European Radiology (2019) 29:806–817
https://doi.org/10.1007/s00330-018-5626-5

MAGNETIC RESONANCE

Chemical shift imaging for evaluation of adrenal masses: a systematic review and meta-analysis

Ivan Platzek[1] · Dominik Sieron[2] · Verena Plodeck[1] · Angelika Borkowetz[3] · Michael Laniado[1] · Ralf-Thorsten Hoffmann[1]

Received: 6 April 2018 / Revised: 13 June 2018 / Accepted: 21 June 2018 / Published online: 16 July 2018

Abstract

Objectives To perform a systematic review and meta-analysis of published data to evaluate the utility of chemical shift imaging (CSI) for differentiating between adrenal adenomas and non-adenomas.

Methods A systematic search of the MEDLINE, Web of Science Core Collection, EMBASE and Cochrane Central Register of Controlled Trials electronic databases was performed. The methodological quality of the included studies was assessed by using the QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies) tool. A bivariate random effect model was used to determine summary and subgroup sensitivity and specificity and calculate summary receiver operating characteristic curves (SROC).

Results Eighteen studies with 1138 patients and 1280 lesions (859 adenomas, 421 non-adenomas) in total were included. In addition to summary analysis, quantitative analyses of the adrenal signal intensity index (SII, 978 lesions, 14 studies), adrenal-to-spleen ratio (ASR; 394 lesions, 7 studies) and visual analysis (560 lesions, 5 studies) were performed. The resultant data showed considerable heterogeneity (inconsistency index I^2 of 94%, based on the diagnostic odds ratio, DOR). The pooled sensitivity of CSI for adenoma was 0.94 [95% confidence interval (CI) 0.88–0.97] and pooled specificity was 0.95 (95% CI 0.89–0.97). The area (AUC) under the SROC curve was 0.98 (95% CI 0.96–0.99). The corresponding AUCs were 0.98, 0.99 and 0.95 for SII, ASR and visual evaluation, respectively.

Conclusion CSI has high sensitivity, specificity and accuracy for adrenal adenoma. Diagnostic performance does not improve when quantitative indices are used.

Key Points

- *Inclusion of CSI in abdominal MRI protocols provides an effective solution for classifying adrenal masses discovered on MR exams*
- *Visual evaluation of adrenal CSI is sufficient; use of quantitative indices does not improve diagnostic accuracy*

Keywords Adrenal gland neoplasms · Adrenocortical adenoma · Chemical shift imaging · Magnetic resonance imaging

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00330-018-5626-5) contains supplementary material, which is available to authorized users.

✉ Ivan Platzek
ivan.platzek@uniklinikum-dresden.de

[1] Department of Radiology, Dresden University Hospital, Fetscherstr. 74, 01307 Dresden, Germany

[2] Department of Radiology, Spital Tiefenau, Tiefenaustr. 112, 3004 Bern, Switzerland

[3] Department of Urology, Dresden University Hospital, Fetscherstr. 74, 01307 Dresden, Germany

Abbreviations

ACR	American College of Radiology
ALR	Adrenal-to-liver ratio
AMR	Adrenal-to-muscle ratio
ASR	Adrenal-to-spleen ratio
CSI	Chemical shift imaging
DCE	Dynamic contrast-enhanced MRI
DOR	Diagnostic odds ratio
ESR	European Society of Radiology
ESS	Effective sample size
QUADAS	Quality assessment of diagnostic accuracy studies

SII Signal intensity index
SROC Summary receiver operating characteristic

Introduction

Adrenal masses are a common incidental imaging finding [1, 2]. The most common adrenal lesions are adenomas [3, 4], benign tumors, which in most cases do not produce hormones and remain asymptomatic. The adrenal glands are also a frequent location of metastases [5]. Even in patients with malignant disease, a large proportion of adrenal masses are adenomas [6, 7]. As non-functioning (endocrine-inactive) adrenal adenomas do usually not require therapy, it is very important to differentiate between adenomas and non-adenomas.

Chemical shift imaging (CSI) is the preferred MR sequence for evaluation of adrenal masses. CSI results in two sets of images, called in-phase and opposed-phase images. On opposed-phase images the signal of fat protons partially cancels the signal of water protons [8]. As a result of their high intracellular fat content [9], most adenomas show a pronounced signal loss on opposed-phase images, when compared to in-phase images.

A number of studies have reported a high sensitivity and specificity of CSI in diagnosing adrenal adenomas.

The aim of this study was to perform a systematic review and meta-analysis of published data to evaluate the utility of CSI to differentiate between adrenal adenomas (lipid-rich adenomas) and non-adenomas.

Materials and methods

Search methods and study selection

The MEDLINE, Web of Science Core Collection, EMBASE and Cochrane controlled trials register electronic databases were searched (from inception through February 2017) using database-specific Boolean search strategies and the terms "adrenal glands", "adrenal adenoma", "chemical shift imaging" and "magnetic resonance imaging". The database-specific search strategies are described in Appendix E1. The final search was performed on 10 September 2017.

The inclusion criteria were as follows: the study used CSI for adrenal imaging, at least 10 lesions were included, the studied condition was having or not having an adrenal adenoma, there was no preselection based on imaging findings, information on the reference standard was provided, articles were in English or German, results were presented in a 2 × 2 contingency table or such a table could be extracted from the article. Exclusion criteria were studies with less than 10 lesions, studies based on non-original data, and the impossibility to retrieve data suitable for meta-analysis.

Retrieved titles and abstracts were independently reviewed for relevance by two investigators (I.P. and D.S., board-certified radiologists with 13 and 10 years of experience in abdominal MRI, respectively). Consequently, the full text of selected articles was independently reviewed by the same two investigators for relevance. Discrepancies between the two investigators were discussed and resolved in consensus.

Data extraction

Two investigators (I.P., D.S) independently extracted data from the previously selected articles. Extracted data included first author, journal, year of publication, study design, field strength of the MRI system(s) used in the study, scanner manufacturer, study population, method(s) of image evaluation, reference standard and lesion size.

The total number of adenomas and non-adenomas and the complete 2 × 2 tables were extracted or calculated from each article and tabulated.

Data synthesis and statistical analysis

Data analysis was performed using the Stata Statistical Software, Release 15 (StataCorp LLC, College Station, TX, USA) and the MIDAS [10] command. MIDAS implements a bivariate approach for meta-analysis of diagnostic studies introduced by Reitsma et al [11].

The sensitivities, specificities, diagnostic odds ratios (DOR) and 95% confidence intervals (CIs) for each study were calculated on the basis of the consistency table. In case of studies which utilized several methods of CSI evaluation, data based on the adrenal signal intensity index (SII, also called ASII) were used for the overall analysis. Summary sensitivity and specificity were calculated for all studies included in the meta-analysis and also for three subgroups defined below (SII, ASR and visual evaluation).

