in our understanding of optimal diagnostic test accuracy study design and study synthesis, it is incumbent on the medical community to provide a solid evidence base to underpin imaging practice in this field. Areas of uncertainty especially include second-line testing for indeterminate adrenal masses and larger adrenal masses, with very limited data on CT washout, MRI, and PET-CT. Further well-designed studies are needed to establish performance and health economic impact of imaging in patients with incidentally discovered adrenal masses.

This meta-analysis has informed the ESE-ENSAT Clinical Guidelines on the management of adrenal incidentalomas (73).

Author contribution statement

Data extraction and quality assessment: J D, I B, L F R, V C, C D; Data analysis: J D, L F R, S B, J J D; Manuscript writing: J D, I B, L F R, A S, P G, M F, J J D, W A; Expert review and advice: A S, P G, J J D, W A.

References

1 Bovio S, Cataldi A, Reimondo G, Sperone P, Novello S, Berruti A, Borasio P, Fava C, Dogliotti L, Scagliotti GV et al. Prevalence of adrenal incidentaloma in a contemporary computerized tomography series. *Journal of Endocrinological Investigation* 2006 **29** 298–302. (doi:10.1007/BF03344099)

2 Mantero F, Terzolo M, Arnaldi G, Osella G, Masini AM, Ali A, Giovagnetti M, Opocher G & Angeli A. A survey on adrenal incidentaloma in Italy. Study Group on Adrenal Tumours of the Italian Society of Endocrinology. *Journal of Clinical Endocrinology & Metabolism* 2000 **85** 637–644. (doi:10.1210/jcem.85.2.6372)

3 Song JH, Chaudhry FS & Mayo-Smith WW. The incidental adrenal mass on CT: prevalence of adrenal disease in 1,049 consecutive adrenal masses in patients with no known malignancy. *American Journal of Roentgenology* 2008 **190** 1163–1168. (doi:10.2214/AJR.07.2799)

4 OECD. OECD Stat (database). 2015.

5 Terzolo M, Stigliano A, Chiodini I, Loli P, Furlani L, Arnaldi G, Reimondo G, Pia A, Toscano V, Zini M et al. AME position statement on adrenal incidentaloma. *European Journal of Endocrinology* 2011 **164** 851–870. (doi:10.1530/EJE-10-1147)

6 Ciftci AO, Senocak ME, Tanyel FC & Buyukpamukcu N. Adrenocortical tumours in children. *Journal of Pediatric Surgery* 2001 **36** 549–554. (doi:10.1053/jpsu.2001.22280)

7 Frilling A, Tecklenborg K, Weber F, Kuhl H, Muller S, Stamatis G & Broelsch C. Importance of adrenal incidentaloma in patients with a history of malignancy. *Surgery* 2004 **136** 1289–1296. (doi:10.1016/j.surg.2004.06.060)

8 Lang BH, Cowling BJ, Li JY, Wong KP & Wan KY. High false positivity in positron emission tomography is a potential diagnostic pitfall in patients with suspected adrenal metastasis. *World Journal of Surgery* 2015 **39** 1902–1908. (doi:10.1007/s00268-015-3035-3)

9 Song JH, Grand DJ, Beland MD, Chang KJ, Machan JT & Mayo-Smith WW. Morphologic features of 211 adrenal masses at initial contrast-enhanced CT: can we differentiate benign from malignant lesions using imaging features alone? *American Journal of Roentgenology* 2013 **201** 1248–1253. (doi:10.2214/AJR.12.10302)

10 Cawood TJ, Hunt PJ, O'Shea D, Cole D & Soule S. Recommended evaluation of adrenal incidentalomas is costly, has high false-positive rates and confers a risk of fatal cancer that is similar to the risk of the adrenal lesion becoming malignant; time for a rethink? *European Journal of Endocrinology* 2009 **161** 513–527. (doi:10.1530/EJE-09-0234)

11 Kloos RT, Gross MD, Francis IR, Korobkin M & Shapiro B. Incidentally discovered adrenal masses. *Endocrine Reviews* 1995 **16** 460–484. (doi:10.1210/edrv-16-4-460)

