



The accuracy of Vesical Imaging-Reporting and Data System (VI-RADS): an updated comprehensive multi-institutional, multi-readers systematic review and meta-analysis from diagnostic evidence into future clinical recommendations

Francesco Del Giudice^{1,2} · Rocco Simone Flammia¹ · Martina Pecoraro³ · Marco Moschini⁴ · David D'Andrea⁵ · Emanuele Messina³ · Lucia Martina Piscioti³ · Ettore De Berardinis¹ · Alessandro Sciarra¹ · Valeria Panebianco³

Received: 31 October 2021 / Accepted: 17 February 2022
© The Author(s) 2022

Abstract

Purpose To determine through a comprehensive systematic review and meta-analysis the cumulative diagnostic performance of vesical imaging-reporting and data system (VI-RADS) to predict preoperative muscle-invasiveness among different institutions, readers, and optimal scoring accuracy thresholds.

Methods PubMed, Cochrane and Embase were searched from inception up to May 2021. Sensitivity (Sn), Specificity (Sp) were first estimated and subsequently pooled using hierarchical summary receiver operating characteristics (HSROC) modeling for both cut-off ≥ 3 and ≥ 4 to predict muscle-invasive bladder cancer (MIBC). Further sensitivity analysis, subgroup analysis and meta-regression were conducted to investigate contribution of moderators to heterogeneity.

Results In total, $n = 20$ studies from 2019 to 2021 with $n = 2477$ patients by $n = 53$ genitourinary radiologists met the inclusion criteria. Pooled weighted Sn and Sp were 0.87 (95% CI 0.82–0.91) and 0.86 (95% CI 0.80–0.90) for cut-off ≥ 3 while 0.78 (95% CI 0.74–0.81) and 0.94 (95% CI 0.91–0.96) for cut-off ≥ 4 . The area under the HSROC curve was 0.93 (95% CI 0.90–0.95) and 0.91 (95% CI 0.88–0.93) for cut-off ≥ 3 and ≥ 4 , respectively. Meta-regression analyses showed no influence of clinical characteristics nor cumulative reader's experience while study design and radiological characteristics were found to influence the estimated outcome.

Conclusion We demonstrated excellent worldwide diagnostic performance of VI-RADS to determine pre-trans urethral resection of bladder tumor (TURBT) staging. Our findings corroborate wide reliability of VI-RADS accuracy also between different centers with varying experience underlining the importance that standardization and reproducibility of VI-RADS may confer to multiparametric magnetic resonance imaging (mpMRI) for preoperative BCa discrimination.

Keywords Bladder cancer · Multiparametric magnetic resonance imaging · Muscle-invasive bladder cancer · Bladder cancer staging

✉ Valeria Panebianco
valeria.panebianco@uniroma1.it

¹ Department of Maternal Infant and Urologic Sciences, "Sapienza" University of Rome, Policlinico Umberto I Hospital, Viale del Policlinico 155, Rome 00161, Italy

² Department of Urology, Stanford Medical Center, Stanford, CA, USA

³ Department of Radiological Sciences, Oncology and Pathology, Policlinico Umberto I Hospital, "Sapienza"

University/Policlinico Umberto I of Rome, Viale del Policlinico 155, 00161 Rome, Italy

⁴ Department of Urology and Division of Experimental Oncology, Urological Research Institute, Vita-Salute San Raffaele University, Milan, Italy

⁵ Department of Urology, Comprehensive Cancer Center, Medical University of Vienna, Vienna General Hospital, Vienna, Austria

Introduction

The adoption of the Vesical Imaging Reporting and Data System (VI-RADS) criteria based on multiparametric magnetic resonance imaging (mpMRI) of the bladder before trans-urethral resection of bladder tumor (TURBT) has been expanding worldwide [1, 2]. The criteria are established on a standardized 5-point scale assessing the preoperative likelihood of muscle invasiveness, but also as potential criterion for predicting final pT stage among bladder cancer (BCa) patients eligible to curative interventions [3]. Despite its relative novel introduction, two diagnostics meta-analysis assessing VI-RADS pooled diagnostic accuracy are already available and showed an excellent promising performance in discriminating non-muscle vs. muscle invasive bladder cancer (NMIBC vs. MIBC) [4, 5]. Still, these preliminary experiences were likely suffering from lack of granularity both in terms of statistical power due to the relatively small number of studies analyzed, and in terms of genito-urinary (GU) readers, MRI scanners and threshold cut-off criteria selected to define MIBC.

VI-RADS is rapidly moving forward into clinical practice. A mounting body of evidence, including published prospective data and ongoing clinical trials, is indeed exploring the wide flexibility of this novel score for new clinical insights to potentially drive the therapeutic algorithm across different BCa stages. This has so far involved selection criteria for high-risk NMIBC candidates for secondary resection [6], complete/partial radiographic response in MIBC undergoing neoadjuvant regimens [7, 8], as well as those locally advanced cases directly addressed to curative interventions avoiding reliance on extended trans-detrusor resections in favor of adequate biopsy sampling [9].

Despite promising implications of internalizing VI-RADS into therapeutic uro-oncologic algorithm, the following relevant issues still remain unsolved: (I) exploring the optimal diagnostic settings of the score; (II) reproducibility between different MRI scanners and readers with varying experience; (III) defining appropriate threshold cut-off criteria ($\text{VI-RADS} \geq 3$ vs ≥ 4) for muscle invasiveness-definition. These unmet needs require a definitive statement before further dedicated clinical trials and investigations can be developed to assess predictive value of VIRADS for determining muscle invasion in bladder cancer patients. With this aim we performed an updated comprehensive systematic review of the literature by including all the available international experiences validating the VI-RADS in the pre-TURBT setting for MIBC determination and we provided pooled estimates regarding the diagnostic performance of the score among all available GU readers, MRI scans and VI-RADS cut-off threshold criteria involved.

Methods

This systematic review and meta-analysis was conducted according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [10]. A research question was established based on the Patient-Index test-Comparator-Outcome-Study design (PICOS) criteria as the following: what is the current cumulative diagnostic performance of VI-RADS scoring criteria for NMIBC vs. MIBC clinical staging discrimination? Furthermore, our goal was to compare current evidence within all available retrospective/prospective and/or single-/multicenter cohort studies applying different MIBC cut-off VI-RADS criteria as compared with histopathological results. In particular, we determined the pooled diagnostic performance estimators among all the available radiologists with different level of expertise and GU-specialized volume institutions.

Evidence acquisition

We performed a systematic review of the literature in PubMed, Web of Science, Embase, and Cochrane from inception to May 2021, without language restriction, to identify studies that examined the implementation of pre-TURBT VI-RADS scoring criteria for BCa staging purposes and evaluated the pooled diagnostic performance between the radiologists involved. The reference lists of the included studies were also screened for relevant articles. Only original prospective and retrospective cohort studies were included and critically evaluated (Level of Evidence: II and III-a). Case reports, abstracts and meeting reports were excluded from the analysis. Search key terms included with primary and secondary fields have been reported in Supplementary Table 1.