Forest plots [12] were generated to visualize the estimated sensitivities and specificities of the studies included in the meta-analysis and the estimated summary sensitivity and specificity. Summary receiver operating characteristic curves (SROC) [13–15] were generated on the basis of the estimated sensitivities and specificities. The area under the curve (AUC) of SROC curves was also calculated. For interpretation of AUC values we used the thresholds proposed by Jones et al [16]: AUC > 0.97, excellent accuracy; AUC 0.93–0.96, very good accuracy; 0.75–0.92, good accuracy and AUC 0.5–0.75, reasonable accuracy.

The inconsistency index (I^2) [17] was used to evaluate heterogeneity. I^2 is calculated on the basis of the diagnostic odds ratio (DOR) which can be used as a summary measure of test performance [18].

Publication bias [19] was assessed with a regression test on the DOR [20]. A funnel plot [21] was generated as a graphical presentation for publication bias. On the basis on the recommendations of Deeks et al [20], the funnel plot was generated by plotting the square root of the inverse of the effective sample size (ESS) against the DOR.

A p value less than 0.05 was considered statistically significant.

The QUADAS-2 [22] (Quality Assessment of Diagnostic Accuracy Studies) tool was used to evaluate the risk of bias and concerns about applicability for each study.

Results

Search results

The initial search revealed 1263 articles from MEDLINE, 849 articles from Web of Science Core Collection, 963 articles from EMBASE and 25 articles from the Cochrane Central Register of Controlled Trials. After duplicates were discarded, 2272 articles remained and the study titles were reviewed for eligibility in this meta-analysis. On the basis of the study title, 1938/2272 studies were excluded (Fig. 1). Subsequently, on the basis of review of the study abstracts, 275/334 studies were excluded. After full text review of the remaining 59 studies, 41 studies were excluded. Thus 18 studies were included in our meta-analysis [23–42].

Included studies

Eight out of 18 included studies were prospective and 9/18 studies were retrospective (Table 1). One study [27] included both a retrospective and prospective component. We only included the prospective cohort (24 patients, 24 lesions) from this study in our meta-analysis, as CSI was only consistently performed in this cohort.

From the 18 studies included in our meta-analysis, data for 1138 (654 female, 484 male) patients with 1280 lesions were available. Average patient age was available for 17/18 studies and varied between 44.8 and 67.5 years. Demographic data of the patients from the studies included in the meta-analysis are summarized in Table 1.

On the basis of the standard of reference used in the respective study, the 1280 lesions included 859 adenomas and 421 non-adenomas. Information about lesion size was available from all studies (Table 2). The mean maximum diameter varied between 1.6 and 2.7 cm for adenomas and between 2.8 and 7.3 cm for non-adenomas.

The standard of reference was combined (imaging follow-up, histopathology or another imaging modality) in 16/18 studies or based only on histopathology in 2/18 studies [23, 41].

Histopathological or cytological correlation was available for 473/1280 lesions (histopathology for 417/1280 lesions and cytology for 56/1280 lesions). The reference standard used in each study is described in Table 3.

Information about known malignant disease in the examined patients was only provided in 4/18 studies [25, 26, 30, 36]. In these four studies (274 patients in total), 121/274 patients had known malignancy and 153/274 patients had no known malignancy. The primary tumor type was only specified in 73/274 patients with known malignancy [25, 26].

In 15/18 studies one or more quantitative indices were used for evaluation of CSI images, while 3/18 studies used qualitative visual evaluation alone and 2/18 studies used both qualitative and visual evaluation. The quantitative indices included the adrenal signal intensity index (SII), adrenal-to-spleen ratio (ASR), adrenal-to-liver ratio (ALR) and adrenal-to-muscle ratio (AMR) [43].

Seven out of 18 studies used SII alone, 5/18 studies used both SII and ASR, 1/18 studies used SII, ASR and ALR, 1/18 studies used SII, ASR, ALR and AMR, 1/18 studies used SII, ASR and visual evaluation, 1/18 studies used ASR, ALR and AMR and 3/18 studies used visual evaluation only.

As a result of the small number of studies which evaluated ALR and AMR, these quantitative indices were not included in the analysis. The meta-analysis includes subgroup analysis for the SII (15/18 studies), ASR (7/18 studies) and visual evaluation methods (5/18 studies).

SII was defined in all studies as follows:

$$\text{SII} = (\text{signal intensity on in-phase imaging–signal intensity on opposed-phase imaging})/(\text{signal intensity on in-phase imaging})$$

SII cut-off values used in each study are summarized in Table E2.

The definition of ASR varied between the included studies. Afifi et al [23], Marin et al [32], Ramalho et al [35] and Ream et al [36] used the following definition:

$$\text{ASR} = \{[(\text{adrenal signal intensity on opposed-phase imaging}/\text{spleen signal intensity on opposed-phase imaging})/(\text{adrenal signal intensity on in-phase imaging}/\text{spleen signal intensity on in-phase imaging})]-1\}$$

Bilbey et al [24], Halefoglu et al [26] and Wu and Tan [40] used a different formula to calculate ASR:

$$\text{ASR} = [(\text{adrenal signal intensity on opposed-phase imaging}/\text{spleen signal intensity on opposed-phase imaging})/(\text{adrenal signal intensity on in-phase imaging}/\text{spleen signal intensity on in-phase imaging})]$$

The thresholds used by the respective authors to differentiate between adenoma and non-adenoma on the basis of ASR are summarized in Table E3.

