Diagnostic Accuracy of MRI for the Detection of Malignant Peripheral Nerve Sheath Tumors: A Systematic Review and Meta-Analysis

Mitchell P. Wilson, MD[1], Prayash Katlariwala, BSc[1], Gavin Low, MBChB, MPhil[1], Mohammad H. Murad, MD[2], Matthew D. F. McInnes, MD, PhD[3], Line Jacques, MD, MSc[4], Andrew S. Jack, MD, MSc[4,5]

Evidence Synthesis and Decision Analysis · Original Research

Keywords

accuracy, malignant, MRI, nerve sheath tumor, peripheral nerve

Submitted: Apr 15, 2020
Revision requested: Jun 4, 2020
Revision received: Jun 7, 2020
Accepted: Jun 10, 2020
First published online: Apr 28, 2021

The authors declare that they have no disclosures relevant to the subject matter of this article.

An electronic supplement is available online at doi.org/10.2214/AJR.20.23403.

OBJECTIVE. This systematic review and meta-analysis evaluates the diagnostic accuracy of MRI for differentiating malignant (MPNSTs) from benign peripheral nerve sheath tumors (BPNSTs).

MATERIALS AND METHODS. A systematic review of MEDLINE, Embase, Scopus, the Cochrane Library, and the gray literature from inception to December 2019 was performed. Original articles that involved at least 10 patients and that evaluated the accuracy of MRI for detecting MPNSTs were included. Two reviewers independently extracted clinical and radiologic data from included articles to calculate sensitivity, specificity, PPV, NPV, and accuracy. A meta-analysis was performed using a bivariate mixed-effects regression model. Risk of bias was evaluated using QUADAS-2.

RESULTS. Fifteen studies involving 798 lesions (252 MPNSTs and 546 BPNSTs) were included in the analysis. Pooled and weighted sensitivity, specificity, and AUC values for MRI in detecting MPNSTs were 68% (95% CI, 52–80%), 93% (95% CI, 85–97%), and 0.89 (95% CI, 0.86–0.92) when using feature combination and 88% (95% CI, 74–95%), 94% (95% CI, 89–96%), and 0.97 (95% CI, 0.95–0.98) using diffusion restriction with or without feature combination. Subgroup analysis, such as patients with neurofibromatosis type 1 (NF1) versus those without NF1, could not be performed because of insufficient data. Risk of bias was predominantly high or unclear for patient selection, mixed for index test, low for reference standard, and unclear for flow and timing.

CONCLUSION. Combining features such as diffusion restriction optimizes the diagnostic accuracy of MRI for detecting MPNSTs. However, limitations in the literature, including variability and risk of bias, necessitate additional methodologically rigorous studies to allow subgroup analysis and further evaluate the combination of clinical and MRI features for MPNST diagnosis.

Malignant peripheral nerve sheath tumors (MPNSTs) are soft-tissue sarcomas originating from peripheral nerve sheath cells, including Schwann cells and fibroblasts, with a 5-year survival rate of only 20–50% [1–4]. They may occur in a de novo (sporadic) fashion or result from the malignant transformation of preexisting benign peripheral nerve sheath tumors (BPNSTs), such as in the context of neurofibromatosis. In fact, approximately 50% of MPNSTs occur in patients with neurofibromatosis type 1 (NF1) [2, 5–9]. Patients with NF1 have a 10–15% lifetime risk of developing MPNSTs, which are a leading cause of morbidity and mortality in this patient group [2, 6, 8, 10–13]. Although NF1 is considered a poor prognostic factor for patients with MPNSTs, the difference in outcome between patients with and those without NF1 has decreased in recent years, most likely due to technologic imaging advancements resulting in better diagnostic capabilities. Earlier detection of MPNSTs and differentiation of these tumors from BPNSTs are clinically important, because surgical resection is the only curative approach for these highly aggressive tumors [14, 15].

Consensus guidelines currently recommend MRI as the imaging modality of choice to evaluate for location, extent of disease, and tumor growth [16]. However, controversy regarding the accuracy of MRI in differentiating MPNSTs from other BPNSTs persists [16–18]. Therefore, current recommendations suggest multidisciplinary treatment planning

[1]Department of Radiology and Diagnostic Imaging, University of Alberta, 2B2.41 WMC, 8440-112 St NW, Edmonton, AB T6G 2B7, Canada. Address correspondence to M. P. Wilson (mitch.wilson@ualberta.ca).

[2]Evidence-Based Practice Center, Mayo Clinic, Rochester, MN.

[3]Departments of Radiology and Epidemiology, University of Ottawa, Ottawa Hospital Research Institute, Ottawa, ON, Canada.

[4]Department of Neurosurgery, University of California, San Francisco, CA.

[5]Department of Surgery, Division of Neurosurgery,University of Alberta, Edmonton, AB, Canada.

doi.org/10.2214/AJR.20.23403

AJR 2021; 217:31–39

ISSN-L 0361–803X/21/2171–31

in addition to biopsy and/or surgery for diagnostic confirmation [19]. The purpose of this systematic review and meta-analysis was to evaluate the diagnostic accuracy of individual and combined MRI features for differentiating MPNSTs from BPNSTs.

Materials and Methods

This systematic review was performed and reported according to current best practices described in the PRISMA Diagnostic Test Accuracy statement and other guidelines [20–22]. The specific methodologic approach has been used in a similar fashion by our group in previous meta-analyses [23, 24]. A preestablished protocol was designed and submitted to the PROSPERO database before initiation of the review (CRD42020163771). Our analysis included only deidentified data, with individual studies acquiring ethical approval through the necessary institutions where required; as such, this study is exempt from ethical approval.