Selection of the studies and criteria of inclusion

Entry into the analysis was restricted to data collected from original articles that examined patients with primary and/or recurrent BCa diagnosis, which assessed final BCa extension through surgical specimen both from TURBT/Re-TURBT or radical/partial cystectomy (RC), and that aimed to report the standard diagnostic indicators of VI-RADS performance for identifying MIBC preoperatively. Moreover, only those studies including sufficient data to reconstruct 2×2 tables with regard to sensitivity and specificity for the outcome of interest (i.e., VI-RADS score cut-off ≥ 3 and cut-off ≥ 4 to predict MIBC) and which assumed “per index lesion” level analysis were considered suitable for further consideration. Additionally, studies were considered eligible if at least one of the involved readers were GU radiologists with at least 5 years’ experience in the GU-MRI imaging, at least one reader was

actively involved in the imaging acquisition and revision, and if MRI images acquisition protocol was consistent with what described in the original VI-RADS document [3].

Articles were excluded if they met one or more of the following criteria: inadequate information for data extraction or quality assessment; inclusion of study population consisting of < 15 index lesions; presented outcomes which dealt with other topics (e.g., MRI used to determine clinical T stage (cT) without VI-RADS, or VI-RADS score for each MRI sequence was available but an overall VI-RADS score was not expressed for MIBC detection).

Six authors (FDG, MP, EM, MLP, SF and EDB) independently screened the titles and abstracts of all articles using predefined inclusion criteria. The full-text articles were examined independently by the five (FDG, MP, EM and VP) to determine whether or not they met the inclusion criteria. Final inclusion was determined by consensus of all investigators. Selected articles were then critically analyzed.

The following data were extracted from the included studies by using a standardized form: origin of study (institution and period of enrollment), size of study population, period of time prospectively/retrospectively covered, gold standard for pathological MIBC definition, technical parameters of MRI acquisition (1.5 vs. 3 Tesla magnet, T2WI slice thickness, b values used for DWI, and temporal resolution of DCE-MRI), details regarding MRI interpretation (number of readers, experience of the senior reader, cumulative GU-MRI experience for all the readers involved, and whether they were blinded or not to clinical history), the VI-RADS cutoff value used for determining MIBC on MRI (i.e., cut-off ≥ 3 and cut-off ≥ 4), outcomes related to diagnostic performance of VI-RADS. Finally, baseline clinical and pathological patients and tumor features (e.g., mean/median age, range of patients, number of tumors, percentage of patients with MIBC and histopathological subtypes of tumors screened).

Assessment of quality for studies included and statistical analysis

To assess the risk of bias (RoB), all included experiences were independently reviewed using the “Quality Assessment of Diagnostic Accuracy Studies” (QUADAS-2) [11], by assessing the potential risk for selection bias, information bias, measurement bias, or confounding bias. Three reviewing authors (FDG, MP and SF) independently assessed the methodological quality based on sequence generation, allocation concealment, blinding of personnel, blinding of outcome assessors, incomplete outcome data, selective outcome reporting, and additional sources of bias. Publication bias was tested both by visual assessment of the Deeks’ funnel plot and calculation of p value using the Deeks’ asymmetry test [13]. We compared diagnostics indicators among the studies at different steps. First, we generated 2×2

contingency tables for each GU radiologist involved in each article based on the reported Sn, Sp or raw data and then we used pooled weighted Sn, Sp, likelihood positive and negative ratio (LR+, LR-) with their computed 95% Confidence Intervals (CI) for determining the overall weighted diagnostic estimate for both VI-RADS threshold cut-off ≥ 3 and ≥ 4 . Secondly, pooled Sn and Sp estimates were calculated with hierarchical logistic regression modeling, including bivariate and hierarchical summary receiver operating characteristic (HSROC) modeling, and then graphically presented using HSROC curves with 95%CI and prediction regions for both threshold cut-off ≥ 3 and ≥ 4 , respectively. Sensitivity analyses were performed to assess the contribution of each study to the pooled estimate by excluding individual trials one at a time and recalculating the pooled estimates for the remaining studies. Evaluation for presence of heterogeneity was done using [12]: (1) Cochran’s *Q* test with $p < 0.05$ signifying heterogeneity; (2) Higgins I^2 test with inconsistency index (I^2) = 0–40%, heterogeneity might not be important; 30–60%, moderate heterogeneity; 50–90%, substantial heterogeneity; and 75–100%, considerable heterogeneity. The pooled weighted Sn, Sp, LR+ and LR- estimates were calculated using random effects model. Our results are graphically displayed as forest plots on a per-single GU reader level, with pooled results indicating overall accuracy to discriminate NMI from MIBC using VI-RADS criteria ≥ 3 and ≥ 4 , respectively. Furthermore, Fagan nomogram was generated to display the post-test probabilities when the pre-test probability was 25%, which corresponds to the prevalence of MIBC over the years screened.

Subgroup analyses were performed looking at differences in categorical confounders (e.g., Magnet strength [T], MIBC proportion, study design, etc.). Meta-regression analyses were performed using available continuous variables retrieved among the studies. Pooled weighted diagnostic estimates were plotted against the following available quantitative variables: mean/median age of the patients, total number of patients/lesions screened, range of study time screened (months retrospectively or prospectively imputed), the relative percentage of MIBC ($\geq T2$) documented and months from original VI-RADS publication release. Calculations were accomplished using the MIDAS command on Stata version 17.1 (Stata Corporation, College Station, TX, USA).

Results

Search results

The initial search yielded $n = 115$ articles (PubMed: 86; Cochrane: 6; and Embase: 23). Forty-four were excluded, as they contained overlapping data or were duplicates

appearing in multiple databases. Of the remaining $n=71$, $n=47$ were further excluded because did not examine VI-RADS ($n=9$), contained only MRI-based sequences information ($n=5$) or were review papers, editorials or abstracts ($n=33$). Full-text articles were then reevaluated and critically analyzed for the remaining $n=24$ journal references. Within this in-depth review, a further $n=4$ did not meet the inclusion criteria. The remaining $n=20$ studies were included in our review (Supplementary Fig. 1). No study was considered to be seriously flawed as per the “Quality Assessment of Diagnostic Accuracy Studies”. Studies’ risk to performance bias was overall low with some attrition bias due to incomplete outcome data across all the studies. Individual RoB as well as visual assessment of the Deeks’ funnel plots are illustrated in Supplementary Table 2 and Fig. 3, respectively.

Location, design, and characteristics of the studies population

Patient, tumor, and study characteristics are summarized in Table 1. Of the 20 included articles, $n=11$ were conducted in Asia ($n=5$ in China [15–17], $n=4$ in Japan [18–21], $n=2$ in Korea [22, 23]), $n=6$ in Europe ($n=4$ in Italy [6, 24–26], $n=1$ in Spain [27]) $n=1$ in Turkey [28] and 3 in other continents (2 in Egypt [29, 30], 1 in Brazil [31]). Range of eligible study time was comprised in between 2019 and 2021 and included patients who had been treated for BCa between 2005 and 2020. All eligible articles were single centered with Metwally et al. [30] representing the only available multicenter experience. Out of these, $n=14$ were retrospective and $n=6$ were prospective with cumulative sample size ranging from 18 up to 340 patients (Table 1).

Across eligible studies, median age was between 57 and 74 years old. Four studies analyzed independently more than one lesion per patients [16, 22–25] and consequently the total amount of lesions investigated exceeds the number of patients included (2609 vs 2477). Nevertheless, all these studies reported VI-RADS accuracy relying on the index lesions identified. All the studies relied on histopathological report from TURBT and/or partial cystectomy and/or RC, performed within 1 to 12 weeks from MRI. MIBC rates retrieved among the experiences included ranged from 10 to 53%. Moreover, the majority of the studies included only urothelial bladder cancer, with other variant histology accounting only for a relative limited proportion in each study (from 1 to 6%) (Table 1).