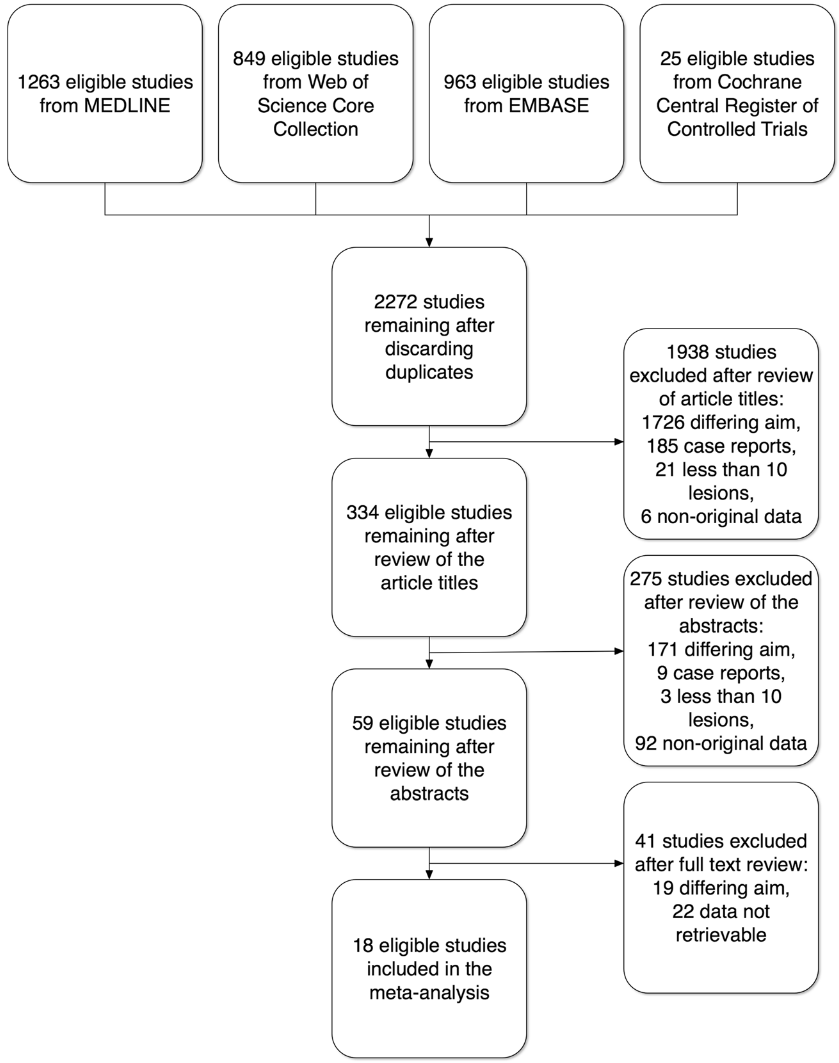


Fig. 1 PRISMA (Preferred Reporting Items for Systematic Reviews) flow diagram of search results and review results

Nine out of 18 studies also evaluated adrenal imaging methods other than CSI, including dynamic contrast-enhanced MRI (DCE) [28, 31, 33, 37, 38], CT histogram analysis [25, 30], delayed contrast-enhanced CT[27] and diffusion-weighted MRI (DWI) [26].

Only 2/18 studies (Sasai et al [38] and Rodacki et al [37]) performed a combined evaluation of CSI and an additional method (CSI and DCE in this case). Sasai et al and Rodacki et al did not compare the accuracy of CSI and CSI + DCE for statistical significance.

Methodologic quality assessment

The question items of the QUADAS-2 tool and the respective results are shown in Appendix E2. No study fulfilled all methodological criteria. The domains index test, reference standard, and flow and timing were most affected by risk of bias, with 9/18, 15/18 and 16/18 studies with high risk of bias in the respective domain. In regard to patient selection, 6/18 studies were rated to have unclear risk and 5/18 studies to have high risk of bias, mainly due to inappropriate exclusions.

In contrast, there were no concerns in regard to the applicability for the majority of studies included in this analysis.

Heterogeneity

The diagnostic odds ratio (DOR) for the pooled data was 250 (95% CI 96–652). Cochran's Q for the DOR was 27.95 and I^2 was 93 (95% CI 86–99). According to the criteria proposed by Higgins et al [17] the I^2 value is consistent with very high heterogeneity.

Publication bias

A funnel plot of the square root of the inverse of the effective sample size (ESS) against the DOR is shown in Fig. 2. The

Table 1 Basic characteristics of the studies included in the meta-analysis

Author, year	Study design	No. of patients	Sex (F/M)	Mean patient age (years)	Age range (years)	No. of lesions	Scanner manufacturer	Field strength (T)	Lesion evaluation
Afifi 2017	Prospective	52	26/26	44.8	21–77	58	Philips	1.5	SII, ASR
Bilbey 1995	Prospective	38	21/17	53	17–78	41	Not specified	0.5	ASR
Halefoglu 2009	Prospective	93	48/45	56.7	22–85	109	GE	1.5	SII
Halefoglu 2012	Prospective	108	63/45	57	30–87	126	GE	1.5	SII, ASR
Hennings 2009	Prospective	24	14/10	67.5	30–87	24	Philips	1.5	Qualitative
Inan 2008	Prospective	50	30/20	61	30–87	64	Philips	1.5	SII
Jhaveri 2006	Retrospective	37	17/20	64	33–77	39	GE	1.5	SII
Korobkin 1995	Prospective	43	24/19	57 in adenoma group, 51 in non-adenoma group	19–78	51	GE	1.5	SII
Marin 2010	Retrospective	34	22/12	57	18–79	37	Siemens	3	SII, ASR
Maurea 2006	Prospective	36	27/9	51.3	Not available	41	Philips	1.5	Visual
Namimoto 2001	Prospective	23	9/14	51.3	15–88	28	Siemens	1.5	SII
Ramalho 2011	Retrospective	44	22/22	59.1		50	Siemens	1.5	SII, ASR
Ream 2014	Retrospective	36	16/20	66.7	Not available	37	Siemens	1.5[a]	SII, ASR
Rodacki 204	Retrospective	205	123/82	58.7	Not available	239	Siemens	1.5/3[b]	SII, visual
Sasai 2003	Retrospective	53	23/30	55.7	30–84	57	Siemens	1.5	SII
Schindera 2008	Retrospective	21	14/7	63	26–88	23	Siemens/GE	3	SII
Wu 2016	Retrospective	49	26/23	Not available	20–81	60	GE/Siemens	1.5/3[c]	SII, ASR, visual
Yip 2010	Retrospective	192	129/63	54	13–83	196	Not specified	Not specified	Visual

[a] This study compared chemical shift MRI at 1.5T and 3T; only the data derived from 1.5T imaging were included in the meta-analysis

[b] The number of patients examined at 1.5T and 3T is not specified

[c] 42/49 patients were examined at 1.5T an 7/49 patients at 3T

Table 2 Number of adenomas and non-adenomas (based on the respective reference standard) and lesion size in each study included in the meta-analysis

