



Meta-analysis of CT and MRI for differentiation of autoimmune pancreatitis from pancreatic adenocarcinoma

Jiyeon Ha¹ · Sang Hyun Choi¹ · Jae Ho Byun¹ · Kyung Won Kim¹ · So Yeon Kim¹ · Jin Hee Kim¹ · Hyoung Jung Kim¹

Received: 4 April 2020 / Revised: 28 August 2020 / Accepted: 13 October 2020 / Published online: 4 November 2020
© European Society of Radiology 2020

Abstract

Objectives To systematically determine the diagnostic performance of computed tomography (CT) and magnetic resonance imaging (MRI) for differentiating autoimmune pancreatitis (AIP) from pancreatic ductal adenocarcinoma (PDAC), with a comparison between the two imaging modalities.

Methods Literature search was conducted using PubMed and EMBASE databases to identify original articles published between 2009 and 2019 reporting the diagnostic performance of CT and MRI for differentiating AIP from PDAC. The meta-analytic sensitivity and specificity of CT and MRI were calculated, and compared using a bivariate random effects model. Subgroup analysis for differentiating focal AIP from PDAC was performed.

Results Of the 856 articles screened, 11 eligible articles are remained, i.e., five studies for CT, four for MRI, and two for both. The meta-analytic summary sensitivity and specificity of CT were 59% (95% confidence interval [CI], 41–75%) and 99% (95% CI, 88–100%), respectively, while those of MRI were 84% (95% CI, 68–93%) and 97% (95% CI, 87–99%). MRI had a significantly higher meta-analytic summary sensitivity than CT (84% vs. 59%, $p = 0.02$) but a similar specificity (97% vs. 99%, $p = 0.18$). In the subgroup analysis for focal AIP, the sensitivity for distinguishing between focal AIP and PDAC was lower than that for the overall analysis. MRI had a higher sensitivity than CT (76% vs. 50%, $p = 0.28$) but a similar specificity (97% vs. 98%, $p = 0.07$).

Conclusion MRI might be clinically more useful to evaluate patients with AIP, particularly for differentiating AIP from PDAC.

Key Points

- MRI had an overall good diagnostic performance to differentiate AIP from PDAC with a meta-analytic summary estimate of 83% for sensitivity and of 97% for specificity.
- CT had a very high specificity (99%), but a suboptimal sensitivity (59%) for differentiating AIP from PDAC.
- Compared with CT, MRI had a higher sensitivity, but a similar specificity.

Keywords Autoimmune pancreatitis · Pancreatic cancer · Multidetector computed tomography · Magnetic resonance imaging · Meta-analysis

Abbreviations

AIP Autoimmune pancreatitis
CT Computed tomography

DWI Diffusion-weighted imaging
HSROC Hierarchical summary receiver operating characteristic
MRI Magnetic resonance imaging
MRP Magnetic resonance pancreatography
PDAC Pancreatic ductal adenocarcinoma

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00330-020-07416-1>) contains supplementary material, which is available to authorized users.

✉ Sang Hyun Choi
edwardchoi83@gmail.com

Introduction

Autoimmune pancreatitis (AIP) is a particular type of chronic pancreatitis characterized by periductal infiltration by IgG4-positive plasma cells leading to periductal and interlobular

¹ Department of Radiology and the Research Institute of Radiology, University of Ulsan College of Medicine, Asan Medical Center, 88 Olympic-Ro 43-Gil, Songpa-Gu, Seoul 138-736, South Korea

fibrosis [1]. AIP can be categorized into diffuse, focal, or multifocal AIP according to the involvement pattern [2]. As 28–48% of all AIP can manifest as a focal pancreatic mass with pancreatic duct stricture, which can also be shown in the case of pancreatic ductal adenocarcinoma (PDAC) [3–6], and both AIP and PDAC have common epidemiologic and clinical manifestations such as preponderance in elderly men and presentation with painless jaundice [7], it is challenging to differentiate AIP from PDAC. Considering the different treatment strategies and prognoses between AIP and PDAC [8, 9], the accurate differential diagnosis of the two diseases is critical to prevent unnecessary surgical resection in patients with AIP, with it being reported that 3–9% of patients who underwent resection for a presumed PDAC actually had AIP [10].

Imaging tests including computed tomography (CT) and magnetic resonance imaging (MRI) have been widely used in the diagnosis and management of patients with AIP. The characteristic imaging findings play an important role in the diagnosis of AIP in most classification systems, including the Japanese consensus guidelines [11], the Mayo clinic HISORt criteria [12], and the International Association of Pancreatology guidelines [13]. Various CT and MRI imaging features are suggested to be valuable for the differential diagnosis between AIP and PDAC, including the morphology of the pancreatic mass, pattern of the pancreatic duct stricture, and enhancement pattern [3–6, 8, 9, 12, 14–24].

Previous studies have reported the diagnostic performance of CT and MRI for differentiating AIP from PDAC when using these imaging features, but the reported values are quite variable, i.e., 18–94% for sensitivity and 85–100% for specificity [8, 9, 12, 16, 18, 21, 23, 25–28]. In addition, as only one study [9] actually compared the diagnostic performance for differentiating AIP from PDAC between CT and MRI, there is only limited information on which to determine the better imaging test for differentiating AIP from PDAC. Considering that MRI has higher soft-tissue contrast than CT, which is useful for detecting focal pancreatic masses [9], we hypothesized that MRI might have the better diagnostic performance for differentiating AIP from PDAC.

Therefore, we aimed to systematically determine the diagnostic performance of CT and MRI for differentiating AIP from PDAC with a meta-analytic comparison between the two tests.

Materials and methods

This meta-analysis followed the Preferred Reporting Items for Systemic Reviews and Meta-Analysis (PRISMA) guidelines for conducting and reporting a study. The analysis was executed using methods advocated by the Diagnostic Test Accuracy Working Group of the Cochrane Collaboration

and the Agency for Healthcare Research and Quality (AHRQ) [29].

Literature search strategy

A comprehensive literature search of PubMed and EMBASE databases was conducted. The query terms were designed for a sensitive literature search and included the search terms of “Pancreas,” “Pancreatitis,” “Pancreatic neoplasm,” “MRI,” “CT,” and “Diagnosis.” The detailed search terms are listed in Supplementary Table 1. Because of the recent fast evolution of imaging techniques, the search included only articles published between January 1, 2009, and December 31, 2019. The studies were limited to English language articles and human patients.

Eligibility criteria

Duplicate articles were removed and the eligibility of articles was assessed by one author according to the following criteria: (a) patients: patients with AIP or PDAC; (b) index test: MRI with or without contrast enhancement; (c) comparison: CT with or without contrast enhancement; (d) outcomes: diagnostic accuracy for differentiating AIP from PDAC; (e) study design: both observational (retrospective or prospective) and clinical trials; and (f) reference standard: both pathological or clinical diagnoses. The listed exclusion criteria were as follows: (a) studies with duplicate patients and data; (b) studies without sufficient information to create a 2×2 diagnostic table; (c) review articles, comments, letters, editorials, case reports, and conference articles; and (d) studies not in the field of interest. Two reviewers (with 9 and 5 years of abdominal radiology) screened the abstracts and titles of the retrieved articles and reviewed the full text of potentially eligible articles. Both review sessions were conducted independently, and articles with definite ineligibility were excluded. Articles with discordant eligibility were reviewed in a consensus meeting with a third reviewer (who had 13 years of experience in abdominal radiology).