Technical imaging modalities and reader characteristics

The MRI parameters and GU reader characteristics of each study are summarized in Table 2. For imaging acquisition

protocols, $n=12$ studies [6, 13–17, 20, 22–26] relied on 3 T scanners, $n=6$ [21, 27–31] on 1.5 T and the remaining $n=2$ [18, 19] on either 1.5 or 3 T scanners, with T2WI reported slice thickness ranging from 2 to 5 mm. In all the enrolled studies, MRI findings were always interpreted blinded to the clinical and pathological patient’s history. Of note, regardless form classic VI-RADS accuracy and reproducibility trials, $n=5$ experiences focused on potential approaches to improve VI-RADS score diagnostic performance. Specifically, Akcay et al. [28] analyzed tumor contact length as a parameter to improve the accuracy of VI-RADS score 3. Moreover, Li et al. and Sakamoto et al. [17, 20] implemented VI-RADS with standardized tumor apparent diffusion coefficients (st-ADC) and volumetric ADC histogram analysis, respectively. Additionally, Arita et al. [21] and Meng et al. [16] relied on 3D fast spin echo (FSE) T2-weighted acquisition, instead of classical 2D FSE T2-weight acquisition and bi-planar reduced field-of-view DWI (rFOV DWI). Finally, Delli Pizzi et al. [26] compared VIRADS score relying on contrast-free bi-parametric vs multiparametric MRI. Among the aforementioned studies, our analysis was, however, focused on the standard VI-RADS outcomes reported and relied on the reporting findings dichotomized for cut-off score ≥ 3 vs. ≥ 4 .

According to single reader diagnostic performance availability, a total of 53 GU radiologists ranging from two up to seven for each study was identified. Only five studies included more than 2 readers [18, 19, 23, 26, 30]. Moreover, two studies included residents among the eligible readers [26, 31] and one reported inexperienced radiologist [19]. However, only $n=7$ [16, 18, 19, 21, 24, 26, 28] of the eligible studies reported diagnostic performance for all included readers [16, 19, 21, 24, 26, 28]. The overall cumulative level of radiologist experience was declared in $n=13$ studies [6, 13, 16, 17, 20–24, 26–28, 30]. It varied from a cumulative GU experience per study of 8 up to 47 years and from a single radiologist expertise varying from 4 up to 20 years. Finally, regarding the VI-RADS score thresholds adopted, $n=12$ studies reported extractable data on both VI-RADS ≥ 3 and VI-RADS ≥ 4 , as cut-off defining the probability of MIBC before staging resection. Conversely, $n=6$ and $n=2$ studies reported only cut-off ≥ 3 or ≥ 4 , respectively. Out of these, only five studies provided raw data for estimating diagnostic performance according to different cut-off.

Diagnostic performance of VI-RADS cut-off ≥ 3 for MIBC detection

All the 20 [13–31] studies enrolled reported diagnostic accuracy information for VI-RADS cut-off ≥ 3 for MIBC. Pooled paired Sn and Sp were 0.87 (95% CI 0.82–0.91) and 0.86 (95% CI 0.80–0.90) (Fig. 1A), while pooled

Table 1 Clinical and demographic characteristics of the studies enrolled in the systematic review and meta-analysis

Study	Year of publication	Study design	Centers	Time (month)	Age (years)	Patients (n)	Lesions (n)	T stage	MIBC (%)	UCUB (%)	From submission (mo)	From acceptance (mo)
Barchetti et al	2019	Retrospective	Single	10	69	75	103	Ta-T3a	29%	100%	20	19
Ueno et al	2019	Retrospective	Single	92	73	74	74	≥Tis	50%	96%	NA	26
Wang H et al	2019	Retrospective	Single	80	64	340	340	Ta-T4	25%	100%	30	26
Makboul et al	2019	Prospective	Single	NA	57	50	50	T1-T4	36%	NA	24	19
Del Giudice et al	2019	Prospective	Single	18	67	231	231	Ta-T4	27%	100%	NA	19
Kim SH	2019	Retrospective	Single	38	66	297	339	Tis-T4	30%	94%	NA	22
Hong et al	2020	Retrospective	Single	6	73	32	66	NA	25%	NA	20	14
Marchioni et al	2020	Prospective	Single	NA	73	68	68	Ta-T3	10%	NA	14	14
Sakamoto et al	2020	Retrospective	Single	67	73	176	176	≥Ta	26%	100%	17	13
Liu et al	2020	Retrospective	Single	17	68	126	126	Ta-T4	39%	100%	19	14
Vaz et al	2020	Retrospective	Single	35	68	30	30	NA	27%	93%	24	22
Wang Z et al	2020	Retrospective	Single	29	66	220	220	Ta-T4	51%	NA	19	13
Arita et al	2020	Retrospective	Single	94	74	66	66	Ta-T3	26%	100%	13	8
Etxano et al	2020	Retrospective	Single	4	70	18	18	Tis-T2	33%	99%	14	6
Ueno et al	2020	Retrospective	Single	96	73	91	91	Ta-T2	53%	100%	12	9
Metwally et al	2021	Prospective	Multi	18	62	331	331	NA	57%	100%	5	3
Akçay et al	2021	Prospective	Single	22	68	73	73	NA	42%	100%	16	2
Meng et al	2021	Retrospective	Single	7	60	61	89	Ta-T4	31%	NA	4	7
Delli Pizzi et al	2021	Prospective	Single	12	73	38	38	Ta-T3	18%	100%	9	7
Li et al	2021	Retrospective	Single	16	61	80	80	Ta-T4	34%	100%	2	1

MIBC muscle-invasive bladder cancer; UCUB urothelial carcinoma of the urinary bladder; NA not applicable

Table 2 Radiological and MRI-specific characteristics of the studies enrolled in the systematic review and meta-analysis

Study	Year	Readers (n)	1st reader experience (years)	Cum readers' experience (years)	Magnetic field (T)	T2WI slice thickness (mm)	DWI b (s/mm ²)	DCE temporal resolution (s)	Interval MRI-surgery (weeks)	Ref. for diagnosis
Barchetti et al	2019	2	10	15	3	3–4	0–88–1000–2000	5"	< 6	TURBT
Ueno et al	2019	5	> 20	NA	1.5/3	4	0–1000	40"–80"–120"–160"	NA	TURBT
Wang H et al	2019	2	32	40	3	NA	NA	(20"–131")×5	< 2	TURBT/Cystectomy
Makboul et al	2019	2	NA	NA	1.5	3	0–400–800–1000	20"–70"–180"	NA	TURBT
Del Giudice et al	2019	2	25	37	3	3–4	0–800–1000–2000	5"	< 6	TURBT/Cystectomy
Kim SH	2019	2	12	18	3	3	0–1000	30" x (4–6)	NA	TURBT/Cystectomy
Hong et al	2020	3	24	47	3	4	0–50–800–1000	30"	2	TURBT/Cystectomy
Marchioni et al	2020	2	> 20	NA	3	3–4	0–600–1000–1500–2000	30"	NA	TURBT
Sakamoto et al	2020	2	15	24	3	4–5/3.5	0–1000/0–2000	NA	NA	TURBT
Liu et al	2020	2	NA	NA	3	4–5	0–50–800–1000	(20"–131")×5	2	TURBT/Cystectomy
Vaz et al	2020	2	13	NA	1.5	NA	50–400–800	60"–300"	1–2	TURBT
Wang Z et al	2020	2	NA	NA	3	NA	NA	NA	2	TURBT/Cystectomy
Arita et al	2020	2	18	30	1.5	2	0–1000	40"	2	TURBT/Cystectomy
Etxano et al	2020	2	4	8	1.5	3	0–800–1000	7"	NA	TURBT
Ueno et al	2020	7	> 20	NA	1.5/3	4	0–1000	30"–60"–90"–120"–180"	12	TURBT/Cystectomy
Metwally et al	2021	4	16	41	1.5	4	0–800–1000	5"	1	TURBT
Akcay et al	2021	2	20	25	1.5	3	0–1800	8"	NA	TURB/Cystectomy
Meng et al	2021	2	10	18	3	4	0–800	NA	2	TURBT
Delli Pizzi et al	2021	4	15	31	3	4	600–1000–1500–2000	NA	NA	TURBT
Li et al	2021	2	16	24	3	4	0–800	NA	NA	TURBT/Cystectomy