Author, year	No. of lesions	Adenomas	Non-adenomas	Mean maximum diameter (cm)	Maximum diameter range (cm)
Afifi 2017	58	24	34	Adenoma: 2.4 ± 0.8 Non-adenoma: 7.3 ± 3.5	Adenoma: 1.4–4.3, Non-adenoma: 2.8–14.5
Bilbey 1995	41	17	24	Adenoma: 2.7 Non-adenoma: 5.7	1.3–12.5
Halefoglu 2009	126	96	30	Adenoma: 2.4 ± 1.3 Non-adenoma: 4.3 ± 1.8	Not available
Halefoglu 2012	109	67	42	3.0 ± 1.8	1.5–13.2
Hennings 2009	24	20	4	3.3	1.0–10
Inan 2008	64	48	16	Adenoma: 2.1 ± 0.8 Non-adenoma: 4.7 ± 2.9	Adenoma: 1.0–5.5 Non-adenoma: 1.4–12.0
Jhaveri 2006	39	28	11	Adenoma: 2.2 Non-adenoma: 2.6	Adenoma: 1.0–5.2 Non-adenoma: 1.4–5.2
Korobkin 1995	51	35	16	Adenoma: 2.7 Non-adenoma: 5.1	Adenoma: 1.0–8.6 Non-adenoma: 1.3–12.3
Marin 2010	37			Adenoma: 2.1 Non-adenoma: 4.1	Adenoma: 1.5–4.5 Non-adenoma: 1.3–7.0
Maurea 2006	41	29	12	Adenoma: 2.0 ± 0.9 Non-adenoma: 5.3 ± 2.9	Adenoma: 0.5–3.5 Non-adenoma: 1–11
Namimoto 2001	28	16	12	3.2	1.0–8.3
Ramalho 2011	50	38	12	1.9 ± 0.8	1.1–3.2
Ream 2014	37	27	10	Adenoma: 1.6 ± 0.6 Non-adenoma: 2.8 ± 2.2	Not available
Rodacki 2014	239	177	62	1.8	0.7–18.0
Sasai 2003	57	38	19	Not available	Adenoma: 0.6–11.0 Non-adenoma: 2.0–7.8
Schindera 2008	23	16	7	Adenoma: 2.5 Non-adenoma: 4.5	Adenoma: 1.1–4.1 Non-adenoma: 1.9–7.5
Wu 2016	60	48	12	Adenoma 1.7 Non-adenoma 4.4	Not available
Yip 2010	196	107	89	4	0.3–15.0

funnel plot shows a deviation of the distribution of the dots from the "ideal" funnel shape, implying the presence of publication bias. However, the results of the Deeks' test were not statistically significant ($p = 0.07$).

Diagnostic performance

The sensitivity and specificity of CSI for the detection of adrenal adenoma ranged from 0.57 to 1.0 and from 0.75 to 1.0, respectively (Table 4 and Fig. 3). The pooled sensitivity and specificity of CSI were 0.94 [95% CI 0.88–0.97] and 0.95 (95% CI 0.89–0.97]. The AUC of the SROC curve for the pooled data was 0.98 (Fig. 4).

The pooled sensitivity and specificity of SII for adrenal adenoma were 0.94 [95% CI 0.90–0.97] and 0.93 [95% CI 0.87–0.97], respectively (Fig. E3). The AUC of the SROC curve for the pooled SII data was 0.98.

The pooled sensitivity and specificity of ASR for adrenal adenoma were 0.93 [95% CI 0.83–0.97] and 0.98 [95% CI 0.87–1.0] (Fig. E4). The AUC of the SROC curve for the pooled ASR data was 0.99.

The pooled sensitivity and specificity of visual CSI evaluation for adrenal adenoma were 0.83 [95% CI 0.66–0.93] and 0.93 [95% CI 0.87–0.96], respectively (Fig. E5). The AUC of the SROC curve for the pooled visual evaluation data was 0.95.

Combined adrenal CSI and DCE

As mentioned above, only two studies performed a combined evaluation of adrenal CSI and DCE. Rodacki et al [37] found that the combination of CSI and DCE has higher sensitivity (94%) and specificity (98%) than standalone CSI (87% and 95%, respectively).

Sasai et al [38] found the positive and negative predictive values of CSI + DCE (0.95 and 0.96) to be higher than those of standalone CSI (0.9 and 0.94).

Table 3 Reference standard for each study included in the meta-analysis

Author, year	No. of lesions	Reference standard				
		Histopathology	Fine needle biopsy	Follow-up	Other	Minimum follow-up (months)
Afifi 2017	58	52	6	0	0	Not applicable
Bilbey 1995	41	17	11	13	0	12
Halefoglu 2009	109	0	6	76	27[a]	6
Halefoglu 2012	126	2	2	116	2[a]	6
Hennings 2009	24	7	1	16	0	12
Inan 2008	64	6	1	57	0	3*
Jhaveri 2006	39	4	0	34	1	6
Korobkin 1995	51	11	9	12	19[b]	6
Marin 2010	34	6	8	17	6[c]	6
Maurea 2006	41	19	3	19	0	Not specified
Namimoto 2001	28	6	4	18	0	12
Ramalho 2011	50	1	0	49	0	12
Ream 2014	37	7	0	28	2[c]	12
Rodacki 2014	239	59	3	177	0	12
Sasai 2003	57	15	0	42	0	12
Schindera 2008	23	4	2	12	5[c]	6
Wu 2016	60	5	0	20	35[c]	6
Yip 2010	196	196	0	0	0	Not applicable

[*] at least 3 months follow-up for malignant lesions, at least 6 months for benign lesions

[a] washout CT

[b] scintigraphy with Iodine-131 labeled NP 59 (6-beta-iodomethyl-19-norcholesterol)

[c] unenhanced CT

Discussion

With contemporary MR scanners, CSI of the upper abdomen is feasible in a single breath-hold and has become a standard component of abdominal MRI. The current meta-analysis shows that CSI has a very high sensitivity and specificity for adrenal adenoma. It demonstrates that the inclusion of CSI in abdominal MRI protocols provides an effective solution for classifying adrenal masses discovered on MR exams.

To our knowledge, this is the first meta-analysis on CSI in adrenal masses. It largely confirms the results of the majority of published studies on this subject, all monocentric.

Our subgroup analysis demonstrates a consistently high sensitivity and specificity of CSI, regardless if quantitative indices like SII and ASR or visual evaluation is used. The 95% confidence regions of the SROC curves in Fig. 4 demonstrate a very substantial overlap, which implies that there is no significant difference in regard to accuracy between the aforementioned CSI evaluation methods. Our results provide an argument for the use of visual evaluation, as there is no evidence that the diagnostic performance of CSI can be improved by using quantitative indices.

Despite the overall excellent accuracy of CSI, lipid-poor adenomas, which cannot be reliably identified with CSI, are still a problem. A possible solution is the use of additional MR sequences or other imaging methods.