Data extraction

The following data were extracted from eligible articles: (a) study characteristics (authors, publication date, study period, study design, and study type); (b) study subject characteristics (total number of patients, age, and the number of patients with the underlying disease); (c) lesion characteristics (total number of lesions, the number of patients with the underlying disease, AIP type [focal or diffuse], AIP diagnostic criteria); (d) CT techniques (use of contrast-enhanced images, multi-phase enhanced images, slice thickness, and number of channels); (e) MRI techniques (magnetic field strength, use of contrast-enhanced images, magnetic resonance pancreatography

[MRP], and diffusion-weighted imaging [DWI]); (f) the method of image review (single reviewer, multiple independent reviewer, or multiple reviewers in consensus); (g) analyzed imaging features; (h) the clarity of blinding to the reference standard during image review; and (i) the reference standard (pathologic diagnosis or clinical diagnosis). Regarding the reference standard, the number of cases determined by pathologic diagnosis and clinical diagnosis was extracted, and the specific criteria for AIP and the method of pathologic diagnosis were investigated. To determine diagnostic accuracy, the numbers of true positives, false positives, false negatives, and true negatives were extracted. In the case of multiple sets of results due to a use of each individual feature, the result set with the highest Youden index value was used. The two reviewers performed data extraction independently, and all disagreements were re-evaluated in a consensus meeting with a third reviewer.

Assessment of study quality

Two reviewers independently assessed the quality of the eligible articles using the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) criteria [30]. Four different domains were assessed: patient selection, index test, reference standard, and flow of patients through the study.

Data synthesis and statistical analysis

Accuracy of CT and MRI for differentiating AIP from PDAC To investigate the effect of the type of imaging test on the diagnostic accuracy, the results of all articles were segregated and analyzed as separate studies according to the type of imaging test (CT vs. MRI). All reported results were analyzed on a per-patient basis. The sensitivity and specificity for differentiating AIP from PDAC and their 95% confidence intervals (CIs) were determined using the relevant extracted data from each individual study. Sensitivity was defined as those patients showing AIP lesions on imaging tests divided by all patients diagnosed as AIP, while specificity was defined as those patients showing PDAC lesions on imaging tests divided by all patients diagnosed as PDAC. The meta-analytic summary sensitivity and specificity were calculated for both CT and MRI using a hierarchical modeling method, i.e., a bivariate random effects model. The hierarchical summary receiver operating characteristic (HSROC) model was used to acquire the summary receiver operating characteristic curve [29]. The diagnostic performance for differentiating AIP from PDAC was compared between CT and MRI using meta-regression analytic methods. In addition, a subgroup analysis was performed for the differentiation of focal AIP from PDAC.

Higgins I^2 statistic was used to examine for the presence of heterogeneity in the sensitivity and specificity of the included studies. When substantial study heterogeneity ($I^2 > 50\%$) was present, the presence of a threshold effect was evaluated using

visual assessment of coupled forest plots of sensitivity and specificity. In addition, the Spearman correlation coefficient between the sensitivity and FP rate was evaluated, with a correlation coefficient > 0.6 being considered to indicate a considerable threshold effect.

Meta-regression analysis Meta-regression was conducted to explore the causes of heterogeneity between the studies. The covariates used in the meta-regression included (a) patient composition (AIP $>$ PDACA vs. PDAC $>$ AIP); (b) publication year (before 2015 vs. after 2015); (c) CT slice thickness (≤ 3 mm vs. > 3 mm); (d) the use of multi-phase enhanced CT (multi-phase vs. single-phase); (e) MR magnet strength (1.5 T only vs. others); (f) the use of MRP (with MRP vs. without MRP); (g) the use of DWI (with DWI vs. without DWI); (h) image review (multiple independent reviewers vs. single reviewer or multiple reviewers with consensus); and (i) the clarity of blinding to the reference standard during review (clear vs. unclear).

Analysis of publication bias Publication bias was determined by visual assessment of a Deeks' funnel plot, and statistical significance was evaluated by Deeks' asymmetry test.

All statistical evaluations were conducted using Stata version 15.0 (StataCorp LP), with $p < 0.05$ being taken to indicate statistical significance.

Results

Literature search

Overall, 856 articles were retrieved after elimination of duplicate articles. Of these potentially eligible articles, 80 were initially identified as possibly being of interest according to their titles and abstracts, and then, a further 69 articles were excluded after full text evaluation. A schematic diagram of the study inclusion process is presented in Fig. 1. Among the remaining 11 articles [8, 9, 12, 16–18, 21, 23, 26–28], five reported the diagnostic accuracy of CT [8, 12, 18, 27, 28], four reported the diagnostic accuracy of MRI [16, 17, 21, 23], and two reported the diagnostic accuracy of both CT and MRI [9, 26].

The characteristics of the included articles are listed in Table 1. All included studies were retrospective case-control studies. Regarding the type of AIP, five studies included only focal AIP [16, 21, 26–28], whereas six included both diffuse and focal AIP [8, 9, 12, 17, 18, 23]. Of the seven studies evaluating CT, one used single-phase contrast-enhanced images [12], whereas six used multi-phase contrast-enhanced images [8, 9, 18, 26–28]. Four studies used thin slices (≤ 3 mm) [9, 18, 27, 28], whereas the others used a slice thickness > 3 mm or did not report the slice thickness [8, 12, 26]. Of