Literature Search

We performed a comprehensive database search of Ovid MEDLINE, Ovid Embase, Scopus, and the Cochrane Library (including the Cochrane Central Register of Controlled Trials, Cochrane Central Register of Protocols, and the Cochrane Database of Systematic Reviews) from inception to December 28, 2019, to identify studies that used MRI for the detection of MPNSTs. Title, abstract, keyword, and medical subject heading terms varied by database and were searched as follows: MRI AND malignant AND peripheral nerve AND (sensitivity OR specificity OR accuracy). Individual database search strategies are outlined in Appendix 1, which can be viewed in the *AJR* electronic supplement to this article at www.ajronline.org. No beginning date or language restrictions were applied, and translation was acquired where necessary. The search was performed according to best practices for electronic search strategies [25]. Articles from each database were then combined, and duplicate articles were removed from the list. Titles and abstracts were screened for relevance, and then two reviewers with 6 and 9 years of imaging experience (M.P.W. and A.S.J., respectively) independently performed a full-text review of potentially relevant studies. Discrepancies were rereviewed to achieve consensus between reviewers. A gray literature search was also performed by one author (M.P.W.), who evaluated the most recent 3 years of major radiologic and peripheral nerve conference proceedings, including Radiological Society of North America, International Society for Magnetic Resonance in Medicine, and Peripheral Nerve Society conferences. Conference abstracts with sufficient information to meet the inclusion criteria were included in the analysis. Finally, reference lists were checked for key studies, and forward searching of these key studies was performed in Google Scholar.

Selection Criteria

All original articles evaluating the diagnostic accuracy of MRI for detecting MPNSTs as compared with a histopathologic reference standard were evaluated with full-text review. To be included in the analysis, studies were required to assess at least 10 patients, assess MPNST as the target lesion, include an MRI index test, use a histopathologic reference standard for MPNST diagnosis and histopathology and/or 1 year of stability (defined as at least 1 year of clinical and/or imaging follow-up showing no tumor change) as a reference standard for BPNST diagnosis, and report sufficient data to reconstruct a 2 × 2 contingency table. Studies using a comparator other than BPNSTs or using a computer algorithm for diagnosis were excluded. Authors were contacted via e-mail if additional information was required to construct a contingency table, and studies were included if sufficient information was supplied. Nonoriginal articles including review articles, guidelines, consensus statements, and letters to the editor were excluded from review.

Data Extraction

Relevant data from included studies were extracted independently by two reviewers (M.P.W. and A.S.J.). Patient characteristics were recorded, including the total number of patients, mean age and range, number and percentage of male patients, number of patients with NF1, total number of lesions, total number of MPNSTs, total number of BPNSTs, number and type of BPNSTs (divided into schwannoma, neurofibroma, and other), combined mean size for all tumors, and location of tumors. Study characteristics including first author, year of publication, prospective versus retrospective study design, single versus multicenter data acquisition, reported reference standard, number of readers, use of consensus reading, and level of training of the reader or readers were recorded. Details of MRI characteristics including scanner brand, magnet strength, sequences used, and sequence-specific data (such as TR, TE, and slice thickness) were recorded when reported in the primary study. Finally, study performance values including true-positives, true-negatives, false-positives, and false-negatives were recorded in relation to the presence of MPNSTs. If a study did not report these performance values directly but reported the sensitivity and specificity in combination with total tumors and total MPNSTs, then the true-positive, true-negative, false-positive, and false-negative results were calculated. For studies with multiple reviewers, a single result was recorded as the mean of all reviewers [26]. When data were inconsistent between the reviewers, the article was reviewed and consensus reached. Data were collected using Excel (version 15.14, Microsoft).

Risk of Bias Assessment

The methodologic quality of each study was evaluated using the QUADAS-2 tool [27]. QUADAS-2 is frequently used in systematic reviews of diagnostic test accuracy to review individual studies for potential sources of bias and concerns regarding applicability by evaluating four individual domains: patient selection, index tests, reference standard, and flow and timing. Studies with a high risk evaluation result for any signaling question in a domain were considered high risk for that respective domain.

Data Analysis

From each study, the number of true-positives, true-negatives, false-positives, and false-negatives was extracted to create a 2 × 2 contingency table. To pool diagnostic accuracy measures, we used a bivariate mixed-effects regression model that allows correlation between sensitivity and specificity. When the regression model did not converge because of a small number of studies (fewer than four) and zero values in a 2 × 2 table, we used a simple random-effects univariate model to pool diagnostic accuracy

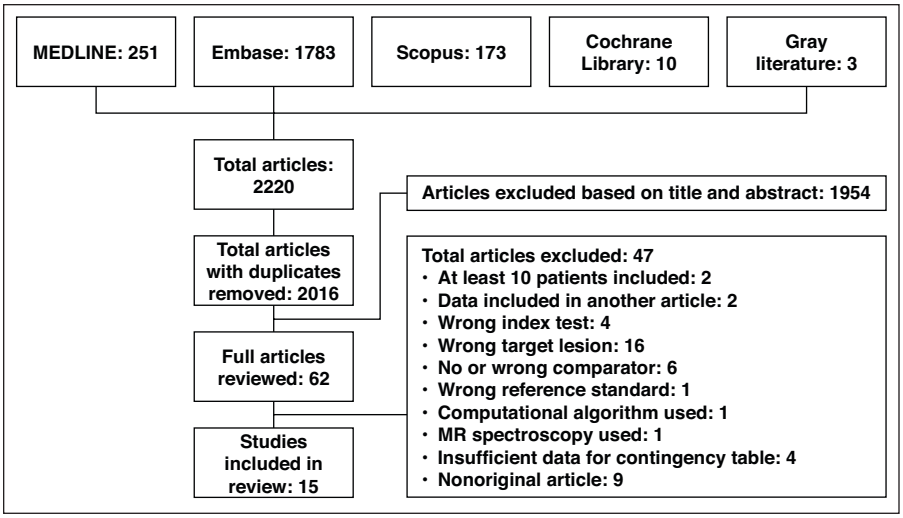

Fig. 1—PRISMA flowchart shows screening and selection of studies included in systematic review.

measures across studies. Results are presented as summary sensitivities, specificities, and ROC AUC values with 95% CIs.

Interstudy variability was assumed, and an exploration of the causes of variability, including clinical features (patients with vs those without NF1), study design (consensus vs nonconsensus), and imaging-specific (1.5 vs 3 T) differences, was planned according to our a priori protocol. Statistical analyses of variability and publication bias are not included in this report according to updated recommendations by the PRISMA Diagnostic Test Accuracy Group [20]. Analysis was conducted using Stata software (version 16, StataCorp).

