Eur Radiol (2018) 28:530–541
DOI 10.1007/s00330-017-4958-x

UROGENITAL

Magnetic resonance imaging for detection of parametrial invasion in cervical cancer: An updated systematic review and meta-analysis of the literature between 2012 and 2016

Sungmin Woo [1] · Chong Hyun Suh [2,3] · Sang Youn Kim [1] · Jeong Yeon Cho [1,4] · Seung Hyup Kim [1,4]

Received: 13 April 2017 / Revised: 18 June 2017 / Accepted: 20 June 2017 / Published online: 19 July 2017

Abstract

Objective To review the diagnostic performance of MRI for detection of parametrial invasion (PMI) in cervical cancer patients.

Methods MEDLINE and EMBASE databases were searched for studies providing diagnostic performance of MRI for detecting PMI in patients with cervical cancer. Studies published between 2012 and 2016 using surgico-pathological results as reference standard were included. Study quality was evaluated using QUADAS-2. Sensitivity and specificity of all studies were calculated. Results were pooled and plotted in a hierarchical summary receiver operating characteristic plot. Meta-regression and subgroup analyses were performed.

Results Fourteen studies (1,028 patients) were included. Study quality was generally moderate. Pooled sensitivity was 0.76 (95% CI 0.67–0.84) and specificity was 0.94 (95% CI 0.91–0.95). The possibility of heterogeneity was considered low: Cochran's Q-test ($p = 0.471$), Tau^2 (0.240), Higgins I^2 (0%). With meta-regression analysis, magnet strength, use of DWI, and antispasmodic drugs were significant factors affecting heterogeneity ($p < 0.01$). Subgroup analysis for studies solely using radical hysterectomy as reference standard yielded pooled sensitivity and specificity of 0.73 (95% CI 0.60–0.83) and 0.93 (95% CI 0.90–0.95), respectively.

Conclusions MRI shows good performance for detection of PMI in cervical cancer. Using 3-T scanners and DWI may improve diagnostic performance.

Key Points

- *MRI shows good performance for detection of parametrial invasion in cervical cancer.*
- *Subgroup of studies using only radical hysterectomy showed consistent results.*
- *Using 3-Tesla scanners and diffusion-weighted imaging may improve diagnostic performance.*

Keywords Cervical cancer · Magnetic resonance imaging · Meta-analysis · Parametrial invasion · Systemic review

Abbreviations

CE	Contrast-enhanced
CT	Computed tomography
DWI	Diffusion-weighted imaging
FIGO	Federation of Gynecology and Obstetrics
FSE	Fast spine echo
HSROC	Hierarchical summary receiver operating characteristic

Sungmin Woo and Chong Hyun Suh contributed equally to this paper.

Electronic supplementary material The online version of this article (doi:10.1007/s00330-017-4958-x) contains supplementary material, which is available to authorized users.

✉ Sang Youn Kim
iwishluv@empas.com

[1] Department of Radiology, Seoul National University Hospital, 101 Daehak-ro, Jongno-gu, Seoul 110-744, Republic of Korea

[2] Department of Radiology and Research Institute of Radiology, University of Ulsan College of Medicine, Asan Medical Center, 86 Asanbyeongwon-Gil, Songpa-Gu, Seoul 138-736, Republic of Korea

[3] Department of Radiology, Namwon Medical Center, 365, Chungjeong-ro, Namwon-si, Jeollabuk-do 590-702, Republic of Korea

[4] Institute of Radiation Medicine and Kidney Research Institute, Seoul National University Medical Research Center, Seoul 110-744, Republic of Korea

MRI	Magnetic resonance imaging
PMI	Parametrial invasion
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QUADAS-2	Quality Assessment of Diagnostic Accuracy Studies-2
SE	Spin echo
TSE	Turbo spin echo
T2WI	T2-weighted imaging

Introduction

Cervical cancer is the second most common malignancy in women [1]. One of the most important aspects in the pretreatment evaluation of cervical cancer is parametrial invasion (PMI) [2]. PMI is known to be associated with prognosis, and patients with suspected PMI usually will be treated with primary chemoradiation or adjuvant treatment after surgery [3]. Therefore, it is crucial to accurately assess PMI in patients with cervical cancer in order to select the optimal treatment.

Currently, the International Federation of Gynecology and Obstetrics (FIGO) staging system is widely used for clinical staging of cervical cancer [4]. FIGO staging is primarily based on physical examination and further evaluation may be performed using modalities of cystoscopy and proctoscopy. Errors in clinical FIGO staging have been consistently reported, with understaging and overstaging up to 40% and 64%, respectively [5]. On the other hand, magnetic resonance imaging (MRI) has shown promising results regarding the staging accuracy of cervical cancer, but it is only recommended, not required according to the FIGO committee of Gynecologic Oncology [6].

Until now, there have been two published meta-analyses assessing the diagnostic performance of MRI for detection of PMI. The sensitivity and specificity of MRI were 74% and 82% in the report by Bipat et al. [7], which compared computed tomography (CT) and MRI using studies published from 1985 to 2002. In a more recent report, MRI showed sensitivity of 84% and specificity of 92% in studies published up to 2011 [8]. Although these two meta-analyses showed that MRI was superior to CT and clinical examination, MRI is still not being used by some groups (up to 30%) according to a recent survey [9]. However, as both meta-analyses included studies in the remote past and with recent technical advances in MRI such as high magnetic field strength (i.e. 3-Tesla) and diffusion-weighted imaging (DWI), one could expect that the diagnostic performance of MRI would have further improved over the years, possibly providing additional evidence for MRI to be incorporated in the FIGO staging system.

Therefore, the purpose of our study was to review the literature published since 2012 to obtain updated diagnostic performance values of MRI for detecting PMI in patients with cervical cancer using surgico-pathological results as the reference standard.

Materials and methods

This meta-analysis was performed according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. For this meta-analysis, we formulated a research question based on the PICOS criteria as the following [10]: What is the diagnostic performance of MRI for detection of PMI in patients with cervical cancer, as compared with surgico-pathological results, in studies published since 2012?

Literature search

A computerised search of MEDLINE and EMBASE databases up to 29 December 2016 was conducted. All synonyms or related terms were included in the search query as the following: ((‘cervical cancer’) OR (‘cervical carcinoma’) OR (‘cervical malignancy’) OR (‘cervical neoplasm’) OR (‘cervical tumor’) OR (‘cervical tumour’) OR (‘cervix cancer’) OR (‘cervix carcinoma’) OR (‘cervix malignancy’) OR (‘cervix neoplasm’) OR (‘cervix tumor’) OR (‘cervix tumour’)) AND ((staging) OR (stage) OR (parametrial invasion) OR (parametrial infiltration)) AND ((magnetic resonance imaging) OR (MRI)). The bibliographies of included articles were screened to identify other eligible studies. We did not limit the search to any particular language.