12 Mansmann G, Lau J, Balk E, Rothberg M, Miyachi Y & Bornstein SR. The clinically inapparent adrenal mass: update in diagnosis and management. *Endocrine Reviews* 2004 **25** 309–340. (doi:10.1210/er.2002-0031)

13 Hamrahian AH, Ioachimescu AG, Remer EM, Motta-Ramirez G, Bogabathina H, Levin HS, Reddy S, Gill IS, Siperstein A & Bravo EL. Clinical utility of noncontrast computed tomography attenuation value (hounsfield units) to differentiate adrenal adenomas/hyperplasias from nonadenomas: Cleveland Clinic experience. *Journal of Clinical Endocrinology & Metabolism* 2005 **90** 871–877. (doi:10.1210/jc.2004-1627)

14 Korobkin M, Brodeur FJ, Francis IR, Quint LE, Dunnick NR & Londy F. CT time-attenuation washout curves of adrenal adenomas and nonadenomas. *American Journal of Roentgenology* 1998 **170** 747–752. (doi:10.2214/ajr.170.3.9490968)

15 Lezoche G, Baldarelli M, Cappelletti Trombettoni MM, Polenta V, Ortenzi M, Tuttolomondo A & Guerrieri M. Two decades of laparoscopic adrenalectomy: 326 procedures in a single-center experience. *Surgical Laparoscopy, Endoscopy & Percutaneous Techniques* 2016 **26** 128–132. (doi:10.1097/SLE.0000000000000249)

16 Young WF Jr. Clinical practice. The incidentally discovered adrenal mass. *New England Journal of Medicine* 2007 **356** 601–610. (doi:10.1056/NEJMcp065470)

17 Zeiger MA, Thompson GB, Duh QY, Hamrahian AH, Angelos P, Elaraj D, Fishman E & Kharlip J. The American Association of Clinical Endocrinologists and American Association of Endocrine Surgeons

medical guidelines for the management of adrenal incidentalomas. *Endocrine Practice* 2009 **15** (Supplement 1) 1–20. (doi:10.4158/EP.15.S1.1)

18 NIH state-of-the-science statement on management of the clinically inapparent adrenal mass ("incidentaloma"). *NIH Consensus and State-of-the-Science Statements* 2002 **19** 1–25.

19 Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA & Bossuyt PM. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine* 2011 **155** 529–536. (doi:10.7326/0003-4819-155-8-201110180-00009)

20 Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, Lijmer JG, Moher D, Rennie D, de Vet HC *et al*. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015 **351** h5527. (doi:10.1148/radiol.2015151516)

21 Willis BH & Quigley M. Uptake of newer methodological developments and the deployment of meta-analysis in diagnostic test research: a systematic review. *BMC Medical Research Methodology* 2011 **11** 27. (doi:10.1186/1471-2288-11-27)

22 Ochodo EA, van Enst WA, Naaktgeboren CA, de Groot JA, Hooft L, Moons KG, Reitsma JB, Bossuyt PM & Leeflang MM. Incorporating quality assessments of primary studies in the conclusions of diagnostic accuracy reviews: a cross-sectional study. *BMC Medical Research Methodology* 2014 **14** 33. (doi:10.1186/1471-2288-14-33)

23 Deeks J, Bossuvt P & Gatsonis C. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. Eds J Deeks, P Bossuvt & C Gatsonis, 2013.

24 Moher D, Liberati A, Tetzlaff J & Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine* 2009 **151** 264–269. (doi:10.7326/0003-4819-151-4-200908180-00135)

25 Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM & Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology* 2005 **58** 982–990. (doi:10.1016/j.jclinepi.2005.02.022)

26 Takwoingi Y, Leeflang MM & Deeks JJ. Empirical evidence of the importance of comparative studies of diagnostic test accuracy. *Annals of Internal Medicine* 2013 **158** 544–554. (doi:10.7326/0003-4819-158-7-201304020-00006)