T2WI T2-weighted imaging; DWI diffusion-weighted imaging; DCE dynamic contrast enhanced; NA not applicable

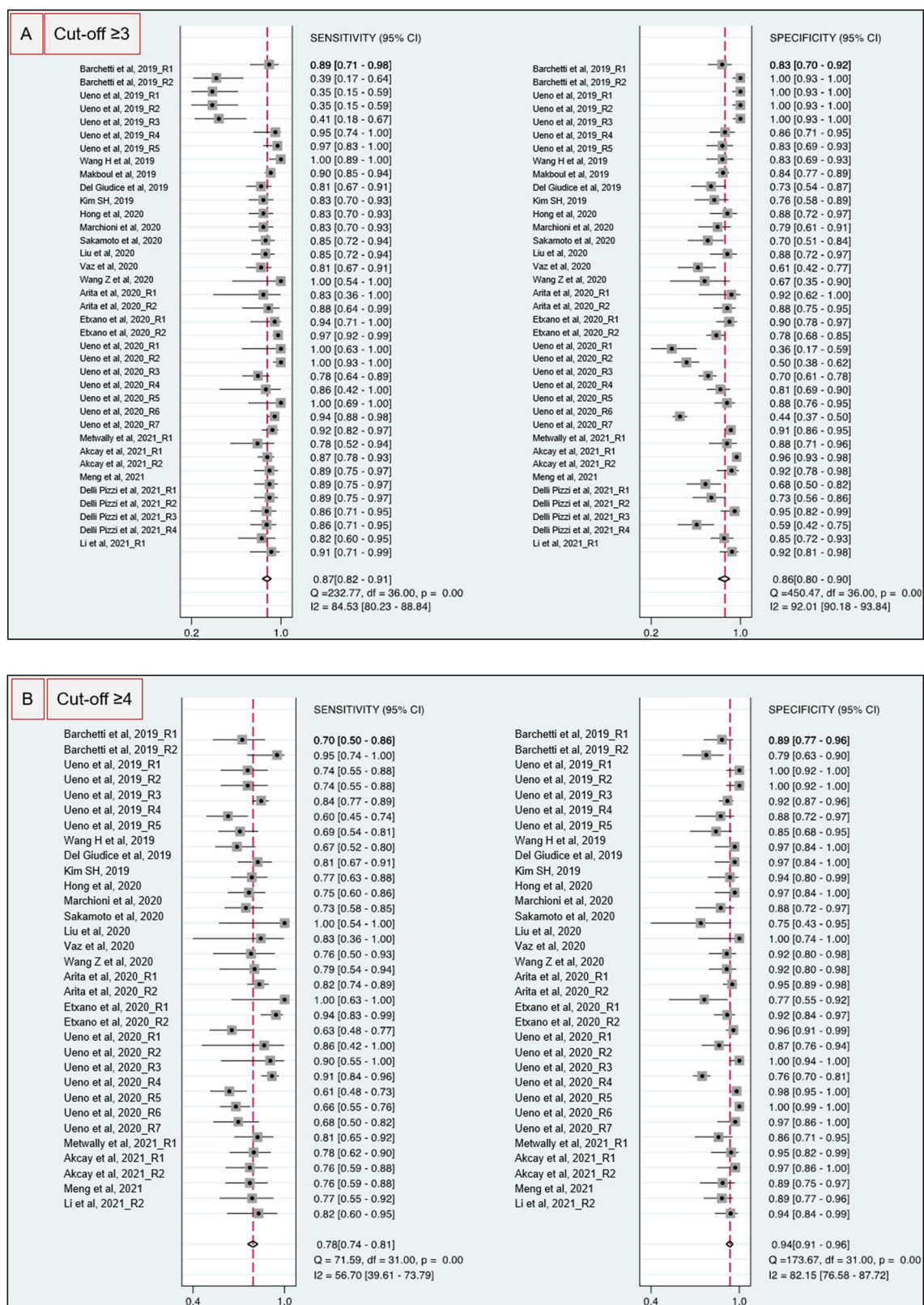


Fig. 1 Forest plot of pooled sensitivity and specificity for VIRADS ≥ 3 (A) or ≥ 4 (B) as the cut-off criterion for MIBC identification

LR + and LR – were 6.1 (95% CI 4.3–8.5) and 0.15 (95% CI 0.11–0.20), respectively. The area under the HSROC curve was 0.93 (95% CI 0.90–0.95; Fig. 2A). Given a pre-test probability of 30%, Fagan nomogram showed a positive and negative post-test probability for detecting MIBC of 72% and 6%, respectively (Fig. 3A). As there was evidence for the presence of substantial heterogeneity within the studies, the analyses reported results according to a random-effect model. Inspection of funnel plot suggested that for all the 20 studies together, there was no significant small-study effect with no study tending to have a higher outlier estimate, as depicted in Supplementary Fig. 3. Indeed, Egger's regression test showed non-significant small-study effect ($p=0.58$) whilst the 'Trim and Fill' method suggested that no 'missing' studies would need to be included to remove asymmetry from the funnel plots. Furthermore, at sub-group analysis, the study design, as well as the magnetic strength and slice thickness adopted revealed to be the source for major heterogeneity among the studies. Meta-regression and sub-analysis plots have been summarized in Supplementary Fig. 4A.

Diagnostic performance of VI-RADS cut-off ≥ 4 for MIBC detection

$N=18$ [6, 13, 14, 16–25, 27, 28, 30] out of 20 studies reported or had extractable information regarding VI-RADS cut-off criteria ≥ 4 . Weighted pooled Sn and Sp were 0.78 (95% CI 0.74–0.81) and 0.94 (95% CI 0.91–0.96) (Fig. 1B), while pooled LR + and LR – were 13 (95% CI 9.2–18.2) and

0.23 (95%CI: 0.20 – 0.27), respectively. The area under the HSROC curve was 0.91 (95% CI 0.88–0.93; Fig. 1B). Given a pre-test probability of 30%, Fagan nomogram showed a positive and negative post-test probability for detecting MIBC of 81% and 7%, respectively (Fig. 2B). Similarly, to cut-off ≥ 3 , the inspection of Funnel's plot revealed absence of asymmetry with no significant small-study effect as depicted in Supplementary Fig. 3. Additionally, Egger's regression test confirmed a non-significant small-study effect ($p=0.26$) whilst the 'Trim and Fill' method revealed that no 'missing' studies would need to be included to remove asymmetry from the funnel plots. Furthermore, the source for heterogeneity was confirmed to be dependent mainly from the study design, together with the use of different magnetic strength and slice thickness as depicted in Supplementary Fig. 4B.