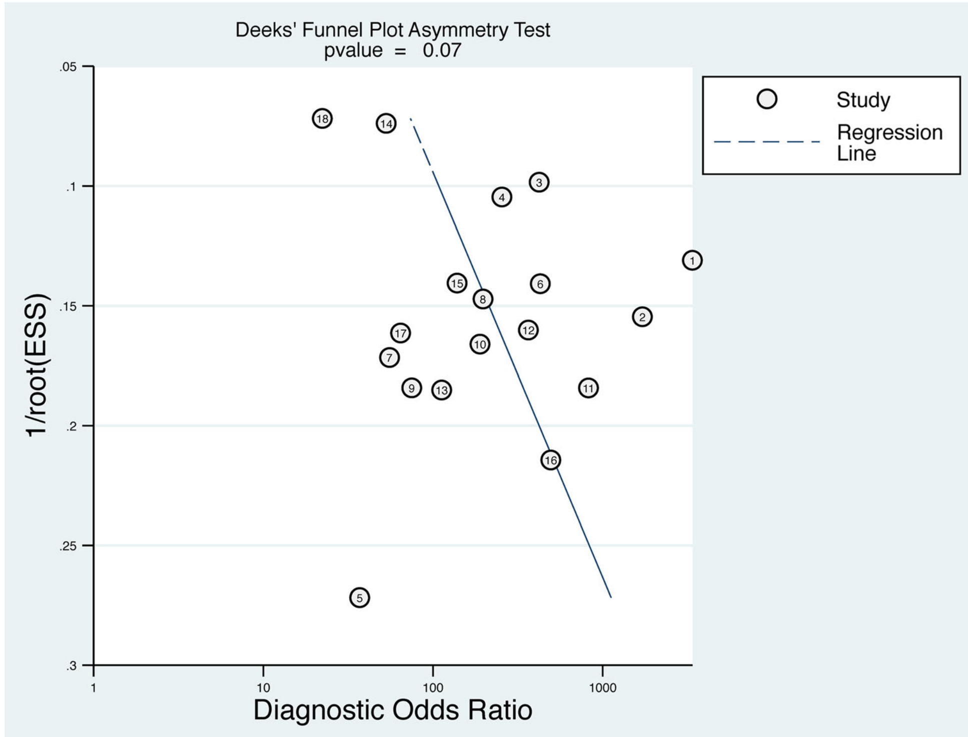
The current analysis focused on CSI and provides limited information on the usefulness of additional imaging. Only two studies evaluated the combined accuracy of CSI and an additional imaging method (DCE). Thus we were not able to extract data for a "combined" subgroup (CSI + additional imaging study) and have to rely on the findings of the individual studies.

The most promising additional imaging method evaluated by some studies in our meta-analysis is DCE. The results of Rodacki et al [37] and Sasai et al [38] suggest that the accuracy of CSI combined with DCE in regard to adrenal adenoma is higher in comparison to standalone CSI.

DWI is not expected to be a useful addition to adrenal CSI at all, as there is a large overlap between the apparent diffusion coefficient (ADC) values of adenomas and non-adenomas [26].

In view of the available data, we would not recommend additional imaging in cases when a lesion has been classified as adenoma on the basis of CSI. This is consistent with the

Fig. 2 Funnel plot as a graphical presentation for publication bias. The square root of the inverse of the effective sample size (ESS) is plotted against the diagnostic odds ratio (DOR). The distribution of the dots (representing the included studies) deviates from a funnel shape, which is suggestive of publication bias. However, the result of the corresponding Deeks' test is not statistically significant ($p = 0.07$)



current guidelines of both the European Society of Endocrinology (ESE) [44] and the American College of Radiology (ACR) [45], which do not recommend additional imaging if CSI (or CT) findings are unequivocal.

Table 4 Estimated sensitivities and specificities of the studies included in the meta-analysis, with 95% confidence intervals (CI) in square brackets

Author, year	TP	TN	FP	FN	Estimated sensitivity [95% CI]	Estimated specificity [95% CI]
Afifi 2017	24	34	0	0	1.00 [0.86–1.00]	1.00 [0.90–1.00]
Bilbey 1995	17	24	0	0	1.00 [0.80–1.00]	1.00 [0.86–1.00]
Halefoglu 2009	65	39	3	2	0.97 [0.90–1.00]	0.93 [0.81–0.99]
Halefoglu 2012	91	28	2	5	0.95 [0.88–0.98]	0.93 [0.78–0.99]
Hennings 2009	18	3	0	3	0.86 [0.64–0.97]	1.00 [0.29–1.00]
Inan 2008	45	16	0	3	0.94 [0.83–0.99]	1.00 [0.79–1.00]
Jhaveri 2006	20	11	0	8	0.71 [0.51–0.87]	1.00 [0.72–1.00]
Korobkin 1995	35	12	4	0	1.00 [0.90–1.00]	0.75 [0.48–0.93]
Marin 2010	25	9	1	3	0.89 [0.72–0.98]	0.90 [0.55–1.00]
Maurea 2006	26	12	0	3	0.90 [0.73–0.98]	1.00 [0.74–1.00]
Namimoto 2001	16	12	0	0	1.00 [0.79–1.00]	1.00 [0.74–1.00]
Ramalho 2011	36	12	0	2	0.95 [0.82–0.99]	1.00 [0.74–1.00]
Ream 2014	25	9	1	2	0.93 [0.76–0.99]	0.90 [0.55–1.00]
Rodacki 2014	157	54	8	20	0.89 [0.83–0.93]	0.87 [0.76–0.94]
Sasai 2003	37	15	4	1	0.97 [0.86–1.00]	0.79 [0.54–0.94]
Schindera 2008	16	7	0	0	1.00 [0.79–1.00]	1.00 [0.59–1.00]
Wu 2016	41	11	1	7	0.85 [0.72–0.94]	0.92 [0.62–1.00]
Yip 2010	61	84	5	46	0.57 [0.47–0.67]	0.94 [0.87–0.98]
Combined					0.94[0.88–0.97]	0.95[0.89–0.97]