27 Angelelli G, Mancini ME, Moschetta M, Pedote P, Pignataro P & Scardapane A. MDCT in the differentiation of adrenal masses: comparison between different scan delays for the evaluation of intralesional washout. *Scientific World Journal* 2013 **2013** 957680. (doi:10.1155/2013/957680)

28 Sandrasegaran K, Patel AA, Ramaswamy R, Samuel VP, Northcutt BG, Frank MS & Francis IR. Characterization of adrenal masses with diffusion-weighted imaging. *American Journal of Roentgenology* 2011 **197** 132–138. (doi:10.2214/AJR.10.4583)

29 Nunes ML, Rault A, Teynie J, Valli N, Guyot M, Gaye D, Belleannee G & Tabarin A. 18F-FDG PET for the identification of adrenocortical carcinomas among indeterminate adrenal tumours at computed tomography scanning. *World Journal of Surgery* 2010 **34** 1506–1510. (doi:10.1007/s00268-010-0576-3)

30 Vilar L, Freitas Mda C, Canadas V, Albuquerque JL, Botelho CA, Egito CS, Arruda MJ, Moura e Silva L, Coelho CE, Casulari LA *et al*. Adrenal incidentalomas: diagnostic evaluation and long-term follow-up. *Endocrine Practice* 2008 **14** 269–278. (doi:10.4158/EP.14.3.269)

31 Maurea S, Caraco C, Klain M, Mainolfi C & Salvatore M. Imaging characterization of non-hypersecreting adrenal masses. Comparison between MR and radionuclide techniques. *Quarterly Journal of Nuclear Medicine and Molecular Imaging* 2004 **48** 188–197.

32 Burt M, Heelan RT, Coit D, McCormack PM, Bains MS, Martini N, Rusch V & Ginsberg RJ. Prospective evaluation of unilateral adrenal masses in patients with operable non-small-cell lung cancer. Impact of magnetic resonance imaging. *Journal of Thoracic and Cardiovascular Surgery* 1994 **107** 584–588.

33 Choi YA, Kim CK, Park BK & Kim B. Evaluation of adrenal metastases from renal cell carcinoma and hepatocellular carcinoma: use of delayed contrast-enhanced CT. *Radiology* 2013 **266** 514–520. (doi:10.1148/radiol.12120110)

34 McNicholas MM, Lee MJ, Mayo-Smith WW, Hahn PF, Boland GW & Mueller PR. An imaging algorithm for the differential diagnosis of adrenal adenomas and metastases. *American Journal of Roentgenology* 1995 **165** 1453–1459. (doi:10.2214/ajr.165.6.7484585)

35 Porte HL, Ernst OJ, Delebecq T, Metois D, Lemaitre LG & Wurtz AJ. Is computed tomography guided biopsy still necessary for the diagnosis of adrenal masses in patients with resectable non-small-cell lung cancer? *European Journal of Cardio-Thoracic Surgery* 1999 **15** 597–601. (doi:10.1016/S1010-7940(99)00047-0)

36 Ream JM, Gaing B, Mussi TC & Rosenkrantz AB. Characterization of adrenal lesions at chemical-shift MRI: a direct intraindividual comparison of in- and opposed-phase imaging at 1.5 T and 3 T. *American Journal of Roentgenology* 2015 **204** 536–541. (doi:10.2214/AJR.14.12941)

37 Uemura S, Yasuda I, Kato T, Doi S, Kawaguchi J, Yamauchi T, Kaneko Y, Ohnishi R, Suzuki T, Yasuda S *et al*. Preoperative routine evaluation of bilateral adrenal glands by endoscopic ultrasound and fine-needle aspiration in patients with potentially resectable lung cancer. *Endoscopy* 2013 **45** 195–201. (doi:10.1055/s-00000012)

38 Villar Del Moral JM, Munoz Perez N, Rodriguez Fernandez A, Olmos Juarez E, Moreno Cortes C, Rodriguez Gonzalez R, Martin Cano FJ, Sanchez Sanchez R & Ferron Orihuela JA. [Diagnostic efficacy and discriminatory capacity of positron emission tomography combined with axial tomography of adrenal lesions]. *Cirugia Espanola* 2010 **88** 247–252. (doi:10.1016/j.ciresp.2010.07.007)