Results

Literature Search

Figure 1 shows the PRISMA flow diagram for the literature search. A total of 62 articles underwent full-text review. Fifteen studies were included in analysis, with 14 studies providing sufficient data in the article to construct a 2 × 2 contingency table [5, 15, 18, 28–38] and one study providing sufficient information for analysis after the authors were contacted [39]. Forty-seven studies were excluded from the review for reasons outlined in Figure 1.

Patient, Study, and MRI Characteristics

Table 1 shows patient characteristics of individual included studies. A total of 798 lesions evaluated with MRI were included in the review. Of these, 252 lesions were MPNSTs and 546 were BPNSTs. The mean age of patients included in each study ranged from 30 to 50 years. Lesions were reported throughout the entire body, including the upper and lower extremities, neck, chest, abdomen, and pelvis.

Table 2 provides the study characteristics of included studies. Fourteen studies were retrospective, with only one study using a prospective design [15]. All studies used a histopathologic reference standard for MPNSTs. Three studies used clinical and imaging follow-up for at least some BPNSTs, which were defined as such if they showed no change for at least 1 year after initial imaging [15, 34, 38]. Most studies used at least two readers and consensus reading. Two studies included only one reader [31, 38]. All studies that described the imaging interpreter used a radiologist with general and/or subspecialty musculoskeletal training.

Table 3 shows the MRI strength and sequences used by individual included studies. Expanded descriptions of MRI protocols by sequence for each study are provided in Appendix 2, which can be viewed in the *AJR* electronic supplement to this article at www.ajronline.org. Most studies used a Siemens, Philips, or GE Healthcare magnet (or a combination of these magnets) with 1.5- and/or 3-T magnet strength. Two studies used magnet strengths ranging from 0.5 to 1.5 T to acquire their images [5, 29]. Sequences were variable but mostly included a combination of T1 weighting, T2 weighting with or without fat saturation and/or STIR, DWI and ADC, and contrast enhancement.

Diagnostic Accuracy of MRI for MPNST

Appendix 3, which can be viewed in the *AJR* electronic supplement to this article at www.ajronline.org, presents the results of individual imaging features reported in the studies. Pooled and weighted sensitivity, specificity, and ROC AUC by imaging criteria are detailed in Table 4. Detection of MPNSTs using diffusion restriction alone showed a sensitivity of 93% (95% CI, 85–96%) for all DWI and ADC assessment values and 100% (95% CI, 28–100%) for minimum ADC (ADC_{min}) alone. Feature combination showed a sensitivity of 68% (95% CI, 52–80%), and diffusion restriction with or without feature combination showed a sensitivity of 88% (95% CI, 74–95%).

The specificity of diffusion restriction for detecting MPNST was 95% (95% CI, 86–98%) for all DWI and ADC assessment values and 74% (95% CI, 51–89%) for ADC_{min} alone. Perilesional edema and irregular margin had specificities of 94% (95% CI, 76–98%) and 90% (95% CI, 81–95%), respectively. The specificity was 93% (95% CI, 85–97%) for feature combination and 94% (95% CI, 89–96%) for diffusion restriction with or without feature combination. The overall AUC values for feature combination and for diffusion restriction with or without feature combination were 0.89 (95% CI, 0.86–0.92) and 0.97 (95% CI, 0.95–0.98), respectively. Figures 2 and 3 show the sensitivity and specificity forest plots for feature combination and for diffusion restriction with or without feature combination, respectively.

For studies evaluating tumor size alone, threshold values for MPNSTs versus BPNSTs ranged between 4.7 and 6.25 cm. Most of the studies evaluating ADC_{min} used a threshold value of 1×10^{-3} mm^2/s

TABLE 1: Characteristics of Patients in Included Studies

First Author [Reference]	Year	Total Patients	Age (y)		Male Sex	Patients With NF1	Total Lesions	Total MPNST	Total BPNST	Schwannoma	Neurofibroma	Other
			Mean	Range								
Ahlawat [38]	2019	21	30	8–53	12 (57)	21	55	19	36	1	35	NR
Broski [37]	2016	38	38	16–79	NR	23	NR	NR	NR	NR	NR	NR
Chhabra [36]	2011	56	50	15–92	30 (54)	14	56	21	35	11	24	NR
Demehri [35]	2014	31	45	13–78	17 (55)	14	31	9	22	14	5	2
Derlin [34]	2013	31	30	2–63.2	13 (42)	31	75	8	67	NR	NR	NR
Furniss [33]	2008	NR	NR	NR	NR	NR	54	16	38	27	10	1
Karsy [18]	2016	127	42	NR	55 (43)	58	127	24	103	17	82	4
Li [32]	2008	26	47	20–82	19 (73)	NR	26	9	17	16	1	NR
Matsumine [31]	2009	37	43	14–80	18 (49)	37	37	19	18	0	18	NR
Razek [30]	2018	34	34	9–64	18 (53)	NR	34	11	23	17	6	NR
Schwabe [39]	2019	41	30	9–62	29 (41)[a]	41	70	36	34	NR	NR	NR
Stull [29]	1991	17	36	14–75	8 (47)	1	22	3	19	12	7	NR
Wasa [5]	2010	61	39[b]	16-83	32 (52)	34	61	41	20	0	20	NR
Well [15]	2019	26	34.2	17–54	13 (50)	26	67	12	55	NR	NR	NR
Yun [28]	2016	50	NR	NR	NR	NR	50	8	42	NR	NR	NR

Note—Values are the number of patients or tumors with the percentage in parentheses unless otherwise indicated. NF1 = neurofibromatosis type 1, MPNST = malignant peripheral nerve sheath tumor, BPNST = benign peripheral nerve sheath tumor, NR = not reported.
[a]Percentage is calculated from the total number of lesions (29/70).
[b]Value is the median.

or less, with mean ADC threshold values between 1.15 and 1.6×10^{-3} mm^2/s. Three studies reported reader agreement in the evaluation of individual imaging features [35, 37, 39]. Agreement ranged from substantial to almost perfect for each feature across all three studies (kappa statistics between 0.64 and 0.94) [40].