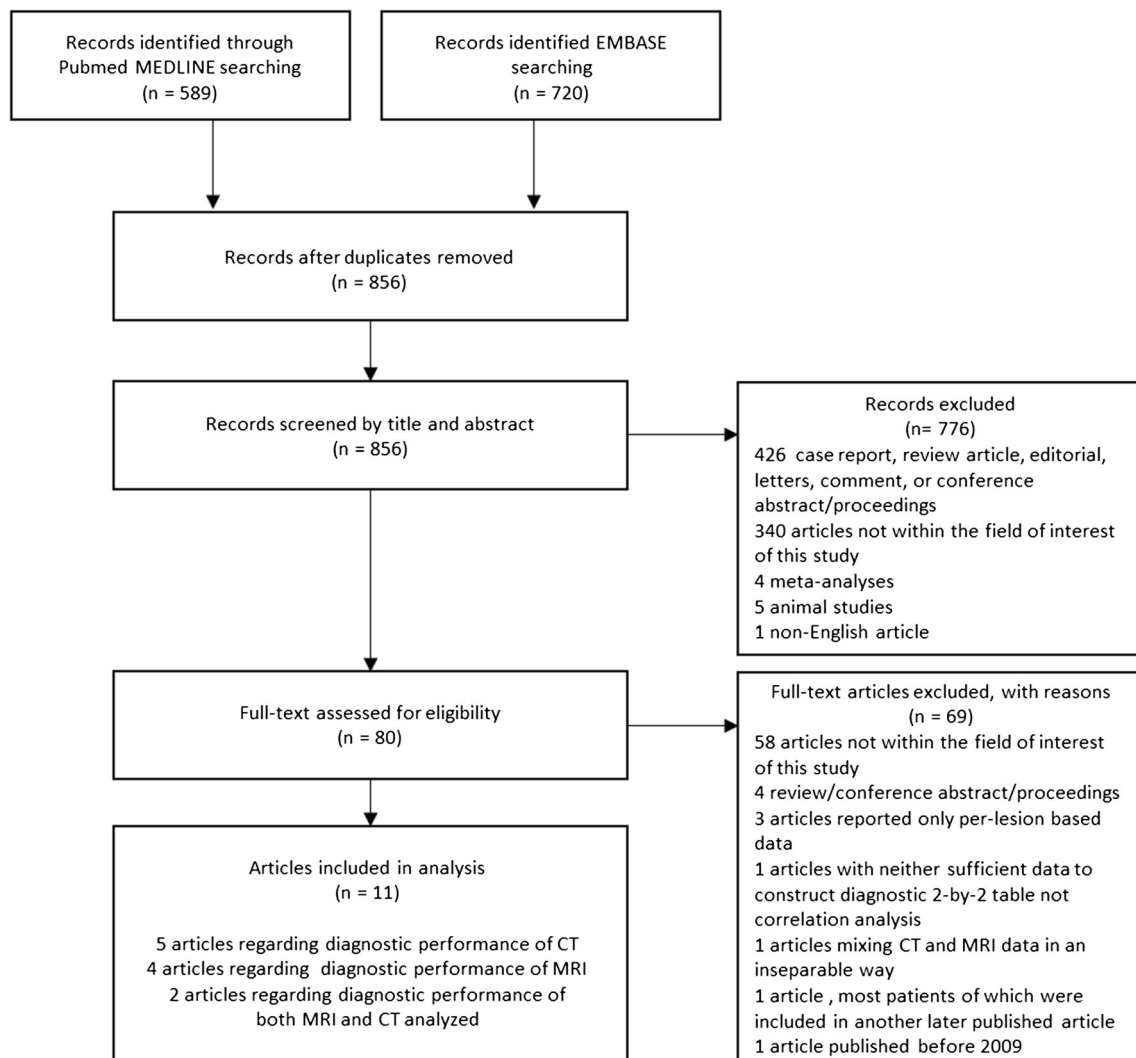


Fig. 1 Flow diagram of the study selection process for the systematic review and meta-analysis

the six studies evaluating MRI, one used both 1.5-T and 3-T MRI [16], whereas five used only 1.5-T MRI [9, 17, 21, 23, 26]. One study did not report the magnetic field strength of the scanner [26]. Five studies used MRP [9, 16, 21, 23, 26], and three studies used DWI [9, 16, 17].

For diagnosing AIP, 10 studies used clinical diagnostic criteria [8, 9, 12, 16–18, 21, 23, 26, 28], and one study used pathology as a reference standard for AIP [27] (Supplementary Table 2). Of the 10 studies using clinical diagnostic criteria, three studies used one diagnostic criterion, and the remaining seven studies used a combination of diagnostic criteria, including the Mayo clinic HISORT criteria, Japanese diagnostic criteria, International Consensus Diagnostic Criteria, Asian diagnostic criteria, and Korean diagnostic criteria. For diagnosing PDAC, nine studies used pathology only [8, 9, 16–18, 21, 23, 27, 28], and two studies used both clinical diagnosis and pathology as a reference standard for PDAC [12, 26]. Of the nine studies using pathology only, two studies used surgical specimens, and the remaining

seven studies used the pathology of both surgical specimens and biopsies.

The detailed imaging features for differentiating AIP from PDAC in 11 studies are summarized in Supplementary Table 3. Nine studies reported the diagnostic performance of each individual imaging feature [12, 16–18, 21, 23, 26–28], but two studies reported the diagnostic performance of CT or MRI considering multiple imaging features together using a 3-point or 5-point scale system [8, 9].

Study quality according to QUADAS-2

The quality of the 11 finally included articles is summarized in Fig. 2. The method for patient selection and the clarity of blinding to the reference standard during review were remarkable areas of quality concern. Ten studies did not include consecutive patients, and all the included studies were of a case-control design. All of the included studies were unclear about blinding to the results of the index test when interpreting

Table 1 Characteristics of the included articles

†Author (year of publication)	Study design	Study type	Number of patients (AIP/PDAC)	‡Patient age (AIP/PDAC)	AIP type	AIP diagnostic criteria	Image test	CT		MR		Image reviewer	Reference standard for PDAC
								CE	Phase Slice thickness	Magnet	CE		
Chari ST (2009)	Retrospective	Case-control	48/100	61/67	F/D	HISORt	CT	CE	Single > 3	N/A	–	–	MC Clinical diagnosis and pathology
Kamisawa T (2010)	Retrospective	Case-control	13/40	58.6/59.8	F/D	HISORt	MR	–	–	–	1.5 T	No CE	Used MC
Park SH (2010)	Retrospective	Case-control	38/40	56.4/60.1	F/D	rJDC, HISORt	MR	–	–	–	1.5 T	No CE	Used MC
Hur BY (2012)	Retrospective	Case-control	9/28	55.4/61.9	F	ADC	MR	–	–	–	1.5 T 3 T	CE	Used MC
Kawai Y (2012)	Retrospective	Case-control	55/50	63.6/65.8	F/D	rJDC, HISORt, KDC	CT	CE	Multi ≤ 3	≥ 64	–	–	– MI
Kim JH (2012)	Retrospective	Case-control	84/73	56.7/58.3	F/D	ADC, HISORt	CT	CE	Multi > 3	< 64	–	–	– MC
Naitoh I (2012)	Retrospective	Case-control	36/60, 26/38	69.0/70.1	F	rJDC, revised HISORt	CT, MR	CE	Multi NA	N/A	NA	No CE	Used Not used
Sun GF (2013)	Retrospective	Case-control	19/30	54/58.2	F	ADC	CT	CE	Multi ≤ 3	< 64	–	–	– MC
Lee S (2018)	Retrospective	Case-control	61/122	58.4/62	F/D	ADC, HISORt, ICDC	CT, MR	CE	Multi ≤ 3	≥ 64	1.5 T	CE	Used MI
Kwon JH (2019)	Retrospective	Case-control	36/72	56.7/60.5	F	ADC, HISORt, ICDC	MR	–	–	–	1.5 T	CE	Used Not used
Ren S (2019)	Retrospective	Case-control	21/47	60.1/64.5	F	No specific criteria	CT	CE	Multi ≤ 3	≥ 64	–	–	– MI

CE, contrast enhanced; NA, not available; F, focal type pancreatitis; D, diffuse type pancreatitis; MI, multiple independent reviewers; MC, multiple reviewers with consensus; HISORt, histology, imaging, serology, other organ involvement, and response to steroid therapy; ADC, Asia diagnostic criteria; rJDC, revised Japanese diagnostic criteria; KDC, Korean diagnostic criteria; ICDC, International Consensus Diagnostic Criteria