Discussion

The first diagnostic meta-analysis by Woo et al. [4] assessing $n=6$ studies with a total of 1170 patients, who underwent mpMRI of the pelvis before TURBT, proved relevant enthusiasm toward the urologic academic community for the VI-RADS scoring criteria and its diagnostic ability to preoperatively provide reliable clinical BCa staging. The article was indeed released after less than 24 months from the original VI-RADS document [3] and reported a pooled sensitivity and specificity for predicting MIBC of 0.83 (95%

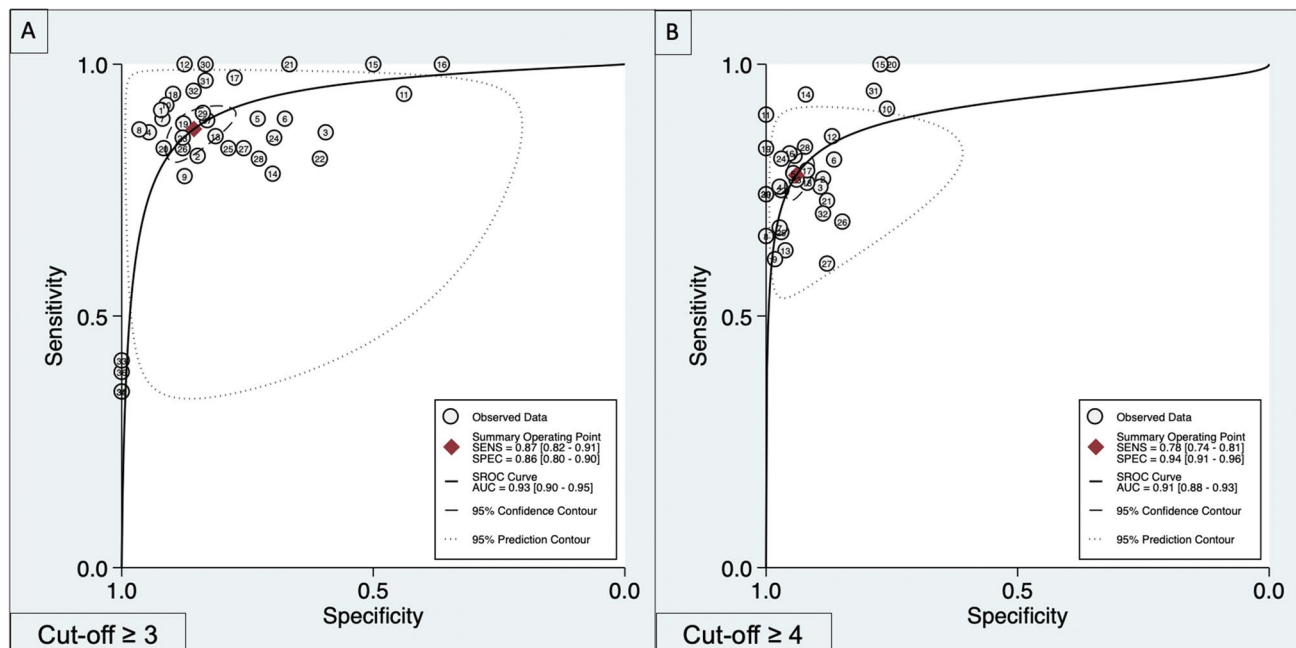


Fig. 2 HSROC for diagnostic performance of studies using VI-RADS predicting MIBC. HSROC hierarchical summary receiver operating characteristics; VI-RADS vesical imaging reporting and data system

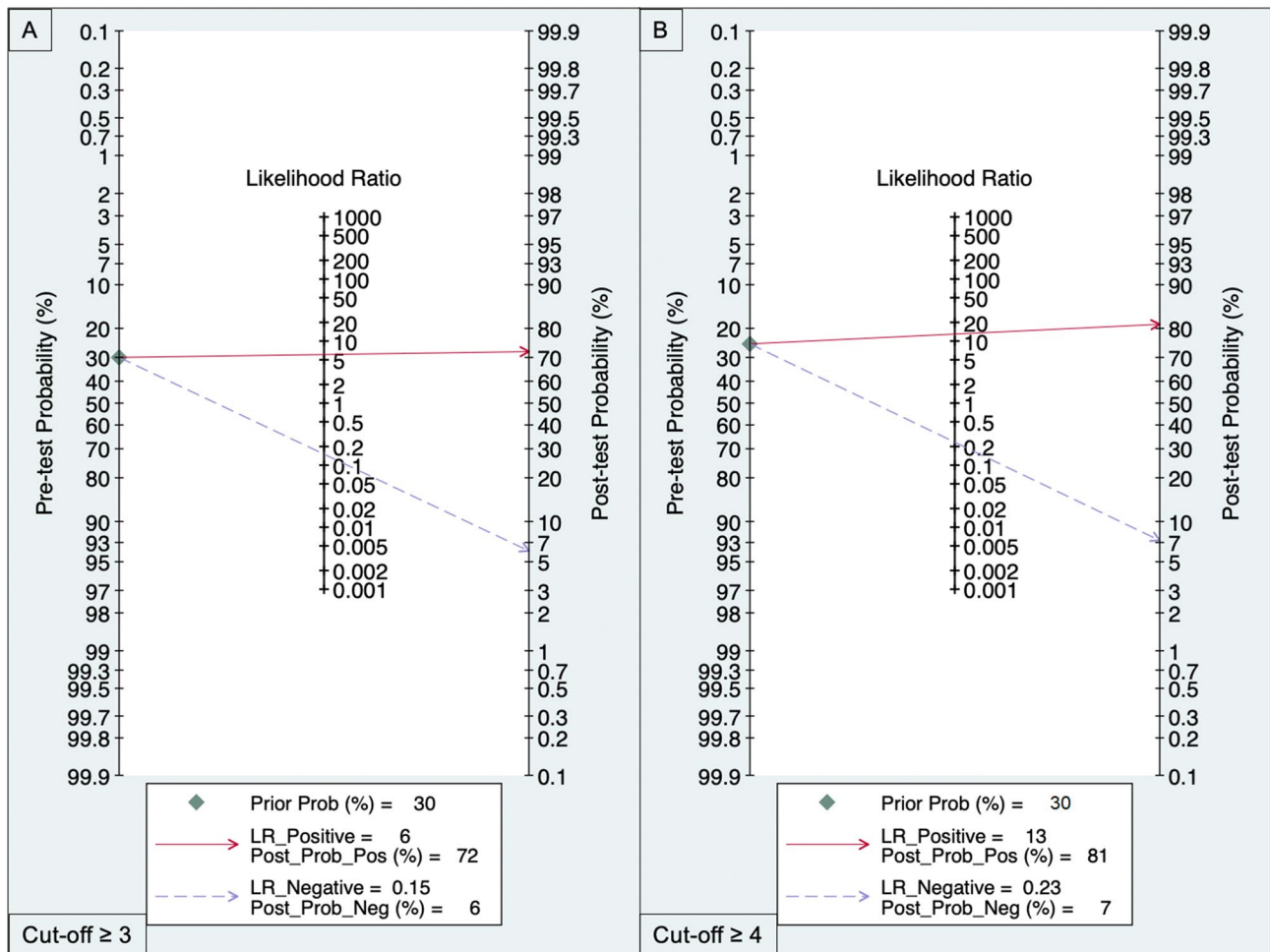


Fig. 3 Fagan nomogram reflecting pre- and post-test probability estimation for clinical utility of VIRADS criterion ≥ 3 (A) or ≥ 4 (B), respectively. *VI-RADS* vesical imaging reporting and data system