Although a robust subgroup analysis evaluating for clinical, study, and imaging differences among studies was planned in an a priori protocol, no meaningful analysis could be performed because of the limited number of individual performance values included in each category.

TABLE 2: Characteristics of Individual Included Studies

First Author [Reference]	Year	Study Design	No. of Hospitals	Reference Standard	No. of Readers	Consensus Reading	Reader Experience
Ahlawat [38]	2019	Retrospective	Single	Histology + follow-up[a]	1	No	MSK radiologist
Broski [37]	2016	Retrospective	Single	Histology	2	Unclear	MSK radiologists (6 and 19 y)
Chhabra [36]	2011	Retrospective	Single	Histology	2	Yes	Radiologists (6 and 14 y)
Demehri [35]	2014	Retrospective	Single	Histology	2	Yes	NR
Derlin [34]	2013	Retrospective	Single	Histology + follow-up[a]	2	Yes	Radiologists
Furniss [33]	2008	Retrospective	Multicenter	Histology	NR	NR	NR
Karsy [18]	2016	Retrospective	Single	Histology	NR	NR	MSK radiologist
Li [32]	2008	Retrospective	Multicenter	Histology	2	Yes	Radiologists
Matsumine [31]	2009	Retrospective	Multicenter	Histology	1	No	MSK radiologist
Razek [30]	2018	Retrospective	Single	Histology	2	Yes	Radiologists (10 and 20 y)
Schwabe [39]	2019	Retrospective	Single	Histology	2	Unclear	MSK radiologist
Stull [29]	1991	Retrospective	Single	Histology	NR	NR	NR
Wasa [5]	2010	Retrospective	Multicenter	Histology	2	Unclear	NR
Well [15]	2019	Prospective	Single	Histology + follow-up[a]	2	No	Radiologists (4 and 7 y)
Yun [28]	2016	Retrospective	Single	Histology	NR	NR	NR

Note—NR = not reported, MSK = musculoskeletal.
[a]Clinical and imaging follow-up for benign lesions with no change for at least 1 year.

TABLE 3: MRI Strength and Sequences Used in Individual Included Studies

First Author [Reference]	Year	MRI Strength	Sequence
Ahlawat [38]	2019	NR	T1, T2 FS, DWI and ADC, enhancement
Broski [37]	2016	1.5 and 3 T	T1, intermediate-weighted FS, enhancement
Chhabra [36]	2011	1.5 and 3 T	T1, T2, T2 FS or STIR, enhancement
Demehri [35]	2014	3T	T1, T2 FS, STIR, DWI and ADC, enhancement or DCE
Derlin [34]	2013	1.5 T	T1, T2 STIR, enhancement
Furniss [33]	2008	NR	NR
Karsy [18]	2016	NR	NR
Li [32]	2008	1.5 T	T1, T2, T2 FS, enhancement
Matsumine [31]	2009	1.5 T	T2, T2 or T2 FS, enhancement
Razek [30]	2018	1.5 T	T1, true FISP, DWI and ADC
Schwabe [39]	2019	NR	NR
Stull [29]	1991	0.5–1.5 T	T1, PD, T2
Wasa [5]	2010	0.5–1.5 T	T1, T2, enhancement
Well [15]	2019	3T	T1, T1 IP or OOP Dixon, T2, T2FS, DWI and ADC
Yun [28]	2016	NR	NR

Note—An expanded description of specific sequence protocols is detailed in Appendix 2, which can be viewed in the *AJR* electronic supplement to this article at www.ajronline.org. FS = fat saturation, DCE = dynamic contrast enhancement, NR = not reported, FISP = fast imaging with steady state precession, PD = proton density, IP = in phase, OOP = out of phase.

Risk of Bias Assessment

Appendix 4, which can be viewed in the *AJR* electronic supplement to this article at www.ajronline.org, outlines the risk of bias by domain for each study. Most studies had high or unclear risk of bias in their patient selection. Seven studies used specific patient cohorts (i.e., only patients with NF1) [15, 18, 31, 33, 34, 38, 39]. One study used a case-control study design [33], and another included six cases in which imaging was performed only at the time of tumor recurrence [32]. Studies showed a mix of low and high risk for index test, a finding that was largely dependent on the type of performance evaluation used in a study. Most studies showed a low risk for dichotomous performance evaluations (e.g., presence

TABLE 4: Pooled and Weighted Sensitivity, Specificity, and ROC AUC of MRI for Detecting Malignant Peripheral Nerve Sheath Tumors by Imaging Criteria

Imaging Features	Sensitivity (%)	Specificity (%)	AUC
Size	64 (52–74)	80 (66–89)	0.74 (0.70–0.78)
Depth[a]	91 (81–96)	22 (12–35)	
Lobulation	50 (33–66)	88 (81–92)	0.84 (0.80–0.87)
Irregular margins	56 (43–69)	90 (81–95)	0.80 (0.76–0.83)
Perilesional edema	60 (39–78)	94 (76–98)	0.84 (0.80–0.87)
Absence of split fat sign[a]	74 (50–89)	52 (36–68)	
Absence of target sign[a]	93 (80–98)	38 (30–48)	
Heterogeneous signal with or without necrosis	63 (44–79)	88 (75–95)	0.84 (0.81–0.87)
DWI and ADC	93 (85–96)	95 (86–98)	0.96 (0.94–0.97)
ADC_{min} alone	100 (28–100)	74 (51–89)	0.88 (0.85–0.91)
Enhancement[b]	72 (46–89)	82 (60–93)	0.84 (0.81–0.87)
Feature combination[c]	68 (52–80)	93 (85–97)	0.89 (0.86–0.92)
ADC_{min} or ADC_{mean} with or without feature combination	88 (74–95)	94 (89–96)	0.97 (0.95–0.98)

Note—Values in parentheses are the 95% CI. ADC_{min} = minimum ADC value, ADC_{mean} = mean ADC value.
[a]Random-effects univariate analysis was used because of a small number of studies.
[b]Enhancement includes a combination of studies evaluating early arterial enhancement [35] and heterogeneous and/or peripheral enhancement [5, 31, 34].
[c]Feature combination includes studies evaluating a combination of clinical and MRI features [28, 33, 37], a radiologist gestalt diagnosis of malignant peripheral nerve sheath tumor [18, 33], and a combination of mean size and ADC_{min} value [35].