† Articles are listed according to year of publication and the alphabetical order of the names of the first authors within the same year of publication

‡ Mean patient age in years

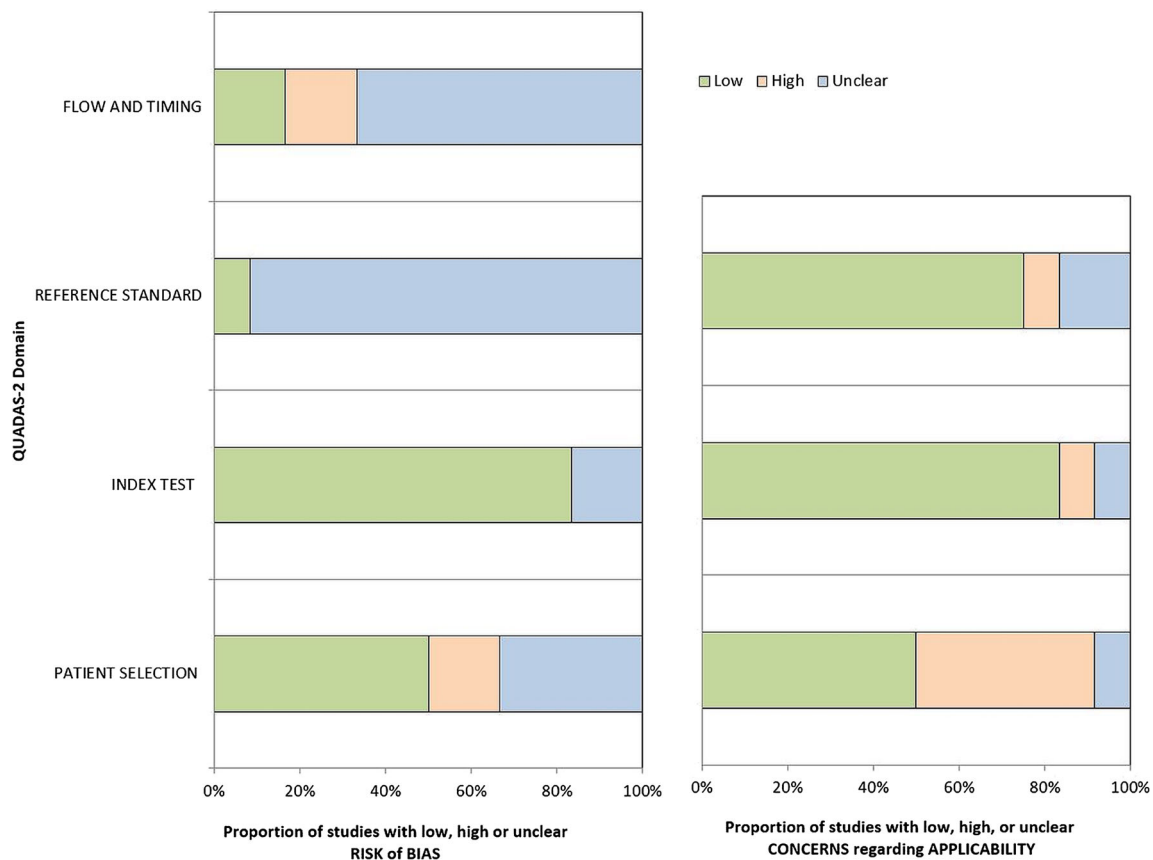


Fig. 2 Quality assessment of the studies according to QUADAS-2 criteria. The methodological quality of the articles is presented as the proportion of articles (0–100%) with low (i.e., high quality), high, or

unclear risk of bias, and the proportion of articles with low (i.e., high quality), high, or unclear concerns regarding the applicability of each domain

the reference standard, and also about blinding to the results of the reference standard while interpreting the index test. Furthermore, a risk of bias was noted in the flow and timing criteria, with ten studies showing uncertainty as to whether the time interval between the index text and reference standard was appropriate.

Diagnostic performance of CT and MRI for differentiating AIP from PDAC

Both the sensitivities (17–86%) and specificities (85–100%) of the individual CT studies (806 patients in seven studies) showed considerable variability (Table 2). The meta-analytic summary sensitivity and specificity of CT were 59% (95% CI, 41–75%) and 99% (95% CI, 88–100%). Substantial study heterogeneity was found in both sensitivity and specificity ($I^2 = 88\%$ and 81% , respectively); however, the coupled forest plot of sensitivity and specificity did not show a threshold effect (Fig. 3a), and the Spearman correlation between sensitivity and FP rate was 0.8 ($p = 0.12$), indicating a positive correlation without a statistical significance. The HSROC curve revealed a quite difference between the 95% confidence and prediction regions, thereby indicating considerable heterogeneity across the studies (Fig. 3b).

For MRI (485 patients in six studies), both sensitivity and specificity showed considerable variability across individual studies (44–94% for sensitivity and 77–100% for specificity), with meta-analytic summary sensitivity and specificity of 84% (95% CI, 68–93%) and 97% (95% CI, 87–99%), respectively (Table 2). Substantial study heterogeneity was found in both sensitivity and specificity ($I^2 = 75\%$ and 81% , respectively), but the coupled forest plot of sensitivity and specificity did not show a threshold effect (Fig. 3c), and the Spearman correlation between sensitivity and FP rate was 0.79 ($p = 0.06$), indicating a positive correlation without a statistical significance (Fig. 3d).

When we compared the diagnostic performance for differentiating AIP from PDAC between CT and MRI, MRI had a significantly higher meta-analytic summary sensitivity than CT (84% vs. 59%, $p = 0.02$) but a similar specificity to CT (97% vs. 99%, $p = 0.18$) (Fig. 4).

Subgroup analysis for differentiating focal AIP from PDAC

The meta-analytic summary sensitivities and specificities for differentiating focal AIP from PDAC were 50% (95% CI, 16–85%) and 98% (95% CI, 93–100%), respectively, for CT, and

Table 2 Diagnostic performance of CT and MRI for differentiating AIP from PDAC

	Author (year of publication)	Total number of patients	Number of patients				Sensitivity (95% CI)	Specificity (95% CI)
			TP	FP	FN	TN		
CT	Chari ST (2009)	148	25	0	23	100	52% (37, 67)	100% (96, 100)
	Kawai Y (2012)	105	37	5	18	45	67% (53, 79)	90% (78, 97)
	Kim JH (2012)	157	49	0	35	73	58% (47, 69)	100% (95, 100)
	Naitoh I (2012)	96	6	0	30	60	17% (6, 33)	100% (94, 100)
	Sun GF (2013)	49	17	0	19	30	47% (30, 65)	100% (88, 100)
	Lee S (2018)	183	49	10	12	112	80% (68, 89)	92% (85, 96)
	Ren S (2019)	68	18	7	3	40	86% (64, 97)	85% (72, 94)
Higgins I^2 for study heterogeneity						88%	81%	
Meta-analytic summary estimate using the bivariate model						59% (41, 75)	99% (88, 100)	
MRI	Kamisawa T (2010)	53	12	9	1	31	92% (64, 100)	77% (62, 89)
	Park SH (2010)	78	27	0	11	40	71% (54, 85)	100% (91, 100)
	Hur BY (2012)	37	4	0	5	28	44% (14, 79)	100% (88, 100)
	Naitoh I (2012)	26	19	2	7	36	73% (52, 88)	95% (82, 99)
	Lee S (2018)	183	55	3	6	119	90% (80, 96)	98% (93, 99)
	Kwon JH (2019)	108	34	9	2	63	94% (81, 99)	88% (78, 94)
	Higgins I^2 for study heterogeneity						75%	82%
Meta-analytic summary estimate using the bivariate model						83% (68, 93)	97% (87, 99)	