TP true positive, *TN* true negative, *FP* false positive, *FN* false negative

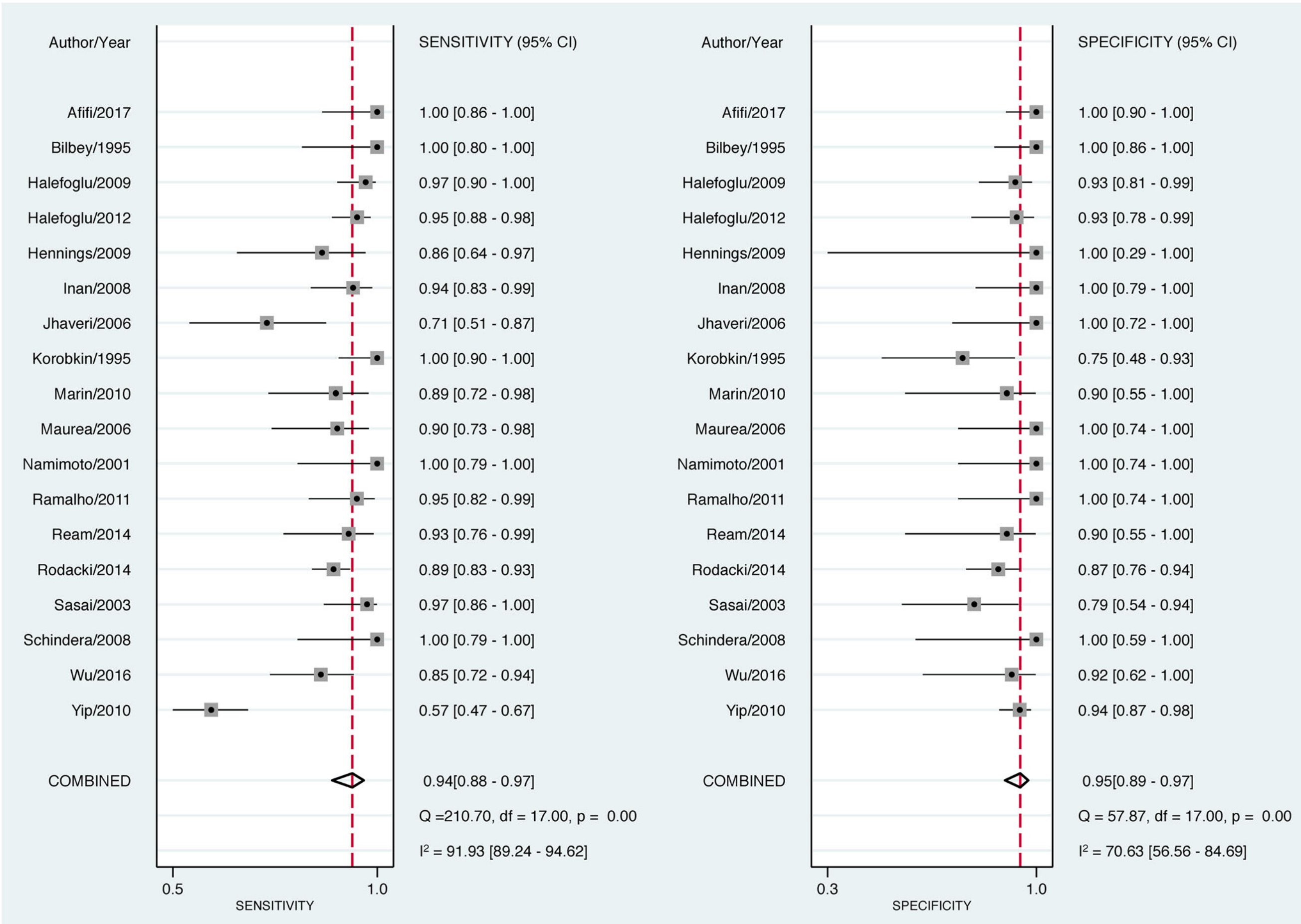


Fig. 3 Forests plots showing individual estimated sensitivities and specificities of studies included in the meta-analysis as well as the estimated pooled sensitivity and specificity, with 95% confidence intervals (95% CIs). The inconsistency index (I^2) and Cochran's Q are provided as measures of heterogeneity

In lesions classified as non-adenoma, we would currently recommend additional DCE, which can help detect lipid-poor adenoma. However, data on adrenal DCE are still scarce and further evaluation of the combined accuracy of CSI and DCE is necessary.

CT histogram analysis and delayed enhanced CT are not expected to be a useful addition to CSI, as both methods had lower sensitivity than CSI [25, 27, 30].

On the other hand, CSI can be used as additional imaging for adrenal masses incidentally detected on CT. As mentioned above, data from studies included in our meta-analysis show that CSI has lower sensitivity than advanced CT-based methods as CT histogram analysis and delayed enhanced CT [25, 27, 30]. Besides that, the sensitivity and specificity of CSI shown in our analysis compare favorably with most available data on non-enhanced CT, the most widely used CT-based method for assessing adrenal masses [46].

The current meta-analysis has several limitations. A very important limitation is the high degree of heterogeneity of the included studies, reflected by the inconsistency index I^2. The heterogeneity is not unexpected, as the included studies implemented a number of differing methods of CSI evaluation, different magnetic field strengths and the images were interpreted by reviewers with different amounts of experience. The subgroup analysis addresses the issue of varying evaluation methods.

In addition, the QUADAS-2 evaluation revealed that a high percentage of the included studies have a high risk of bias, which means that the design and/or the execution of the respective study is a possible source of systematic error, which in turn will influence the results of the meta-analysis. The QUADAS-2 domains reference standard, patient selection, and flow and timing were most affected by a high risk of bias. The non-uniform reference standard (imaging follow-up in some cases vs. histopathology in other patients) presented a widespread source of possible bias. The inconsistency of the thresholds used to differentiate between adenoma and non-adenoma on the basis of quantitative indices (ASII and ASR) is a further limitation and a possible source of bias. In some cases these