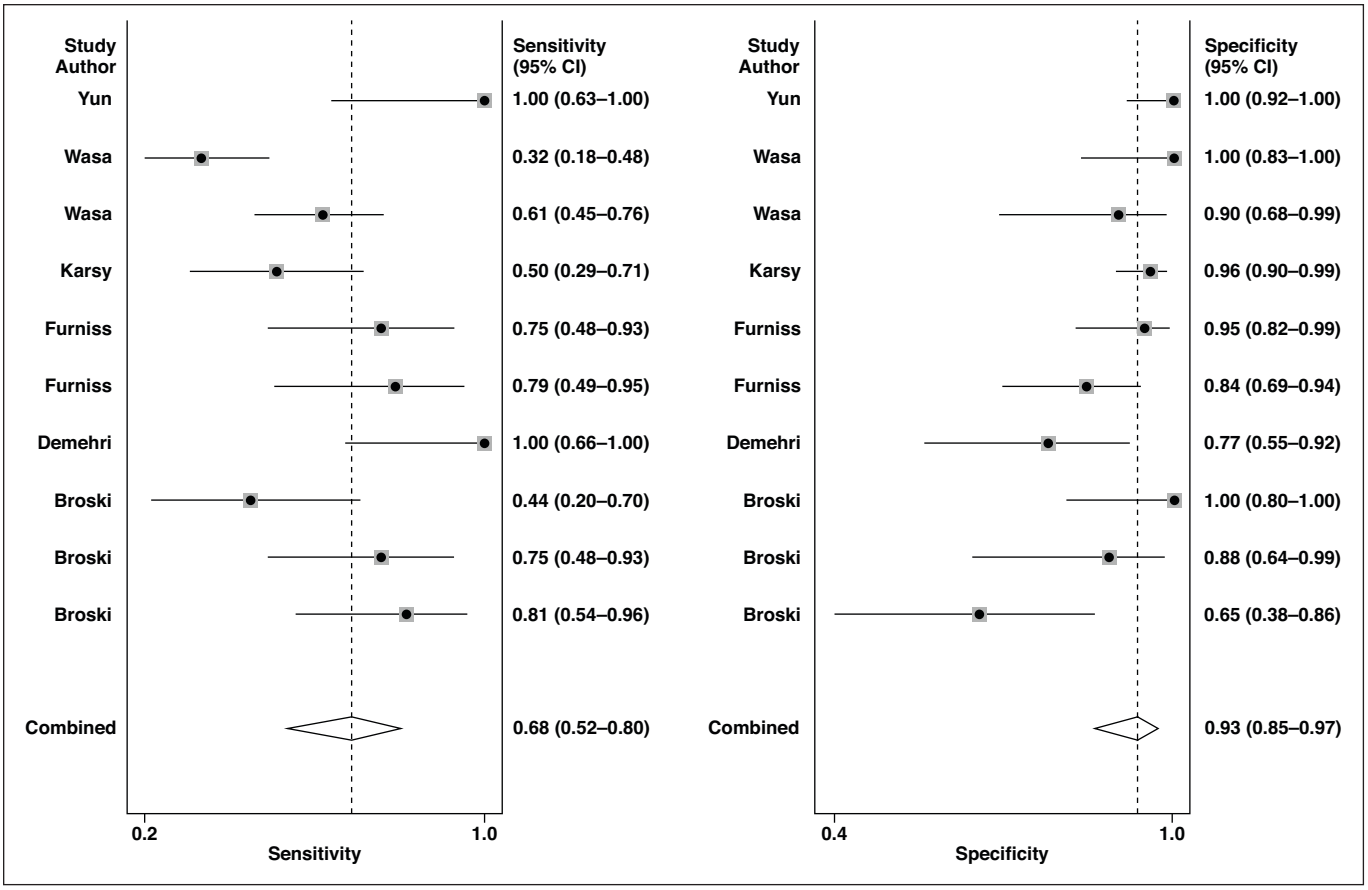

Fig. 2—Pooled sensitivity and specificity of combined features on MRI for differentiating malignant peripheral nerve sheath tumors from other peripheral nerve sheath tumors. Multiple entries for individual studies represent different feature combinations evaluated within an individual study (see Appendix 3, which can be viewed in the *AJR* electronic supplement to this article at www.ajronline.org). Circles show sensitivity or specificity of individual studies, horizontal lines represent 95% CIs, diamonds denote pooled sensitivity or specificity of all studies, and vertical lines represent line of no effect.

of perilesional edema) but a high risk for performance evaluations with continuous variables requiring a threshold value (e.g., tumor size and ADC value). Risk of bias was predominantly low for reference standard and unclear for flow and timing.

Discussion

This review shows that combining features on MRI may increase the diagnostic accuracy of MRI in differentiating MPNSTs and BPNSTs. When studies assessing diffusion restriction and those assessing combined features were evaluated together, the AUC value was 0.97 (95% CI, 0.95–0.98). Few studies, however, evaluated the accuracy of combining multiple imaging features. In a study by Broski et al. [37], the progressive combination of perilesional edema, cystic degeneration or necrosis, and irregular margins sequentially increased the specificity from 65% with any single feature to 100% when all three features were included. However, sensitivity also decreased from 82% with any single feature to 44% with all three. Conversely, in a retrospective single-surgeon study, Yun and Winfree [28] reported that a scoring system using a combination of clinical features (pain, neurologic defects, schwannomatosis, and entrapment) and imaging features (perilesional edema, irregular margins, and contrast enhancement) resulted in perfect performance (100% sensitivity and specificity). Similarly, another study found that the diagnostic accuracy increased from 83% to 89% when the interpreting radiologist used clinical and imaging features compared with imaging features alone [33].