Articles are listed according to year of publication and in alphabetical order of the names of the first authors within the same year of publication
AIP, autoimmune pancreatitis; *TP*, true positive; *FP*, false positive; *FN*, false negative; *TN*, true negative; *CI*, confidence interval

76% (95% CI, 54–99%) and 97% (95% CI, 91–100%) for MRI, with MRI showing a higher meta-analytic summary sensitivity than CT, although the difference was not statistically significant (76% vs. 50% respectively, $p = 0.28$) (Fig. 5). Both MRI and CT had similar specificities (97% vs. 98% respectively, $p = 0.07$).

Meta-regression analysis

The meta-regression analysis results are summarized in Table 3. For both CT and MRI, the year of publication was a significant factor associated with study heterogeneity ($p \leq 0.05$). Studies published after 2015 showed a higher sensitivity than those published before 2015 for both CT (83% vs. 49%) and MRI (93% vs. 74%). In the CT studies, slice thickness, the method of image review, and the clarity of blinding during the review were also significantly associated with study heterogeneity ($p \leq 0.03$). Studies with a CT slice thickness ≤ 3 mm had a higher sensitivity than those with a slice thickness > 3 mm (72% vs. 42%), but the specificities were similar (93% vs. 100%). In addition, studies with multiple independent reviewers and studies with clear blinding during the review had higher sensitivities than those with single or multiple reviewers with consensus, and those with unclear blinding during the review.

A remarkable publication bias was noted for MRI ($p = 0.02$) but not for CT ($p = 0.28$; Supplementary Fig. 1).

Discussion

This study showed very high meta-analytic summary specificities for both CT (99%) and MRI (97%) for the differentiation of AIP from PDAC. However, the meta-analytic summary sensitivity of CT was 59%, whereas that of MRI was 84%. Compared with CT, MRI had a significantly higher sensitivity for differentiating AIP from PDAC ($p = 0.02$) but a similar specificity ($p = 0.18$).

The higher sensitivity of MRI found in this study is in agreement with a recent study that reported MRI as having a higher sensitivity than CT (88.5–90.2% vs. 77–80.3%, $p \leq 0.07$) [9]. Although this previous study did not show a statistically significant difference between CT and MRI, our meta-analysis showed MRI to have a significantly higher sensitivity than CT in the diagnosis of AIP. The higher sensitivity of MRI might be due to the higher soft-tissue contrast of MRI, which enables the detection of focal pancreatic masses [9]. Although the detection of a subtle mass or small lesion might be difficult on CT because of the poor soft-tissue contrast between focal mass-like AIP lesions and the normal pancreatic parenchyma, MRI including fat-suppressed pre-contrast T1-weighted images can demonstrate even small focal pancreatic masses with a high lesion conspicuity [31]. Given previous findings that multiplicity of AIP was more frequently observed on MRI than on CT (33–44% vs. 6%), the higher sensitivity of MRI for detecting AIP in comparison with CT seems reasonable [8, 16, 32].

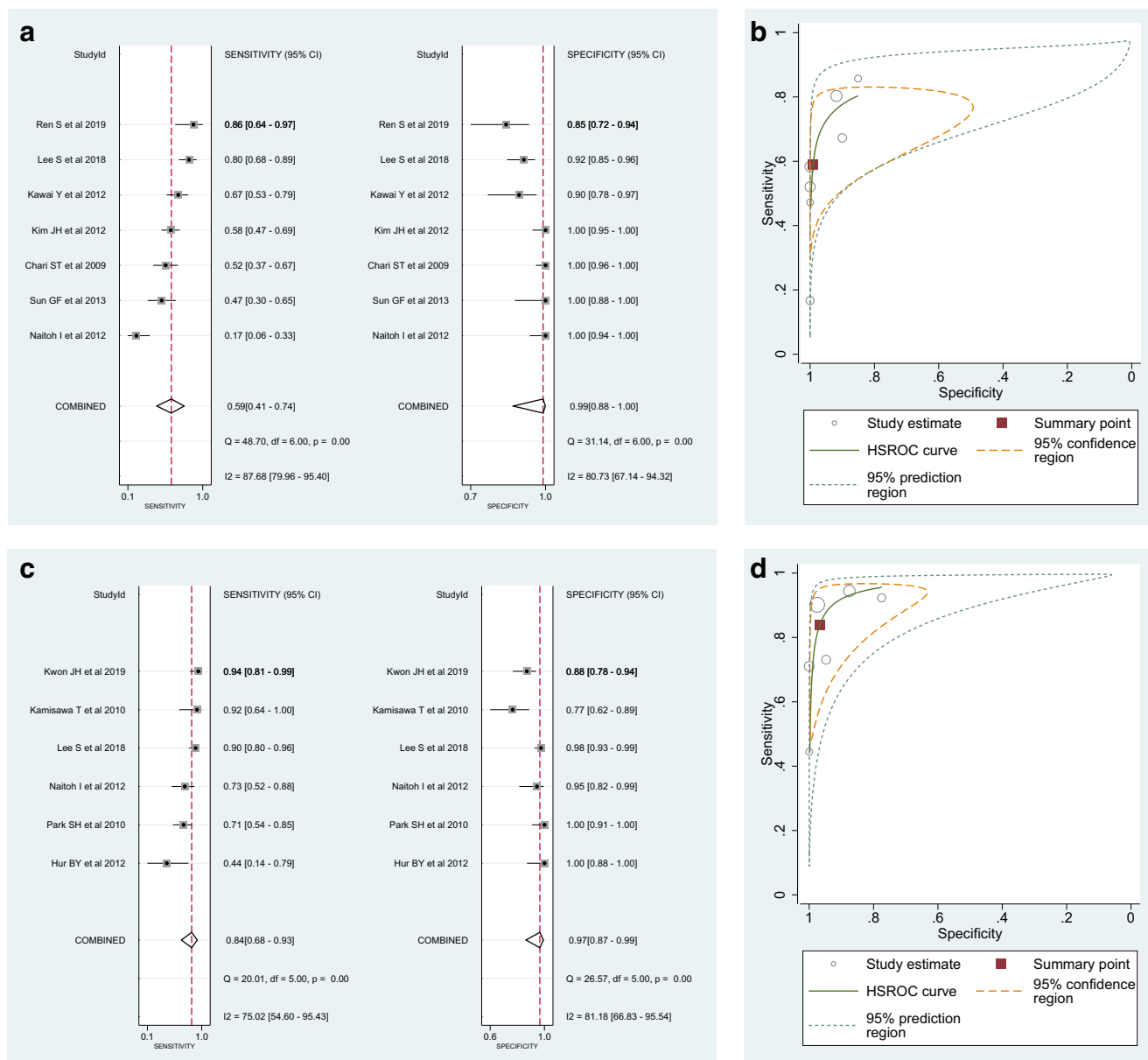
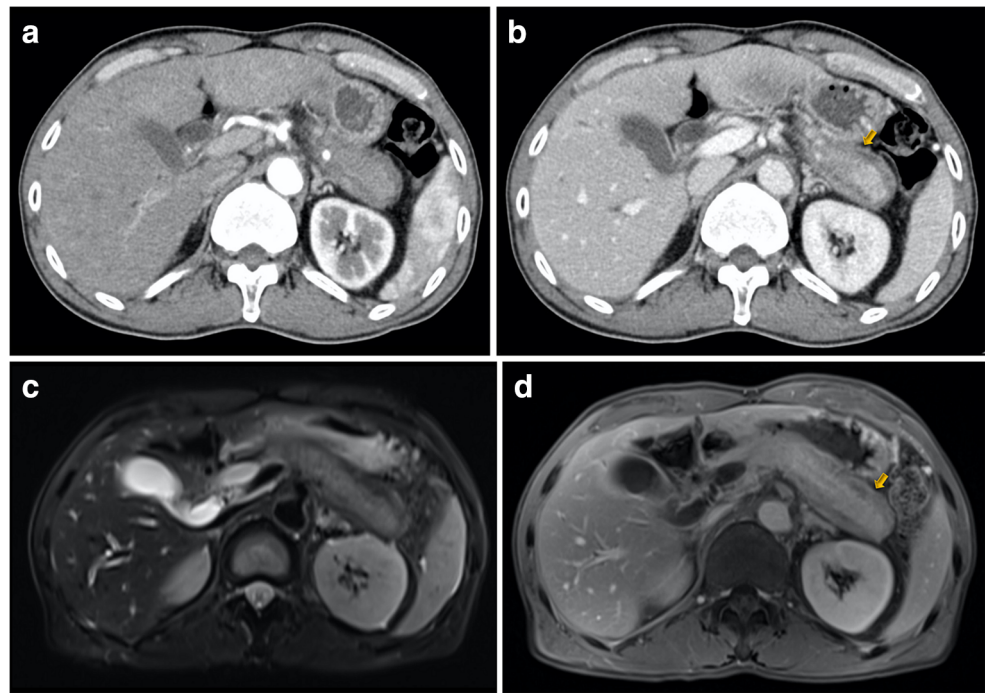