39 Ichikawa T, Fujimoto H, Murakami K, Tauchi M, Mochizuki S, Ohtomo K & Uchiyama G. Adrenal tissue characterization with 0.5-T MR imaging: value of T2*-weighted images. *Journal of Magnetic Resonance Imaging* 1993 **3** 742–745. (doi:10.1002/(ISSN)1522-2586)

40 Nwariaku FE, Champine J, Kim LT, Burkey S, O'Keefe G & Snyder WH 3rd. Radiologic characterization of adrenal masses: the role of computed tomography – derived attenuation values. *Surgery* 2001 **130** 1068–1071. (doi:10.1067/msy.2001.119189)

41 Kebapci M, Kaya T, Gurbuz E, Adapinar B, Kebapci N & Demirustu C. Differentiation of adrenal adenomas (lipid rich and lipid poor) from nonadenomas by use of washout characteristics on delayed enhanced CT. *Abdominal Imaging* 2003 **28** 709–715. (doi:10.1007/s00261-003-0015-0)

42 Blake MA, Kalra MK, Sweeney AT, Lucey BC, Maher MM, Sahani DV, Halpern EF, Mueller PR, Hahn PF & Boland GW. Distinguishing benign from malignant adrenal masses: multi-detector row CT protocol with 10-minute delay. *Radiology* 2006 **238** 578–585. (doi:10.1148/radiol.2382041514)

43 Remer EM, Motta-Ramirez GA, Shepardson LB, Hamrahian AH & Herts BR. CT histogram analysis in pathologically proven adrenal masses. *American Journal of Roentgenology* 2006 **187** 191–196. (doi:10.2214/AJR.05.0179)

44 Park SH, Kim MJ, Kim JH, Lim JS & Kim KW. Differentiation of adrenal adenoma and nonadenoma in unenhanced CT: new optimal threshold value and the usefulness of size criteria for differentiation. *Korean Journal of Radiology* 2007 **8** 328–335. (doi:10.3348/kjr.2007.8.4.328)

45 Kamiyama T, Fukukura Y, Yoneyama T, Takumi K & Nakajo M. Distinguishing adrenal adenomas from nonadenomas: combined use of diagnostic parameters of unenhanced and short 5-minute dynamic enhanced CT protocol. *Radiology* 2009 **250** 474–481. (doi:10.1148/radiol.2502080302)

46 Marin D, Dale BM, Bashir MR, Ziemlewicz TJ, Ringe KI, Boll DT & Merkle EM. Effectiveness of a three-dimensional dual gradient echo two-point Dixon technique for the characterization of adrenal lesions at 3 Tesla. *European Radiology* 2012 **22** 259–268. (doi:10.1007/s00330-011-2244-x)

47 Aksakal N, Sahbaz A, Ozcinar B, Ozemir A, Caglayan K, Agcaoglu O, Barbaros U, Salmaslioglu A & Erbil Y. Nonfunctional adrenal lesions without loss of signal intensity on MRI: whose problem is it? The

patient's? The surgeon's? *International Journal of Surgery* 2013 **11** 169–172. (doi:10.1016/j.ijsu.2012.12.014)

48 Park SY, Park BK, Park JJ & Kim CK. CT sensitivities for large (>/=3 cm) adrenal adenoma and cortical carcinoma. *Abdominal Imaging* 2015 **40** 310–317. (doi:10.1007/s00261-014-0202-1)

49 Petersenn S, Richter PA, Broemel T, Ritter CO, Deutschbein T, Beil FU, Allolio B & Fassnacht M. Computed tomography criteria for discrimination of adrenal adenomas and adrenocortical carcinomas: analysis of the German ACC registry. *European Journal of Endocrinology* 2015 **172** 415–422. (doi:10.1530/EJE-14-0916)

50 Bilbey JH, McLoughlin RF, Kurkjian PS, Wilkins GE, Chan NH, Schmidt N & Singer J. MR imaging of adrenal masses: value of chemical-shift imaging for distinguishing adenomas from other tumours. *American Journal of Roentgenology* 1995 **164** 637–642. (doi:10.2214/ajr.164.3.7863885)