Diffusion restriction may represent a particularly important feature for identifying malignant transformation. When all DWI and ADC performance values were evaluated together, the AUC was 0.96 (95% CI, 0.94–0.97). Although the AUC value of ADC_{min} alone was slightly lower (0.88; 95% CI, 0.85–0.91), sensitivity was high (100%; 95% CI, 28–100%). Given the prognosis of MPNSTs, this high sensitivity represents a preferred test characteristic to avoid a missed diagnosis. Moreover, AUC values of 0.84 were found for lobulation, perilesional edema, heterogeneous signal with or without necrosis, and enhancement, which is in keeping with a previous study of adult soft-tissue tumors that identified irregular or blurred margins and central necrosis as predictors of malignancy [41].

Size alone showed an accuracy of only 0.74 (95% CI, 0.70–0.78). However, both large size and increasing size have been correlated with malignancy in soft-tissue tumors [41]. Demehri et al. [35] combined a size threshold of 4.2 cm with diffusion restriction ($ADC_{min} \leq 1 \times 10^{-3}$ mm^2/s), yielding a sensitivity of 100%.

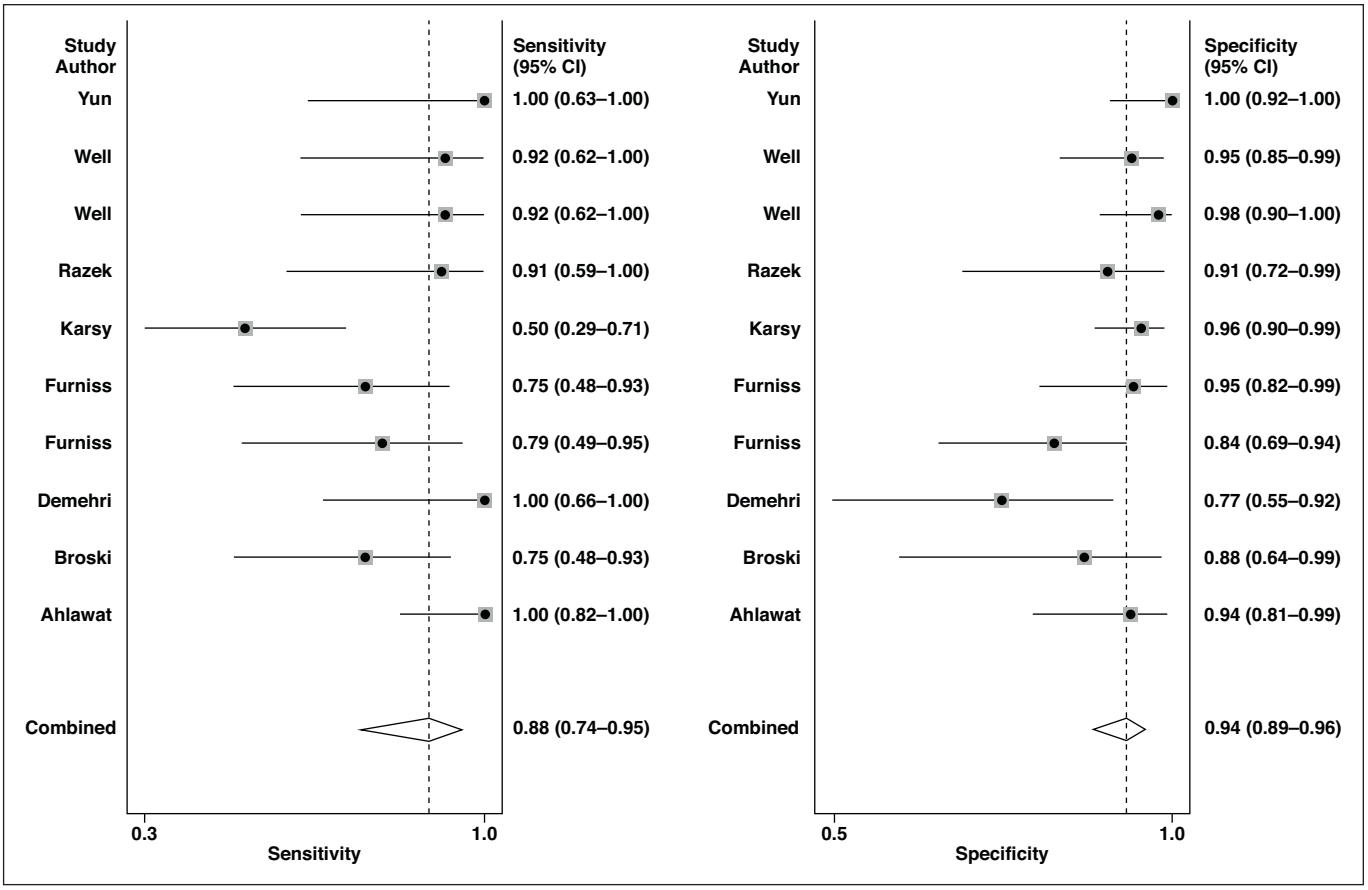

Fig. 3—Pooled sensitivity and specificity of diffusion restriction with or without combined features on MRI for differentiating malignant peripheral nerve sheath tumors from other peripheral nerve sheath tumors. Multiple entries for individual studies represent different feature combinations evaluated within an individual study (see Appendix 3, which can be viewed in the *AJR* electronic supplement to this article at www.ajronline.org). Circles show sensitivity or specificity of individual studies, horizontal lines represent 95% CIs, diamonds denote pooled sensitivity or specificity of all studies, and vertical lines represent line of no effect.