Inclusion criteria

We included studies that met the following PICOS criteria [10]: (1) patients diagnosed with cervical cancer; (2) index test used MRI for detection of PMI; (3) for comparison, surgico-pathological results were available as the reference standard; (4) the study provided the sensitivity and specificity of MRI, or the corresponding raw data for constructing a 2×2 contingency table; and (5) publication type had to be original articles.

Exclusion criteria

The exclusion criteria were: (1) published before 1 January 2012; (2) less than ten patients; (3) publication type other than original articles; (4) MRI was used for evaluation of cervical cancer, but focused on topics other than detection of PMI; (5) overlapping patient population; and (6) insufficient data for reconstruction of 2×2 tables (even after attempts to contact

the authors). If multiple publications with an overlapping study population were identified, we only included the study with the largest patient cohort.

Two reviewers (S.W. and C.H.S.) independently performed the literature search and study selection. When disagreement was present, consensus was reached after discussion with a third reviewer (S.Y.K.).

Data extraction and quality assessment

The following data regarding patient, study and MRI characteristics were extracted using a standardised form: (1) patient characteristics – number of patients, median age and range of patients, prevalence of PMI, histological subtypes of included tumours, and FIGO stages; (2) study characteristics – origin of study (authors, institution and country), publication year, duration of patient recruitment, study design (prospective vs. retrospective and whether enrolment was consecutive or not), reference standard, interval between MRI and the reference standard, blinding to surgico-pathological results, level of analysis (per-patient or separately analysed for each parametria); and (3) MRI characteristics – magnet field strength (3- vs. <3-Tesla), scanner model and manufacturer, coil type, spin echo (SE) technique (fast SE [FSE] or turbo SE [TSE] vs. SE), slice thickness (≤5 mm vs. >5 mm), acquired imaging planes for T2-weighted imaging (T2WI), inclusion of DWI and contrast-enhanced (CE) MRI, and use of an antispasmodic drug (i.e. scopolamine butylbromide).

We assessed the methodological quality of the selected studies using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool [11].

Both data extraction and quality assessment were performed independently by two previously noted reviewers (S.W. and C.H.S.) followed by discussion with a third reviewer (S.Y.K.) in case of disagreement.

Data synthesis and analysis

Data from the included studies were reconstructed in 2 × 2 tables (true positive, false negative, false positive and true negative) and their sensitivity and specificity were calculated. If diagnostic performance of various MRI protocols was separately provided, then the one including the most advanced and comprehensive MRI protocol was selected (T2WI + DWI > T2WI; endovaginal coil > phased array coil; and 3D > 2D). If results by multiple independent readers were available, we chose the one with higher sensitivity. If diagnostic performance had been assessed both on a per-patient basis and for both parametria, we used the results of the per-patient analysis, as the treatment decision making (primary chemoradiation vs. radical surgery) is based on whether PMI is present on at least one side.

Pooled estimates of sensitivity and specificity were calculated with hierarchical logistic regression modelling including bivariate modelling and hierarchical summary receiver operating characteristic (HSROC) modelling [12]. An HSROC curve with 95% confidence and prediction regions was plotted to graphically present the results. Publication bias was evaluated by visual inspection of the Deeks' funnel plot and calculating the *p*-value from Deeks' asymmetry test [13].

Heterogeneity, or in other words the variation in study outcomes between the included studies, was determined using various statistical methods. First, Cochran's Q-test was performed with $p < 0.05$ indicating heterogeneity. Second, Higgins I^2 test was performed and interpreted using the following criteria: inconsistency index (I^2), 0–40%, heterogeneity might not be important; 30–60%, moderate heterogeneity may be present; 50–90%, substantial heterogeneity may be present; and 75–100%, considerable heterogeneity [14]. Third, we looked for a threshold effect in terms of a positive correlation between the sensitivity and false-positive rate among the selected studies. Fourth, tau squared (τ^2), which is considered the most informative expression of heterogeneity in a meta-analysis, was calculated [15].

Meta-regression analysis was performed to investigate the cause of heterogeneity using the following variables: (1) study design (prospective vs. retrospective); (2) ethnicity (Asian vs. non-Asian); (3) prevalence of PMI (≥16.7% [median value of study population] vs. <16.7%); (4) magnet field strength (3- vs. <3-Tesla); (5) coil type (phased-array or endovaginal coil vs. others); (6) SE technique (FSE/TSE vs. SE); (7) slice thickness (≤5 mm vs. >5 mm); (8) T2WI planes (included both axial oblique and sagittal planes vs. not included), (9) inclusion of DWI; (10) inclusion of CE MRI; and (11) use of an antispasmodic drug. In addition, subgroup analysis was planned for studies solely using radical hysterectomy as the reference standard.

The 'midas' module in Stata 10.0 (StataCorp LP, College Station, TX, USA) and 'mada' package in R software version 3.2.1 (R Foundation for Statistical Computing, Vienna, Austria) were used for statistical analyses with $p < 0.05$ indicating statistical significance.

Results

Literature search

The systematic literature search yielded 1,195 articles. Among them 379 were duplicates and 785 were excluded based on the review of the abstract alone. Full-text reviews were performed for the remaining 31 articles and 17 were excluded for the following reasons (see Online Supplementary Table 1): (1)

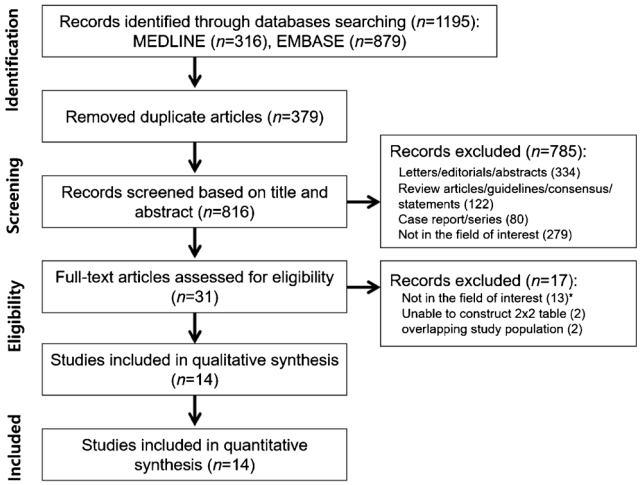

Fig. 1 Flow diagram showing study selection process. * = MRI was used for staging but parametrial invasion was not separately assessed (n = 6), criteria for determination of parametrial invasion was tumour size (n = 3) or visibility of tumour (n = 2), MRI was used to assess parametrial invasion but was correlated only with clinical FIGO staging (n = 2)

not in the field of interest (n = 13); (2) overlapping population (n = 2); and (3) insufficient data to reconstruct 2 × 2 contingency tables (n = 2). Ultimately, 14 original articles including a total of 1,436 patients were included [16–29]. Figure 1 shows the detailed study selection process.