51 Mayo-Smith WW, Lee MJ, McNicholas MM, Hahn PF, Boland GW & Saini S. Characterization of adrenal masses (<5 cm) by use of chemical shift MR imaging: observer performance versus quantitative measures. *American Journal of Roentgenology* 1995 **165** 91–95. (doi:10.2214/ajr.165.1.7785642)

52 Boraschi P, Braccini G, Grassi L, Campatelli A, Di Vito A, Mosca F & Perri G. Incidentally discovered adrenal masses: evaluation with gadolinium enhancement and fat-suppressed MR imaging at 0.5 T. *European Journal of Radiology* 1997 **24** 245–252. (doi:10.1016/S0720-048X(97)01046-2)

53 Chung JJ, Semelka RC & Martin DR. Adrenal adenomas: characteristic postgadolinium capillary blush on dynamic MR imaging. *Journal of Magnetic Resonance Imaging: JMRI* 2001 **13** 242–248. (doi:10.1002/1522-2586(200102)13:2<242::AID-JMRI1035>3.3.CO;2-R)

54 Boraschi P, Braccini G, Gigoni R, Perri G, Campatelli A, Di Vito A & Bonadio AG. Diagnosis of adrenal adenoma: value of central spot of high-intensity hyperintense rim sign and homogeneous isointensity to liver on gadolinium-enhanced fat-suppressed spin-echo MR images. *Journal of Magnetic Resonance Imaging* 1999 **9** 304–310. (doi:10.1002/(ISSN)1522-2586)

55 Groussin L, Bonardel G, Silvera S, Tissier F, Coste J, Abiven G, Libe R, Bienvenu M, Alberini JL, Salenave S *et al.* 18F-Fluorodeoxyglucose positron emission tomography for the diagnosis of adrenocortical tumours: a prospective study in 77 operated patients. *Journal of Clinical Endocrinology & Metabolism* 2009 **94** 1713–1722. (doi:10.1210/jc.2008-2302)

56 Gust L, Taieb D, Beliard A, Barlier A, Morange I, de Micco C, Henry JF & Sebag F. Preoperative 18F-FDG uptake is strongly correlated with malignancy, Weiss score, and molecular markers of aggressiveness in adrenal cortical tumours. *World Journal of Surgery* 2012 **36** 1406–1410. (doi:10.1007/s00268-011-1374-2)

57 Launay N, Silvera S, Tenenbaum F, Groussin L, Tissier F, Audureau E, Vignaux O, Dousset B, Bertagna X & Legmann P. Value of 18-F-FDG PET/CT and CT in the diagnosis of indeterminate adrenal masses. *International Journal of Endocrinology* 2015 **2015** 213875. (doi:10.1155/2015/213875)

58 Zettinig G, Mitterhauser M, Wadsak W, Becherer A, Pirich C, Vierhapper H, Niederle B, Dudczak R & Kletter K. Positron emission tomography imaging of adrenal masses: (18)F-fluorodeoxyglucose and the 11beta-hydroxylase tracer (11)C-metomidate. *European Journal of Nuclear Medicine and Molecular Imaging* 2004 **31** 1224–1230. (doi:10.1007/s00259-004-1575-0)

59 Zielonko J, Studniarek M, Rzepko R, Babinska A & Siekierska-Hellmann M. Value of MRI in differentiating adrenal masses: quantitative analysis of tumour signal intensity. *Polish Journal of Radiology* 2008 **73** 7–12.