Limitations

Variability is a well-recognized and expected limitation of diagnostic test accuracy studies [20–22]. To account for this, we used a random-effects model that calculates the sensitivity and specificity with a mean and variance rather than estimating a setting-specific sensitivity and specificity. Subgroup analysis, though also planned, was not possible given that few studies evaluated the effect of combining features versus assessing individual features alone. Despite our efforts, variability is likely present and multifactorial in nature. Variability may relate to clinical features (e.g., proportion of included patients with NF1, follow-up, or type of comparator BPNST), study design (e.g., number of hospitals included, consensus reading among radiologists, or reader experience), differences in imaging specifics (e.g., MRI strength or sequence parameters), or differences in data analysis and reporting (e.g., prespecified vs retrospective threshold values). Differences between patient groups with NF1 and those without NF1 may represent a particularly important source of variability (different prognoses as well as MPNST prevalence and pretest probability) [35]; however, one study of 127 patients (58 with NF1 and 69 without NF1) found no difference in diagnostic accuracy when comparing combined features between the two patient groups [18]. Because of a relatively small total number of available cases in which data were only reported for patients with NF1, a meaningful subgroup analysis of this cohort alone could not be performed.

Other potential limitations relating to the included studies are as follows: comparison of lesions with other BPNSTs only and not with other soft-tissue lesions, predominance of consensus reading (which is not common in a clinical setting), the high risk for bias among patient selection and index test domains, and the lack of granularity on histopathologic reference standards. More specifically, other histopathologic classification schemes have differentiated low-grade MPNSTs, high-grade MPNSTs, and atypical neurofibromatous neoplasms of uncertain biologic potential, which can all have different imaging characteristics, outcomes, and management [42]. Furthermore, histopathologic classification also relates to sampling error (particularly if the reference standard for BPNSTs was a biopsy and not resection) as a source of diagnostic variability.

Future Research and Clinical Application

Future research can further evaluate MRI accuracy using feature combinations (e.g., size, perilesional edema, and diffusion restriction). Considering our study's findings, including diffusion restriction may be of particular importance, with an ADC_{min} value

of 1×10^{-3} mm^2/s or less representing a reasonable threshold for detecting MPNSTs. Clinical information would also be helpful when diffusion restriction is included, such as in the context of using NF1 plexiform neurofibromas to monitor for MPNST occurrence or recurrence. Previous studies suggest that adding such clinical information improves the diagnostic accuracy of MRI for differentiating MPNSTs [28, 33]. One study found a high sensitivity for MPNST in symptomatic patients [38]. Another study found a high sensitivity for MPNST in patients with NF1 [35]. Combining features to yield the highest accuracy while maintaining perfect or near-perfect sensitivity is ideal to avoid delayed diagnosis of MPNSTs. Patients with imaging and clinical features indicative of a perfect sensitivity and near-perfect accuracy could be considered for active surveillance rather than invasive interventions such as biopsy and/or surgery. Regardless, a multidisciplinary approach to workup and management remains appropriate for most patients, with individualized approaches modified depending on posttest probabilities. In this regard, combining clinical and imaging features to create a validated patient scoring system would be useful. Because of the relative importance of these different features, a total score could be calculated from the sum of the weighted features, and a decision tree could be constructed to guide individualized management. For example, a low score (indicating BPNST) could suggest conservative management, a high score (indicating MPNST) could suggest operative management, and an intermediate score could suggest biopsy. In summary, this systematic review suggests that combining clinical and MRI features, such as diffusion restriction, will optimize the diagnostic accuracy of MRI for detecting MPNSTs.