Characteristics of included studies

The patient characteristics are described in Table 1. The size of the study population ranged from 25 to 298 patients and the prevalence of PMI ranged from 4.0% to 43.3%. The patients had a median age of 34.4–57.8 years. The studies mostly included only adeno- or squamous subtypes, but other histological subtypes constituted 0.9–6.6% in five studies, and in one study the subtype was not mentioned. The FIGO stages varied among the studies, including only stages IIA or lower in eight, including advanced stages of IIB or greater in five and not explained in one.

The study characteristics are described in Table 2. Articles originated from Asian countries in seven studies, and from non-Asian countries in the other seven. Regarding study design, six studies were prospective and eight were retrospective. Patient recruitment was consecutive in seven studies, but was not explicitly mentioned in the other seven. Eight studies solely used radical hysterectomy specimens as the reference standard; three used radical hysterectomy or trachelectomy; and others were based on surgery (without details of the type of operation), radical hysterectomy or biopsy, and histopathological correlation or multidisciplinary decision based on imaging (initial and follow-up), clinical examination and treatment change. Only one study analysed the right and left parametria separately, whereas all other studies performed analysis on a per-patient basis.

The MRI characteristics are described in Table 3. Six studies used only 3-Tesla scanners; six used only 1.5-Tesla scanners; one used either 1.5- or 3-Tesla scanners; and one used a 1-Tesla scanner. With regard to coil type, one study used an endovaginal coil, two used phased array or body coils, and the remaining studies used only phased-array coils. All studies, except for one that was not explicit, used TSE or FSE sequences. Only two studies used a slice thickness >5 mm; it was ≤5 mm in the remaining studies. Both axial oblique and sagittal planes were included in the MRI protocol in half of the studies. DWI was used in two studies, CE-MRI in five, and both DWI and CE-MRI were used in three. Antispasmodic drugs were used in eight studies, not used in one, and use was not reported in five.

Quality assessment

The distribution of QUADAS-2 scores in the 14 included studies is shown in Fig. 2. The quality of the studies was generally moderate, with 12 (86%) studies satisfying more than four of the seven domains [30]. The details for each domain are provided in the Online Supplementary Material.

Heterogeneity among the included studies

Based on the Q-test, heterogeneity was not likely to be present among the 14 studies (p = 0.127). The Higgins I^2 statistics showed that there may be moderate heterogeneity in terms of the sensitivity (I^2 = 49.29%) and specificity (I^2 = 51.16). However, the coupled forest plot of the sensitivity and specificity demonstrated no threshold effect through visual assessment (Fig. 3). In addition, no threshold effect was demonstrated between the sensitivity and false-positive rate (Spearman correlation coefficient = −0.042 [95% CI −0.560–0.500]). When heterogeneity was assessed in terms of the diagnostic odds ratio (DOR), Cochran's Q-test (p = 0.471), Tau^2 (0.240), and Higgins I^2 (0%) all suggested that heterogeneity is not likely to be present.

Diagnostic accuracy of MRI for detection of parametrial invasion

For all 14 studies, the pooled sensitivity was 0.76 (95% CI 0.67–0.84) with a specificity of 0.94 (95% CI 0.91–0.95). In the HSROC curve, there was only a small difference between the 95% confidence and prediction regions, again implying that the heterogeneity among the included studies was low (Fig. 4). The area under the HSROC curve was 0.94 (95% CI 0.92–0.96). The Deeks' funnel plot and the results of the Deeks' asymmetry test showed that the likelihood of publication bias was low (p = 0.31) (Fig. 5).

Table 1 Patient characteristics

Study	Institution	Duration of patient recruitment (year.month)	No. of patients	No. of patients with PMI	Prevalence of PMI (%)	Age, y		Other† histology (%)	FIGO stage
						Median	Range		
Bleker et al. [16]	Academic Medical Center	2003.1–2011.1	203	34	16.7	43*	23-69	2.5	IB1-IIA
Bourgioti et al. [17]	Aretaieion Hospital, University of Athens	2009.4–2014.3	115	15	13.0	44.7*	NR	0.9	<IIB
Downey et al. [18]	The Institute of Cancer Research and Royal Marsden NHS Foundation Trust	2013.5–2014.8	25	1	4.0	34.4*	24-53	0	<IIB
Epstein et al. [19]	Lund University Hospital, Catholic University of the Sacred Heart, Charles University, Lithuanian University of Health Sciences Hospital	2007.9–2010.4	182	13	7.1	46.3	NR	0	IA2-IIA
Kitajima et al. [20]	Kobe University Graduate School of Medicine	2011.12–2013.2	30	13	43.3	57.8*	27-88	0	IB1-IV
Kong et al. [21]	Ajou University Hospital	2000.2–2015.3	298	64	21.5	NR	24-72	0	IB
Kraljević et al. [22]	»Sestre milosrdnice« University Hospital Center, University of Zagreb	2006.1–2008.12	33	4	12.1	54.5	35-75	3	IIA-IIB
Lee et al. [23]	Seoul National University Hospital	2003–2011	190	19	10.0	49*	25-78	1.1	IB1
Moloney et al. [24]	Cork University Hospital	2011.1–2013.12	33	5	15.2	44	NR	0	IB-IIA
Park et al. [25]	Samsung Medical Center	2010.1–2012.12	152	37	24.3	51	26-80	0	IA-IIA
Shin et al. [26]	Seoul St. Mary's Hospital	2009.8–2010.11	45	18	40.0	54.5*	29-88	NR	≥IB
Shweel et al. [27]	Minia University	2009.2–2010.8	30	8	26.7	45*	40-65	6.6	1B-IVA
Wei et al. [28]	Anhui Provincial Tumor Hospital	2012.1–2013.3	29	8	27.6	56.6*	30-83	0	IA-IIB
Yu et al. [29]	Peking Union Medical College Hospital	2009.4–2010.9	71	3	4.2	47*	28-71	0	NR