Fig. 3 Coupled forest plots and HSROC curve of the sensitivity and specificity for the differential diagnosis of AIP from PDAC on CT (**a, b**) and MRI (**c, d**)

Although there was wide heterogeneity among the multiple proposed criteria for AIP [11–13], which was likely due to regional and ethnic differences in pathologic and clinical manifestations of AIP [7], they commonly include imaging evidence as a key feature, i.e., diffuse or focal enlargement of the pancreas with a diffuse or segmental pancreatic duct stricture. Most guidelines recommend CT, MRI, or endoscopic retrograde pancreatography for the evaluation of AIP; however, there is a lack of information about the choice of imaging tests. In this study, MRI had a significantly higher sensitivity than CT for detecting AIP. In addition, previous studies reported that MRI can be more useful than CT for evaluating pancreatic duct stricture [5, 9, 33], with reported detection rates for pancreatic duct stricture of 85.2% for MRI and 54.1% for CT [9].

This is because MRI using heavily T2-weighted cross-sectional imaging or MRP can illustrate minimal or mild pancreatic duct strictures with absent or minimal upstream pancreatic duct dilatation, whereas CT has limitations when it comes to showing subtle pancreatic duct abnormalities [9]. In the case of differential diagnosis between AIP and PDAC, MRI may be more useful than CT, considering the higher detection of multiple lesions and the better ductal imaging in MRI. Although the differential diagnosis of focal AIP from PDAC is challenging, combining these imaging features with serum IgG4 levels, serum tumor markers, and responsiveness to steroid therapy would be helpful to differentiate between the two diseases in the clinical practice [7, 9].

Fig. 4 A 62-year-old man with diffuse AIP (author’s own collection). **a, b** CT shows diffuse parenchymal swelling on the arterial-phase image and a low density halo (arrow) surrounding the pancreas on the delayed-phase image. **c, d** MRI shows T2 hyperintensity parenchymal change with delayed enhancing rim (arrow)



Of the 11 included studies, Naitoh et al reported low sensitivity of 17% in CT [26]. Because this study evaluated only three CT findings including mass size, delayed enhancement, and capsule-like rim, the sensitivity of this study might be limited. In addition, Hur et al reported low sensitivity of

44% in MRI [16]. Given the lower sensitivity of focal AIP than diffuse AIP, the low sensitivity in this study might be explained by patients consisting of only focal AIP. In addition, both studies reported the diagnostic performance of each CT or MRI feature, but the reported data varied according to each