CI 0.70–0.90) and 0.90 (95% CI 0.83–0.95) while the area under the HSROC curve was 0.94 (95% CI 0.91–0.95). Similarly, Luo et al. [5] in a separate and independent meta-analysis obtained an overlapping area under the HSROC curve of 0.93 (95% CI 0.91–0.95). Following these preliminary but enthusiastic findings, van der Heijden and Witjes [32] have reasonably and rightly addressed the lack of data regarding reproducibility of the scoring system among different centers, linked to both radiologists' varying expertise and the MRI scanners (e.g., magnetic field, vendor etc.), thus warranting the need for larger updated analysis and cautiously advocating the need for more data and experience before providing definitive statement, especially for the potential inclusion of VI-RADS in the diagnostic algorithm of BCa.

These considerations appear now particularly timely given a complementary growing body of evidence is emerging, exploring not solely the diagnostic performance indexes of VI-RADS itself, but also the potential novel implications, which could derive from the internalization of a reliable

preoperative staging tool in the decision-making process of daily urological practice. For example, Del Giudice et al. [6] have reliably demonstrated the ability of VI-RADS score 2 as a clinical predictor of non-invasive disease at Re-TURBT in those high-risk NMIBCs candidate for eventually avoiding early repeated resection. At the same time VI-RADS cut-off ≥ 3 was considered as for predicting under-staged MIBC at TURBT with the intent for future identification of those false-negative cases that should definitely not miss Re-TURBT.

Finally, the most ambitious available but still ongoing Bladder Path Trial (www.birmingham.ac.uk/bladderpath) is testing whether TURBT can be substituted by mpMRI, after bladder biopsy has proven the presence of bladder cancer, for determining if the patient will be submitted to conservative rather radical interventions.

All these emerging landscapes have led us to develop an updated comprehensive meta-analysis pooling the cumulative diagnostic performance of both VI-RADS cut-off

criteria for MIBC discrimination and furthermore to investigate the contribution of clinical and radiological confounders to the diagnostic accuracy with the aim to provide useful recommendations for future trials and investigations.

The first reassuring finding from our analysis was that VI-RADS score confirmed its excellent performance regardless of the cut-off adopted (HROC: 0.93, 95% CI 0.90–0.95 and 0.91, 95% CI 0.88–0.93, respectively; $p=0.16$). Nevertheless, the implementation of one threshold criterion over another demonstrated specific diagnostic applicability, which could be adapted and implemented in different clinical spectrum. In particular, MIBC defined by VI-RADS cut-off ≥ 4 was clearly associated with greater specificity and LR+ (0.94, 95% CI 0.91–0.96 and 13, 95% CI 9.2–18.2 respectively) reaching a post-test positive probability of 81% compared to 72% for cut-off ≥ 3 . Such difference should therefore translate into recommendations for future research where VI-RADS would be utilized to address patients directly into invasive and radical interventions, such as NAC \pm RC, as for example proposed in the Bladder Path trial. In line with these findings, recently, Del Giudice et al. [9] reported VI-RADS score 5 having the highest ever reported diagnostic performance in identifying locally advanced BCa with extravesical involvement (sensitivity 90.2%; specificity 98.1%; AUC 94.2%, 95% CI 88.7–99.7%), thus delineating an even more reliable imaging-based risk profile of patients, who could avoid the reliance and morbidity of invasive diagnostic TUR, and therefore could receive definitive treatments without time-consuming sequelae related to the staging procedures. On the contrary, VI-RADS cut-off ≥ 3 , given its broader inclusivity of even those VI-RADS score 3 suspicious lesions, delineates the shape of a predictive tool with better sensitivity and LR–, which would decrease the misdiagnosis of MIBC. For these properties, the utilization of such criterion should be preferred in those trials evaluating patients undergoing pathologic confirmation of muscle invasiveness (e.g., selection of high-risk NMIBCs candidate for Re-TURBT). All these considerations find an appropriate support from the maturity and robustness of the data presented in our analysis. We were indeed able to compare the highest ever reported number of experiences validating VI-RADS for MIBC diagnosis in the pre-TURBT setting. More interestingly, our analysis was based on a per-single reader contribution, thus expanding the heterogeneity of different cumulative level of GU radiological expertise worldwide and therefore closely mirroring daily VI-RADS clinical reproducibility.

Our study is, however, not devoid of limitations. First and more importantly, we would readily acknowledge the existence of variable heterogeneity and the risk of bias deemed by the quality of the included studies which exhibit some differences in terms of study designs (retrospective, prospective, single institution, multi-center, etc.), magnetic

strength and MR imaging characteristics. Nevertheless, we deeply investigated through sensitivity, subgroup and meta-regression analyses all the possible available confounders on a per single-reader level in order to balance and model the contribution of clinical and radiological variables to the overall effect size. Of note, differently from the two previous existing meta-analysis, the sample size enrolled in each study was resized as non-significant confounder. Conversely, we confirmed a trend toward greater diagnostic performance in those studies adopting 3 T MR magnet with thinner slices (3 mm).

Finally, according to our updated findings together with the already available data regarding high GU inter-reader variability [33], the shape of this innovative and versatile imaging tool seems confirmed, with potential useful indications which may be further internalized in upcoming clinical trials.

Conclusion

VI-RADS is a diagnostic tool characterized by excellent diagnostic performance regardless its score cut-off criterion for defining MIBC in the pre-TURBT setting. Our analysis further supports the adoption of VI-RADS ≥ 3 or ≥ 4 in certain clinical settings, spanning from accurate diagnostic staging purpose to patient-level therapeutic algorithm personalization.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00345-022-03969-6>.

Author contributions FDG: protocol/project development, data collection or management, manuscript writing/editing, and data analysis. SF: manuscript writing/editing, data collection or management, and data analysis. MP: data collection or management and manuscript writing/editing. MM: manuscript writing/editing and data collection or management. DD'A: manuscript writing/editing and data collection or management. EM: manuscript writing/editing and data collection or management. LMP: manuscript writing/editing and data collection or management. EDB: protocol/project development and manuscript writing/editing. AS: protocol/project development and manuscript writing/editing. VP: protocol/project development and manuscript writing/editing.

Declarations

Conflict of interest Regarding the content of this paper, the authors declare that they have no conflict of interest to disclose.

Research involving human participants and/or animals The present systematic review and meta-analysis do not involve human participants and/or animals.