FIGO International Federation of Gynecology and Obstetrics, *PMI* parametrial invasion, *NR* not reported

* Mean

† Other than squamous cell carcinoma or adenocarcinoma

Heterogeneity exploration using meta-regression and subgroup analyses

The results of meta-regression analysis are shown in Table 4. Among the different variables evaluated, only magnet field strength, use of DWI and administration of antispasmodic drugs were significant factors affecting the heterogeneity ($p < 0.01$ for all three variables). Regarding magnet field strength, studies using 3-T MRI scanners showed higher sensitivity (0.84 [95% CI 0.76–0.93]) but similar specificity (0.94 [95% CI 0.91–0.98]) compared with studies using MRI scanners with 1.5-T or lower (sensitivity of 0.66 [95% CI 0.55–0.77] and specificity of 0.94 [95% CI 0.91–0.97]). Studies that used DWI demonstrated higher sensitivity (0.82 [95% CI 0.70–0.94]) and specificity (0.97 [95% CI 0.95–0.99]) compared with studies that did not (sensitivity of 0.72 [95% CI 0.62–0.82] and specificity of 0.91 [95% CI 0.89–0.93]). A statistical comparison was limited regarding the use of antispasmodic drugs, as there was only one study that did not use this drug [20]. Other factors, including study design ($p = 0.97$), ethnicity ($p = 0.17$), prevalence of PMI ($p = 0.15$), coil type ($p = 0.11$), slice thickness ($p = 0.17$), T2WI planes ($p = 0.42$) and inclusion of CE MRI ($p = 0.61$) were not significant factors affecting the heterogeneity. SE technique (FSE/TSE vs SE) was not included as a covariate in the meta-regression analysis as there was no study that did not use FSE/TSE sequences.

Table 2 Study characteristics

Study	Year	Country	Study design	Consecutive enrolment	MRI blinded to reference standard	Reference standard	MRI-reference standard interval	Reference standard blinded to MRI	Level of analysis
Bleker et al. [16]	2013	The Netherlands	Retrospective	NR	Yes	RH	NR	NR	Patient
Bourgioti et al. [17]	2016	Greece	Prospective	NR	Yes	RH (94/115) or RT (21/115)	NR	NR	Patient
Downey et al. [18]	2016	UK	Prospective	Yes	Yes	RH or RT	NR	NR	Parametrium
Epstein et al. [19]	2013	Sweden, Italy, Czech Republic, Lithuania	Prospective	Yes	Yes	RH (164/182) or RT (18/182)	<2 wks	NR	Patient
Kitajima et al. [20]	2014	Japan	Retrospective	NR	Yes	Histopathological correlation* (13/30) or multidisciplinary† (17/30)	NR	NR	Patient
Kong et al. [21]	2016	South Korea	Retrospective	NR	NR	RH	NR	NR	Patient
Kraljević et al. [22]	2013	Croatia	Prospective	Yes	Yes	RH	NR	NR	Patient
Lee et al. [23]	2014	South Korea	Retrospective	Yes	Yes	RH	<4 wks	Yes	Patient
Moloney et al. [24]	2015	Ireland	Prospective	Yes	Yes	RH	NR	NR	Patient
Park et al. [25]	2015	South Korea	Retrospective	Yes	Yes	RH	1-30 d	NR	Patient
Shin et al. [26]	2013	South Korea	Retrospective	NR	Yes	RH	1-38 d	Yes	Patient
Shweel et al. [27]	2012	Egypt	Prospective	Yes	Yes	RH(16/30) or colposcopic vaginal biopsy (14/30)	<30 d	Yes	Patient
Wei et al. [28]	2013	China	Retrospective	NR	NR	Surgical treatment‡	NR	No	Patient
Yu et al. [29]	2015	China	Retrospective	Yes	Yes	RH	3-7 d	No	Patient

MRI magnetic resonance imaging, *NR* not reported, *PMI* parametrial invasion, *RH* radical hysterectomy, *RT* radical trachelectomy

* No details provided on histopathological correlation

† Multidisciplinary decision based on imaging (initial and follow-up), clinical examination and treatment change

‡ No details provided on surgical procedures

As some studies used methods other than radical hysterectomy as the reference standard, additional subgroup analysis was performed to obtain the diagnostic performance values using studies that solely used radical hysterectomy as the reference standard. The pooled sensitivity estimates for the eight included studies was 0.73 (95% CI 0.60–0.83) with specificity of 0.93 (95% CI 0.90–0.95). The area under the HSROC curve was 0.94 (95% CI 0.91–0.96).

Discussion

In our meta-analysis, we assessed the diagnostic performance of MRI for detection of PMI in patients with cervical cancer. The pooled sensitivity and specificity of the included studies were 0.76 (95% CI 0.67–0.84) and 0.94 (95% CI 0.91–0.95), respectively. Moreover, the summary estimates using a subgroup of studies (n = 8) that solely used radical hysterectomy

Table 3 MRI characteristics

Study	Magnet strength (T)	Vendor	Machine	Coil	FSE /TSE	Imaging planes for T2WI	Slice thickness (mm)	DWI		CE-MRI		Antispasmodic agent
								Used	B-value (mm^2/s)	Used	Temporal resolution (s)	
Bleker et al. [16]	1.5	GE, Siemens	Signa Horizon Echospeed, Magnetom Avanto	PA	Yes	AO/S/CO	4	No	N/A	No	N/A	NR
Bourgioti et al. [17]	1.5	Philips	NR	PA	Yes	A/AO/S	4.5/4/3.5	Yes	0,400,1000	Yes	17	Yes
Downey et al. [18]	3	Philips	Achieva	Endovaginal	Yes	AO/SO/CO	3/2/2	Yes	0,100,300,500,800	No	N/A	Yes
Epstein et al. [19]	1.5 or 3	Philips, GE, Siemens	Intera, Achieva, Vectra, Magnetom Avanto	PA or Body	NR	AO/S	3-5	No	N/A	Yes	NR	Yes
Kitajima et al. [20]	1.5	GE	Signa Echo Speed plus Excite	PA	Yes	A/S/C	4-5	Yes	0,1000	Yes	30,60,120,180 s	No
Kong et al. [21]	3	Philips	Achieva	PA or Body	Yes	A/AO/S	5	No	N/A	No	N/A	NR
Kraljević et al. [22]	1.5	Siemens	Magnetom Avanto Harmony	PA	Yes	AO/SO/CO	5	No	N/A	Yes	1,2,5 min	Yes
Lee et al. [23]	1.5	GE	Signa	PA	Yes	A/S	5/3	No	N/A	No	N/A	NR
Moloney et al. [24]	1.5	Siemens	Avanto	PA	Yes	A/C/SO	3.5/4/5	No	N/A	Yes	NR	NR
Park et al. [25]	3	Philips	Intera Achieva	PA	Yes	A/S/C	4	Yes	0,1000	No	N/A	Yes
Shin et al. [26]	3	Siemens	Verio	PA	Yes	A/AO/S/C	3/5/5/1	No	N/A	No	N/A	Yes
Shweel et al. [27]	1	Philips	Gyroscan	PA	Yes	A/S/C	5-7/4-6/4-6	No	N/A	Yes	NR	NR
Wei et al. [28]	3	GE	Signa HDxT	PA	Yes	A/S	6	No	N/A	Yes	25,50,180 s	Yes
Yu et al. [29]	3	GE	Signa excite HD	PA	Yes	A/S/C	3	Yes	NR	Yes	NR	Yes