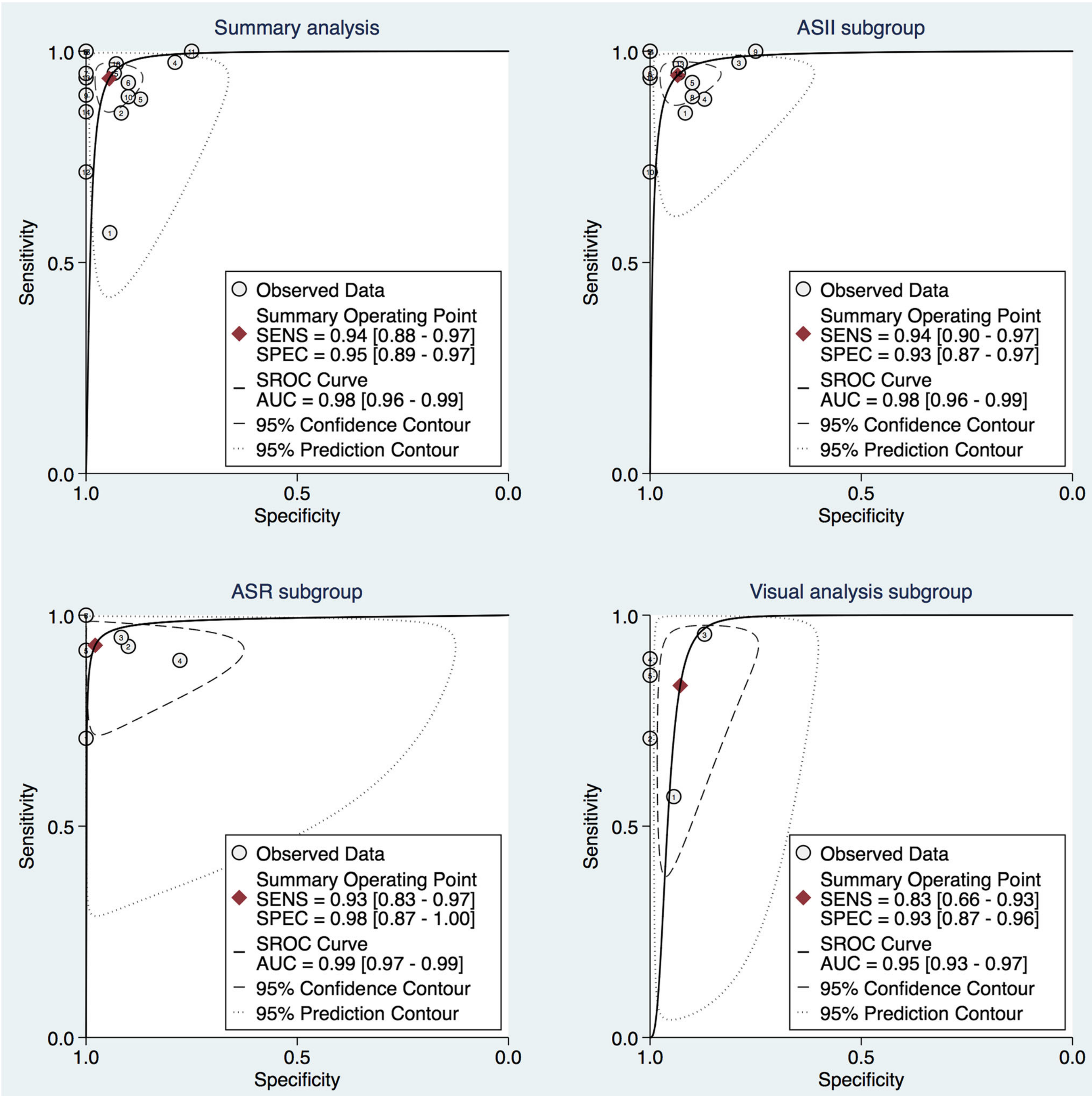


Fig. 4 Summary receiver operating characteristic (SROC) curves for summary analysis, adrenal signal intensity index (SII), adrenal-to-spleen ratio (ASR) and visual analyses, with estimated combined sensitivity and specificity values. AUC area under the ROC curve

thresholds were predefined on the basis of the results of earlier studies, while other authors determined the optimal thresholds on the basis of their own findings, e.g. using ROC curves.

The inclusion criteria for the meta-analysis constitute a further limitation. We included studies in which the studied condition was having or not having an adrenal adenoma. In contrast, a number of studies have evaluated the ability of MRI to differentiate between benign and malignant adrenal lesions. As adenomas are the most common benign lesions, we can assume that there are a lot of similarities between the results of such studies and the results of our meta-analysis. However, we chose to keep the inclusion criteria relatively stringent in order to avoid too much heterogeneity in regard to patient selection.

The majority of included studies do not provide information about possible malignant disease in the included patients, which constitutes a further limitation.

In addition, publication bias, which can influence the results of the meta-analysis, cannot be ruled out entirely. While the results of the Deeks' funnel plot asymmetry test did not show significant evidence of publication bias, the corresponding p value ($p = 0.07$) is not far above the significance threshold.

In conclusion, CSI allows for identifying adrenal adenoma with a very high sensitivity and specificity. Visual evaluation of adrenal CSI should be preferred, as the use of quantitative indices does not improve diagnostic accuracy.

Funding The authors state that this work has not received any funding.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Ivan Platzek, MD

Conflict of interest The authors of this manuscript declare no relationships with any companies whose products or services may be related to the subject matter of the article.

Statistics and biometry Mrs. Uta Schwanebeck, from the Kompetenzzentrum für Klinische Studien (KKS), Technical University of Dresden, kindly provided statistical advice for this manuscript.

Informed consent Written informed consent was not required for this study because the study is a systematic review and meta-analysis. The authors of this study did not have access to primary data, just metadata extracted from previously published studies. Because of this, the current study does not influence the well-being or privacy of patients.

Ethical approval Institutional review board approval was not required because the study is a systematic review and meta-analysis. The authors of this study did not have access to primary data, just metadata extracted from previously published studies. Because of this, the current study does not influence the well-being or privacy of patients.

Study subjects or cohorts overlap No study subjects or cohorts have been previously reported, except as a part of the primary studies included in this meta-analysis, which are listed in the reference list.

Methodology
• retrospective
• meta-analysis
• performed at one institution

References

1. Herrera MF, Grant CS, van Heerden JA, Sheedy PF, Ilstrup DM (1991) Incidentally discovered adrenal tumors: an institutional perspective. Surgery 110:1014–1021
2. Belldegrun A, Hussain S, Seltzer SE, Loughlin KR, Gittes RF, Richie JP (1986) Incidentally discovered mass of the adrenal gland. Surg Gynecol Obstet 163:203–208
3. Mantero F, Terzolo M, Arnaldi G et al (2000) A survey on adrenal incidentaloma in Italy. Study Group on Adrenal Tumors of the Italian Society of Endocrinology. J Clin Endocrinol Metab 85:637–644
4. Song JH, Chaudhry FS, Mayo-Smith WW (2008) The incidental adrenal mass on CT: prevalence of adrenal disease in 1,049 consecutive adrenal masses in patients with no known malignancy. AJR Am J Roentgenol 190:1163–1168
5. Lam KY, Lo CY (2002) Metastatic tumors of the adrenal glands: a 30-year experience in a teaching hospital. Clin Endocrinol (Oxf) 56:95–101
6. Candel AG, Gattuso P, Reyes CV, Prinz RA, Castelli MJ (1993) Fine-needle aspiration biopsy of adrenal masses in patients with extraadrenal malignancy. Surgery 114:1132–1136 discussion 1136-1137
7. Silverman SG, Mueller PR, Pinkney LP, Koenker RM, Seltzer SE (1993) Predictive value of image-guided adrenal biopsy: analysis of results of 101 biopsies. Radiology 187:715–718
8. Pokharel SS, Macura KJ, Kamel IR, Zaheer A (2013) Current MR imaging lipid detection techniques for diagnosis of lesions in the abdomen and pelvis. Radiographics 33:681–702
9. Korobkin M, Giordano TJ, Brodeur FJ et al (1996) Adrenal adenomas: relationship between histologic lipid and CT and MR findings. Radiology 200:743–747
10. Dwamena BA (2007) MIDAS: Stata module for meta-analytical integration of diagnostic test accuracy studies, S456880. Boston College Department of Economics, Boston
11. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH (2005) Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 58:982–990
12. Lewis S, Clarke M (2001) Forest plots: trying to see the wood and the trees. BMJ 322:1479–1480
13. Doebler P, Holling H (2015) Meta-analysis of diagnostic accuracy and ROC curves with covariate adjusted semiparametric mixtures. Psychometrika 80:1084–1104
14. Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JA (2007) A unification of models for meta-analysis of diagnostic accuracy studies. Biostatistics 8:239–251
15. Rosman AS, Korsten MA (2007) Application of summary receiver operating characteristics (sROC) analysis to diagnostic clinical testing. Adv Med Sci 52:76–82
16. Jones CM, Athanasiou T (2005) Summary receiver operating characteristic curve analysis techniques in the evaluation of diagnostic tests. Ann Thorac Surg 79:16–20
17. Higgins JP, Thompson SG, Deeks JJ, Altman DG (2003) Measuring inconsistency in meta-analyses. BMJ 327:557–560
18. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PM (2003) The diagnostic odds ratio: a single indicator of test performance. J Clin Epidemiol 56:1129–1135
19. Dickersin K (1990) The existence of publication bias and risk factors for its occurrence. JAMA 263:1385–1389
20. Deeks JJ, Macaskill P, Irwig L (2005) The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J Clin Epidemiol 58:882–893
21. Egger M, Davey Smith G, Schneider M, Minder C (1997) Bias in meta-analysis detected by a simple. graphical test. BMJ 315:629–634
22. Whiting PF, Rutjes AW, Westwood ME et al (2011) QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 155:529–536
23. Afifi AH, Abdel Shafy HL, Ramadan AA, Ataa MA, Assad SN (2017) Role of quantitative chemical shift magnetic resonance imaging and chemical shift subtraction technique in discriminating adenomatous from non adenomatous adrenal solid lesions. Egypt J Radiol Nucl Med 48:271–284
24. Bilbey JH, McLoughlin RF, Kurkjian PS et al (1995) MR imaging of adrenal masses: value of chemical-shift imaging for