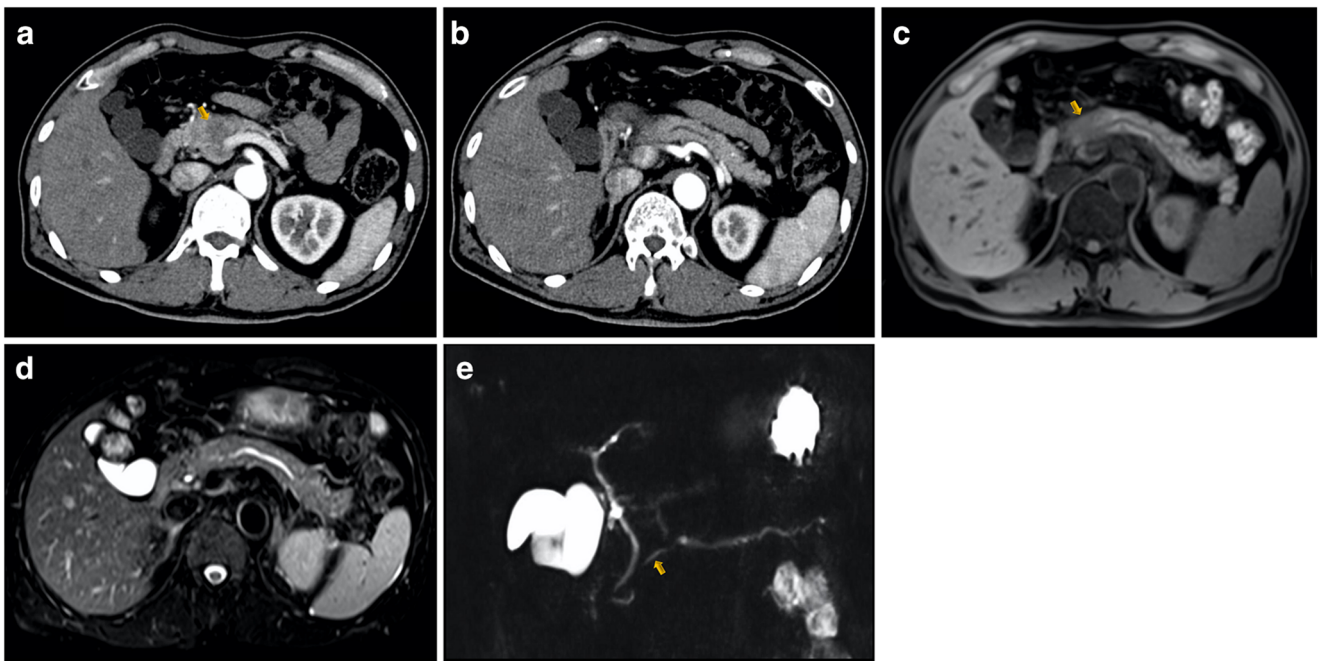


Fig. 5 A 54-year-old man with focal AIP (author’s own collection). **a, b** CT shows a 2.0-cm ill-defined low density lesion (arrow) in the pancreatic head portion with upstream pancreatic duct dilatation. **c, d** This lesion appears more conspicuous on MRI (arrow) than on CT, demonstrating

pancreatic ductal dilatation and stricture more clearly. **e** MRCP shows multifocal stricture and dilatation with the duct penetrating sign (visible duct within a mass) (arrow)

Table 3 Results of the meta-regression analysis

Covariate	Subgroup	Sensitivity (95% CI)	Specificity (95% CI)	<i>p</i> value	
CT	Patient composition	AIP > PDAC (<i>n</i> = 1)	58% (13, 100)	100% (95, 100)	0.41
		AIP < PDAC (<i>n</i> = 6)	59% (40, 79)	98% (95, 100)	
	Publication year	Before 2015 (<i>n</i> = 5)	49% (34, 63)	100% (99, 100)	0.03
		After 2015 (<i>n</i> = 2)	83% (70, 97)	89% (61, 100)	
	CT slice thickness	≤ 3 mm (<i>n</i> = 4)	72% (56, 87)	93% (87, 98)	0.01
		> 3 mm (<i>n</i> = 3)	42% (21, 63)	100% (100, 100)	
	CT multi-phase	Multi-phase (<i>n</i> = 6)	51% (28, 74)	98% (94, 100)	0.45
		Single-phase (<i>n</i> = 1)	52% (6, 98)	100% (100, 100)	
	Image reviewer	Multiple independent reviewers (<i>n</i> = 3)	77% (64, 91)	90% (86, 94)	< 0.01
		Single or multiple reviewers with consensus (<i>n</i> = 4)	44% (30, 58)	100% (100, 100)	
Clarity of review blinding	Clear (<i>n</i> = 6)	61% (55, 76)	98% (94, 100)	0.03	
	Unclear (<i>n</i> = 1)	16% (0, 34)	100% (100, 100)		
MR	Patient composition	NA	NA	NA	
	Publication year	Before 2015 (<i>n</i> = 4)	74% (60, 87)	97% (91, 100)	0.05
		After 2015 (<i>n</i> = 2)	93% (86, 99)	94% (85, 100)	
	MR magnet	1.5 T only (<i>n</i> = 4)	90% (82, 98)	95% (88, 100)	0.07
		Others (<i>n</i> = 2)	62% (1, 87)	98% (94, 100)	
	MRCP	With MRCP (<i>n</i> = 5)	81% (68, 94)	97% (94, 100)	0.17
		Without MRCP (<i>n</i> = 1)	93% (78, 100)	78% (64, 91)	
	DWI	With DWI (<i>n</i> = 2)	73% (44, 100)	93% (81, 100)	0.35
		Without DWI (<i>n</i> = 4)	86% (74, 97)	97% (93, 100)	
	Image reviewer	Multiple independent reviewers (<i>n</i> = 1)	90% (75, 100)	98% (92, 100)	0.17
		Single or multiple reviewers with consensus (<i>n</i> = 5)	81% (67, 94)	95% (89, 100)	
	Clarity for blinding review	Clear (<i>n</i> = 4)	83% (67, 98)	98% (95, 100)	0.10
		Unclear (<i>n</i> = 2)	84% (64, 100)	89% (72, 100)	

The results were obtained using meta-regression analysis with the bivariate model

AIP, autoimmune pancreatitis; MR, magnetic resonance

individual feature. Therefore, further studies using combined imaging features will be needed to determine the diagnostic performance of CT or MRI.

Publication year was a significant factor affecting study heterogeneity in both CT and MRI. As technical advances such as thin-slice images and three-dimensional imaging lead to improvements in the diagnostic performance of imaging tests [34], diagnostic performance may differ according to publication year. In addition, given that the diagnostic test accuracy is based upon a comparison between the results of the index test and those of the reference standard, knowledge of the reference standard may influence the interpretation of index test results [35], and the clarity of blinding during review might be a factor influencing study heterogeneity.

This study has several limitations. First, substantial study heterogeneity was noted in both the sensitivity and specificity of CT and MRI. Therefore, to explore the causes of study heterogeneity, we evaluated the presence of a threshold effect between sensitivity and specificity, and performed meta-

regression analysis. Second, we could not show a statistically significant difference between CT and MRI in the subgroup analyses for differentiating focal AIP from PDAC, even though MRI had a significantly higher sensitivity than CT in all eligible studies. This may be due to the small number of studies with focal AIP only (three studies for CT and four studies for MRI). Further studies are needed to verify that MRI has a higher sensitivity than CT in the differentiation of focal AIP from PDAC.

In conclusion, both CT and MRI had very good diagnostic performance for the diagnosis of PDAC, but both showed suboptimal moderate diagnostic performance for AIP. Compared with CT, MRI had a significantly higher sensitivity for the detection of AIP lesions. Therefore, MRI may be more useful for evaluating patients with AIP in the clinic, especially for the differentiation of AIP from PDAC.

Funding This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (grant number: NRF-2019R1G1A1099743).

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Sang Hyun Choi.

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 One of the authors (Sang Hyun Choi) has significant statistical expertise.

Informed consent Written informed consent was not required for this study because this study was a meta-analysis.

Ethical approval Institutional review board approval was not required because this study was a meta-analysis.