A axial, *AO* axial oblique, *C* coronal, *CE* contrast-enhanced, *CO* coronal oblique, *DWI* diffusion-weighted imaging, *FSE* fast spin echo, *NR* not reported, *N/A* not available, *PA* phased-array, *S* sagittal, *SO* sagittal oblique, *TSE* turbo spin echo, *T2WI* T2-weighted imaging

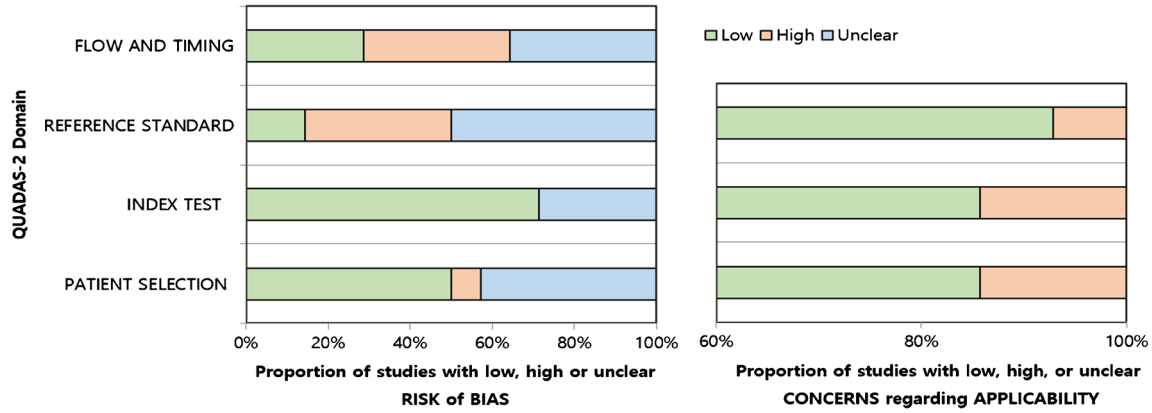

Fig. 2 Grouped bar charts show risk of bias (left) and concerns for applicability (right) of 14 included studies assessed with QUADAS-2

as the reference standard showed consistent results with a pooled sensitivity and specificity of 0.73 (95% CI 0.60–0.83) and 0.93 (95% CI 0.90–0.95), respectively. The overall sensitivity estimates for MRI in detecting PMI have not shown substantial improvement from those in previous meta-analyses by Bipat et al. [7] published in 2002 and Thomeer et al. [8] published in 2013, which reported sensitivities of 74% and 84%, respectively. Furthermore, in the study by Bipat et al. [7], the publication period (1985–1991 vs. 1992–1997 vs. 1998–2002) was demonstrated not to have influenced the sensitivity of MRI. The studies included in our study (n = 14) do not overlap with the studies included in the prior meta-analysis, and therefore represent the performance of MRI using more recent techniques. For instance, FSE or