Informed consent All the studies analyzed in the present article involve experiences where the informed consent was obtained from all individual participants included in the respective analyses.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Panebianco V, Pecoraro M, Del Giudice F, Takeuchi M, Muglia VF, Messina E, Cipollari S, Giannarini G, Catalano C, Narumi Y (2020) VI-RADS for bladder cancer: current applications and future developments. *J Magn Reson Imaging*. <https://doi.org/10.1002/jmri.27361> (Epub ahead of print. PMID: 32939939)
2. Del Giudice F, Campa R, Bicchetti M, De Berardinis E, Panebianco V (2020) Vesical imaging-reporting and data system (VI-RADS) incorporated into bladder cancer clinical practice: what's the perspectives beyond diagnostic accuracy? *Transl Androl Urol* 9(5):2320–2322. <https://doi.org/10.21037/tau-2020-07> (PMID: 3216837; PMCID: PMC7658160)
3. Panebianco V, Narumi Y, Altun E, Bochner BH, Efstathiou JA, Hafeez S, Huddart R, Kennish S, Lerner S, Montironi R, Muglia VF, Salomon G, Thomas S, Vargas HA, Witjes JA, Takeuchi M, Barentsz J, Catto JWF (2018) Multiparametric Magnetic Resonance Imaging for Bladder Cancer: Development of VI-RADS (Vesical Imaging-Reporting And Data System). *Eur Urol* 74(3):294–306. <https://doi.org/10.1016/j.eururo.2018.04.029> (Epub 2018 May 10. PMID: 29755006; PMCID: PMC6690492)
4. Woo S, Panebianco V, Narumi Y, Del Giudice F, Muglia VF, Takeuchi M, Ghafoor S, Bochner BH, Goh AC, Hricak H, Catto JWF, Vargas HA (2020) Diagnostic performance of vesical imaging reporting and data system for the prediction of muscle-invasive bladder cancer: a systematic review and meta-analysis. *Eur Urol Oncol*. 3(3):306–315. <https://doi.org/10.1016/j.euo.2020.02.007> (Epub 2020 Mar 19. Erratum in: *Eur Urol Oncol*. 2020 Dec;3(6):811. PMID: 32199915; PMCID: PMC7293940)
5. Luo C, Huang B, Wu Y, Chen J, Chen L (2020) Use of vesical imaging-reporting and data system (VI-RADS) for detecting the muscle invasion of bladder cancer: a diagnostic meta-analysis. *Eur Radiol* 30(8):4606–4614. <https://doi.org/10.1007/s00330-020-06802-z> (Epub 2020 Apr 2 PMID: 32242273)
6. Del Giudice F, Barchetti G, De Berardinis E, Pecoraro M, Salvo V, Simone G, Sciarra A, Leonardo C, Gallucci M, Catalano C, Catto JWF, Panebianco V (2020) Prospective assessment of vesical imaging reporting and data system (VI-RADS) and its clinical impact on the management of high-risk non-muscle-invasive bladder cancer patients candidate for repeated transurethral resection. *Eur Urol* 77(1):101–109. <https://doi.org/10.1016/j.eururo.2019.09.029> (Epub 2019 Nov 5 PMID: 31699526)
7. Necchi A, Bandini M, Calareso G, Raggi D, Pederzoli F, Farè E, Colecchia M, Marandino L, Bianchi M, Gallina A, Colombo R, Fossati N, Gandaglia G, Capitanio U, Dehò F, Giannatempo P, Lucianò R, Salonia A, Madison R, Ali SM, Chung JH, Ross JS, Briganti A, Montorsi F, De Cobelli F, Messina A (2020) Multiparametric magnetic resonance imaging as a noninvasive assessment of tumor response to neoadjuvant pembrolizumab in muscle-invasive bladder cancer: preliminary findings from the PURE-01 study. *Eur Urol* 77(5):636–643. <https://doi.org/10.1016/j.eururo.2019.12.016> (Epub 2019 Dec 25 PMID: 31882281)
8. Bandini M, Calareso G, Raggi D, Marandino L, Colecchia M, Gallina A, Giannatempo P, Pederzoli F, Gandaglia G, Fossati N, Capitanio U, Colombo R, Salonia A, Briganti A, Montorsi F, De Cobelli F, Messina A, Necchi A (2020) The value of multiparametric magnetic resonance imaging sequences to assist in the decision making of muscle-invasive bladder cancer. *Eur Urol Oncol* S2588–9311(20):30082–30091. <https://doi.org/10.1016/j.euo.2020.06.004> (Epub ahead of print. PMID: 32605888)
9. Del Giudice F, Leonardo C, Simone G, Pecoraro M, De Berardinis E, Cipollari S, Flammia S, Bicchetti M, Busetto GM, Chung BI, Gallucci M, Catalano C, Panebianco V (2020) Preoperative detection of Vesical Imaging-Reporting and Data System (VI-RADS) score 5 reliably identifies extravesical extension of urothelial carcinoma of the urinary bladder and predicts significant delayed time to cystectomy: time to reconsider the need for primary deep transurethral resection of bladder tumour in cases of locally advanced disease? *BJU Int* 126(5):610–619. <https://doi.org/10.1111/bju.15188> (Epub 2020 Aug 17 PMID: 32783347)
10. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D (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(10):e1–34. <https://doi.org/10.1016/j.jclinepi.2009.06.006> (Epub 2009 Jul 23 PMID: 19631507)
11. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM, QUADAS-2 Group (2011) QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 155(8):529–536. <https://doi.org/10.7326/0003-4819-155-8-201110180-00009> (PMID: 22007046)
12. Higgins JP, Thompson SG (2002) Quantifying heterogeneity in a meta-analysis. *Stat Med* 21(11):1539–1558. <https://doi.org/10.1002/sim.1186> (PMID: 12111919)
13. Wang H, Luo C, Zhang F, Guan J, Li S, Yao H, Chen J, Luo J, Chen L, Guo Y (2019) Multiparametric MRI for bladder cancer: validation of VI-RADS for the detection of detrusor muscle invasion. *Radiology* 291(3):668–674. <https://doi.org/10.1148/radiol.2019182506> (Epub 2019 Apr 23 PMID: 31012814)
14. Wang Z, Shang Y, Luan T, Duan Y, Wang J, Wang H, Hao J (2020) Evaluation of the value of the VI-RADS scoring system in assessing muscle infiltration by bladder cancer. *Cancer Imaging* 20(1):26. <https://doi.org/10.1186/s40644-020-00304-3> (PMID: 32252816; PMCID: PMC7137481)
15. Liu S, Xu F, Xu T, Yan Y, Yao X, Tang G (2020) Evaluation of vesical imaging-reporting and data system (VI-RADS) scoring system in predicting muscle invasion of bladder cancer. *Transl Androl Urol* 9(2):445–451. <https://doi.org/10.21037/tau.2020.02.16> (PMID: 32420150; PMCID: PMC7215047)
16. Meng X, Hu H, Wang Y, Hu D, Li Z, Feng C (2021) Application of bi-planar reduced field-of-view DWI (rFOV DWI) in the assessment of muscle-invasiveness of bladder cancer. *Eur J Radiol* 136:109486. <https://doi.org/10.1016/j.ejrad.2020.109486> (Epub 2020 Dec 31 PMID: 33434861)
17. Li S, Liang P, Wang Y, Feng C, Shen Y, Hu X, Hu D, Meng X, Li Z (2021) Combining volumetric apparent diffusion coefficient histogram analysis with vesical imaging reporting and data system to predict the muscle invasion of bladder cancer. *Abdom Radiol (NY)* 46(9):4301–4310. <https://doi.org/10.1007/s00261-021-03091-y> (Epub 2021 Apr 28 PMID: 33909091)
18. Ueno Y, Takeuchi M, Tamada T, Sofue K, Takahashi S, Kamishima Y, Hinata N, Harada K, Fujisawa M, Murakami T (2019) Diagnostic accuracy and interobserver agreement for the vesical