Methodology

- Meta-analysis
- Diagnostic study
- Performed at one institution

References

- Ectors N, Mailliet B, Aerts R et al (1997) Non-alcoholic duct destructive chronic pancreatitis. *Gut* 41:263–268
- Vlachou PA, Khalili K, Jang HJ, Fischer S, Hirschfield GM, Kim TK (2011) IgG4-related sclerosing disease: autoimmune pancreatitis and extrapancreatic manifestations. *Radiographics* 31:1379–1402
- Manfredi R, Frulloni L, Mantovani W, Bonatti M, Graziani R, Pozzi Mucelli R (2011) Autoimmune pancreatitis: pancreatic and extrapancreatic MR imaging-MR cholangiopancreatography findings at diagnosis, after steroid therapy, and at recurrence. *Radiology* 260:428–436
- Rehmitz C, Klauss M, Singer R et al (2011) Morphologic patterns of autoimmune pancreatitis in CT and MRI. *Pancreatology* 11:240–251
- Takahashi N, Fletcher JG, Hough DM et al (2009) Autoimmune pancreatitis: differentiation from pancreatic carcinoma and normal pancreas on the basis of enhancement characteristics at dual-phase CT. *AJR Am J Roentgenol* 193:479–484
- Wakabayashi T, Kawaura Y, Satomura Y et al (2003) Clinical and imaging features of autoimmune pancreatitis with focal pancreatic swelling or mass formation: comparison with so-called tumor-forming pancreatitis and pancreatic carcinoma. *Am J Gastroenterol* 98:2679–2687
- Khandelwal A, Shanbhogue AK, Takahashi N, Sandrasegaran K, Prasad SR (2014) Recent advances in the diagnosis and management of autoimmune pancreatitis. *AJR Am J Roentgenol* 202:1007–1021
- Kim JH, Kim MH, Byun JH et al (2012) Diagnostic strategy for differentiating autoimmune pancreatitis from pancreatic cancer: is an endoscopic retrograde pancreatography essential? *Pancreas* 41:639–647
- Lee S, Kim JH, Kim SY et al (2018) Comparison of diagnostic performance between CT and MRI in differentiating non-diffuse-type autoimmune pancreatitis from pancreatic ductal adenocarcinoma. *Eur Radiol* 28:5267–5274
- Abraham SC, Wilentz RE, Yeo CJ et al (2003) Pancreaticoduodenectomy (Whipple resections) in patients without malignancy: are they all ‘chronic pancreatitis’? *Am J Surg Pathol* 27:110–120
- Kawa S, Kamisawa T, Notohara K et al (2020) Japanese clinical diagnostic criteria for autoimmune pancreatitis, 2018: revision of Japanese clinical diagnostic criteria for autoimmune pancreatitis, 2011. *Pancreas* 49:e13–e14
- Chari ST, Takahashi N, Levy MJ et al (2009) A diagnostic strategy to distinguish autoimmune pancreatitis from pancreatic cancer. *Clin Gastroenterol Hepatol* 7:1097–1103
- Shimosegawa T, Chari ST, Frulloni L et al (2011) International consensus diagnostic criteria for autoimmune pancreatitis: guidelines of the International Association of Pancreatology. *Pancreas* 40:352–358
- Bodily KD, Takahashi N, Fletcher JG et al (2009) Autoimmune pancreatitis: pancreatic and extrapancreatic imaging findings. *AJR Am J Roentgenol* 192:431–437
- Furuhashi N, Suzuki K, Sakurai Y, Ikeda M, Kawai Y, Naganawa S (2015) Differentiation of focal-type autoimmune pancreatitis from pancreatic carcinoma: assessment by multiphase contrast-enhanced CT. *Eur Radiol* 25:1366–1374
- Hur BY, Lee JM, Lee JE et al (2012) Magnetic resonance imaging findings of the mass-forming type of autoimmune pancreatitis: comparison with pancreatic adenocarcinoma. *J Magn Reson Imaging* 36:188–197
- Kamisawa T, Takuma K, Anjiki H et al (2010) Differentiation of autoimmune pancreatitis from pancreatic cancer by diffusion-weighted MRI. *Am J Gastroenterol* 105:1870–1875
- Kawai Y, Suzuki K, Itoh S, Takada A, Mori Y, Naganawa S (2012) Autoimmune pancreatitis: assessment of the enhanced duct sign on multiphase contrast-enhanced computed tomography. *Eur J Radiol* 81:3055–3060
- Kim HJ, Kim YK, Jeong WK, Lee WJ, Choi D (2015) Pancreatic duct “icicle sign” on MRI for distinguishing autoimmune pancreatitis from pancreatic ductal adenocarcinoma in the proximal pancreas. *Eur Radiol* 25:1551–1560
- Kim SS, Choi GC, Jou SS (2018) Pancreas ductal adenocarcinoma and its mimics: review of cross-sectional imaging findings for differential diagnosis. *J Belg Soc Radiol* 102:71
- Kwon JH, Kim JH, Kim SY et al (2019) Differentiating focal autoimmune pancreatitis and pancreatic ductal adenocarcinoma: contrast-enhanced MRI with special emphasis on the arterial phase. *Eur Radiol* 29:5763–5771
- Muhi A, Ichikawa T, Motosugi U et al (2012) Mass-forming autoimmune pancreatitis and pancreatic carcinoma: differential diagnosis on the basis of computed tomography and magnetic resonance cholangiopancreatography, and diffusion-weighted imaging findings. *J Magn Reson Imaging* 35:827–836
- Park SH, Kim MH, Kim SY et al (2010) Magnetic resonance cholangiopancreatography for the diagnostic evaluation of autoimmune pancreatitis. *Pancreas* 39:1191–1198
- Sugiyama Y, Fujinaga Y, Kadoya M et al (2012) Characteristic magnetic resonance features of focal autoimmune pancreatitis useful for differentiation from pancreatic cancer. *Jpn J Radiol* 30:296–309
- Ichikawa T, Sou H, Araki T et al (2001) Duct-penetrating sign at MRCP: usefulness for differentiating inflammatory pancreatic mass from pancreatic carcinomas. *Radiology* 221:107–116
- Naitoh I, Nakazawa T, Hayashi K et al (2012) Clinical differences between mass-forming autoimmune pancreatitis and pancreatic cancer. *Scand J Gastroenterol* 47:607–613
- Ren S, Chen X, Cui W et al (2019) Differentiation of chronic mass-forming pancreatitis from pancreatic ductal adenocarcinoma using contrast-enhanced computed tomography. *Cancer Manag Res* 11:7857–7866

28. Sun GF, Zuo CJ, Shao CW, Wang JH, Zhang J (2013) Focal autoimmune pancreatitis: radiological characteristics help to distinguish from pancreatic cancer. *World J Gastroenterol* 19:3634–3641
29. Agency for Healthcare Research and Quality. Chapter 8: Meta-analysis of test performance when there is a “gold standard” 2020. Available at <https://effectivehealthcare.ahrq.gov/products/methods-guidance-tests-metaanalysis/methods/>. Accessed June 10
30. 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
31. Bollen TL (2012) Imaging of acute pancreatitis: update of the revised Atlanta classification. *Radiol Clin North Am* 50:429–445
32. Choi SY, Kim SH, Kang TW, Song KD, Park HJ, Choi YH (2016) Differentiating mass-forming autoimmune pancreatitis from pancreatic ductal adenocarcinoma on the basis of contrast-enhanced MRI and DWI findings. *AJR Am J Roentgenol* 206:291–300
33. Negrelli R, Manfredi R, Pedrinolla B et al (2015) Pancreatic duct abnormalities in focal autoimmune pancreatitis: MR/MRCP imaging findings. *Eur Radiol* 25:359–367
34. Sahni VA, Morteke KJ (2008) Magnetic resonance cholangiopancreatography: current use and future applications. *Clin Gastroenterol Hepatol* 6:967–977
35. Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J (2004) Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med* 140:189–202

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.