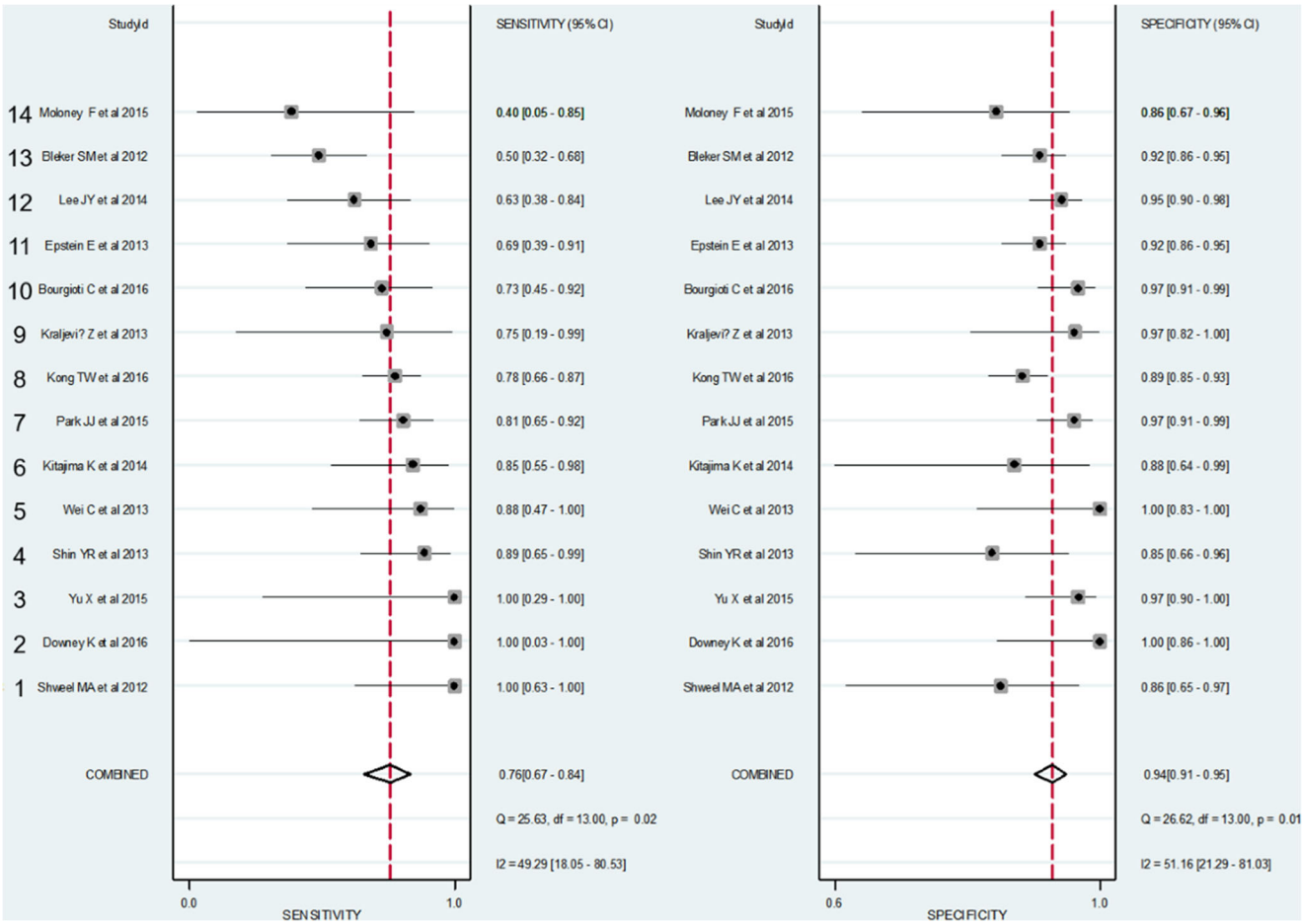

Fig. 3 Coupled forest plots of pooled sensitivity and specificity. Numbers are pooled estimates with 95% confidence intervals (CIs) in parentheses. Corresponding heterogeneity statistics are provided at the bottom right corners. Horizontal lines indicate 95% CIs. Studies are number (1–14) from bottom to top in descending order of sensitivity

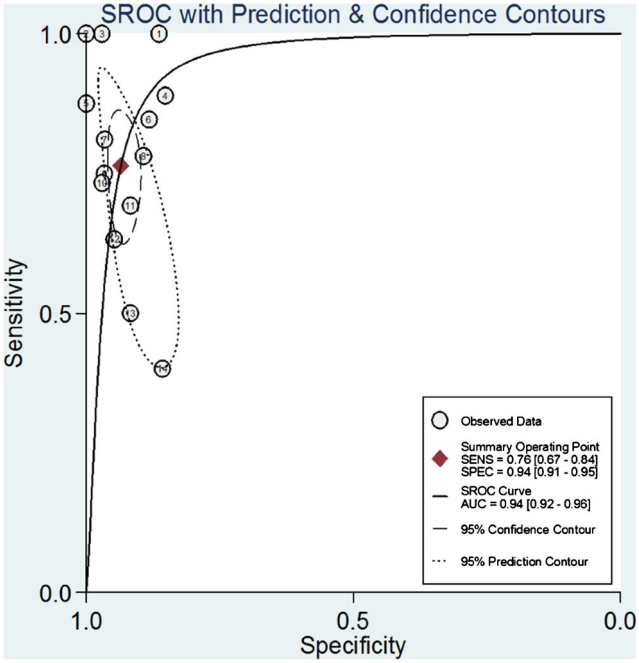

Fig. 4 Hierarchical summary receiver operating characteristic curve of the diagnostic performance of MRI for detection of parametrial invasion in cervical cancer. Each numbered circle represents each included study in order of descending sensitivity, as annotated in Fig. 3

TSE sequences were used in 13 of 14 studies in our meta-analysis, while only 17 of 36 used them in the analysis by Thomeer et al. [8]. In addition, all but one of the studies used 1.5- or 3-Tesla MRI scanners in our study, while only 24 of 36 studies in the meta-analysis by Thomeer et al. [8] used such

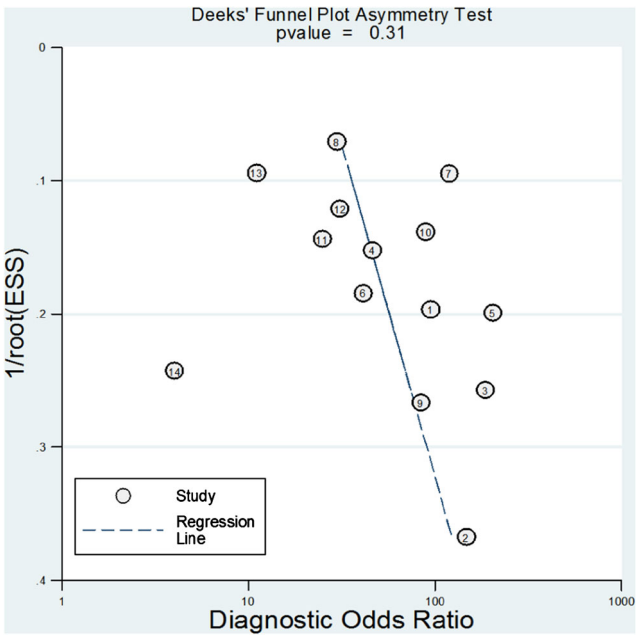

Fig. 5 Deeks' funnel plot. A *p*-value of 0.31 suggests that the likelihood of publication bias is low. Each numbered circle represents each included study in order of descending sensitivity, as annotated in Fig. 3. *ESS* effective sample size

scanners. Although the sensitivities of 74–84% reported in the previous two studies and ours can be considered good, it is still discouraging that there has been no remarkable improvement over decades, given the rapid advancement in MRI technology. However, the updated performance values of MRI in our study are consistent with previously reported values, showing better diagnostic accuracy than CT (sensitivity of 55%) and clinical examination (sensitivity and specificity of 40% and 93%, respectively), and on the basis of these results, MRI should be the preferred modality for detection of PMI in patients with cervical cancer.

One of the strengths of the current meta-analysis is the relatively low degree of heterogeneity between the included studies. Except for the results from Higgins I^2 statistics, which suggested that there may be moderate heterogeneity (I^2 = 49.29% and 51.16% for sensitivity and specificity, respectively), all other results from various statistical methods indicated that the possibility of heterogeneity was low. On the other hand, in the earlier meta-analysis by Thomeer et al. [8], substantial heterogeneity was thought to be present among the studies assessing the diagnostic performance of MRI (I^2 = 72.93% and 70.94%, respectively). The degree of heterogeneity in meta-analyses is important, as it may affect the general applicability of the results [14, 15]. Therefore, the small degree of heterogeneity among the studies in the current meta-analysis suggest that the good performance of MRI for detection of PMI may be generally applicable, and provide additional evidence for MRI to be used as a crucial modality in the FIGO staging system.