distinguishing adenomas from other tumors. AJR Am J Roentgenol 164:637–642
25. Halefoglu AM, Yasar A, Bas N, Ozel A, Erturk SM, Basak M (2009) Comparison of computed tomography histogram analysis and chemical-shift magnetic resonance imaging for adrenal mass characterization. Acta Radiol 50:1071–1079
26. Halefoglu AM, Altun I, Disli C, Ulusay SM, Ozel BD, Basak M (2012) A prospective study on the utility of diffusion-weighted and quantitative chemical-shift magnetic resonance imaging in the distinction of adrenal adenomas and metastases. J Comput Assist Tomogr 36:367–374
27. Hennings J, Hellman P, Ahlstrom H, Sundin A (2009) Computed tomography, magnetic resonance imaging and 11C-metomidate positron emission tomography for evaluation of adrenal incidentalomas. Eur J Radiol 69:314–323
28. Inan N, Arslan A, Akansel G, Anik Y, Balci NC, Demirci A (2008) Dynamic contrast enhanced MRI in the differential diagnosis of adrenal adenomas and malignant adrenal masses. Eur J Radiol 65: 154–162
29. Israel GM, Korobkin M, Wang C, Hecht EN, Krinsky GA (2004) Comparison of unenhanced CT and chemical shift MRI in evaluating lipid-rich adrenal adenomas. Am J Roentgenol 183:215–219
30. Jhaveri KS, Wong F, Ghai S, Haider MA (2006) Comparison of CT histogram analysis and chemical shift MRI in the characterization of indeterminate adrenal nodules. AJR Am J Roentgenol 187: 1303–1308
31. Korobkin M, Lombardi TJ, Aisen AM et al (1995) Characterization of adrenal masses with chemical shift and gadolinium-enhanced MR imaging. Radiology 197:411–418
32. Marin D, Soher BJ, Dale BM, Boll DT, Youngblood RS, Merkle EM (2010) Characterization of adrenal lesions: comparison of 2D and 3D dual gradient-echo MR imaging at 3 T–preliminary results. Radiology 254:179–187
33. Maurea S, Imbriaco M, D'Angelillo M, Mollica C, Camera L, Salvatore M (2006) Diagnostic accuracy of chemical-shift MR imaging to differentiate between adrenal adenomas and non adenoma adrenal lesions. Radiol Med 111:674–686
34. Namimoto T, Yamashita Y, Mitsuzaki K et al (2001) Adrenal masses: quantification of fat content with double-echo chemical shift in-phase and opposed-phase FLASH MR images for differentiation of adrenal adenomas. Radiology 218:642–646
35. Ramalho M, de Campos RO, Heredia V et al (2011) Characterization of adrenal lesions with 1.5-T MRI: preliminary observations on comparison of three in-phase and out-of-phase gradient-echo techniques. AJR Am J Roentgenol 197:415–423
36. Ream JM, Gaing B, Mussi TC, Rosenkrantz AB (2015) 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. Am J Roentgenol 204:536–541
37. Rodacki K, Ramalho M, Dale BM et al (2014) Combined chemical shift imaging with early dynamic serial gadolinium-enhanced MRI in the characterization of adrenal lesions. AJR Am J Roentgenol 203:99–106
38. Sasai N, Togami I, Tsunoda M, Sei T, Akaki S, Hiraki Y (2003) Differential diagnosis of adrenal masses by chemical shift and dynamic gadolinium enhanced MR imaging. Acta Med Okayama 57: 163–170
39. Schindera ST, Soher BJ, Delong DM, Dale BM, Merkle EM (2008) Effect of echo time pair selection on quantitative analysis for adrenal tumor characterization with in-phase and opposed-phase MR imaging: initial experience. Radiology 248:140–147
40. Wu YW, Tan CH (2016) Determination of a cutoff attenuation value on single-phase contrast-enhanced CT for characterizing adrenal nodules via chemical shift MRI. Abdom Radiol (NY) 41:1170–1177
41. Yip L, Tublin ME, Falcone JA et al (2010) The adrenal mass: correlation of histopathology with imaging. Ann Surg Oncol 17: 846–852
42. Ichikawa T, Fujimoto H, Uchiyama G, Murakami K, Ohtomo K (1994) Distinction between adrenal adenomas and metastases using 0.5 Telsa MR imaging: diagnosis with out-of-phase T2*-weighted gradient-field-echo image. Nihon Igaku Hoshasen Gakkai Zasshi 54:229–234
43. Fujiyoshi F, Nakajo M, Fukukura Y, Tsuchimochi S (2003) Characterization of adrenal tumors by chemical shift fast low-angle shot MR imaging: comparison of four methods of quantitative evaluation. AJR Am J Roentgenol 180:1649–1657
44. Fassnacht M, Arlt W, Bancos I et al (2016) Management of adrenal incidentalomas: European Society of Endocrinology Clinical Practice Guideline in collaboration with the European Network for the Study of Adrenal Tumors. Eur J Endocrinol 175:G1–G34
45. Mayo-Smith WW, Song JH, Boland GL et al (2017) Management of incidental adrenal masses: a white paper of the ACR Incidental Findings Committee. J Am Coll Radiol 14:1038–1044
46. Boland GW, Lee MJ, Gazelle GS, Halpern EF, McNicholas MM, Mueller PR (1998) Characterization of adrenal masses using unenhanced CT: an analysis of the CT literature. Am J Roentgenol 171:201–204