References

1. Luzar B, Falconieri G. Cutaneous malignant peripheral nerve sheath tumor. *Surg Pathol Clin* 2017; 10:337–343
2. Kolberg M, Høland M, Agesen TH, et al. Survival meta-analyses for >1800 malignant peripheral nerve sheath tumor patients with and without neurofibromatosis type 1. *Neuro-oncol* 2013; 15:135–147
3. Ducatman BS, Scheithauer BW, Piepgras DG, Reiman HM, Ilstrup DM. Malignant peripheral nerve sheath tumors: a clinicopathologic study of 120 cases. *Cancer* 1986; 57:2006–2021
4. Ducatman BS, Scheithauer BW, Piepgras DG, Reiman HM. Malignant peripheral nerve sheath tumors in childhood. *J Neurooncol* 1984; 2:241–248
5. Wasa J, Nishida Y, Tsukushi S, et al. MRI features in the differentiation of malignant peripheral nerve sheath tumors and neurofibromas. *AJR* 2010; 194:1568–1574
6. Staedtke V, Bai RY, Blakeley JO. Cancer of the peripheral nerve in neurofibromatosis type 1. *Neurotherapeutics* 2017; 14:298–306
7. Mautner VF, Asuagbor FA, Dombi E, et al. Assessment of benign tumor burden by whole-body MRI in patients with neurofibromatosis 1. *Neuro-oncol* 2008; 10:593–598
8. Longo JF, Weber SM, Turner-Ivey BP, Carroll SL. Recent advances in the diagnosis and pathogenesis of neurofibromatosis type 1 (NF1)-associated peripheral nervous system neoplasms. *Adv Anat Pathol* 2018; 25:353–368
9. Carroll SL. Molecular mechanisms promoting the pathogenesis of Schwann cell neoplasms. *Acta Neuropathol* 2012; 123:321–348
10. Pasmant E, Sabbagh A, Spurlock G, et al.; members of the NF France Network. NF1 microdeletions in neurofibromatosis type 1: from genotype to phenotype. *Hum Mutat* 2010; 31:E1506–E1518
11. McGaughran JM, Harris DI, Donnai D, et al. A clinical study of type 1 neurofibromatosis in north west England. *J Med Genet* 1999; 36:197–203
12. McCaughan JA, Holloway SM, Davidson R, Lam WW. Further evidence of the increased risk for malignant peripheral nerve sheath tumour from a Scottish cohort of patients with neurofibromatosis type 1. *J Med Genet* 2007; 44:463–466
13. Evans DG, Baser ME, McGaughran J, Sharif S, Howard E, Moran A. Malignant peripheral nerve sheath tumours in neurofibromatosis 1. *J Med Genet* 2002; 39:311–314
14. Yuan Z, Xu L, Zhao Z, et al. Clinicopathological features and prognosis of malignant peripheral nerve sheath tumor: a retrospective study of 159 cases from 1999 to 2016. *Oncotarget* 2017; 8:104785–104795
15. Well L, Salamon J, Kaul MG, et al. Differentiation of peripheral nerve sheath tumors in patients with neurofibromatosis type 1 using diffusion-weighted magnetic resonance imaging. *Neuro-oncol* 2019; 21:508–516
16. Ferner RE, Gutmann DH. International consensus statement on malignant peripheral nerve sheath tumors in neurofibromatosis. *Cancer Res* 2002; 62:1573–1577
17. Kubiena H, Entner T, Schmidt M, Frey M. Peripheral neural sheath tumors (PNST): what a radiologist should know. *Eur J Radiol* 2013; 82:51–55
18. Karsy M, Guan J, Ravindra VM, Stilwill S, Mahan MA. Diagnostic quality of magnetic resonance imaging interpretation for peripheral nerve sheath tumors: can malignancy be determined? *J Neurol Surg A Cent Eur Neurosurg* 2016; 77:495–504
19. Bhattacharyya AK, Perrin R, Guha A. Peripheral nerve tumors: management strategies and molecular insights. *J Neurooncol* 2004; 69:335–349
20. McInnes MDF, Moher D, Thombs BD, et al.; PRISMA-DTA Group. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. *JAMA* 2018; 319:388–396
21. McGrath TA, Bossuyt PM, Cronin P, et al. Best practices for MRI systematic reviews and meta-analyses. *J Magn Reson Imaging* 2019; 49:e51–e64
22. Macaskill P, Gatsonis C, Deeks J. Analyzing and presenting results. In: Deeks J, Bossuyt P, Gatsonis C, eds. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. Cochrane Collaboration, 2010
23. Singh R, Wilson MP, Katlariwala P, Murad MH, McInnes MDF, Low G. Accuracy of liver and spleen stiffness on magnetic resonance elastography for detecting portal hypertension: a systematic review and meta-analysis. *Eur J Gastroenterol Hepatol* 2021; 32:237–245
24. Wilson MP, Katlariwala P, Murad MH, Abele J, McInnes MDF, Low G. Diagnostic accuracy of ^{99m}Tc-sestamibi SPECT/CT for detecting renal oncocytomas and other benign renal lesions: a systematic review and meta-analysis. *Abdom Radiol (NY)* 2020; 45:2532–2541
25. Sampson M, McGowan J, Cogo E, Grimshaw J, Moher D, Lefebvre C. An evidence-based practice guideline for the peer review of electronic search strategies. *J Clin Epidemiol* 2009; 62:944–952
26. McGrath TA, McInnes MDF, Langer FW, Hong J, Korevaar DA, Bossuyt PMM. Treatment of multiple test readers in diagnostic accuracy systematic reviews-meta-analyses of imaging studies. *Eur J Radiol* 2017; 93:59–64
27. Whiting PF, Rutjes AW, Westwood ME, et al.; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; 155:529–536
28. Yun J, Winfree CJ. A clinical and radiographic score to assess malignant potential of peripheral nerve sheath tumors. *Neurosurgery* 2016; 63(suppl 1):169
29. Stull MA, Moser RP Jr, Kransdorf MJ, Bogumill GP, Nelson MC. Magnetic resonance appearance of peripheral nerve sheath tumors. *Skeletal Radiol* 1991; 20:9–14
30. Razek AAKA, Ashmalla GA. Assessment of paraspinal neurogenic tumors with diffusion-weighted MR imaging. *Eur Spine J* 2018; 27:841–846

31. Matsumine A, Kusuzaki K, Nakamura T, et al. Differentiation between neurofibromas and malignant peripheral nerve sheath tumors in neurofibromatosis 1 evaluated by MRI. *J Cancer Res Clin Oncol* 2009; 135:891–900
32. Li CS, Huang GS, Wu HD, et al. Differentiation of soft tissue benign and malignant peripheral nerve sheath tumors with magnetic resonance imaging. *Clin Imaging* 2008; 32:121–127
33. Furniss D, Swan MC, Morritt DG, et al. A 10-year review of benign and malignant peripheral nerve sheath tumors in a single center: clinical and radiographic features can help to differentiate benign from malignant lesions. *Plast Reconstr Surg* 2008; 121:529–533
34. Derlin T, Tornquist K, Münster S, et al. Comparative effectiveness of ^{18}F-FDG PET/CT versus whole-body MRI for detection of malignant peripheral nerve sheath tumors in neurofibromatosis type 1. *Clin Nucl Med* 2013; 38:e19–e25
35. Demehri S, Belzberg A, Blakeley J, Fayad LM. Conventional and functional MR imaging of peripheral nerve sheath tumors: initial experience. *AJNR* 2014; 35:1615–1620
36. Chhabra A, Soldatos T, Durand DJ, Carrino JA, McCarthy EF, Belzberg AJ. The role of magnetic resonance imaging in the diagnostic evaluation of malignant peripheral nerve sheath tumors. *Indian J Cancer* 2011; 48:328–334
37. Broski SM, Johnson GB, Howe BM, et al. Evaluation of ^{18}F-FDG PET and MRI in differentiating benign and malignant peripheral nerve sheath tumors. *Skeletal Radiol* 2016; 45:1097–1105
38. Ahlawat S, Blakeley JO, Rodriguez FJ, Fayad LM. Imaging biomarkers for malignant peripheral nerve sheath tumors in neurofibromatosis type 1. *Neurology* 2019; 93:e1076–e1084
39. Schwabe M, Spiridonov S, Yanik EL, et al. How effective are noninvasive tests for diagnosing malignant peripheral nerve sheath tumors in patients with neurofibromatosis type 1? Diagnosing MPNST in NF1 patients. *Sarcoma* 2019; 2019:4627521
40. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med* 2005; 37:360–363
41. Lisson CS, Lisson CG, Beer M, Schmidt SA. Radiological diagnosis of soft tissue tumors in adults: MRI imaging of selected entities delineating benign and malignant tumors. *Rofo* 2019; 191:323–332
42. Miettinen MM, Antonescu CR, Fletcher CDM, et al. Histopathologic evaluation of atypical neurofibromatous tumors and their transformation into malignant peripheral nerve sheath tumor in patients with neurofibromatosis 1: a consensus overview. *Hum Pathol* 2017; 67:1–10