Several variables with regard to the MRI techniques were evaluated as potential sources of variation in the diagnostic performance of MRI. Magnet field strength, use of DWI sequences and administration of antispasmodic drugs were statistically significant factors, whereas coil type, slice thickness, imaging planes and use of CE-MRI sequences were not. Regarding magnet field strength, a previous meta-analysis reported that higher magnetic field (≥1.5- compared with <1.5-T) had a positive influence on detecting PMI in cervical cancer [8]. In our meta-analysis, we found that even at a greater threshold (3- vs. <3-T), higher field strength still demonstrates incremental value in the detection of PMI ($p < 0.01$). The pooled sensitivity was higher in studies using 3-T scanners than those using <3-T machines (0.84 vs. 0.66, respectively). We speculate that the improved performance when using 3-T scanners may be attributed to the higher spatial resolution, greater signal-to-noise ratio (for tumour and for cervical stroma), and greater tumour-to-cervical stroma contrast-to-noise ratio compared with using 1.5-T [31]. In addition, the pooled sensitivity and specificity were higher in studies that used DWI than in those that did not (0.82 vs. 0.72 for sensitivity; and 0.97 vs. 0.91 for specificity; $p < 0.010$). Although T2WI provides a high contrast between cervical cancer (high signal intensity) and cervical stroma (low signal intensity) for

Table 4 Results of meta-regression analysis of MRI for detection of parametrial invasion (PMI) in cervical cancer

Parameter	Category	No. of studies	LRT Chi-square	*p* value
Study design	Prospective	6	0.06	0.97
	Retrospective	8		
Race	Asian	7	3.51	0.17
	Non-Asian	7		
Prevalence of PMI (%)	≥16.7	7	3.78	0.15
	<16.7	7		
Magnet field strength (T)	3	6	14.35	<0.01
	≤1.5	7		
Coil type	Phased-array or endovaginal	12	4.33	0.11
	Others	2		
Slice thickness (mm)	≤5	12	3.49	0.17
	>5	2		
T2WI planes	Includes AO/S	7	1.74	0.42
	Others	7		
DWI	Used	5	10.66	<0.01
	Not used	9		
CE-MRI	Used	8	0.61	0.74
	Not used	6		
Antispasmodic agents	Used	8	61.87	<0.01
	Not used	1		

LRT likelihood ratio test, *DWI* diffusion-weighted imaging, *T2WI* T2-weighted imaging, *CE* contrast-enhanced

evaluation of PMI, DWI has been considered to have the potential for added value, as cervical cancer manifests with higher signal intensity on DWI with corresponding lower apparent diffusion coefficient values compared with the normal cervical stroma [32]. However, it is crucial to note that no study included our meta-analysis used DWI alone, but interpretation of PMI in studies using DWI was based on comprehensive evaluation of both T2WI and DWI. DWI by itself suffers from poor spatial resolution and anatomical detail, but using DWI as an adjunct to T2WI may improve the diagnostic performance of detecting PMI in cervical cancer. The use of antispasmodic drugs was also shown to be a significant factor affecting heterogeneity ($p < 0.01$). However, further statistical analysis was limited due to the fact that only one study reported that they did not use antispasmodic drugs. Therefore, caution is needed when interpreting the effect of antispasmodic drugs. Nevertheless, antispasmodic drugs are well known to decrease bowel motion artefacts from peristalsis, and are used in the majority of ESUR members for this reason [33].

Our meta-analysis had some limitations. First, it included only a relatively small number of articles (n = 14). However, this was mainly due to the fact that we only included studies since 2012 so that there was no overlap with previous meta-analyses in order to obtain updated diagnostic performance values. Nevertheless, we were able to acquire pooled estimates with relatively low heterogeneity from the included studies. Second, there was a lack of patients with advanced disease in most studies. Specifically, only five (35.7%) of the 14 studies included patients with FIGO stage IIB or higher. Pooling studies mostly including patients with low or intermediate stage disease may have led to a bias toward decreased sensitivity. This bias may have been more evident as the five studies that included high-stage disease had a relatively smaller number of patients (n = 29–45). Third, there were six studies that included patients that did not undergo radical hysterectomy. Therefore, we performed a subgroup analysis for studies that solely used radical hysterectomy as the reference standard, and obtained consistent results. Fourth, we used the performance values from the reader with highest experience when there were multiple readers. However, the inter-reader agreement was substantial or almost perfect with kappa values of 0.735 (between 20 and 11 years' experience by Yu et al. [29]), 0.82 (between 14 and 4 years by Shin et al. [26]) and 0.86 (between 9 and 3 years by Park et al. [25]) [34]. Fifth, it should be noted that the ethnicity of the study population was based on the nationality of the institution. Although this may be generally correct, there could have been a minor population of non-Asian patients who underwent MRI examinations in hospitals in Asia, or vice versa.

Conclusion

MRI shows good performance for detection of PMI in patients with cervical cancer with a pooled sensitivity of 0.76 and specificity of 0.94. The use of 3-T scanners and DWI may further improve diagnostic performance.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Sang Youn Kim.

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.

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

Statistics and biometry One of the authors (Chong Hyun Suh) has significant statistical expertise.

Informed consent Written informed consent was not required for this study due to the nature of the study, which was a systematic review and meta-analysis.

Ethical approval Institutional Review Board approval was not required for this study due to the nature of the study, which was a systematic review and meta-analysis.