60 Schwartz LH, Panicek DM, Koutcher JA, Brown KT, Getrajdman GI, Heelan RT & Burt M. Adrenal masses in patients with malignancy: prospective comparison of echo-planar, fast spin-echo, and chemical shift MR imaging. *Radiology* 1995 **197** 421–425. (doi:10.1148/radiology.197.2.7480686)

61 Tessonnier L, Sebag F, Palazzo FF, Colavolpe C, De Micco C, Mancini J, Conte-Devolx B, Henry JF, Mundler O & Taieb D Does. 18F-FDG PET/CT add diagnostic accuracy in incidentally identified non-secreting adrenal tumours? *European Journal of Nuclear Medicine and Molecular Imaging* 2008 **35** 2018–2025. (doi:10.1007/s00259-008-0849-3)

62 Kunikowska J, Matyskiel R, Toutounchi S, Grabowska-Derlatka L, Koperski L & Krolicki L. What parameters from 18F-FDG PET/CT are useful in evaluation of adrenal lesions? *European Journal of Nuclear Medicine and Molecular Imaging* 2014 **41** 2273–2280. (doi:10.1007/s00259-014-2844-1)

63 Leeflang MM, Rutjes AW, Reitsma JB, Hooft L & Bossuyt PM. Variation of a test's sensitivity and specificity with disease prevalence. *Canadian Medical Association Journal* 2013 **185** E537–E544. (doi:10.1503/cmaj.121286)

64 Mulherin SA & Miller WC. Spectrum bias or spectrum effect? Subgroup variation in diagnostic test evaluation. *Annals of Internal Medicine* 2002 **137** 598–602. (doi:10.7326/0003-4819-137-7-200210010-00011)

65 Brenner H & Gefeller O. Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. *Statistics in Medicine* 1997 **16** 981–991. (doi:10.1002/(SICI)1097-0258(19970515)16:9<981::AID-SIM510>3.3.CO;2-E)

66 Lau J, Balk E, Rothberg M, Ioannidis JP, DeVine D, Chew P, Kupelnick B & Miller K. Management of Clinically Inapparent Adrenal Mass. Summary, Evidence Report/Technology Assessment: Number 56. AHRQ Publication No. 02-E013, February 2002. Agency for Healthcare Research and Quality, Rockville, MD, USA (http://hstat.nlm.nih.gov/hq/Hquest/screen/DirectAccess/db/local.epc.ersum.adrensum).

67 Boland GW, Dwamena BA, Jagtiani Sangwaiya M, Goehler AG, Blake MA, Hahn PF, Scott JA & Kalra MK. Characterization of adrenal masses by using FDG PET: a systematic review and meta-analysis of diagnostic test performance. *Radiology* 2011 **259** 117–126. (doi:10.1148/radiol.11100569)

68 Fontela PS, Pant Pai N, Schiller I, Dendukuri N, Ramsay A & Pai M. Quality and reporting of diagnostic accuracy studies in TB, HIV and malaria: evaluation using QUADAS and STARD standards. *PLos ONE* 2009 **4** e7753. (doi:10.1371/journal.pone.0007753)

69 Morris RK, Selman TJ, Zamora J & Khan KS. Methodological quality of test accuracy studies included in systematic reviews in obstetrics and gynaecology: sources of bias. *BMC Women's Health* 2011 **11** 7. (doi:10.1186/1472-6874-11-7)

70 Henschke N, Keuerleber J, Ferreira M, Maher CG & Verhagen AP. The methodological quality of diagnostic test accuracy studies for musculoskeletal conditions can be improved. *Journal of Clinical Epidemiology* 2014 **67** 416–424. (doi:10.1016/j.jclinepi.2013.11.008)

71 Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D & De Vet HC. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *American Journal of Roentgenology* 2003 **181** 51–55. (doi:10.2214/ajr.181.1.1810051)

72 Korevaar DA, van Enst WA, Spijker R, Bossuyt PM & Hooft L. Reporting quality of diagnostic accuracy studies: a systematic review and meta-analysis of investigations on adherence to STARD. *Evidence-Based Medicine* 2014 **19** 47–54. (doi:10.1136/eb-2013-101637)

73 Fassnacht M, Arlt W, Bancos I, Dralle H, Newell-Price J, Sahdev A, Tabarin A, Terzolo M, Tsagarakis S & Dekkers OM. Management of adrenal incidentalomas: ESE clinical practice guideline in collaboration with the ENSAT. *European Journal of Endocrinology* 2016 **175** G1–G34. (doi:10.1530/EJE-16-0467)