- imaging-reporting and data system for muscle-invasive bladder cancer: a multireader validation study. *Eur Urol* 76(1):54–56. <https://doi.org/10.1016/j.eururo.2019.03.012> (Epub 2019 Mar 26 PMID: 30922688)
19. Ueno Y, Tamada T, Takeuchi M, Sofue K, Takahashi S, Kamishima Y, Urabe Y, Kido A, Hinata N, Harada K, Fujisawa M, Miyaji Y, Murakami T (2021) VI-RADS: multiinstitutional multireader diagnostic accuracy and interobserver agreement study. *AJR Am J Roentgenol* 216(5):1257–1266. <https://doi.org/10.2214/AJR.20.23604> (Epub 2020 Jul 29 PMID: 32755215)
 20. Sakamoto K, Ito M, Ikuta S, Nakanishi Y, Kataoka M, Takemura K, Suzuki H, Tobisu K-I, Kamai T, Koga F (2020) Detection of muscle-invasive bladder cancer on biparametric MRI using vesical imaging-reporting and data system and apparent diffusion coefficient values (VI-RADS/ADC). *Bladder Cancer* 6:161–169
 21. Arita Y, Shigeta K, Akita H, Suzuki T, Kufukihara R, Kwee TC, Ishii R, Mikami S, Okuda S, Kikuchi E, Oya M, Jinzaki M (2021) Clinical utility of the vesical imaging-reporting and data system for muscle-invasive bladder cancer between radiologists and urologists based on multiparametric MRI including 3D FSE T2-weighted acquisitions. *Eur Radiol* 31(2):875–883. <https://doi.org/10.1007/s00330-020-07153-5> (Epub 2020 Aug 23 PMID: 32829418)
 22. Kim SH (2020) Validation of vesical imaging reporting and data system for assessing muscle invasion in bladder tumor. *Abdom Radiol (NY)* 45(2):491–498. <https://doi.org/10.1007/s00261-019-02190-1> (PMID: 31422440)
 23. Hong SB, Lee NK, Kim S, Son IW, Ha HK, Ku JY, Kim KH, Park WY (2020) Vesical imaging-reporting and data system for multiparametric MRI to predict the presence of muscle invasion for bladder cancer. *J Magn Reson Imaging* 52(4):1249–1256. <https://doi.org/10.1002/jmri.27141> (Epub 2020 Mar 27 PMID: 32220105)
 24. Barchetti G, Simone G, Ceravolo I, Salvo V, Campa R, Del Giudice F, De Berardinis E, Buccilli D, Catalano C, Gallucci M, Catto JWF, Panebianco V (2019) Multiparametric MRI of the bladder: inter-observer agreement and accuracy with the vesical imaging-reporting and data system (VI-RADS) at a single reference center. *Eur Radiol* 29(10):5498–5506. <https://doi.org/10.1007/s00330-019-06117-8> (Epub 2019 Mar 18 PMID: 30887202)
 25. Marchioni M, Primiceri G, Delli Pizzi A, Basilico R, Berardinelli F, Mincuzzi E, Castellucci R, Sessa B, Di Nicola M, Schips L (2020) Could bladder multiparametric MRI be introduced in routine clinical practice? Role of the new VI-RADS score: results from a prospective study. *Clin Genitourin Cancer* 18(5):409–415. e1. <https://doi.org/10.1016/j.clgc.2020.03.002> (Epub 2020 Mar 13 PMID: 32273236)
 26. Delli Pizzi A, Mastrodicasa D, Marchioni M, Primiceri G, Di Fabio F, Cianci R, Seccia B, Sessa B, Mincuzzi E, Romanelli M, Castellan P, Castellucci R, Colasante A, Schips L, Basilico R, Caulo M (2021) Bladder cancer: do we need contrast injection for MRI assessment of muscle invasion? A prospective multi-reader VI-RADS approach. *Eur Radiol* 31(6):3874–3883. <https://doi.org/10.1007/s00330-020-07473-6> (Epub 2020 Nov 19. PMID: 33211143; PMCID: PMC8128729)
 27. Etxano J, Rodríguez-Vigil B, Pérez F, Beardo P, Viguri A, Tolosa E, Martínez de Guereñu B (2021) Vesical imaging-reporting and data system (VI-RADS®): initial experience in the classification of muscle-invasive bladder cancer. *Actas Urol Esp (Engl Ed)* 45(4):320–325. English, Spanish. <https://doi.org/10.1016/j.acuro.2020.10.004>. Epub 2021 Jan 30. PMID: 33531283.
 28. Akcay A, Yagci AB, Celen S, Ozlulderden Y, Turk NS, Ufuk F (2021) VI-RADS score and tumor contact length in MRI: A potential method for the detection of muscle invasion in bladder cancer. *Clin Imaging* 77:25–36. <https://doi.org/10.1016/j.clinimag.2021.02.026> (Epub 2021 Feb 23 PMID: 33639497)
 29. Makboul M, Farghaly S, Abdelkawi IF (2019) Multiparametric MRI in differentiation between muscle invasive and non-muscle invasive urinary bladder cancer with vesical imaging reporting and data system (VI-RADS) application. *Br J Radiol* 92(1104):20190401. <https://doi.org/10.1259/bjr.20190401> (Epub 2019 Oct 8. PMID: 31573328; PMCID: PMC6913364)
 30. Metwally MI, Zeed NA, Hamed EM, Elshetry ASF, Elfwakhry RM, Alaa Eldin AM, Sakr A, Aly SA, Mosallam W, Ziada YMA, Balata R, Harb OA, Basha MAA (2021) The validity, reliability, and reviewer acceptance of VI-RADS in assessing muscle invasion by bladder cancer: a multicenter prospective study. *Eur Radiol* 31(9):6949–6961. <https://doi.org/10.1007/s00330-021-07765-5> (Epub 2021 Feb 19 PMID: 33606105)
 31. Vaz A, Zapparoli M (2020) Diagnostic accuracy of retrospective application of the vesical imaging-reporting and data system: preliminary results. *Radiol Bras* 53(1):21–26. <https://doi.org/10.1590/0100-3984.2019.0063>
 32. van der Heijden AG, Witjes JA (2020) Vesical imaging-reporting and data system (VI-RADS) for bladder cancer diagnostics: the replacement for surgery? *Eur Urol Oncol* 3(3):316–317. <https://doi.org/10.1016/j.euo.2020.03.001> (Epub 2020 Apr 14 PMID: 32303479)
 33. Del Giudice F, Pecoraro M, Vargas HA, Cipollari S, De Berardinis E, Bicchetti M, Chung BI, Catalano C, Narumi Y, Catto JWF, Panebianco V (2020) Systematic review and meta-analysis of vesical imaging-reporting and data system (VI-RADS) inter-observer reliability: an added value for muscle invasive bladder cancer detection. *Cancers (Basel)* 12(10):2994. <https://doi.org/10.3390/cancers12102994> (PMID:33076505;PMCID:PMC7602537)

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