Methodology
• Meta-analysis
• Multicentre study

References

1. Siegel R, Naishadham D, Jemal A (2012) Cancer statistics, 2012. CA Cancer J Clin 62:10–29
2. Landoni F, Bocciolone L, Perego P, Maneo A, Bratina G, Mangioni C (1995) Cancer of the cervix, FIGO stages IB and IIA: patterns of local growth and paracervical extension. Int J Gynecol Cancer 5: 329–334
3. Peters WA 3rd, Liu PY, Barrett RJ 2nd et al (2000) Concurrent chemotherapy and pelvic radiation therapy compared with pelvic radiation therapy alone as adjuvant therapy after radical surgery in high-risk early-stage cancer of the cervix. J Clin Oncol 18:1606–1613
4. Pecorelli S, Benedet JL, Creasman WT, Shepherd JH (1999) FIGO staging of gynecologic cancer. 1994–1997 FIGO Committee on Gynecologic Oncology. International Federation of Gynecology and Obstetrics. Int J Gynaecol Obstet 65:243–249
5. Motta-Ramirez GA, Remer EM, Herts BR, Gill IS, Hamrahian AH (2005) Comparison of CT findings in symptomatic and incidentally discovered pheochromocytomas. AJR Am J Roentgenol 185:684–688
6. Pecorelli S, Zigliani L, Odicino F (2009) Revised FIGO staging for carcinoma of the cervix. Int J Gynaecol Obstet 105:107–108
7. Bipat S, Glas AS, van der Velden J, Zwinderman AH, Bossuyt PM, Stoker J (2003) Computed tomography and magnetic resonance imaging in staging of uterine cervical carcinoma: a systematic review. Gynecol Oncol 91:59–66
8. Thomeer MG, Gerestein C, Spronk S, van Doorn HC, van der Ham E, Hunink MG (2013) Clinical examination versus magnetic resonance imaging in the pretreatment staging of cervical carcinoma: systematic review and meta-analysis. Eur Radiol 23:2005–2018
9. Hori M, Kim T, Onishi H et al (2011) Uterine tumors: comparison of 3D versus 2D T2-weighted turbo spin-echo MR imaging at 3.0 T: initial experience. Radiology 258:154–163
10. Liberati A, Altman DG, Tetzlaff J et al (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol 62:e1–e34
11. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J (2003) The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 3:25
12. Suh CH, Park SH (2016) Successful publication of systematic review and meta-analysis of studies evaluating diagnostic test accuracy. Korean J Radiol 17:5–6
13. 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
14. Higgins J, Green S. Cochrane handbook for systematic reviews of interventions version 5.1.0. The Cochrane Collaboration. http://handbook.cochrane.org/chapter_9/9_5_2_identifying_and_measuring_heterogeneity.htm. Updated March 2011. Accessed January 3, 2017
15. McInnes MDF, Bossuyt PMM (2015) Pitfalls of systematic reviews and meta-analyses in imaging research. Radiology 277:13–21
16. Bleker SM, Bipat S, Spijkerboer AM, van der Velden J, Stoker J, Kenter GG (2013) The negative predictive value of clinical examination with or without anesthesia versus magnetic resonance imaging for parametrial infiltration in cervical cancer stages IB1 to IIA. Int J Gynecol Cancer 23:193–198
17. Bourgioti C, Chatoupis K, Rodolakis A et al (2016) Incremental prognostic value of MRI in the staging of early cervical cancer: a prospective study and review of the literature. Clin Imaging 40:72–78
18. Downey K, Attygalle AD, Morgan VA et al (2016) Comparison of optimised endovaginal vs external array coil T2-weighted and diffusion-weighted imaging techniques for detecting suspected early stage (IA/IB1) uterine cervical cancer. Eur Radiol 26:941–950
19. Epstein E, Testa A, Gaurilcikas A et al (2013) Early-stage cervical cancer: tumor delineation by magnetic resonance imaging and ultrasound - a European multicenter trial. Gynecol Oncol 128:449–453
20. Kitajima K, Suenaga Y, Ueno Y et al (2014) Fusion of PET and MRI for staging of uterine cervical cancer: comparison with contrast-enhanced (18)F-FDG PET/CT and pelvic MRI. Clin Imaging 38:464–469
21. Kong TW, Kim J, Son JH et al (2016) Preoperative nomogram for prediction of microscopic parametrial infiltration in patients with FIGO stage IB cervical cancer treated with radical hysterectomy. Gynecol Oncol 142:109–114
22. Kraljevic Z, Viskovic K, Ledinsky M et al (2013) Primary uterine cervical cancer: correlation of preoperative magnetic resonance imaging and clinical staging (FIGO) with histopathology findings. Coll Antropol 37:561–568
23. Lee JY, Youm J, Kim TH et al (2014) Preoperative MRI criteria for trials on less radical surgery in Stage IB1 cervical cancer. Gynecol Oncol 134:47–51
24. Moloney F, Ryan D, Twomey M, Hewitt M, Barry J (2016) Comparison of MRI and high-resolution transvaginal sonography for the local staging of cervical cancer. J Clin Ultrasound 44:78–84
25. Park JJ, Kim CK, Park SY, Park BK (2015) Parametrial invasion in cervical cancer: fused T2-weighted imaging and high-b-value diffusion-weighted imaging with background body signal suppression at 3 T. Radiology 274:734–741
26. Shin YR, Rha SE, Choi BG, Oh SN, Park MY, Byun JY (2013) Uterine cervical carcinoma: a comparison of two- and three-dimensional T2-weighted turbo spin-echo MR imaging at 3.0 T for image quality and local-regional staging. Eur Radiol 23:1150–1157
27. Shweel MA, Abdel-Gawad EA, Abdel-Gawad EA, Abdelghany HS, Abdel-Rahman AM, Ibrahim EM (2012) Uterine cervical malignancy: diagnostic accuracy of MRI with histopathologic correlation. J Clin Imaging Sci 2:42
28. Wei C, Dong JN, Li NY, Wei SH, Gao F (2013) Value of 3.0T MR LAVA-Flex multi-phase enhancement techniques combined with DWI in preoperative evaluation and staging of cervical cancer. Chin J Med Imag Technol 29:2015–2019
29. Yu X, Lin M, Ye F et al (2015) Comparison of contrast-enhanced isotropic 3D-GRE-T1WI sequence versus conventional non-isotropic sequence on preoperative staging of cervical cancer. PLoS One 10, e0122053
30. Lee J, Kim KW, Choi SH, Huh J, Park SH (2015) Systematic review and meta-analysis of studies evaluating diagnostic test accuracy: A practical review for clinical researchers–Part II. Statistical Methods of Meta-Analysis. Korean J Radiol 16:1188–1196

31. Hori M, Kim T, Murakami T et al (2009) Uterine cervical carcinoma: preoperative staging with 3.0-T MR imaging: comparison with 1.5-T MR imaging. Radiology 251:96–104
32. Naganawa S, Sato C, Kumada H, Ishigaki T, Miura S, Takizawa O (2005) Apparent diffusion coefficient in cervical cancer of the uterus: comparison with the normal uterine cervix. Eur Radiol 15:71–78
33. Balleyguier C, Sala E, Da Cunha T et al (2011) Staging of uterine cervical cancer with MRI: guidelines of the European Society of Urogenital Radiology. Eur Radiol 21: 1102–1110
34. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33:159–174